PROCEEDINGS OF SPIE

International Conference on Computer Graphics, Artificial Intelligence, and Data Processing (ICCAID 2022)

Ruishi Liang
Jing Wang
Editors

23–25 December 2022
Guangzhou, China

Organized by
Open University of Guangdong (China)

Sponsored by
Guangdong Polytechnic Institute (China)
Academic Exchange Information Centre (China)

Published by
SPIE

Volume 12604

Part One of Two Parts

Proceedings of SPIE 0277-786X, V. 12604

International Conference on Computer Graphics, Artificial Intelligence, and Data Processing (ICCAID 2022)
edited by Ruishi Liang, Jing Wang, Proc. of SPIE Vol. 12604, 1260401
© 2023 SPIE · 0277-786X · doi: 10.1117/12.2685295

The papers in this volume were part of the technical conference cited on the cover and title page. Papers were selected and subject to review by the editors and conference program committee. Some conference presentations may not be available for publication. Additional papers and presentation recordings may be available online in the SPIE Digital Library at SPIEDigitalLibrary.org.

The papers reflect the work and thoughts of the authors and are published herein as submitted. The publisher is not responsible for the validity of the information or for any outcomes resulting from reliance thereon.

Please use the following format to cite material from these proceedings:
Author(s), "Title of Paper," in *International Conference on Computer Graphics, Artificial Intelligence, and Data Processing (ICCAID 2022)*, edited by Ruishi Liang, Jing Wang, Proc. of SPIE 12604, Seven-digit Article CID Number (DD/MM/YYYY); (DOI URL).

ISSN: 0277-786X
ISSN: 1996-756X (electronic)

ISBN: 9781510663350
ISBN: 9781510663367 (electronic)

Published by
SPIE
P.O. Box 10, Bellingham, Washington 98227-0010 USA
Telephone +1 360 676 3290 (Pacific Time)
SPIE.org

Printed in the United States of America by Curran Associates, Inc., under license from SPIE.

Publication of record for individual papers is online in the SPIE Digital Library.

SPIE. DIGITAL LIBRARY
SPIEDigitalLibrary.org

Paper Numbering: A unique citation identifier (CID) number is assigned to each article in the Proceedings of SPIE at the time of publication. Utilization of CIDs allows articles to be fully citable as soon as they are published online, and connects the same identifier to all online and print versions of the publication. SPIE uses a seven-digit CID article numbering system structured as follows:

- The first five digits correspond to the SPIE volume number.
- The last two digits indicate publication order within the volume using a Base 36 numbering system employing both numerals and letters. These two-number sets start with 00, 01, 02, 03, 04, 05, 06, 07, 08, 09, 0A, 0B … 0Z, followed by 10-1Z, 20-2Z, etc. The CID Number appears on each page of the manuscript.

Contents

Part One

COMPUTER GRAPHICS AND VISUALIZATION TECHNOLOGY

BIG DATA TECHNOLOGY AND INFORMATION PROCESSING

COMPUTER CONTROL AND NETWORK SECURITY MANAGEMENT

Part Two

-

ARTIFICIAL INTELLIGENCE AND IMAGE VISION PROCESSING

INTELLIGENT ROBOTICS AND ADVANCED ALGORITHM MODELING

Conference Committee

Conference Chair

Wenqing Liu, Open University of Guangdong (China)

Technical Program Committee Chair

Ruishi Liang, Zhongshan Institute (China)

Local Organizing Chair

Gang Liu, Harbin Engineering University (China)

Publication Chair

Zhisheng Bi, Guangzhou Medical University (China)

Organizing Committees

Ling Cen, South China University of Technology (China)
Yijun Yang, Xi'An Jiaotong University (China)
Jing Wang, Open University of Guangdong (China)
Qi Zhou, Open University of Guangdong (China)
Bin Cai, Open University of Guangdong (China)
Yiqun Chen, Guangdong University of Education (China)
Jinping Liu, Hunan Normal University (China)
Wei Liu, Universidade de Macau University of Macau (China)
Yanping Chen, Guizhou University (China)
Ziyan Zhang, Hainan Tropical Ocean University (China)
Chengyuan He, Asia University (China)
Rajeev Tiwari, University of Petroleum and Energy Studies (India)
Azim Zaliha Abd Aziz, Universiti Sultan Zainal Abidin (Malaysia)
Dimitrios Kollias, University of Greenwich (United Kingdom)
Aslina Baharum, Universiti Malaysia Sabah (Malaysia)
Marina Yusoff, Universiti Teknologi MARA (Malaysia)
Prateek S. Srivastav, University of Chinese Academy of Sciences (China)

Technical Program Committees

Noreddine Gherabi, Sultan Moulay Slimane University (Morocco)
Sahil Verma, Lovely Professional University (India)

P. C. Srinivasa Rao, Koneru Lakshmaiah University (India)
Rahul Vishwanath Dandage, Rajendra Mane College of Engineering and Technology (India)
Surej Rajan C., Toc H Institute of Science and Technology (Arakunnam)
Attlee Munyaradzi Gamundani, Namibia University of Science and Technology (Namibia)
Julia Qing Zheng, California Baptist University (United States)
Shivani Dhall, DAV college (India)
Khaja Mohiddin, Bhilai Institute of Technology, Raipur (India)
Wan Nor Shuhadan Wan Nik, Universiti Sulatan Zainal Abidin (Malaysia)

Modeling of Stereoscopic Images in 3D Environmental Art Design

Yue Yuan[*a]

[a]Shandong Institute of Commerce and Technology, Jinan, Shandong, 250090, China

*Corresponding author: YY23140527@163.com

Abstract

In order to meet the requirements of environmental design, this paper adopts a three-dimensional projection oriented dioramma. In this paper, a method of minimizing matching cost is proposed to find the geometric model of the best model, so as to achieve better results. By using the geometric correlation of image elements, the coordinates of three-dimensional space are reversed, and the accuracy of reconstruction is optimized. Through the simulation and data analysis, it is proved that the proposed method is much better than the existing reconstruction methods in the accuracy and complexity of multi-dimensional space model, and the accuracy and complexity of reconstruction are better than the existing reconstruction methods in the same situation.

Keywords: environmental design; Three-dimensional modeling; Stereo imaging; Matching cost function; Reconstruction error; Verification by simulation

1. INTRODUCTION

With the rapid development of virtual technology, 3D environment design and modeling technology has attracted more and more attention from people [1], and some achievements have been made in the fields of network virtual assistance, urban planning, 3D games and so on [2-3]. However, the current 3D model reconstruction algorithm has a lot of errors, especially in the practical application, and the impact on the environment [4-6]. The three-dimensional modeling of environment requires high three-dimensional reducibility. On this basis, the accuracy of reconstruction algorithm using a single Angle is limited, while 3D modeling using multidimensional data from double angles has become the main research direction at present. In the multi - directional 3D model, the texture mapping technique is used to restore the environment. To further improve the accuracy of the model, the learn-based diorama has been widely applied [10-13]. According to the requirement of current environment, three-dimensional model is established by using stereo image technology. Based on the three-dimensional non-parallel stereoscopic images, this paper uses the stereoscopic matching algorithm based on the least matching cost to optimize the three-dimensional model. On this basis, this paper also proposes an inverse method of 3D reconstruction using image depth extraction technology, and makes the optimal correction. Experiments show that the error rate of the proposed method is significantly higher than that of the existing methods. Model of system

According to the actual demand of environment, a modeling method of three-dimensional environment design using stereo image technology is proposed.

International Conference on Computer Graphics, Artificial Intelligence, and Data Processing (ICCAID 2022)
edited by Ruishi Liang, Jing Wang, Proc. of SPIE Vol. 12604, 1260402
© 2023 SPIE · 0277-786X · doi: 10.1117/12.2674870

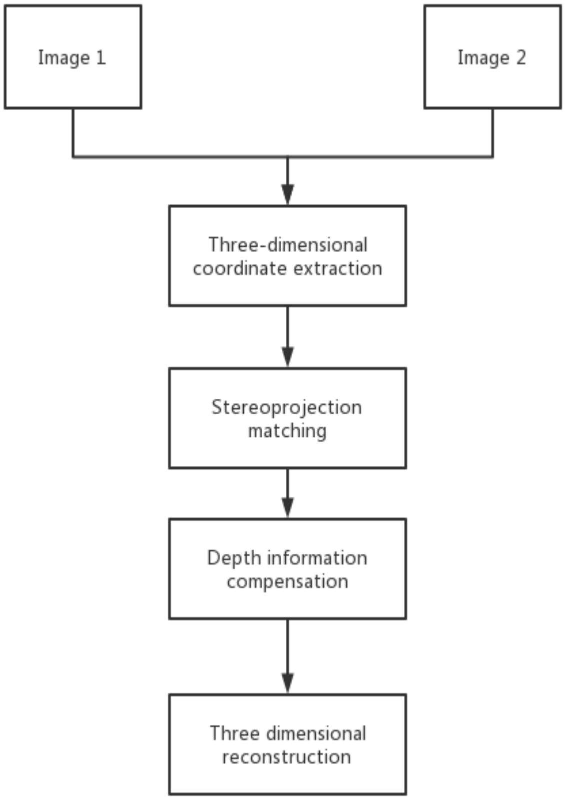


Image 1
Image 2
Three-dimensional coordinate extraction
Stereoprojection matching
Depth information compensation
Three dimensional reconstruction

Figure 1. Principle of bidirectional three-dimensional imaging method

The images are extracted on two non-parallel planes respectively, and the three-dimensional coordinates of the objects in multiple coordinate systems can be obtained through the projection of triangles. Then, a method based on stereoscopic projection is proposed to locate pixel points in 3D scenes. In view of the three-dimensional distortion existing in the 3D reconstruction model, this paper proposes a 3D model based on the traditional, which can achieve a higher restoration effect by compensating the depth of the extracted image in stereo. Figure 2 shows a three-dimensional model of a non-parallel secondary stereoscopic image.

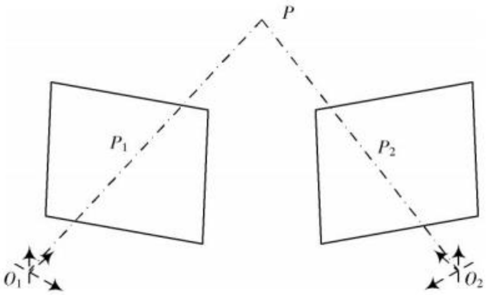


Figure 2. Schematic diagram of non-parallel bidirectional stereoscopic imaging 3D modeling

In FIG. 2, P carries out stereoscopic projection in the coordinate systems O1 and O2. The projection points of the projection plane are P1 and P2 respectively, and the observation coordinates of P1 and P2 in the coordinate system with O1 and O2 as the origin are P1()x1,y1, P2()x2,y2 respectively. Let Xt represent the real coordinates of P, and use Xl and Xr to represent the coordinates of P1 and P2 in the observation coordinate system, then the corresponding relation can be written as:

$$\begin{cases} X_1 = k_1 X_t + t_1 \\ X_t = k_r X_t + t_r \end{cases} \tag{1}$$

Where, kl, kt, tl and tt are the parameters of the stereoscopic projection transformation between the two observation coordinate systems and the real three-dimensional coordinate systems. By transforming equation (1), it can be obtained that: $X_t = KX_1 + T$ (2) in equation, K and T are stereoscopic projection transformation parameter matrices, which are defined as:

$$\begin{cases} K = k_r k_1^{-1} \\ T = t_r - K_{t1} \end{cases} \tag{2}$$

The stereo projection transformation parameters at different points are different, and the stereo three-dimensional matching is to determine the optimal stereo projection transformation parameter matrix by nonlinear optimization.

2. Stereo matching

In this paper, a stereo matching method is proposed for 3D reconstruction of stereo projection. Based on the non-parallel bidirectional stereoscopic imaging model, pixel matching is carried out according to the coordinates of the projection points in two directions to realize the polar line correction. The coordinates of the corrected projection points are:

$$\begin{cases} y_1 = y_2 \\ d = x_1 - x_2 \end{cases} \tag{3}$$

Here d is the pixel difference. The method proposed in this paper is a three-dimensional space matching method based on image. Firstly, the projection difference is used to construct a similar cost function for matching. With the increase of pixel similarity, pixel difference also decreases. By using the constructed cost function for local optimization, a cluster window is selected and the local matching is performed in the set window. Through the local matching of the image and the minimization of the image, the three-dimensional projection coordinates are obtained.

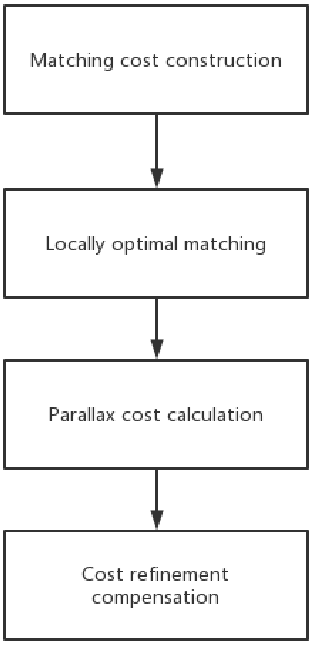

Matching cost construction
Locally optimal matching
Parallax cost calculation
Cost refinement compensation

Figure 3. Steps of stereo matching algorithm

The global matching cost function can be expressed as:

$$E_t = E_d + E_s \tag{4}$$

Where, Ed and Es respectively represent the data cost and smoothing cost of the matching cost function, which are defined as:

$$E_d = \sum_N C_p(d_p) \tag{5}$$

$$E_d = \sum_N V_p(d_p,\ d_s) \tag{6}$$

Where: N represents the number of pixels in the image; p is one pixel; Cp()dp represents the parallax matching cost at p; Vp()dp,ds represents the smoothing cost at p. The smaller the difference between adjacent pixels, the smaller the smoothing cost, then stereo matching is to minimize the matching cost function.

3. Depth information extraction

By comparing the stereoscopic projection, the 3D model can be built preliminarily. However, when 3D image is used for 3D reconstruction, stereo distortion will inevitably occur, so the depth information must be extracted to compensate for the distortion of the model. The 3D model is established by using 2D non-parallel stereoscopic imaging technology. The projection points P2 on the left and right side of the projection point P are P1 and P2.

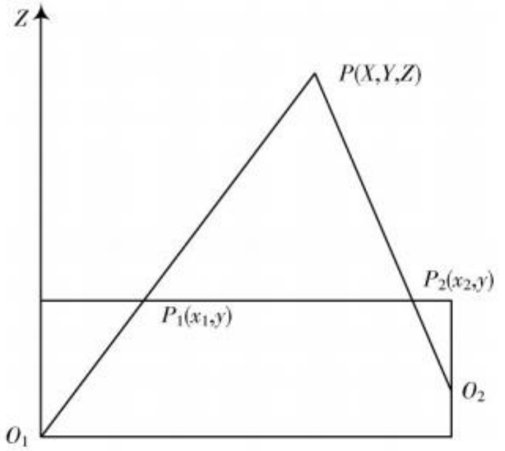


Figure 4. Principle of parallax ranging

Let f be the distance between the origin of coordinates and the projection plane, and B be the horizontal distance between the two observation origins. Then, according to the geometric relationship, we can get:

$$\begin{cases} \frac{X}{x_1} = \frac{Z}{f} \\ \frac{B-X}{-x_2} = \frac{Z}{f} \end{cases} \tag{7}$$

Let F represent the projection size of f in the pixel dimension, which is defined as:

$$F = \frac{f}{d} \tag{8}$$

Then, according to equations (8) and (9), we can get:

$$\begin{cases} Z = \frac{BF}{d} \\ X = \frac{Bx_1}{d} \\ Y = \frac{By}{d} \end{cases} \tag{9}$$

The size of the pixel difference is inversely proportional to the real distance. The 3D depth information of the image can be extracted according to the size of the pixel difference and the projection parameters, and the stereo modeling can be fused and compensated.

4. Simulation verification and performance analysis

The correctness of the model is tested by three-dimensional modeling of the real image in the real scene. On this basis, the three-dimensional image is reconstructed by multiple points and compared with the conventional reconstruction algorithm.

As can be seen from Table 1, three-dimensional reconstruction is carried out in different four Spaces using the method proposed in the paper. By comparing the root mean square value in 3D matching with the reconstruction results, it can be seen that with the increase of the mean difference, the reconstruction effect will be better. The average deviation is not proportional to the error matching rate, which indicates the rationality of the method. In general, the reconstruction method proposed in this paper has great deviation in the reconstruction process.

Table 1 Comparison of spatial multi-point 3D matching reconstruction errors

Number of points	Rootmeansquare error/cm	Mismatching rate/%	Reconstruction error percm
1	2.4	2.2	2.0
2	1.5	0.9	1.4
3	1.2	1.0	1.2
4	1.3	1.2	1.2

To verify the correctness of the proposed method, 3D reconstruction of multiple scenes under the same scene is carried out. According to the reconstruction results in Table 2, the error degree of the 3-D reconstruction method based on depth sensation is similar to that of the method proposed in this paper, but its complexity is much greater than that given in this paper.

Table 2 Comparison cm of 3D reconstruction errors of different methods in the same environment

Number of points	Depth perception method	Texture mapping method	Algorithm of this paper
1	1.8	2.6	2.0
2	1.5	1.6	1.4
3	1.3	1.4	1.2
4	1.0	1.2	1.2

5. Conclusion

Based on stereo image technology, this paper has modeled the three-dimensional space environment. Based on the 3D model of 2D non-parallel stereo imaging, a geometric model based on 3D reconstruction model is constructed by using the stereo matching method. In this paper, an image depth extraction algorithm based on geometric relations is proposed to achieve the optimal correction by solving the three-dimensional space coordinates in reverse. Therefore, 3D environment can be optimized for modeling, and the accuracy of reconstruction can be effectively improved. The correctness of the method is verified by simulation experiments.

Acknowledgment

Project Fund: Environmental Art Design major credit bank pilot project, project number D101

REFERENCES

[1] Zheng Chaoxin, Dong Chen, He Guorong, et al. Dynamic 3D real-time Modeling based on improved Particle Swarm Optimization Algorithm [J]. Computer Engineering and Applications, 2019,55 (5) : 65-71. (in Chinese)

[2] NYSETVOLDJ, SALMONJ. Evaluation of user preferences for 3D modeling and design review sin virtual reality[C]//International CADC on ference and Exhibi tion.Barcelona: IEEE, 202:203-205.

[3] Qin Qianxin, Luo Jianli. Free stereoscopic display technology and its development [J]. Journal of Image and Graphics,2009(10):1934-1941.

[4] ParkJH, HongK, LeeB. Recentprogressinthree-dimensionalinformationprocessingbasedonintegralimaging[J]. AppliedOptics, 2009, 48 (34): 77-94.

[5] Dong Lei, Yang Fugui, Chang Hong, et al. Research progress of stereoscopic display technology [J]. Optics & Optoelectronics Technology,2010,8(2):88-92. (in Chinese)

[6] Kong Lingsheng, Nan Jingshi, Xun Xianchao. Research status of planar 3D display technology [J]. Chinese Optics and Applied Optics,2009,2(2):112-118. (in Chinese)

[7] Wang Yang, Wang Yuanqing. Multi-user free stereoscopic display technology [J]. Chinese Journal of Liquid Crystals and Displays,2009,24(3):434-437. (in Chinese)

[8] Feng Jianping, Wu Lihua. Construction of 3D Panoramic Roaming System based on panoramic image [J]. Computer and Digital Engineering,2013,41(1):115-117. (in Chinese)

[9] KimC, HornungA HeinzleS, etal., Multi - perspectivestereoscopyfromlightfields [J]. J ACMTransactionsonGraphics (TOG), 2011, 30 (6) : 190.

[10] Wang Hongxia, Wu Chunhong, Yang Yang, et al. Research status and development of computer-generated 3D panoramic image [J]. Computer Science,2008,35(6):11-14. (in Chinese)

[11] Li Jia, Sheng Yehua, Zhang Ka, et al. Panoramic Image Mosaic Method Based on Uncalibrated Ordinary Camera [J]. Journal of System Simulation,2013,25(009):2070-2074.

[12] Li Shilei, Du Mingyi, Liu Yanwei, et al. Research on integrated technology of panoramic image and GIS [J]. Urban Survey,2013(4):12-16. (in Chinese)

[13] Li Xiaofang, Wang Qionghua, Li Dahai, et al. The Relationship between the parallax Range of cylindrical lens Grating 3D Display and Stereoscopic Viewing Fatigue [J]. Journal of Optoelectronics · Laser,2012,23(5)873-877

[14] Song Xiaowei, Yang Lei. A grating universal stereo image synthesis method for LCD multi-view image [J]. Computer Applications,2008,28(1):195-198.

[15] Wang Qionghua and Wang Aihong. Review of three-dimensional display [J]. Journal of Computer Applications,2010,30(3):579-581.

[16] Wang Aihong, Wang Qionghua, Li Dahai, et al. Three-dimensional display technology [J]. Electronic Devices,2008,31(1):299-301. (in Chinese)

Video detection method of pointer instrument based on improved FCOS

Xi Chen[1], Pengfei Zhang[1], Wei Xu[2*], Yongjuan Chang[1], Mingshuo Liu[1], Zhenyuan Zhao[2]
[1] State Grid Hebei Information &telecommunication Branch, Shijiazhuang 050000, China
[2] North China Electric Power University, Baoding 071003, Hebei, China
*Email: 1850508590@qq.com

ABSTRACT

Aiming at the problem that pointer instrument detection algorithm has slow locating speed and low real time performance in edge equipment, this paper proposes a pointer instrument video detection method based on improved FCOS. The algorithm is based on FCOS model and uses lightweight network ShuffleNetV2 to extract image features. Using PAN structure to strengthen the original feature fusion network, a bidirectional feature fusion network is formed. The attention module with global context information is introduced to reduce the information attenuation in the process of feature fusion. The two parameters of pixel utilization PUR and relative time increase RIT are introduced to test the influence of images with different image pixels on the detection effect in a more intuitive form. Through experiments, when the resolution of the input image is 1 280×1 280, compared with the baseline model, the detection time of the pointer instrument video detection method based on improved FCOS is reduced by 91.60% when the detection accuracy is similar.

Key words: Pointer instrument; Target detection; FCOS; Lightweight; ShuffleNetV2

1 INTRODUCTION

Distribution station is one of the important hub stations in the power system. In the distribution room, due to the special electromagnetic environment [1], a large number of pointer meters with low cost, high sensitivity and easy use and maintenance are often required [2-3]. Since these pointer meters have no electrical signal output port, they can only be read manually on the spot, which has the problems of high labor cost and low efficiency, as well as certain safety risks [4-5]. With the advancement of smart power grid and the rapid development of computer vision, robotics and other technologies, the inspection of power distribution rooms is gradually being replaced by robots [6-7]. Instead of humans, robots enter the power distribution room, collect video information through the camera mounted on the robot, use computer vision technology to locate the position of the pointer instrument, and then get the instrument reading [8]. Since the identification of pointer meter reading depends heavily on the positioning of the meter dial, how to locate the pointer meter quickly and accurately in the video taken by the roving robot has become a key problem.

To solve the above problems, Li Jun et al. proposed a pointer instrument automatic detection method based on YOLOV4, which improved the detection accuracy and algorithm robustness [9]. Xin Zhang et al. tested the instrument with Faster-RCNN and used the traditional method for correction and other pretreatment, and finally conducted readings on multiple types of instruments to improve the reading accuracy [10]. Liu Jiale et al. proposed an improved YOLOv3 detection algorithm by expanding the data set, using the lightweight network MobileNet framework and designing a new loss function, which improved the detection accuracy and speed of the pointer instrument to a certain extent [11]. Hu Xin et al. proposed a method of correcting the pointer first and then introducing the deep learning yolov5 model to realize the direct positioning of the pointer and the final reading, which solved the limitation of the traditional identification method and improved the detection speed [12]. Li Huihui et al. adopted Hough transform to solve the interference problem of non-circular area in complex scenes and proposed a circular pointer instrument detection algorithm based on the combination of improved pre-trained MobileNetV2 network model and circular Hough transform, which ensured the detection accuracy and reduced the detection time [13]. The above algorithms can solve the problems of slow positioning speed and low detection accuracy of pointer instrument to some extent. However, limited by the power consumption and volume of edge equipment, the above algorithms still cannot solve the defect of slow reasoning speed in edge equipment, resulting in the identification and reading tasks of pointer instrument cannot be applied to the actual production environment. At the same time, with the development of the chip manufacturing industry, ultra-high-definition cameras, even 2K and 4K precision cameras, are widely used in the inspection robots of the distribution room. For the previous

International Conference on Computer Graphics, Artificial Intelligence, and Data Processing (ICCAID 2022)
edited by Ruishi Liang, Jing Wang, Proc. of SPIE Vol. 12604, 1260403
© 2023 SPIE · 0277-786X · doi: 10.1117/12.2674694

algorithms, the image is usually directly subsampled when processing the high-resolution image, so the pixel utilization rate of the original image is very low, which seriously waste the information in the original image. To some extent, it will also affect the accuracy of instrument positioning.

In order to reduce the inference time of the model on the edge equipment and make it applicable to the actual production environment, a pointer instrument video detection method based on the improved FCOS is proposed in this paper. Firstly, by replacing ResNet network with a lightweight feature extraction network ShuffleNetV2, the feature extraction speed of the input image was accelerated. Secondly, PAN module is used to strengthen the original feature fusion network and form a bidirectional feature fusion network, which improves the ability of the model to locate the pointer instrument under complex background while hardly increasing the model training parameters. Finally, an attention module with contextual information is introduced to highlight the significant features in the process of feature fusion and make up for the decline in positioning accuracy caused by the use of ShuffleNetV2. The loss function is improved to solve the problem that the centrality branch is not easy to converge on the lightweight model.

2 APPROACH

2.1 Overall framework of the model

The overall framework of pointer instrument video detection method based on improved FCOS is shown in Figure 1. Firstly, the image is input into the model for convolution, and the feature map is obtained. Feature maps with different resolutions are fed into the bidirectional feature fusion network with attention mechanism to fuse semantic information and location information of different layers. Finally, different scale targets are detected on the fused features. In order to increase the robustness of detection, the sliding window mechanism is used to make statistics on the adjacent frame frequency images, and the detection criterion is designed. When the adjacent frames in the sliding window detect the target at the same position, the detection target of the current frame will be determined. Therefore, during the training, the image cut by the video stream can be directly input as the training set.

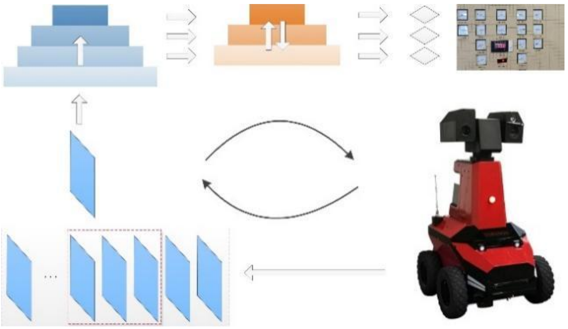


Fig.1 System structure diagram

2.2 FCOS network framework

FCOS (Fully Convolutional One-Stage Object Detection) refers to full-convolution single-stage target detection. Proposed by Zhi Tian[14] equals in 2019, FCOS is a full-convolution based single-stage target detection algorithm, which predicts the category of each spatial position on the feature graph by pixel by pixel prediction. Different from traditional target detection networks based on anchor frames, FCOS does not rely on pre-defined anchor frames and suggested regions, so the network training avoids complex calculations related to anchor frames, saves memory consumption during training, and avoids hyperparameters related to anchor frames that are very sensitive to final detection results. Compared with the target detection algorithm based on anchor frame, it can achieve higher accuracy and reduce the detection time.

Structurally, the FCOS detection method is divided into three parts: backbone network, feature fusion module and detection head. The network structure of FCOS is shown in Figure 2.

(1) The backbone network is responsible for receiving the input image and extracting the image features. It is generally composed of the artificially designed classical feature extraction network (such as ResNet) or the network automatically searched by the Neural Architecture Search algorithm (NAS). This part is the basis of the whole network. But its training parameters are also a large part of the whole network.

(2) Feature fusion module fuses multi-layer features extracted from the backbone network. When extracting image features from backbone network, the extracted features contain richer semantic information with the deepening of backbone network, but the boundary details are less and less due to repeated downsampling. Generally, the backbone network will output the image features of different layers. In order to make full use of the image features of different layers, it is necessary to add the feature fusion module for fusion, so as to give consideration to the semantic information and the boundary details.

(3) The detection head is mainly responsible for detecting targets of different scales on different feature layers after fusion. For each layer of features, there are three branches, namely classification, centrality and regression, which are respectively responsible for predicting the confidence degree of each spatial position on the feature map belonging to the target, the deviation degree from the target center and the position information of the boundary box.

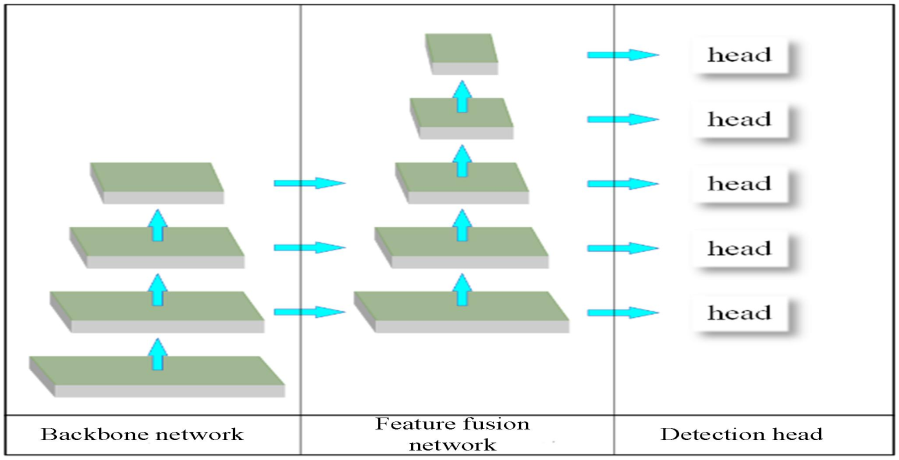


Fig.2 Structure diagram of FCOS detection method

2.3 ShuffleNetV2 network

The original FCOS model uses ResNet[15] as the backbone network to extract image features. However, due to the high computational complexity of ResNet, the deployment requires high computing power of hardware devices, and the realization of real-time target detection of video streams on edge devices and low-power devices is poor. Therefore, ShuffleNetV2[16] with better real-time performance is used in this paper to replace the original ResNet as the feature extraction network of FCOS model. The network structure of ShuffleNetV2 is shown in Table 1, and the backbone network structure of ShuffleNetV2 is shown in Figure 3.

Tab.1 ShuffleNetV2 Network structure

layer	KSize	Stride	Repeat
Image	--	--	--
Convolution	3×3/3×3	2/2	1
Stage2	3×3/3×3	2/1	1/3
Stage3	--	2/1	1/7
Stage4	--	2/1	1/3
Conv5	1×1	1	1
GlobalPool	40×40	--	--

Note: -- indicates no operation.

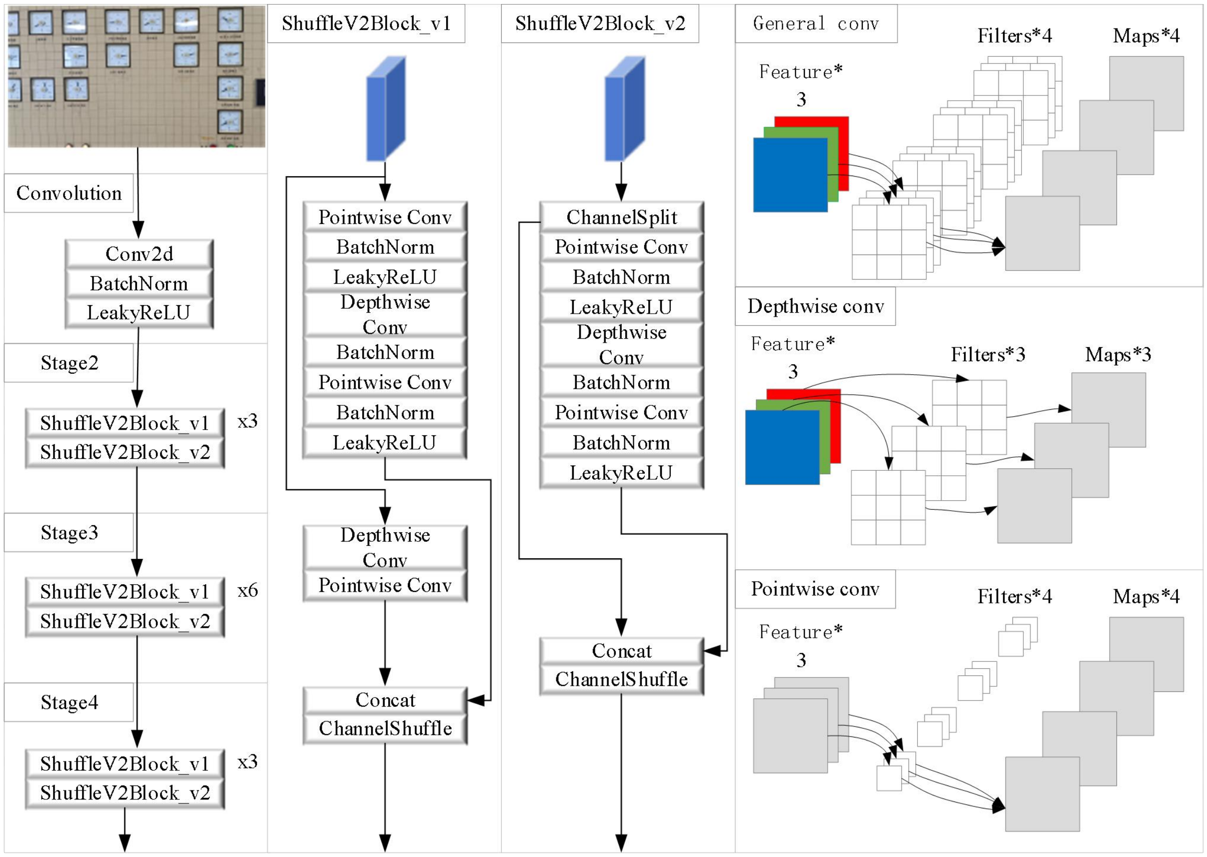


Fig.3 ShuffleNetV2 Backbone network

In ShuffleNetV2 network, 1×1 and 3×3 convolution combinations are repeatedly used to extract image features, and short-circuit connection design is used to increase the depth of the network and obtain better semantic information. At the same time, the operation of adding two features in the short-circuit branch is replaced by the operation of splicing, which realizes the reuse of features. The operation of Channelsplit and Channelshuffle are added before and after the design of short-circuit connection respectively. The former is used to optimize the computational efficiency and reduce the huge training parameters, while the latter is used to fuse features and improve the detection accuracy of convolutional networks. The application of channel separation and channel scrambling, while maintaining the detection accuracy, reduces the computational complexity of the network, reduces the memory consumption and detection time, and greatly improves the computational efficiency of the network. In addition, for the subsampling module, channel separation operation is no longer used, but the input is first subsampled with a deep convolution step of 2, and then directly joined with the original input of the branch after adjusting the channel with 1×1 convolution. So the size of the feature graph is halved, but the number of channels is doubled.

Based on the above design, compared with the traditional feature extraction network with high density convolution operation, ShuffleNetV2 is used to obtain a great improvement in execution efficiency at a small cost of accuracy, which is of great value from the perspective of engineering.

2.4 Integrate the attention mechanism module

2.4.1 PAN structure

In order to solve the multi-scale problem in object detection, Lin T Y et al. [17] introduced Feature Pyramid Networks (FPN), whose structure is shown in Figure 4.

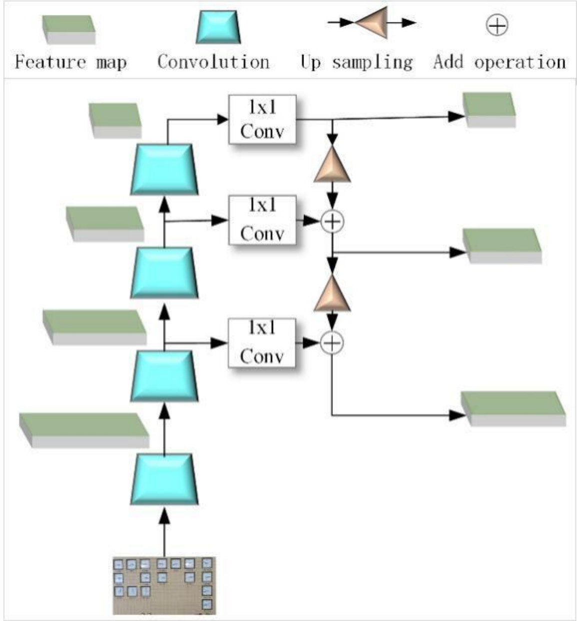


Fig.4 FPN structure

For the feature map extracted through the backbone network, the semantic information of low-level features is less, but the target location information is accurate. On the contrary, the semantic information of high-level features is rich, but the target location information is sketchy. Therefore, the performance of the detection network is greatly improved without increasing the calculation amount of the original model.

The FPN structure can be divided into three parts, namely, bottom up (convolution operation), top down (upsampling) and horizontal join. The bottom-up process represents the forward propagation of the backbone network. The input image is downsampled by the convolution of great lengths to generate features with different resolutions. Top-down process means that the features obtained from the upper level are transmitted to the lower level through upsampling, and the semantic features from the upper level are transmitted to the lower level in turn. The horizontal joining process means that features of the same size are combined by convolution with the convolution kernel size of 1×1 for feature fusion.

However, in the FPN structure, the top-down process only transmits the upper-level semantic information, but not the positioning information. Moreover, the propagation path from lower-level features to higher-level features is too long, resulting in the subsequent prediction based on a single perspective.

Pointer instrument background is complex, there are many interference factors, when there is a visually presented circular object on the background, it is easy to be confused with pointer instrument, resulting in positioning error. In view of this, a Pyramid Attention Network (PAN)[18] under the attention mechanism is introduced behind the FPN structure to build a bottom-up pyramid, which is used to enhance the original FPN structure and transfer the strong positioning features of the lower layers. Thus enhancing localization capability at multiple scales. The PAN structure is shown in Figure 5.

Through the PAN structure, the bottom-up path enhancement is realized to shorten the information transmission path. At the same time, the precise positioning information of low-level features is utilized to form a bidirectional feature fusion network. Through the bidirectional feature fusion network, the targets of different scales can be predicted on the feature maps of different layers after fusion, which greatly improves the detection accuracy of the pointer instrument.

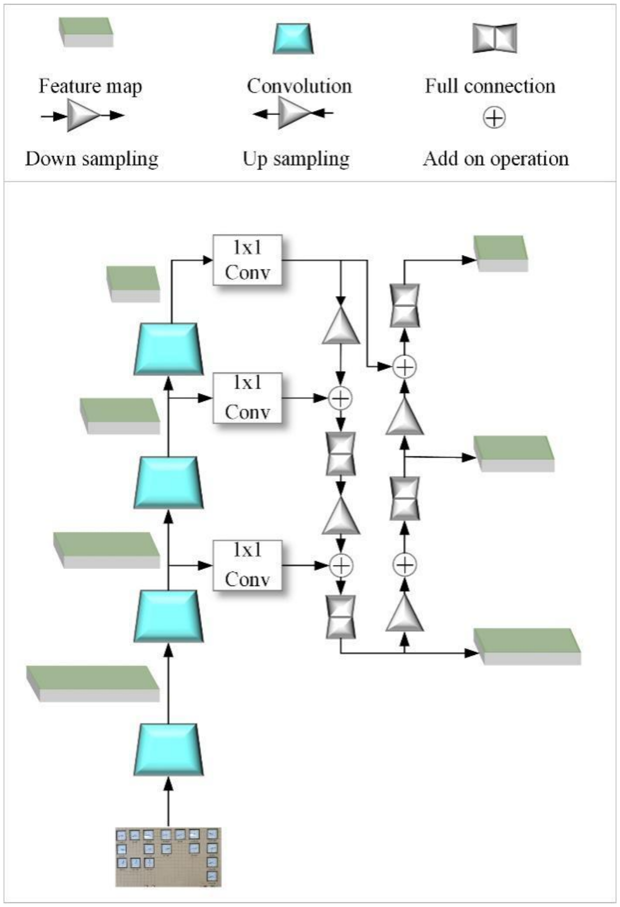


Fig.5 PAN structure

2.4.2 Module of attention

In order to compensate the declining detection accuracy caused by the use of the lightweight feature extraction network ShuffleNetV2, the CBAM (Convolutio-nal Block Attention Module) is introduced into each feature layer of PAN [19]. The module is composed of channel attention and spatial attention, as shown in Figure 6.

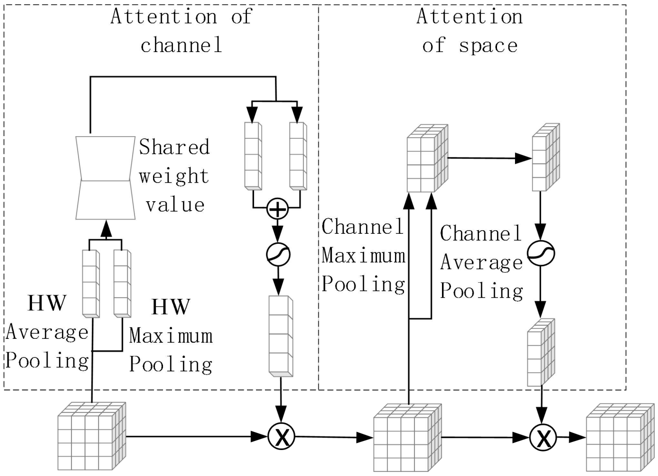


Fig.6 CBAM structure diagram

In each feature layer, attention is added to the two dimensions of space and channel to highlight the significant features of each scale, and the adaptive refinement of the multi-scale feature map. Since CBAM is a lightweight general purpose module, the performance loss caused by its addition can be ignored, so that it can be seamlessly integrated into any CNN architecture and efficiently improve the detection accuracy. CBAM module input feature mapping, then first through the channel attention module M_C, get weighted results and then through the spatial attention mapping module M_S, finally output significant feature mapping. CMAB is calculated as shown in the equation.

$$F' = M_{\mathrm{C}}(F) \otimes F \tag{1}$$

$$F'' = M_{\mathrm{S}}(F') \otimes F' \tag{2}$$

Where: F is input feature mapping; F' represents the features mapped by channel attention; F'' means that the salient features are mapped; ⊗ said dot product.

The channel attention module mainly focuses on the specific content of input features. Firstly, the average pooling and maximum pooling are applied to compress the input feature mapping in the length and width dimensions to aggregate the spatial information of the input feature mapping. Then the aggregated features are sent into the shared network to further compress the spatial dimension of the features to reduce the parameters. Finally, the channel attention diagram is obtained by summing element by element and nonlinear processing. According to the feedback during the calculation of gradient back propagation, the average pooling has feedback for every pixel on the input feature map, while the maximum pooling only has feedback in the feature map with the maximum response. The channel attention module can be expressed as:

$$\begin{aligned} M_{\mathrm{C}}(F) &= \sigma(MLP(AvgPool(F)) + MLP(MaxPool(F))) \\ &= (W_1(W_0(F_{\mathrm{avg}}^{\mathrm{C}})) + W_1(W_0(F_{\mathrm{max}}^{\mathrm{C}}))) \end{aligned} \tag{3}$$

Where: MLP represents multi-layer perceptron; $AvgPool$ and $MaxPool$ represent average pooling and maximum pooling, respectively; σ means the nonlinear function sigmod.

Spatial attention module is mainly used to highlight the location information of input features. Firstly, the input features are compressed by applying average pooling and maximum pooling in the channel dimension successively. After that, the two pooled output results were spliced and sent to a convolution layer with a convolution kernel size of 7×7 to compress the channel dimension. Finally, the spatial attention diagram is obtained by nonlinear processing. The spatial attention module can be expressed as:

$$\begin{aligned} M_{\mathrm{S}}(F) &= \sigma(f([AvgPool(F); MaxPool(F)])) \\ &= \sigma(f(F_{\mathrm{avg}}^{\mathrm{S}}; F_{\mathrm{max}}^{\mathrm{S}})) \end{aligned} \tag{4}$$

Where: f stands for convolution layer.

By introducing the attention-mechanism module CBAM into PAN structure, the representation ability of feature graph can be improved by continuous channel attention and spatial attention, and the detection ability of pointer instrument can be effectively improved. As shown in the blue rectangle box in Figure 1, CBAM is closely connected with the feature map, connected by the up-sampled higher-order feature map, and then weighted fusion operation is performed with the original feature map.

2.5 Function of loss

The detection head of the baseline model contains three branches, namely category, centrality and position regression. Therefore, the loss function of the baseline model is divided into three parts, as shown in the equation.

$$loss = loss_{\mathrm{cls}} + loss_{\mathrm{cnt}} + loss_{\mathrm{reg}} \tag{5}$$

Where: $loss_{\mathrm{cls}}$ is classified loss; $loss_{\mathrm{cnt}}$ is center degree loss; $loss_{\mathrm{reg}}$ is positional regression loss.

In order to reduce the computation amount of the model, and considering that the centrality branch does not converge easily in lightweight model, Generalized Focal Loss (GFL) is used to replace the original loss function, so that the new loss only contains two terms. As shown in the equation:

$$loss_{\text{new}} = loss_{\text{cls+cnt}} + loss_{\text{reg}} \tag{6}$$

Where： $loss_{\text{cls+cnt}}$ is a classification Loss, and it is realized by Quality Focal Loss.$loss_{\text{reg}}$ is a regression Loss and is implemented using Distribution Focal Loss.

3 EXPERIMENTS

3.1 Experimental environment

In order to demonstrate the effectiveness of the proposed algorithm qualitatively and quantitatively, images including pointer meters are collected from the actual production environment of the distribution house to form a data set, and the proposed algorithm is trained and verified. The data set adopts the frame-by-frame cropping of the video collected by the power distribution room to obtain a total of 2387 images including various pointer meter images. The training set and verification set are divided according to the ratio of 7:3. In order to avoid the influence of accidental factors on the results, the data set was divided into the proportion five times by random sampling, and five kinds of data sets containing different data but the same proportion were divided. For each group of experiments, the training and verification were carried out on five data sets respectively, and the average value was taken as the final result. In the verification process, the video stream was input and the output result was the data result of each frame.

The training platform of this paper is a universal server with Ubuntu18.04 LTS operating system installed. Python is used as the programming language, Pytorch is used as the neural network framework, and CUDA10.1 is used to call NVIDIA 1080Ti graphics card for accelerated training. Batchsize is set to 4, and 100 epochs are trained. Stochastic Gradient Descent (SGD) is used as the optimizer, and dynamic learning rate adjustment strategies are adopted. The initial learning rate is set to 0.005, the momentum parameter is set to 0.9, and the learning rate attenuation is 5. The verification platform is Jetson XAVIER NX, an embedded development board with 384 NVIDIA VoltaTM GPU cores and 48 tensor cores, and a CPU with 6 NVIDIA Carmel ARM 64-bit cores, with a power consumption of only 15W. Install the Ubuntu18.04 LTS OS. Compared with the general server platform, the embedded platform is limited by power consumption and volume, so its performance is much weaker, but it is more in line with the actual working situation. The strategy adopted in this paper is to first train the model on the universal server platform, and then deploy the model on the embedded platform for verification, and compare the performance of the model by comparing the detection time consumption and detection accuracy.

In order to quantify the accuracy and efficiency of the algorithm in this paper, since the specific categories of pointer meters are not distinguished in the algorithm, Average Precision (AP) is used to evaluate the accuracy of the model. The rapidity of the model was evaluated using the time T of an image detected on an embedded platform. Pixel Utilization Rate 2K (PUR2K) and Relatively Increased Time (RIT) were simultaneously introduced. PUR2K represents the ratio of the number of pixels of the image actually processed by the model to the number of pixels of the image to be processed (2K resolution); RIT represents the relative increase of the detection time consumed for the same model under other input resolution images compared with the detection time consumed for the image with input resolution of 224x224. The calculation is as follows:

$$\text{PUR2K} = \frac{\text{pixelsum}(I_{rsl})}{\text{pixelsum}(I_{\text{2K}})} \times 100\% \tag{7}$$

Where: pixelsum() represents the pixel summation function, which is used to calculate the total number of pixels in the image; I_rsl and I_2K represent images with a resolution of rsl and images with a resolution of 2K respectively. rsl is 224, 512, 960, 1 280, and 2k is 2 048.

$$PIT = \frac{T_{rsl} - T_{224}}{T_{224}} \times 100\% \tag{8}$$

Where:T_{rsl} and T_{224} respectively model in the input image resolution for $r\ s\ l$ and 224 at the time of the testing time.

3.2 Experiment of contrast

In order to fully illustrate the effectiveness of the improved FCOS model, the detection effects of using FCOS and improving FCOS were compared respectively for four different resolution images (224x224, 512x512, 960x960, 1 280x1

280). Table 2 shows the test results of the pointer instrument before and after improving FCOS under different input resolutions.

Tab.2 Pointer meter detection results before and after fusion similarity metric loss (Average of 10 times)

model	PUR(2K)/%	AP/%	T/ms	PIT/%
*(224)	1.20	98.63	124.3	0
*(512)	6.25	98.81	326.8	162.9
*(960)	21.97	99.02	777.7	525.7
*(1280)	39.06	99.08	1 271.6	923.0
Imp*(224)	1.20	98.55	59.4	0
Imp*(512)	6.25	98.74	61.9	4.2
Imp*(960)	21.97	98.89	67.0	12.8
Imp*(1280)	39.06	99.04	106.8	79.8

Note：* stands for FCOS; Imp stands for Improved。

First of all, under the same resolution, the improved FCOS model has little difference in detection accuracy, and the detection speed is much better than the baseline model. From the perspective of accuracy, under the four different resolutions, AP has a certain degree of decline compared with the baseline model, but the decline is very small and has almost no impact on engineering applications. When the input resolution is 1 280x1 280, AP decreases the least, only 0.04%. From the perspective of detection time, the detection time T is significantly smaller than that of the baseline model, with a decrease range of 52.21%, 81.06%, 91.38% and 91.60%, respectively. When the input resolution is 1 280x1 280, the detection time T decreases the most.

Secondly, for the improved FCOS model, with the increase of the input image resolution, the input image pixel utilization rate increases, thus increasing the AP value and detection time. For example, in the improved FCOS (224) model, the input image resolution is set to 224x224 and the pixel utilization rate is 1.20%, and then the AP and detection time are 98.55% and 59.4ms respectively. In the improved FCOS (1 280) model, the input image resolution is set to 1 280x1 280. The pixel utilization becomes 39.06%, and the AP and detection time are 99.04% and 106.8ms, respectively. Compared with the model with low pixel utilization rate, the model with high pixel utilization rate will have better detection accuracy, but also have longer detection time, which also has a similar rule for the original FCOS model. Therefore, we can draw a conclusion that the improvement of pixel utilization can improve the detection accuracy. Meanwhile, the detection time of the model with high pixel utilization rate will be longer, so it needs to rely on a lighter model.

Finally, combined with Table 2, it can be seen that compared with the baseline model, the improved FCOS model has a smaller relative increase time with the increase of input image resolution, indicating that the time complexity of the improved FCOS model is smaller than that of the original FCOS model, and the detection consumption time changes more slowly with the increase of input image resolution. It can be shown that the improved FCOS model is more suitable for processing high-resolution images, so as to obtain more accurate detection results.

To comprehensively evaluate the performance of a target detection model that incorporates similarity measurement losses, This paper compared the model performance of Faster R-CNN[20], Cascade RCNN[21], YOLOV3[22], YOLOV5[23], SSD[24], RetinaNet-101[25], and the proposed algorithm. For a more scientific comparison, the input image resolution of all models is set to 1 280x1 280, and the comparison results are shown in Table 3.

Tab.3 Performance comparison for detection

methods (1 280)

Model	AP/%	T/ms
SSD	98.71	601.4
Faster R-CNN	98.77	2 322.1
Cascade RCNN	98.56	1 666.3
YOLOV3	98.82	225.8
RetinaNet-101	98.65	1 916.4
YOLOV5	98.98	332.8
Ours	**99.04**	**106.8**

As can be seen from Table 3, compared with other models, the improved FCOS (1 280) model is superior to other models in terms of detection accuracy and time. Compared with SSD model, AP increases by 0.33% and detection speed increases by 463.1%. Compared with the Fster R-CNN model, the AP increased by 0.27%, and the detection speed was more than 20 times. Compared with Cascade R-CNN, AP increases by 0.48%, and the model detection speed in this paper is 14.6 times that of Cascade R-CNN. Compared with YOLOV3 model, AP increased by 0.22% and detection speed increased by 111.4%. Compared with the model of RetinaNet-101, the AP increase is 0.39%, and the detection speed of this algorithm is 18 times that of it. Compared with YOLOV5 model, AP is improved by 0.06%, and the detection speed is 3 times as fast. Compared with Cascade RCNN model, AP has the most obvious increase, with an increase of 0.48%. Compared with Faster R-CNN model, the detection time T decreases most obviously, and the reduction rate can reach 95.40%. It can be seen that the proposed algorithm has advantages in video detection in edge equipment.

3.3 Ablation experiment

In order to deeply understand the function of each module in the algorithm presented in this paper, ablation experiments were conducted on each important module. It includes backbone network, PAN module and CBAM attention module.

Table 4 and Figure 7 show the ablation results of backbone network (ResNet, ShuffleNetV2) and feature fusion network (FPN, PAN, PAN+CBAM) when the input image size is 1 280x1 280.

Tab.4 Performance comparison for detection

methods (1 280)

model	AP/%	T/ms
FCOS_ResNet_FPN(I)	99.08	1271.6
FCOS_ResNet_PAN(II)	99.11	280.2
FCOS_ResNet_PAN+CBAM(III)	99.15	1291.5
FCOS_ShuffleNetV2_FPN(IV)	96.13	88.3
FCOS_ShuffleNetV2_PAN(V)	97.20	96.7
FCOS_ShuffleNetV2_PAN+CBAM(VI)	99.04	**106.8**

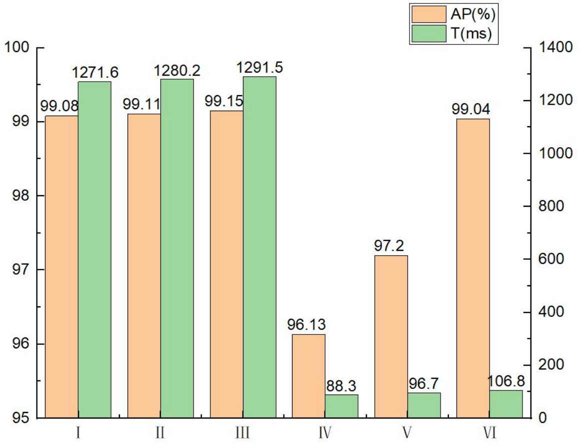


Fig.7 Ablation study results

The I-VI in the figure is respectively the six models in Table 4. As shown in Table 4 and Figure 7, compared with FCOS_ResNet_FPN and FCOS_ShuffleNet-V2_FPN, ShuffleNetV2 is used to replace ResNet as the backbone network, which reduces the AP of the whole model by 2.95%. However, its detection speed was greatly improved by 93.05%. Compared with FCOS_ShuffleNetV2_FPN and FCOS-_ShuffleNetV2_PAN, compared with FPN as the feature fusion network, PAN can improve the detection accuracy to a certain extent with less increase in detection time, and the AP increase rate is 1.07%. FCOS_ShuffleNetV2_PAN and FCOS_ShuffleNetV2_P-AN+CBAM are compared in the feature fusion network. Adding CBAM attention module into the network can further improve the accuracy of the model, and the increase of AP is 1.84%. The above experiments show that a video detection model for pointer meters with high accuracy and low detection time can be obtained by organizing the network structure reasonably. FIG. 8 shows the instrument test results.

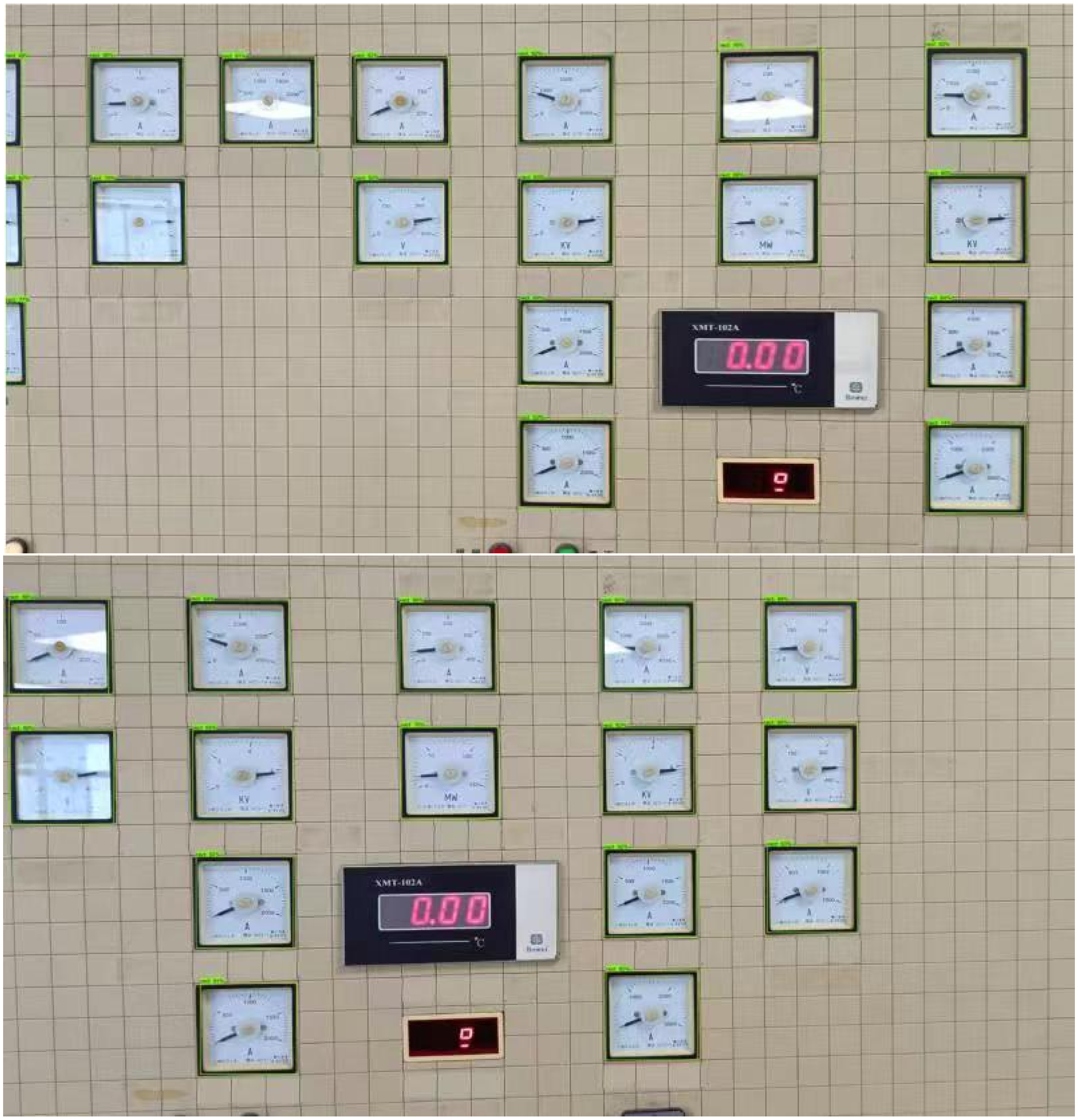


Fig.8 results of instrument detection

4 CONCLUSION

On the basis of full investigation of actual production environment demand, combined with the characteristics of deep learning, a pointer instrument video detection method based on improved FCOS was proposed. By using the lightweight feature extraction network ShuffleNetV2, the detection time of the model is greatly reduced. Meanwhile, the bidirectional feature fusion network module with the fusion attention mechanism is used to recover the location accuracy decrease caused by switching to ShuffleNetV2. Through comparative experiments, the video detection method of pointer instrument based on the improved FCOS in the video detection task of pointer instrument, compared with the baseline model, decreases by 0.04%AP, but the detection speed increases by more than 90%. The algorithm framework can be well applied to the actual power plant and power distribution room, and realize the automatic identification of various kinds of instruments of handheld instrument. The high precision detection of pointer type instrument is also a necessary prerequisite for the realization of the following functions, such as the identification of instrument reading, which fully explains the engineering application significance of the method in this paper.

Funding: This paper was supported by the Science and Technology Project of State Grid Hebei Information and Telecommunication Branch: Research projects on key technologies and typical application scenarios based on “5G+AR+AI” (No. SGHEXT00SJJS2100036)

Chen Xi, Email: 1247683744@qq.com

Zhang Pengfei, Email: 362158294@qq.com

Xu Wei, Email: x_237344650@163.com

Chang Yongjuan, Email: 94268979@qq.com

Liu Mingshuo, Email: liumgsh@163.com

Zhao Zhenyuan, Email: zhaozhenyuan@ncepu.edu.cn

REFERENCES

[1] ZHANG X F， HUANG S， Research on pointer identifying and number reading algorithm of multi-class pointer instruments[J]. Electrical Measurement & Instrumentation, 2020, 57(16): 147-152.

[2] HAN S C, XU Z Y, YIN Z C, et al. Research review and development for automatic reading recognition technology of pointer instruments[J]. Computer Science, 2018, 45(S1): 54-57.

[3] LIU Y, LIU J, KE Y. A detection and recognition system of pointer meters in substations based on computer vision[J]. Measurement, 2020, 152: 107333.

[4] WAN J L, WANG H F, GUAN M Y, et al. An automatic identification for reading of substation pointer-type meters using Faster R-CNN and U-Net[J]. Power System Technology, 2020, 44(08): 3097-3105.

[5] WANG W, HONG X F, LEI S G. Intelligent inspection and maintenance of mechanical and electrical equipment based on MR[J]. Journal of Graphics, 2022, 43(1): 141.

[6] XU L, SHI W, FANG T. Pointer meter reading recognition system used in patrol robot[J]. Chinese Journal of Scientific Instrument, 2017, 38(07): 1782-1790.

[7] PENG X Y, JIN L, WANG R, et al. Substation robot intelligent inspection technology and its application[J]. High Voltage Apparatus, 2019, 55(04): 223-232.

[8] ZHU B B, FAN S S. Reading method of substation pointer meter in rain-fog environment[J]. Laser & Optoelectronics Progress, 2021, 58(24): 221-230.

[9] LI J, YUAN L, RAN T. Automatic detection and reading method of pointer instrument based on YOLOv4[J]. Journal of Mechanical & Electrical Engineering, 2021, 38(07): 912-917.

[10] ZHANG X, DANG X, LV Q, et al. A pointer meter recognition algorithm based on deep learning[C]//2020 3rd International Conference on Advanced Electronic Materials, Computers and Software Engineering (AEMCSE). IEEE, 2020: 283-287.

[11] LIU J L, WU H Y, CHEN Z H, et al. Improved YOLOv3 industrial pointer instrument detection method[J]. Computer Engineering and Design, 2021, 42(07): 2001-2008.

[12] FAN X N, HUANG W S, SHI P C, et al. Embedded substation instrument detection algorithm based on improved YOLOv4[J]. Journal of Graphics, 2022, 43(3): 396-403.

[13] LI H H, YAN K, ZHANG L X, et al. Circular pointer instrument recognition system based on MobileNetV2[J]. Journal of Computer Applications, 2021, 41(04): 1214-1220.

[14] TIAN Z, SHEN C, CHEN H, et al. Fcos: Fully convolutional one-stage object detection[C]//Proceedings of the IEEE/CVF international conference on computer vision. 2019: 9627-9636.

[15] HE K, ZHANG X, REN S, et al. Deep residual learning for image recognition[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. 2016: 770-778.

[16] LIN T Y, DOLLAR P, GIRSHICK R, et al. Feature pyramid networks for object detection[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. 2017: 2117-2125.

[17] MA N, ZHANG X, ZHENG H T, et al. Shufflenet v2: Practical guidelines for efficient cnn architecture design[C]//Proceedings of the European conference on computer vision (ECCV). 2018: 116-131.

[18] WANG W, XIE E, SONG X, et al. Efficient and accurate arbitrary-shaped text detection with pixel aggregation network[C]//Proceedings of the IEEE/CVF International Conference on Computer Vision. 2019: 8440-8449.

[19] WOO S, PARK J, LEE J Y, et al. Cbam: Convolutional block attention module[C]//Proceedings of the European conference on computer vision (ECCV). 2018: 3-19.

[20] REN S, HE K, GIRSHICK R, et al. Faster r-cnn: Towards real-time object detection with region proposal networks[J]. Advances in neural information processing systems, 2015, 28.

[21] CAI Z, VASCONCELOS N. Cascade r-cnn: Delving into high quality object detection[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. 2018: 6154-6162.

[22] REDMON J, FARHADI A. Yolov3: An incremental improvement[J]. arXiv preprint arXiv:1804.02767, 2018.

[23] BOCHKOVSKIY A, WANG C Y, LIAO H Y M. Yolov4: Optimal speed and accuracy of object detection[J]. arXiv preprint arXiv:2004.10934, 2020.

[24] LIU W, ANGUELOY\\ D, ERHAN D, et al. Ssd: Single shot multibox detector[C]//European conference on computer vision. Springer, Cham, 2016: 21-37.

[25] LIN T Y, GOYAL P, GIRSHICK R, et al. Focal loss for dense object detection[C]//Proceedings of the IEEE international conference on computer vision. 2017: 2980-2988.

A Path Planning Method for Unmanned Surface Vehicles based on Riemannian Geometry

Jiping Yan*[a], Baoan Li[b], Zilong Lu[a]
[a]Beihang University, 37 College Road, Haidian District, Beijing 100191, China
[b]Hefei Innovation Research Institute, Beihang University, Hefei 230013, China
*Corresponding author: 15901265770@163.com

ABSTRACT

An Unmanned Surface Vehicle (USV) is an autonomous marine unmanned vehicle. USV navigation control has been extensively studied as a key technology. However, for USVs in complex environments, there are many difficulties in modelling and controlling under the wind, wave, and current conditions. At present, path planning mainly focuses on the realization of obstacle avoidance function and the rapid generation of feasible paths. The main evaluation criteria depend on the calculation time and convergence of the algorithm, while the evaluation criteria of the optimal path itself and the selection of the optimal path are less involved. This paper does not use the existing path planning methods for USV. Instead, the environment is equivalently modeled in the Riemannian space, and the geodesic between two points on the surface is calculated. In this paper, two geodesic algorithms based on triangular mesh models are studied. By comparing the calculation accuracy and running time of the two, a more suitable path-planning method for the USV system is determined.

Keywords: Unmanned Surface Vehicle, geometric space, Riemann surface, path planning, geodesic algorithm

1. INTRODUCTION

Unmanned Surface Vehicles (USVs) are equipped with advanced sensing systems, communication systems, control systems and mission payloads to perform tasks that are dangerous and unsuitable for manned vessels [1]. It combines the characteristics of high speed, high safety, multi-functional integration, etc., and can meet the requirements of completing tasks quickly, efficiently, and accurately.

In addition to traditional boat technology, USV also involves multi-sensor intelligent perception and fusion technology, path planning/automatic obstacle avoidance technology [2], autonomous navigation and control technology, etc. Due to the interference of wind, waves, currents and various obstacles, the autonomous navigation and control of the USV is very complicated. Especially when USV sails under high sea conditions, the motion force changes with time, which is a typical nonlinear time-varying motion process. In studying such problems, we wish to apply nonlinear mathematical theory [3].

The success of Riemannian geometry [4] and general relativity provides a new idea for the study of this paper. It is believed that the gravitational force on celestial bodies is the geometric property of space-time curvature caused by the existence of material mass. In curved spacetime, objects still move along the shortest path, that is, the motion of objects must satisfy the geodesic equation in curved spacetime [5]. Therefore, when dealing with mechanical system problems, it is no longer limited to the force analysis and solution of nonlinear mechanical system equations, but uses the Riemannian geometry manifold space to represent the force of the mechanical system. The trajectory of the mechanical system is no longer solved by the Lagrangian equation of analytical mechanics. Instead, geodesic equations are discussed and focus on obtaining the integral curve of motion of the mechanical system.

The influence of external factors on the USV can be equivalent to the obstruction of an object in the Riemannian manifold space, and the track of the USV is the solution of the geodesic equation in the manifold space. Therefore, to analyze and solve the force of the unmanned vehicle system, we turn to discuss the establishment of Riemannian geometry manifold space and the solution of the geodesic equations [6], and then obtain the position, velocity, and track curves of the unmanned vehicle motion.

Inspired by the study of such problems, this paper will not use the previous path-planning methods for unmanned systems, but use geodesics to predict the trajectory of USV navigation.

International Conference on Computer Graphics, Artificial Intelligence, and Data Processing (ICCAID 2022)
edited by Ruishi Liang, Jing Wang, Proc. of SPIE Vol. 12604, 1260404
© 2023 SPIE · 0277-786X · doi: 10.1117/12.2674514

2. RULE OF EQUIVALENCE

Using the geometric thinking method, various environmental factors (sea wind, sea waves, sea currents), engine driving force, water flow resistance, rudder propeller torque, and obstacles are equivalent to the representation of Riemannian geometric manifolds.

2.1 Sea wind

Sea wind is divided into two categories based on the time constant relative to USV time. The first type of sea wind steady flow has a relatively large time period, and its amplitude generally obeys the Rayleigh distribution. Its wind speed is high and the duration is long, so the parameters of the boat are much greater than the parameters of the movement of the ship. For a certain period of time, it can be seen as a wind blowing in one direction at a steady speed. The second type of turbulent sea wind generally consists of an infinite number of harmonics that vary randomly in amplitude, direction, and frequency.

The influence of steady wind on the motion of the ship is related to the windward angle of the ship. For different windward angles, the steady-flow wind can be equivalent to slopes with different inclinations in Riemannian geometry. The equivalence principle is that in Riemannian geometry, the trajectory change of the USV caused by the inclined surface is consistent with the change caused by the sea wind.

Figure 1 shows the equivalent simulation results for sea wind.

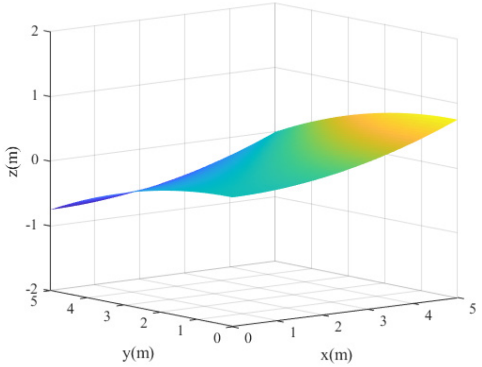


Figure 1. Equivalent simulation of sea wind

2.2 Sea waves

Under the dual action of gravity field and wind, seawater fluid is deformed to form sea waves. The Riemannian space representation is the shape of the ocean waves itself, which can be numerically simulated using spectrum analysis methods.

Ocean waves are generally generated by sea winds, as well as some other factors such as tidal surges. These waves are superimposed on a patch of hills of various sizes and directions on the sea level. The amplitude varies with the strength of the sea wind and the amplitude of the swell and cannot be represented by an accurate function. After statistical analysis of a large number of observation data, it is found that ocean waves can be regarded as a smooth random process. In common numerical simulations, ocean waves are generally formed by the superposition of sine waves and cosine waves, where the phase, amplitude, and period are randomly generated within a certain range [7].

Taking x-y as the horizontal plane, the wave height expression is shown in Equation (1).

$$z(t)=\sum_{i=0}^{\infty} a_i \cos\left(k_i x \cos\mu + k_i y \sin\mu - \omega_i t + \varepsilon_i\right) \tag{1}$$

Where a_i denotes the amplitude of the i^{th} harmonic, ε_i denotes the initial phase of the i^{th} harmonic, ω_i denotes the angular frequency of the i^{th} harmonic, k_i represents the wave index of the i^{th} harmonic, and μ represents the wave direction angle.

Figure 2 shows the equivalent simulation results for sea waves.

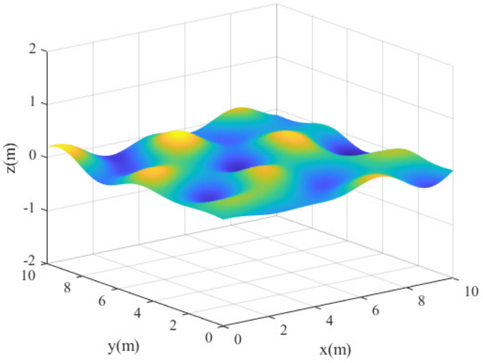


Figure 2. Equivalent simulation of sea waves

2.3 Obstacles

During the voyage of the USV in the sea area, it will encounter various ships and obstacles, which will affect the safety of navigation. Considering the safety of USV navigation, the obstacles are marked out within a safe range for unfolding, and then made into a three-dimensional surface, which is equivalent to a convex manifold in Riemannian geometry. The equivalent convex manifold of a moving boat also needs to obtain its orientation, position, and velocity information through GPS. At different times, it appears in different directions in the Riemannian geometric space. Further, the unmanned ship will calculate the movement information of dynamic obstacles through AIS and ARPA and predict whether there is a danger of collision with the ship according to the ship's sailing speed and azimuth, so as to avoid a collision.

In this paper, the obstacle is equivalent to a surface represented by a Gaussian function, which satisfies the functional relationship of Equation (2).

$$\mathrm{f}(x,y)=A\exp\left(-\left(\frac{(x-x_0)^2}{2\sigma_x^2}+\frac{(y-y_0)^2}{2\sigma_y^2}\right)\right) \tag{2}$$

Where A denotes the amplitude of the surface, (x_0, y_0) denotes the center point coordinates of the surface, σ_x denotes the standard variance of x, and σ_y denotes the standard variance of y.

Figure 3 shows the equivalent simulation results for obstacles.

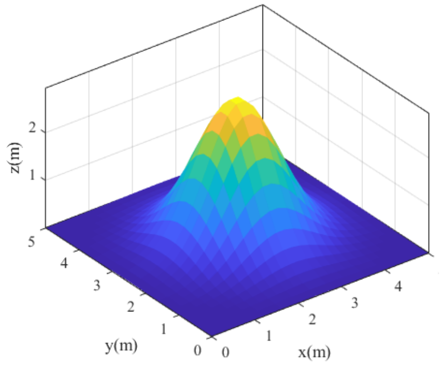


Figure 3. Equivalent simulation of obstacles

2.4 Rule of composition

Nonlinear mechanical systems are represented in geometric space. The state (position and velocity) of a mechanical system can be represented by the manifold position space and tangent vector velocity space of the geometry. The control input can be represented by the input space of the manifold, and its motion trajectory (translation and rotation) is represented by the integral curve of the tangent vector in the manifold space [8].

In the motion space of USV, there are various factors affecting a particle. The tangent vector on the particle satisfies the principle of vector synthesis, and a new space is generated after synthesis.

Figure 4 shows the synthetic simulation results of the equivalent surfaces of sea wind, sea waves, and obstacles.

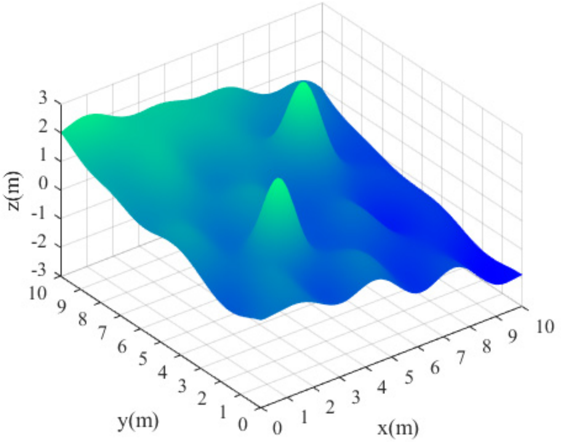


Figure 4. Synthesis results of equivalent surfaces of wind, wave, and obstacle

The following is the optimal trajectory of the movement of the unmanned ship, that is, the geodesic line between two points on the curved surface formed by environmental factors such as wind, waves, and obstacles.

3. THE SOLUTION TO GEODESIC

For the navigation control system of USV, the input of rudder angle rotation and engine speed is a typical nonlinear input, and the influence of sea wind, sea wave and sea current on the ship is time-varying. Therefore, this kind of nonlinear mechanical system is non-deterministic and time-varying, which makes the force of the system very complicated, and it is impossible to solve the contact coefficient term in the geodesic equation. Therefore, the numerical calculation method of a geodesic is adopted.

According to the calculation accuracy, the geodesic algorithms can be divided into two types. One is the approximate algorithm, which saves running space and improves calculation efficiency at the cost of certain precision. The classic Dijkstra algorithm is the most representative one [9]. The other precise algorithm focuses on the accuracy of geodesic distance, which requires plenty of operation space and consumes plenty of computing time, resulting in a certain degree of waste of resources. The representative algorithm is the MMP (Mitchell, Mount, Papadimitrious) algorithm [10].

In this paper, a triangular mesh model is used to represent the equivalent three-dimensional surface of the environmental factors in the previous section, and two geodesic algorithms are studied according to the accuracy, both of which are introduced separately below.

The core idea of the Dijkstra algorithm is that the length of the shortest path increases from the source point to the destination point, and the shortest path is generated in turn. The core idea of the MMP algorithm is to gradually cover all the surfaces of the model through window propagation, so as to calculate the geodesic distance and geodesic path from any point in the model to the source point.

Figure 5 shows the simulation of a 50*50 square-meters sea area.

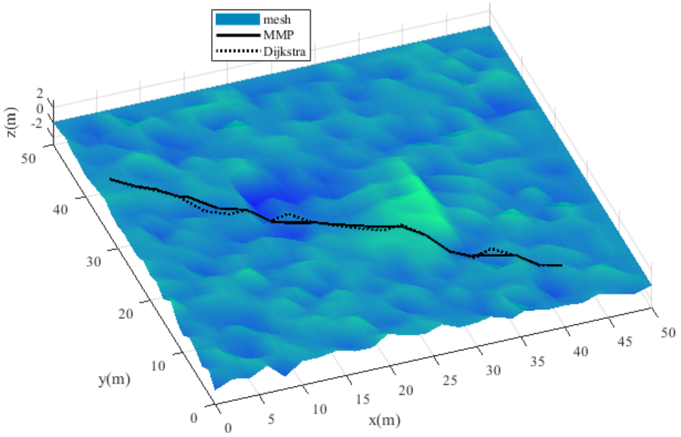


Figure 5. Simulation of 50*50 square-meters sea area

It can be seen that the Dijkstra algorithm can only pass through the edge of the triangular mesh, and the error is relatively large. The MMP algorithm can make the geodesic line pass through the triangular surface, and the path is relatively smooth. However, when there are a large number of triangular vertices on the mesh model, the MMP algorithm takes a longer time. In practice, geodesics can often pass through triangular meshes, which often requires better path smoothness. Therefore, in the case of a small amount of data, it is more accurate to choose the MMP algorithm as the path-planning method for USV.

4. CONCLUSION

The theory and method of USV track prediction based on Riemann surface synthesis and geodesic numerical calculation method are given, as well as in the case of obstacles on the water surface. However, USV did not obtain verification data when navigating on high seas, and only simulation verification was carried out. It is hoped that the geodesic algorithm can be applied to the high sea state USV navigation trajectory prediction in future research. The geodesic line is used as the target path of the unmanned ship, and the rudder is used to control and correct the direction of the ship. Using the rudder equivalent surface to control and correct the trajectory still needs further research.

REFERENCES

[1] Liu Z, Zhang Y, Yu X, et al. Unmanned surface vehicles: An overview of developments and challenges[J]. Annual Reviews in Control, 2016, 41: 71-93.

[2] Song L, Mao Y, Xiang Z, et al. Path Planning Research of Unmanned Surface Vehicle Based on Electronic Chart[J]. Journal of information and computational science, 2014, 11(17): 6245-6254.

[3] Brockett R W. Nonlinear systems and differential geometry[J]. Proc. of IEEE, 1976, 12(3): 371-393.

[4] Scholz E. Riemanns frühe Notizen zum Mannigfaltigkeitsbegriff und zu den Grundlagen der Geometrie[J]. Archive for History of Exact Sciences, 1982, 27(3): 213-232.

[5] Kumar G, Srinivasan P, Holla V D, et al. Geodesic curve computations on surfaces[J]. Computer Aided Geometric Design, 2003, 20(2): 119-133.

[6] Hotz I, Hagen H. Visualizing geodesics[C]. In: Proceedings IEEE Visualization. Salt LakeCity, 2000: 311-318.

[7] Premoe S, Ashikhmin M. Rendering natural waters[J]. Computer Graphics Forum, 2001, 20(4): 189-200.

[8] Rodnay G, Rimon E. Isometric visualization of configuration spaces of two-degrees-of-freedom mechanisms[J]. Mechanism and Machine Theory, 2001, 36(4): 523-545.

[9] Barbehenn M. A note on the complexity of Dijkstra's algorithm for graphs with weighted vertices[J]. Computers IEEE Transactions on, 1998, 47(2): 263.

[10] Surazhsky V, Surazhsky T, Kirsanov D, et al. Fast exact and approximate geodesics on meshes[J]. ACM Transactions on Graphics, 2005, 24(3): 553-560.

Lateral stability analysis and equivalent circuit simulation of double-axis bogie

Zheng Wang [*], Sheng Wang ,Yuanpeng Lei.
Lanzhou Jiaotong University, Lanzhou, Gansu,China.
[*]Corresponding author's email: 402099248@qq. com

Abstract

The stability of the bogie limits the maximum safe operating speed of the vehicle. The analytical methods are diverse and have their own characteristics. In this paper, the dynamics equation of the double-axis bogie model is established. The critical velocity is calculated by Matlab, and the reduced order is converted into an equation of state according to the characteristics of the mathematical model. The analog circuits of addition and subtraction, differentiation and integration, which are composed of resistive and capacitive elements and operational amplifiers, are mathematically equivalent. The equivalent circuit simulation and numerical simulation established by Multisim are compared and analyzed, and the simulation results are basically the same, which indicates that the equivalent circuit not only has the advantages of real-time parameter adjustment and fast calculation speed, but also has good equivalence and feasibility, which provides a reference to the research and analysis of mechanical system dynamics.

Key words: double-axis bogie; numerical simulation; circuit simulation; dynamics

1. Introduction

The maximum train speed is closely associated with the lateral stability of the bogie. In the process of high-speed running, due to the influence of its own suspension parameters, track irregularities and other factors, the wheelset, frame and car body will appear vibration, lateral movement and other phenomena. The key to exploring this kind of problem is to analyze the stability of the dynamic system[1]. For the dynamic model of the vehicle bogie with multiple degrees of freedom, the application of Lyapunov direct method, describing function method, Nyquist criterion and other theoretical methods to judge the stability is relatively complicated. Therefore, the stability of the coefficient matrix can be judged and analyzed by the eigenvalue method after the order reduction. This paper takes the 6-DOF double-axis bogie model as the research object, carries out a numerical simulation of the bogie system in the stable and unstable state, and analyzes the transverse and head-shaking motion of the frame and the two-wheel pair[2]. In the process of numerical simulation, the relevant parameters of the image cannot be adjusted in real time, and the phenomenon of instability occurs when the step size is large, the idea of transforming the dynamic equation into the circuit equation is proposed in this paper, and the corresponding equivalent electronic circuit is designed and built for simulation. At the same time, it is compared with the numerical results to verify the feasibility and correctness of using the circuit simulation to analyze the multi-DOF system.

2. Dynamic model and numerical analysis of two-axis bogie

2.1 Dynamics model and equation

Fig.1 is a typical double-axis bogie dynamics model, considering the creep between wheelsets and the coupling between wheelsets and bogies. A dynamic model that can be built[3][4]. Let the transverse displacement of the frame be y_t and the shaking Angle be Ψ_t; The transposition of the one-bit wheelset is y_{w1}; and the shaking Angle is Ψ_{w1}; The transposition of the two-bit wheelset is y_{w2}; and the shaking Angle is Ψ_{w2}.

International Conference on Computer Graphics, Artificial Intelligence, and Data Processing (ICCAID 2022)
edited by Ruishi Liang, Jing Wang, Proc. of SPIE Vol. 12604, 1260405
© 2023 SPIE · 0277-786X · doi: 10.1117/12.2674532

Proc. of SPIE Vol. 12604 1260405-39

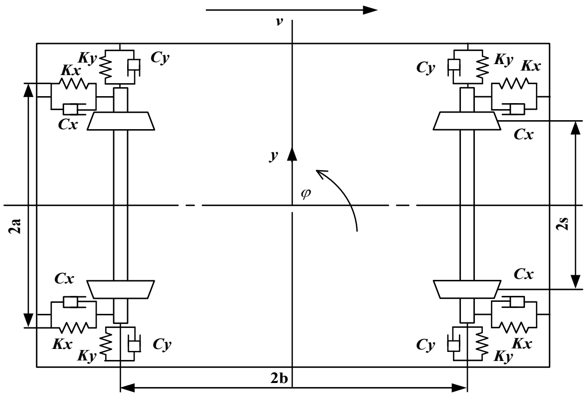


Fig. 1 Dynamics model of double-axis bogie

The bogie is the core part of the vehicle, which plays an important role in improving the damping characteristics of the vehicle and mitigating the impact of line irregularity of the vehicle. According to the force balance condition, the dynamics equation of the bogie is obtained.

1 wheelset on the left:

$$\begin{cases} m_w\ddot{y}_{w1} + 2C_y(\dot{y}_{w1} - \dot{y}_t - b\dot{\psi}_t) + 2K_y(y_{w1} - y_t) \\ \quad -2K_y\psi_t b = -\dfrac{2f_{22}}{V}\dot{y}_{w1} + 2f_{22}\psi_{w1} - \dfrac{W\lambda}{s}y_{w1}, \\ I_w\ddot{\psi}_{w1} + 2C_x a^2(\dot{\psi}_{w1} - \dot{\psi}_t) + 2K_x a^2(\psi_{w1} - \psi_t) \\ \quad = -\dfrac{2f_{11}s^2}{V}\dot{\psi}_{w1} - \dfrac{2f_{11}s\lambda}{r_0}y_{w1}. \end{cases} \tag{1}$$

2 wheelsets on the right:

$$\begin{cases} m_w\ddot{y}_{w2} + 2C_y(\dot{y}_{w2} - \dot{y}_t + b\dot{\psi}_t) + 2K_y(y_{w2} - y_t) \\ +2K_y\psi_t b + \dfrac{W\lambda}{s}y_{w2} = -\dfrac{2f_{22}}{V}\dot{y}_{w2} + 2f_{22}\psi_{w2}, \\ I_w\ddot{\psi}_{w2} + 2C_x a^2(\dot{\psi}_{w2} - \dot{\psi}_t) + 2K_x a^2(\psi_{w2} - \psi_t) \\ \quad = -\dfrac{2f_{11}s^2}{V}\dot{\psi}_{w2} - \dfrac{2f_{11}s\lambda}{r_0}y_{w2}. \end{cases} \tag{2}$$

Architecture:

$$\begin{cases} m_t\ddot{y}_t + 2C_y(\dot{y}_{w1} - \dot{y}_{w2} + 2\dot{y}_t) \\ \quad = 2K_y(\dot{y}_{w1} - \dot{y}_{w2} - 2\dot{y}_t), \\ I_t\ddot{\psi}_t - 2bC_{\mathrm{y}}(\dot{y}_{w1} - \dot{y}_{w2} - 2b\dot{\psi}_t) \\ \quad - 2C_x a^2(\dot{\psi}_{w1} + \dot{\psi}_{w2} - 2a\dot{\psi}_t) \\ = 2bK_{\mathrm{y}}(y_{w1} - y_{w2} - 2b\psi_t) \\ \quad - 2K_x a^2(\psi_{w1} + \psi_{w2} - 2a\psi_t). \end{cases} \tag{3}$$

See Table 1 for the meanings of parameters in the above equations (1), (2) and (3).

Table 1 Vehicle system parameters and values

Parameter	Sign	numerical value	Unit
Mass of wheelset	M_w	1670	kg
Quality of construction	M_t	4060	kg
Rotational inertia of wheelset shake head	I_w	1068	kgm^2
Moment of inertia of frame shaking head	I_t	4591	kgm^2
Wheelset transverse damping	C_y	2. 1×10^4	Ns/m
Wheelset longitudinal damping	C_x	2. 1×10^4	Ns/m
Lateral stiffness of wheelset	K_y	7. 8	MN/m
Longitudinal stiffness of wheelset	K_x	7. 8	MN/m
Lateral creep coefficient	f_{11}	6. 728	MN/m
Longitudinal creep coefficient	f_{22}	6. 728	MN/m
Half of the wheelbase	b	1. 25	m
Half of the transverse spacing of the longitudinal spring	a	1. 02	m
Half of the wheel-rail contact spacing	s	0. 75	m
Axle load	w	15. 6	t
Equivalent taper	λ	0. 3	
Forward speed of vehicle	v	108	km/h
Radius of wheel rolling circle	r_0	0. 42	m

2.2 Numerical Analysis and simulation

The dynamic equations of wheelset and frame are 6-dimensional second-order differential equations, and considering the coupling effect between wheelset and bogies, the analytical method is relatively complicated, and the numerical method is generally used to convert them into twelve-dimensional first-order differential equations for solving by reducing the order. To:

$$\begin{aligned} &y_1 = y_{w1}, y_2 = \psi_{w1}, y_3 = y_{w2}, y_4 = \psi_{w2}, \\ &y_5 = y_t\ , y_6 = \psi_t, y_7 = \dot{y}_1, y_8 = \dot{y}_2, \\ &y_9 = \dot{y}_3, y_{10} = \dot{y}_4, y_{11} = \dot{y}_5, y_{12} = \dot{y}_6. \end{aligned}$$

Then the state vector is:

$$y = \left[y_1, y_2, y_3, y_3, y_4, y_5, y_6, y_7, y_8, y_9, y_{10}, y_{11}, y_{12}\right]^T. \tag{4}$$

Thus, the equation of motion of the bogie can be converted into an equation of state:

$$\dot{y} = Ay. \tag{5}$$

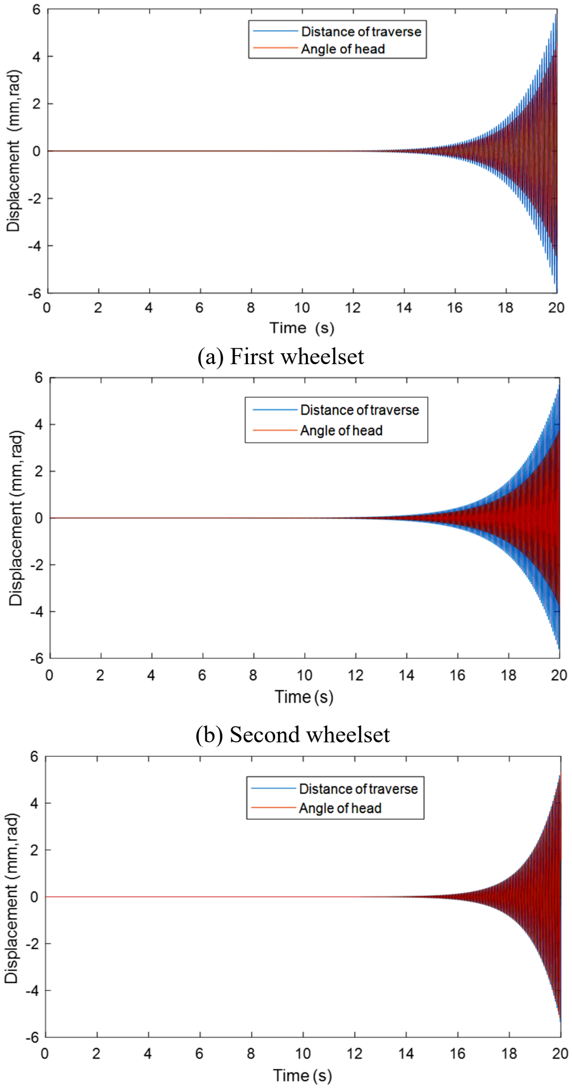
Among them,

$$A = \begin{bmatrix} 0 & E \\ -M^{1}K & -M^{1}C \end{bmatrix} . \tag{6}$$

In Equation (6), E, M, K and C are the sixth-order identity matrix, mass matrix, stiffness matrix and damping matrix respectively.

Eigenvalues and eigenvectors are important mathematical tools for studying and analyzing the stability of vehicle system and solving engineering vibration problems. In order to calculate the eigenvalues of the higher-order matrix quickly and precisely, this paper uses the built-in function eig(*) in MATLAB to solve the eigen roots of the equation of state. When the velocity is 101km/h, the real parts of the eigenvalues are all negative, the system can be judged to be stable[5].

While velocity is 103km/h, the real part of the eigenvalue appears positive, the coefficient matrix diverges, and the system changes from stable to unstable. Therefore, it can be judged that the critical speed of the vehicle in this state is around 102km/h. The lateral displacement at 108km/h and 98 km/h is simulated by Matlab. The results are shown as fig.2 and fig.3.

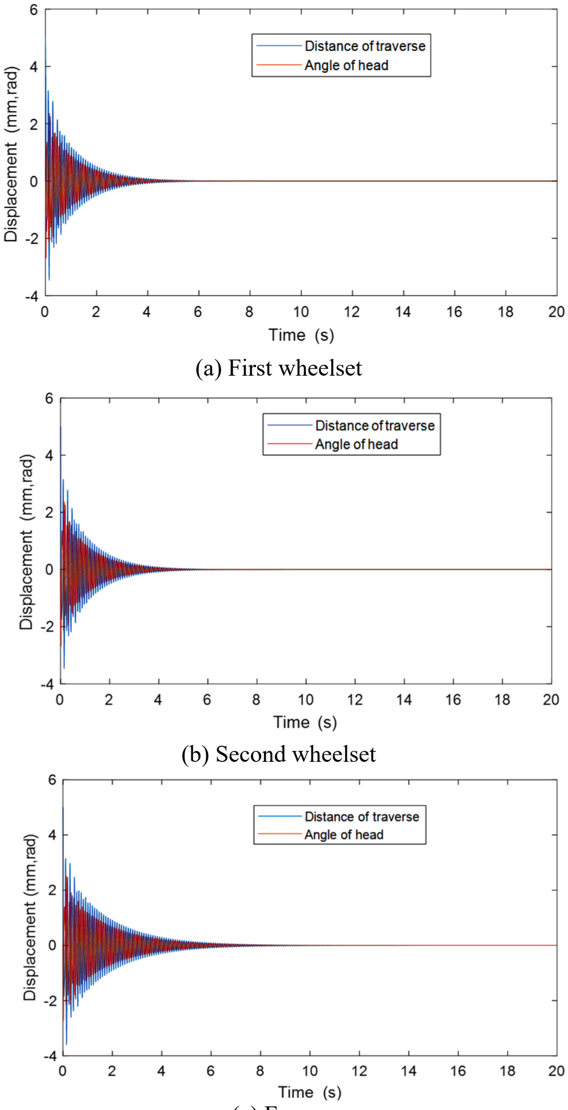


(a) First wheelset

(b) Second wheelset

(c) Frame

Fig. 2 Numerical simulation results when speed V=108 km/ m



(a) First wheelset

(b) Second wheelset

(c) Frame

Fig. 3 Numerical simulation results when speed V=90 km/h

The results show that when the vehicle speed is greater than the critical stable speed, the small oscillations of wheelset and frame caused by the initial disturbance continue for a period of time, and then gradually expand with the time. In addition, due to the coupling effect between the head-shaking motion of the wheelset and the sideways motion, the vehicle will appear the phenomenon of serpentine motion when driving for a long time in this state[7]. Otherwise, both of them are in a stable state, and the oscillation caused by the initial excitation will quickly disintegrates. The simulation results are consistent with the theoretical analysis.

3. Equivalent circuit simulation

3.1 Mathematical analysis of equivalent circuit

The equivalent electronic circuit design of the dynamic system mainly applies the integral circuit and the addition circuit. Therefore, when establishing the equivalent circuit model of the two-axis bogie, the equation of state (5) should be converted into the form of the integral equation, namely:

$$y = A\int ydt \tag{7}$$

On the basis of the above analysis and calculation, in order to facilitate the selection of the corresponding active integral circuit combination, Equation (7) is further converted into the form of circuit equation:

$$\left\{\begin{aligned}
y_i &= \frac{1}{R_m C_n}\int y_{6+i}dt \ \left(i=1,2...6\right),\\
y_7 &= -\frac{1}{R_7C_7}\int y_1dt+\frac{1}{R_8C_7}\int y_2dt+\frac{1}{R_9C_7}\int y_5dt+\frac{1}{R_{10}C_7}\int y_6dt\\
&\quad -\frac{1}{R_{11}C_7}\int y_7dt+\frac{1}{R_{12}C_7}\int y_{11}dt+\frac{1}{R_{13}C_7}\int y_{12}dt,\\
y_8 &= -\frac{1}{R_{14}C_8}\int y_1dt-\frac{1}{R_{15}C_8}\int y_2dt\\
&\quad +\frac{1}{R_{16}C_8}\int y_6dt-\frac{1}{R_{17}C_8}\int y_8dt+\frac{1}{R_{18}C_8}\int y_{12}dt,\\
y_9 &= -\frac{1}{R_{19}C_9}\int y_3dt+\frac{1}{R_{20}C_9}\int y_4dt+\frac{1}{R_{21}C_9}\int y_5dt\\
&\quad -\frac{1}{R_{22}C_9}\int y_6dt-\frac{1}{R_{23}C_9}\int y_9dt+\frac{1}{R_{24}C_9}\int y_{11}dt-\frac{1}{R_{25}C_9}\int y_{12}dt,\\
y_{10} &= -\frac{1}{R_{26}C_{10}}\int y_3dt-\frac{1}{R_{27}C_{10}}\int y_4dt\\
&\quad +\frac{1}{R_{28}C_{10}}\int y_6dt-\frac{1}{R_{29}C_{10}}\int y_{10}dt+\frac{1}{R_{30}C_{10}}\int y_{12}dt,\\
y_{11} &= \frac{1}{R_{31}C_{11}}\int y_1dt+\frac{1}{R_{32}C_{11}}\int y_3dt-\frac{1}{R_{33}C_{11}}\int y_5dt\\
&\quad +\frac{1}{R_{34}C_{11}}\int y_7dt+\frac{1}{R_{35}C_{11}}\int y_9dt-\frac{1}{R_{36}C_{11}}\int y_{11}dt,\\
y_{12} &= \frac{1}{R_{37}C_{12}}\int y_1dt+\frac{1}{R_{38}C_{12}}\int y_2dt-\frac{1}{R_{39}C_{12}}\int y_3dt\\
&\quad +\frac{1}{R_{40}C_{12}}\int y_4dt-\frac{1}{R_{41}C_{12}}\int y_6dt+\frac{1}{R_{42}C_{12}}\int y_7dt\\
&\quad +\frac{1}{R_{43}C_{12}}\int y_8dt-\frac{1}{R_{44}C_{12}}\int y_9dt+\frac{1}{R_{45}C_{12}}\int y_{10}dt-\frac{1}{R_{46}C_{12}}\int y_{12}dt.
\end{aligned}\right. \tag{8}$$

In order to realize that the circuit equation (8) is mathematically equivalent to the integral equation (7), C_{1}~$_{6}$=1000nF and C_{7}~$_{12}$=10nF are selected, and the corresponding relation between the equations in the equations (7) and (8)1/RC=A, the value of R can be calculated. In order to ensure the accuracy and stability of the circuit simulation, the equivalent parameters were calculated by Matlab[8]. The calculated results are shown in Fig.4. The modules of *UjA, UjB, UjC* and *UjD(j*=1,2,3,4,5,6) in Fig.4 are used to realize Equation (7). (*j*=1 implements equation 1, 7; Similarly, *j*=2 implements equation 2, 8; *j*=3 implements equation 3, 9; *j*=4 implements equation 4, 10; *j*=5 implements Equation 5, 11; *j*= 6 implements equation 6, 12).

3.2 Equivalent circuit design

The dynamic equivalent circuit model of the two-axis bogie is built through multiple modules, such as multi-input addition circuit and integrator circuit, which are composed of operational amplifier, resistor, capacitor and other components[9]. When simulating the motion of each component in the stable state, the resistance value at the corresponding speed should be adjusted, (R_{11}=286. 9kΩ, R_{17}=309. 7kΩ, R_{23}=484. 3kΩ, R_{29}=183. 5kΩ)

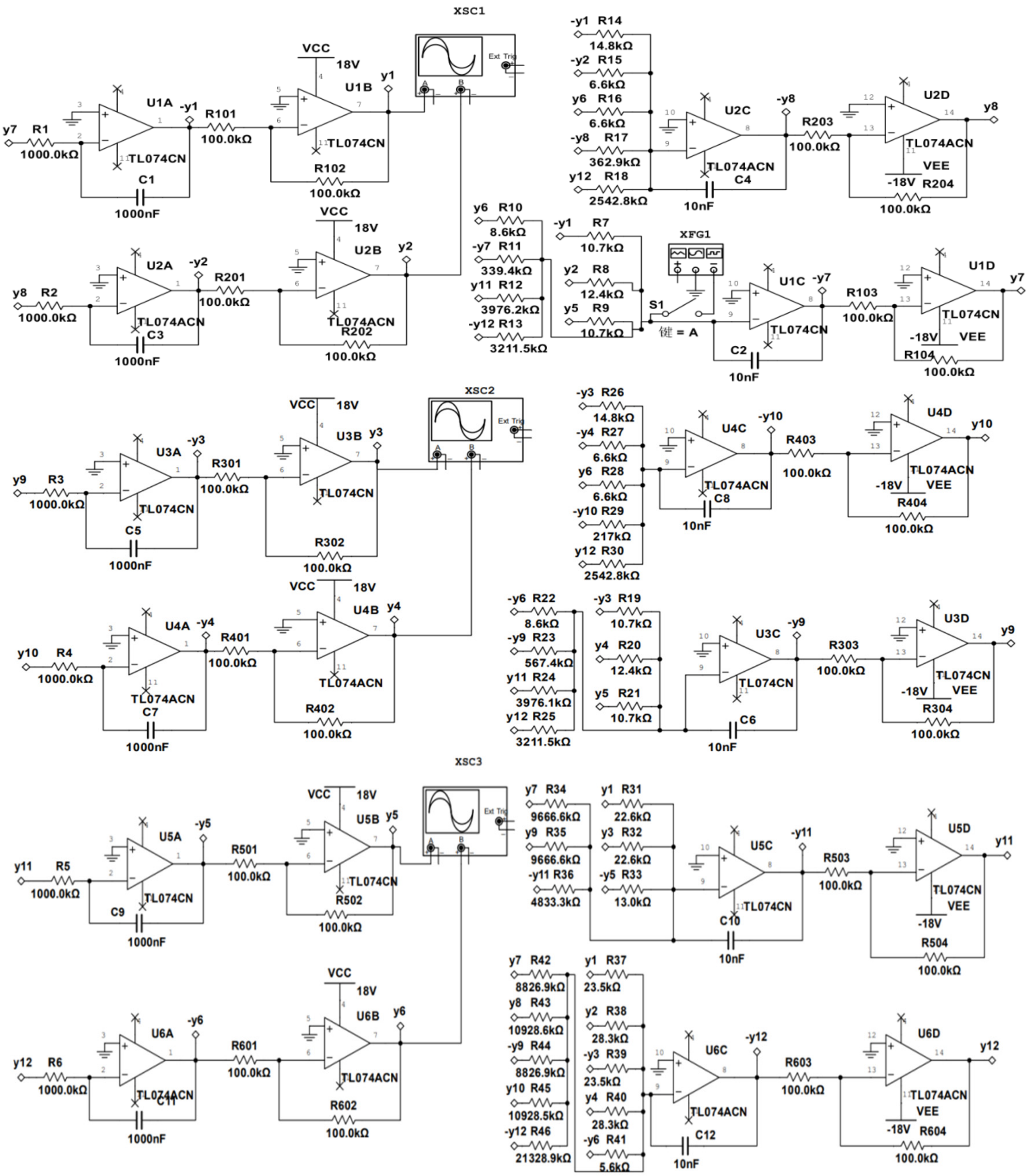
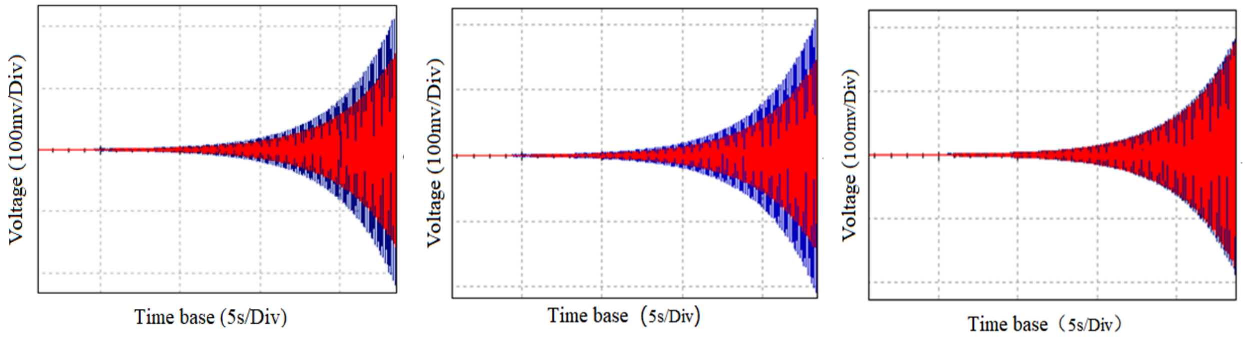


Fig. 4 Equivalent circuit model

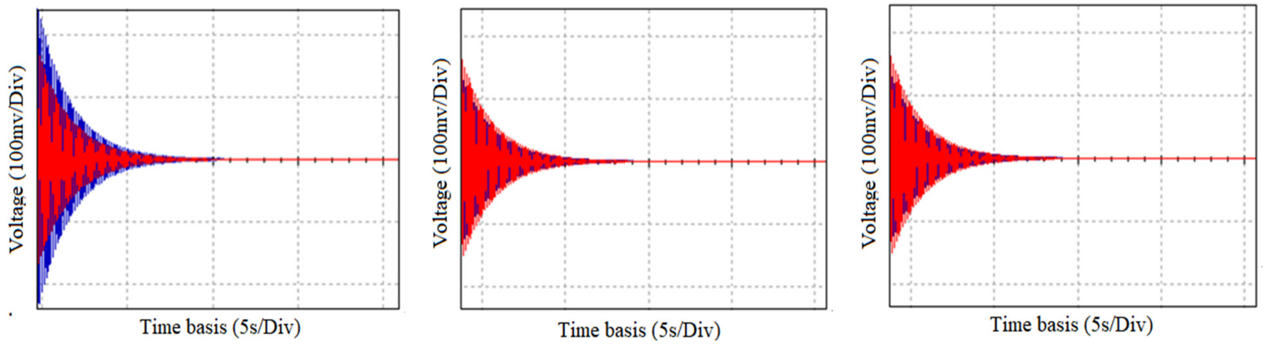
3.3 Multisim12. 0 Simulation

The vehicle running speed *V*=108Km/h and *V*=90Km/h. The circuit model shown in fig.4 is operated, and the simulation results are directly displayed on the oscilloscope as shown in fig.5 and fig.6. (The abscissa represents the time base scale, and each cell represents 5s/Div; The vertical coordinate represents the voltage, and each cell represents 100mv/Div; Where blue represents the amount of sideways movement, red represents the shaking of the head). The simulation results agree well with the numerical simulation results.



(a) First wheelset (b) Second wheelset (c) Frame

Fig. 5 Circuit simulation results at speed V=108 km/ h



(a) First wheelset (b) Second wheelset (c) Frame

Fig. 6 Circuit simulation results at speed V=90 km/ h

4. Conclusion

In this paper, the transverse motion and head motion of a two-axis bogie frame and a two-wheel pair are studied and the second order dynamic system is transformed into a general form of the equation of state. The eigenvalue method is used to determine the critical velocity of the vehicle when the system is stable, and the numerical simulation method is designed to analyze the linear stability of the vehicle system in different states. In view of the shortcomings of numerical analysis, such as the parameter cannot be adjusted in real time, the programming is complicated, and the simulation speed is slow. The equivalent circuit model is designed by using electronic components such as function generator, oscilloscope, capacitor and resistor. In Multisim 12.0, the equivalent simulation results are basically consistent with the numerical simulation results, and are in line with the theoretical analysis, which further confirms the reliability, efficiency and accuracy of the circuit simulation analysis of linear dynamics problems. On the basis of this study, the advantages of applying analog circuit simulation to analyze nonlinear dynamics problems and complex mechanical system kinematics problems need to be further studied.

References

[1] Zhang. X. X, Wu. G. S, Li. G etal. Actuator optimal placement studies of high-speed power bogie for active hunting stability[J]. Vehicle System Dynamics, 2020, 58(1):233-241.

[2] Gan. F, Dai. H. Y, Luo. G. B. Railway vehicle flexible bogie sinusoidal frequency analysis method [J]. Journal of dalian jiaotong university, 2021 and (01) : 1-8. DOI: 10. 13291 / j. carol carroll nki djdxac. 2021. 01. 001.

[3] Shen. G, Rail Vehicle System Dynamics [M]. China Railway Publishing House, 2014. 9.

[4] Lei. X. Y, Wang. Z. G, and Luo. K, Research on the dynamic performance of Nanchang metro vehicles[J]. Journal of Railway Science and Engineering, 2017, 14(11): 2460-2466.

[5] Li. R,He. Z. H. Stability analysis of time-delay differential systems with impulsive effect suffered by logic choice[J]. Results in Control and Optimization,2021,49(4):112-116.

[6] Yuan. Y, Wu. G. S ,Yousef Sardahi. Hunting stability analysis of high-speed train bogie under the frame lateral vibration active control[J]. Vehicle System Dynamics,2018,56(2):69-76.

[7] Yu. N Stroganov, Bolev. V ,Mikheev. G. M. Model for assessing the road train stability movement[J]. IOP Conference Series: Earth and Environmental Science,2020,604(1):332-344.

[8] Chen. X. J. Circuit analysis simulation platform of GUI design [J]. Journal of equipment management and maintenance, 2021 (9) : 135-137. The DOI: 10. 16621 / j. carol carrollnki issn10 01-0599. 2021. 05. 6.

[9] Wang. Z, Pan. L H, Liu. W. H. Equivalent Circuit Simulation and Experiment of Symm etrical Clearance Single-DOF Vibration System [J]. Journal of Vibration and Shock, 2017,36 (1) :141-145.

[10] Chang. F. L, Wang. Z, Tao. Y. M. Circuit simul- ation of two-degree-of-freedom unilateral impact dynamics system with gap. [J]. Journal of Physics: Conference Series, 2021, 1827(1):2-11.

[11] Wang. Z, Liu. X. P, Cui. J. T.Nonlinear Vibration Equivalent Circuit Model and Simulation of Moving Pair with Clearance [J]. Journal of Lanzhou Jiaotong University,2018,37(2):79-83.

Fine-grained insulator defect detection method based on vision-transformer

Jiani Yang [1a*], Libo Yang[1b], Lanlan Liu[1c], Fuli Wan[1d], Wanxia Deng[1e]

[1]Hunan Province Key Laboratory of Intelligent Live Working Technology and Equipment (Robot),
Live Inspection and Intelligent Operation Technology State Grid Corporation Laboratory, 410004,
Changsha, China

[a*]ilwte_hn@163.com, [b]82468299@qq.com, [c]603050307@163.com, [d]13317335777@163.com,
[e]657983554@qq.com

Abstract

Insulator defects are unavoidable due to long-term exposure and harsh natural environment. With the development of the computer vision and deep learning, researchers have been drawn to automatic unmanned aerial vehicle-based insulator inspection to improve power transmission safety. However, achieving full automation of fine-grained insulator defect detection is still very challenging due to the visual complexity of defects and the high-resolution image computation complexity. This study focuses on fine-grained insulator defect detection by Vision Transformer based on deep learning. The proposed method is based on a Swin-Transformer framework, which focuses on learning hierarchical image feature representation computed with shifted windows scheme. Experiments were carried out on high-resolution image datasets to evaluate the performance of the proposed method for fine-grained insulator defect detection tasks. The results show that the method takes advantage of Vision Transformer's capabilities and outperforms the state-of-the-art method in terms of mean average precision (mAP) at 94.2% when the intersection threshold over union is set to 0.5.

Keywords- insulator defect detection; object detection; vision transformer; deep learning; aerial image

1. INTRODUCTION

The insulator is widely used to achieve conductor support and ground insulation, both of which are critical to the safe and reliable operation of the power grid. Defect issues, such as broken components, flashover damage, missing caps, and other faults, will significantly impact the power transmission system [1]. Various sensors are included in the inspection platforms as a result of the development of the smart grid. The inspection tech based on sensor systems such as thermal and radar-based detection has quickly replaced manual patrols [2] and voltage distribution methods [3]. However, the complicated surface of insulators limits these processors' ability to detect the defect. Due to their intuitive and efficient properties, visible images are always essential for inspecting insulator anomalies [4]. Because of its low-cost, high efficiency, and high-precision qualities, UAV-based inspection has attracted the attention of researchers and power grid corporations. Nevertheless, in an actual detection environment, the images captured by UAVs often face cluttered backgrounds, varying illumination conditions, and different views of changing environments; robust and effective insulator inspection remains a significant challenge.

Traditional insulator defect detection studies principally finish the task of image extraction by manual feature descriptors, such as SVM [5], DPM [6], and HOG [7]. These methods have high computational costs and are not robust in the real-world images captured by UAVs. Several practical approaches for insulator defect detection have been developed with the development of a deep learning network. Liu X et al. [8] use the Faster R-CNN (Faster Regions with Convolutional Neural Network) [9] to detect insulators and their defect. Xia et al. [10] proposed a detection method by a hybrid network model fused with the CNNs (convolutional neural networks) and RNNs (Recurrent Neural Networks). Tao X et al. [11] convert the defect inspection to a two-stage object detection problem by cascading CNN. To identify insulators, Chen Z et al. [12], Qiu Z et al. [13], and Li Q et al. [14] show the possibility of one-stage detection methods accurately and efficiently by using YOLOv3, YOLOv4, and YOLOv5 in this task. Furthermore, with the big success of Vision Transformer [15], Xu W et al. [16] applied a self-attention mechanism and proposed a transformer-based insula- tor defect detection network. However, due to the small scale of insulators in high-resolution images, which causes low inter-class variance and high inner-class variance, most previous works serve the defect problems as single-class detection. The research on fine-grained insulator defect detection is still very limited.

International Conference on Computer Graphics, Artificial Intelligence, and Data Processing (ICCAID 2022)
edited by Ruishi Liang, Jing Wang, Proc. of SPIE Vol. 12604, 1260406
© 2023 SPIE · 0277-786X · doi: 10.1117/12.2674512

Instead of detecting individual defect classes, this paper addresses the problem of fine-grained insulator defect detection using high-resolution aerial images. Vision transforms [15] have garnered much interest in recent research on many computer vision problems. Still, they struggle with data efficiency because their computational complexity is quadratic to image size. Inspired by the idea of the Swin Transformer [17], we designed a robust and efficient insulator defect detection framework named Swin-Faster-RCNN. The main contributions of this study are threefold:

1) We propose a new method to detecting fine-grained insulator defects by Swin-Transformer based on Faster-RCNN, which differs from previous methods that rely on single-class insulator defects.
2) The experimental results show that our proposed method outperforms others. The model detects the insulator string and fine-grained insulator defects with an mAP (mean average precision) of 94.2% under 50% intersection over the union (IoU) threshold.
3) The method not only keeps the advantages of the high recall rate of the Faster R-CNN method but also reduces the computational cost of the self-attention mechanism by using the Swin Transformer.

2. METHODOLOGY

2.1. Overview of Swin-Faster-RCNN

The method aims to detect fine-grained insulator defects in high-resolution aerial images efficiently and accurately. The model comprises a Swin Transformer-based backbone [17] for extracting fine-grained image features and a faster R-CNN network for detecting bounding boxes and their corresponding labels. First, the images are split into various patches using the patch partition method. The features of Each patch are set as a concatenation of the original pixel RGB values. The image patches are then routed through four sequential transformer stages. Each stage conclude a linear embedding or patch embedding layer and several swin transformer blocks. A linear embedding layer is applied in the first stage to increase the feature dimension. The output of linear embedding or patch merging layers is added during the computation of the swin transformer blocks for feature transformation. In the end, the hierarchical representation feeds into a faster R-CNN network to give the defect object bounding boxes with the corresponding class. In Figure 1, we show the entire pipeline of our model.

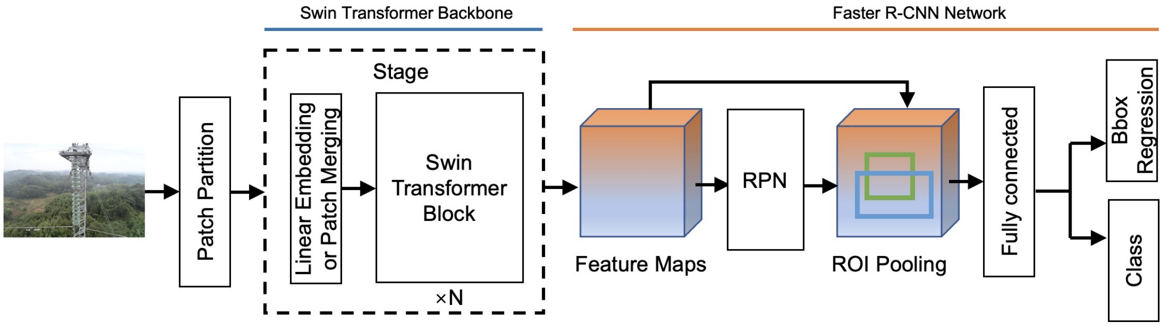


Figure 1: Network architecture of Swin-Faster-RCNN

2.2. Swin-Transformer-based backbone

Patch partition and linear embedding. The insulator images are split into various patches using the patch partition method, like Vit [15]. Every 4 × 4 adjacent pixel is combined into a patch. After the patch partition, the image size changes from $[H, W, 3]$ to $[\frac{H}{4}, \frac{W}{4}, C]$, where $H\ and\ W$ refer to the height and width, respectively. The channel is then projected into an arbitrary dimension by a linear embedding layer, denoted by C.

Patch merging layer. Hierarchical feature maps allow the Swin Transformer to be applied in fine-grained detection tasks. Patch merging layers produce a hierarchical representation by concatenating the features of neighbouring patches. This effectively downsamples the input by a factor of n transforming the input 2 from a shape of $H \times W \times C$ to $(\frac{H}{n}) \times (\frac{H}{n}) \times (\frac{n^2}{C})$. In our work, the patch size n is set to 2. With the patch merging layers, the succession of stages' output resolutions was kept at $\frac{H}{8} \times \frac{W}{8}$, $\frac{H}{16} \times \frac{W}{16}$ and $\frac{H}{32} \times \frac{W}{32}$ respectively. This hierarchical architecture has linear computational complexity concerning image size, improving the ability to process high-resolution images.

Swin Transformer block. A window-based self-attention module replaces the MSA (multi-head self-attention module) used in ViT. The Swin Transformer block consists of two successive sub-units. Each sub-unit consists of a Layer Normalization (LN) layer [18], followed by an attention module based on a window or shifted window. The first sub-unit uses a W-MSA (Window MSA) module, while the second sub-unit uses a SW-MSA (Shifted Window MSA) module. Following the module is another Layer Normalization layer and a Multilayer Perceptron Layer (MLP). After each module, a residual connection is applied. The Swin Transformer block is illustrated in Figure 2.

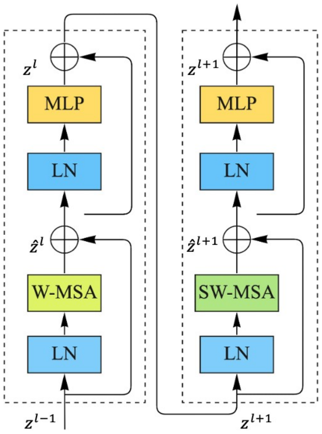


Figure 2: Two Successive Swin Transformer Blocks

Shifted window-based self-attention. The MSA employed in ViT enables global self-attention, and the relationship between each patch is computed against all other patches. This results in a quadratic complexity concerning the number of patches, making it impractical for high-resolution images. Swin Transformer addresses this by employing a window-based MSA approach (W-MSA). The window is a collection of non-overlap image patches. Self-attention is computed exclusively within each local window, decreasing the model's computational complexity compared with the standard Vit.

Restricting self-attention limits the modeling power of the network. To address this problem, SW-MSA shifts the windows towards the bottom right corner by a factor of $M/2$, where M denotes the window size, then moves the patches that do not belong to any window into windows with incomplete patches. SW-MSA overcomes W-MSA's limitation of focusing only on adjacent patches. With the W-MSA and SW-MSA, consecutive Swin Transformer blocks are computed as:

$$\begin{aligned} \hat{z}^l &= W - MSA\left(LN\left(z^{l-1}\right)\right) + z^{l-1} \\ z^l &= MLP(LN(\hat{z}^l)) + \hat{z}^l \\ \hat{z}^{l+1} &= SW - MSA\left(LN\left(z^l\right)\right) + z^l \\ z^{l+1} &= MLP\left(LN\left(\hat{z}^{l+1}\right)\right) + \hat{z}^{l+1} \end{aligned} \tag{1}$$

where $\hat{z}^l$and z^l indicate the output features of the (S)W-MSA module and the MLP module for block l, respectively.

The self-attention in (S)W-MSA is computed as:

$$Attention(Q, K, V) = SoftMax(QK^T/\sqrt{d} + B)V \tag{2}$$

Where $Q, K, V \in \mathbb{R}^{M^2 \times d}$ are the *query, key, value* metrics; d denotes the dimension; M denotes the window size; Besides, the relative position bias B to each head compute the similarity.

2.3. Faster R-CNN

After extracting the hierarchical feature maps by Swin Transformer, the proposed method employs Faster R-CNN [9] to detect insulator string and the fine-grained insulator defect. The RPN (region proposal network) calculate region proposals of the feature maps by CNNs (convolutional neural networks) for obtaining a set of possible relevant objects with bounding boxes. Next, the region of interest pooling layer takes the region proposals and feature maps as inputs and extracts the fixed-length RoI features. Finally, the output passes through a sequence of FC (fully connected layers) to obtain the bounding box coordinates and the corresponding confidence scores. It then goes into two sibling output branches, the multi-class classification layer and the bounding box regression layer, as shown in Figure 1.

3. EXPERIMENTS

3.1. Dataset

We construct a fine-grined insulator defect image dataset to show the robustness of the proposed method. Part of the dataset comes from the UAV inspection images captured by the In-Vehicle Control UAVs insulator inspection systems, and part is collected from the public data site[1]. The dataset includes 1684 insulator images, with a high resolution of 2144 × 1424. All the insulator shells in the dataset are carefully labeled to meet the needs of multi-class fine-grained insulator defeat detection. Insulators are classified into 5 categories: insulator strings; good insulator shell; broken insulator shell; flashover damage insulator shell; and another damaged insulator shell, as shown in Table 1.

Table 1: Category distribution of the dataset

Category	Train set	Test set
Total number of images	1596	88
Total labeled assets	7568	403
Insulator strings	1788	103
Good insulator shell	2636	147
Broken insulator shell	1140	64
Flashover damage insulator shell	2004	89

3.2. Implementation Details

The raw images are resized into 1000 × 600; each training iteration has a batch size of 16 and a maximum number of training iterations of 10000. the initial learning rate is set at 0.01; and the weight decay and momentum are 0.0001 and 0.9, respectively.

4. RESULTS AND ANALYSIS

4.1. Comparison of Different Methods and Detection Results

Current dominant deep learning object detection framework is used to evaluate in our dataset: RetinaNet, SSD, YOLOv3, Cascade RCNN, Faster RCNN (with ResNet-50), and the proposed Swin-Faster-RCNN, as shown in Table 2.

Table 2: The results of different algorithms.

Methods	**Recall /%**					**AP@IoU=0.5 /%**					**mAP /%**
	Insulator strings	Good	Broken	Flashover	Others	Insulator strings	Good	Broken	Flashover	Others	
RetinaNet [19]	98.1	98.1	87.5	96.6	100.0	82.0	72.5	39.9	52.0	50.8	59.4
SSD [20]	98.9	96.9	95.2	98.0	96.6	90.8	82.0	72.4	65.7	61.5	74.5
YOLOv3 [21]	98.8	98.4	92.1	98.5	97.1	83.7	86.4	79.6	81.0	58.3	77.8
Cascade RCNN [22]	98.4	98.4	94.4	96.6	94.8	90.9	90.3	90.9	90.8	90.2	90.6
Faster R-CNN [9]	**100.0**	**98.6**	96.8	97.8	**100.0**	**99.6**	90.0	90.7	90.4	93.6	92.9
Ours	**100.0**	**98.6**	**98.0**	**99.0**	**100.0**	**99.6**	**90.5**	**90.9**	**90.9**	**99.1**	**94.2**

We evaluate the proposed method from the perspective of recall, average precision (AP), and mean average precision (mAP). Furthermore, the mAP has reached 94.2%, with state-of-the-art performance. Figure 3 compares the average training loss between the proposed Swin-Faster-RCNN and the Faster-RCNN with a backbone of ResNet-50. The training losses tend to be stable as the number of iterations increases. The training loss of the proposed method has a faster convergence speed than the original algorithm. It demonstrated that Swin-Faster-RCNN keeps the advantages of the high performance of the Faster RCNN and the robustness of self-attention mechanisms.

[1] https://www.kaggle.com/competitions/insulator-defect-detection/data

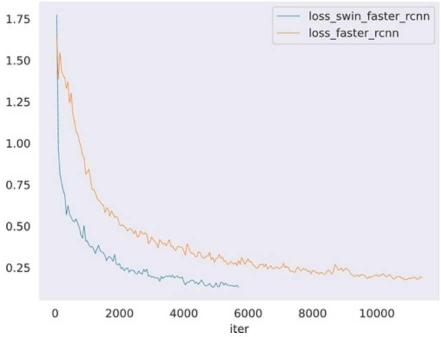


Figure 3: Comparison of average train loss of Swin-Faster-RCNN and Faster-RCNN

We compared the computational complexity and model size of four different backbones based on Faster-RCNN. Table 3 shows that Swin Transformer has both the 2^{nd}smallest model size (44.77 million parameters) and the 2^{nd} smallest FLOPs (207.07 GFLOPs), which is not much different from the smallest model ResNet-50. Furthermore, the Swin Transformer is more robust than other versions of ResNet, which cause detection overfitting.

Table 3: Comparisons of model size and complexity between different backbones based on Faster-RCNN. FLOPs: the number of floating-point operations; Params: total training parameter.

Backbone	FLOPS	Params (million)	mAP /%
ResNet-50	207.07	41.53	92.9
ResNet-101	282.75	60.14	92.2
ResNet-152	358.9	75.78	92.1
Swin	210.33	44.77	94.2

The qualitative detection results are shown in Figure 4. Under complex backgrounds such as multiple insulator strings, occlusion occurs, and the various colors of foreground insulator shells, the proposed model in this paper still accurately and effectively detect different insulators and small insulator shells. In summary, the proposed method in this paper outperformed current state-of-art object detection methods, particularly when considering the advantages of densely distributed insulators against complex backgrounds.

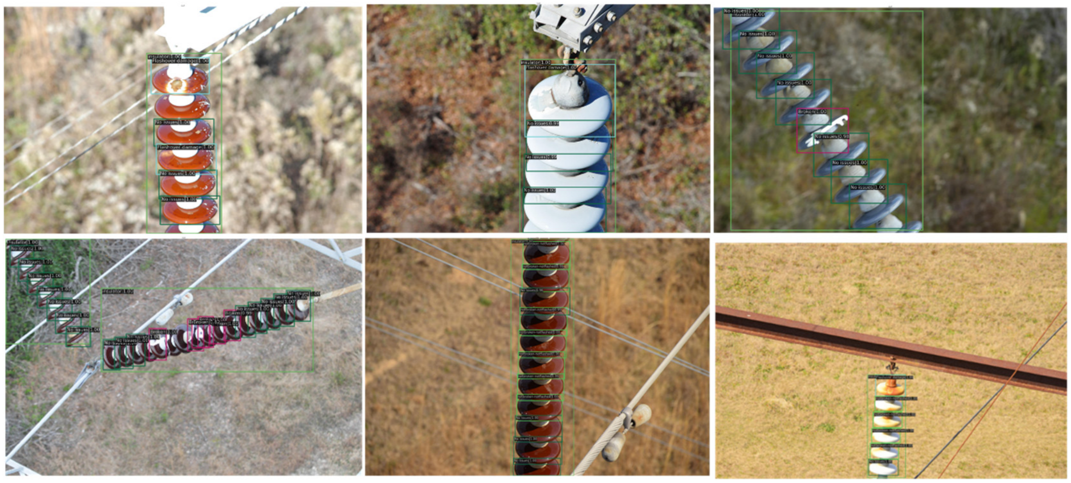


Figure 4: Qualitative detection results of Swin-Faster-RCNN.

5. CONCLUSIONS

We design an insulator defect detection framework named Swin-Faster-RCNN and explore its effectiveness in multi-class insulator defect detection. The proposed method encourages the network to learn hierarchical image feature representation computed with a self-attention scheme. Experiments show that our method improved the capability of fine-grained insulator defeat detection based on high-resolution aerial images and outperforms the state-of-art object detection methods.

It preserves the advantages of the high-recall of the Faster R-CNN and the robustness of self-attention mechanism. Further study may consider more characteristics of insulator images and study on the larger dataset with more diversity of defect categories and backgrounds.

Acknowledgments

This work was supported by Hunan Province Key Laboratory of Intelligent Live Working Technology and Equipment (Robot), Changsha, China; Live Inspection and Intelligent Operation Technology State Grid Corporation Laboratory, Changsha, China; Live Inspection and Intelligent Operation Technology State Grid Corporation Laboratory, Changsha, China. (Grant No: 5216A5210019)

References

[1] Yang, L., Fan, J., Liu, Y., Li, E., Peng, J., & Liang, Z. (2020). A review of state-of-the-art power line inspection techniques. IEEE Transactions on Instrumentation and Measurement, 69(12), 9350-9365.

[2] Liu, Z., Wang, X., & Liu, Y. (2019). Application of unmanned aerial vehicle hangar in transmission tower inspection considering the risk probabilities of steel towers. IEEE Access, 7, 159048-159057.

[3] Jiang, Z., Wu, W., Wang, B., Xie, P., Li, H., & Lin, F. (2019). Design and test of 500-kV lightning protection insulator. IEEE Access, 7, 135957-135963.

[4] Jenssen, R., & Roverso, D. (2018). Automatic autonomous vision-based power line inspection: A review of current status and the potential role of deep learning. International Journal of Electrical Power & Energy Systems, 99, 107-120.

[5] Lin, C. F., & Wang, S. D. (2002). Fuzzy support vector machines. IEEE transactions on neural networks, 13(2), 464-471.

[6] Felzenszwalb, P. F., Girshick, R. B., McAllester, D., & Ramanan, D. (2010). Object detection with discriminatively trained part-based models. IEEE transactions on pattern analysis and machine intelligence, 32(9), 1627-1645.

[7] Dalal, N., & Triggs, B. (2005, June). Histograms of oriented gradients for human detection. In 2005 IEEE computer society conference on computer vision and pattern recognition (CVPR'05) (Vol. 1, pp. 886-893). IEEE.

[8] Liu, X., Jiang, H., Chen, J., Chen, J., Zhuang, S., & Miao, X. (2018, June). Insulator detection in aerial images based on faster regions with a convolutional neural network. In 2018 IEEE 14th International Conference on Control and Automation (ICCA) (pp. 1082-1086). IEEE.

[9] Ren, S., He, K., Girshick, R., & Sun, J. (2015). Faster r-CNN: Towards real-time object detection with region proposal networks. Advances in neural information processing systems, 28.

[10] Xia, Y., Lu, J., Li, H., & Xu, H. (2018, October). Deep learning-based image recognition and processing model for electric equipment inspection. In 2018 2nd IEEE Conference on Energy Internet and Energy System Integration (EI2) (pp. 1-6). IEEE.

[11] Tao, X., Zhang, D., Wang, Z., Liu, X., Zhang, H., & Xu, D. (2018). Detection of power line insulator defects using aerial images analyzed with convolutional neural networks. IEEE Transactions on Systems, Man, and Cybernetics: Systems, 50(4), 1486-1498.

[12] Chen, Z., Xiao, Y., Zhou, Y., Li, Z., & Liu, Y. (2020, November). Insulator recognition method for distribution network overhead transmission lines based on modified YOLOv3. In 2020 Chinese Automation Congress (CAC) (pp. 2815-2820). IEEE.

[13] Qiu, Z., Zhu, X., Liao, C., Shi, D., & Qu, W. (2022). Detection of Transmission Line Insulator Defects Based on an Improved Lightweight YOLOv4 Model. Applied Sciences, 12(3), 1207.

[14] Li, Q., Zhao, F., Xu, Z., Wang, J., Liu, K., & Qin, L. (2022, February). Insulator and damage detection and location based on YOLOv5. In 2022 International Conference on Power Energy Systems and Applications (ICoPESA) (pp. 17-24). IEEE.

[15] Dosovitskiy, A., Beyer, L., Kolesnikov, A., Weissenborn, D., Zhai, X., Unterthiner, T., ... & Houlsby, N. (2020). An image is worth 16x16 words: Transformers for image recognition at scale. arXiv preprint arXiv:2010.11929.

[16] Xu, W., Zhong, X., Luo, M., Weng, L., & Zhou, G. (2022). End-to-End Insulator String Defect Detection in a Complex Background Based on a Deep Learning Model. Frontiers in Energy Research, 10, 928162.

[17] Liu, Z., Lin, Y., Cao, Y., Hu, H., Wei, Y., Zhang, Z., ... & Guo, B. (2021). Swin transformer: Hierarchical vision transformer using shifted windows. In Proceedings of the IEEE/CVF International Conference on Computer Vision (pp. 10012-10022).

[18] Ba, J. L., Kiros, J. R., & Hinton, G. E. (2016). Layer normalization. arXiv preprint arXiv:1607.06450.

[19] Lin, T. Y., Goyal, P., Girshick, R., He, K., & Dollár, P. (2017). Focal loss for dense object detection. In Proceedings of the IEEE international conference on computer vision (pp. 2980-2988).

[20] Liu, W., Anguelov, D., Erhan, D., Szegedy, C., Reed, S., Fu, C. Y., & Berg, A. C. (2016, October). SSD: Single shot multibox detector. In European conference on computer vision (pp. 21-37). Springer, Cham.

[21] Redmon, J., & Farhadi, A. (2018). Yolov3: An incremental improvement. arXiv preprint arXiv:1804.02767.

[22] Cai, Z., & Vasconcelos, N. (2018). Cascade r-CNN: Delving into high-quality object detection. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 6154-6162).

The Optimal Path of New Media City Image Communication Strategy Based on Data Mining Technology

JiaYin Li
Xi'an Eurasia University, Xi'an,710065, Shaanxi, China
* Corresponding author: lijiayin@eurasia.edu

ABSTRACT

With the wider application of data mining technology, new media has been accepted by people for its high communication efficiency, good communication quality, and strong interaction. The city is an important factor of modernization, and the use of new media to spread the image of the city can promote the rapid development of the city. Based on this, the purpose of this article is to research and analyze the optimization path of new media city image communication strategy based on data mining technology. This article first summarizes the status quo of urban image communication, and then extends the theory of urban image communication and analyzes it in detail. Then through the analysis of new media communication methods and their advantages in urban communication. This article elaborates on the new media communication based on data mining technology, and uses questionnaire survey method, field survey method and other research methods to carry out experimental research on the theme of this article. Research shows that new media has become an important channel for city image communication.

Keywords: Data Mining, New Media, City Image, Communication Analysis

1.INTRODUCTION

The in-depth development of globalization has promoted the process of domestic urbanization, and the competition pressure between cities has intensified. People attach importance to urban branding and urban marketing. With the rapid development of digital technology and mobile Internet, "big data" has entered people's eyes. In the era of big data, the audience's interests, preferences and attention are preserved in the form of unstructured data. By collecting this part of data, we can accurately describe user needs, more accurately understand user behavior and habits, and effectively support the quantification of communication effects. Based on data mining technology, this paper analyzes the optimization path of new media city image communication strategy.

With the continuous development of social economy and times, the study of urban image has attracted more and more scholars' attention since the middle of the last century, but most researchers still write some related papers and studies from the perspective of aesthetics and architecture. Walters t's urban image emphasizes that "managing our homes will have a far-reaching impact on our lives". Based on this, the concept of urban image is included in the discipline of urban planning. It shows not only the external image of the city, but more importantly, the external audience can deeply explore the culture of the city's internal spirit through the impression of the external factors of the city, laying a foreshadowing for the audience to "remember and spread widely" [1]. Foreign Studies on the city image from the beginning of architecture, design art, marketing management, public relations to later communication, they pay more attention to the personalized development of the city, that is, try to eliminate the phenomenon of "one side of a thousand cities". In recent years, most of the theories and views on the research of urban image communication in China are based on foreign research theories, and then put forward more practical ways and Strategies of urban image communication in combination with the unique characteristics of our respective cities [2].

This paper first deeply analyzes the current situation of urban image communication, and then summarizes some of the existing problems. Finally, the new media urban image communication strategy optimization path countermeasures based on data mining technology, hoping to provide some guidance and reference for the future image communication activities of cities in China [3-4].

International Conference on Computer Graphics, Artificial Intelligence, and Data Processing (ICCAID 2022)
edited by Ruishi Liang, Jing Wang, Proc. of SPIE Vol. 12604, 1260407
© 2023 SPIE · 0277-786X · doi: 10.1117/12.2674703

2.OPTIMAL PATH APPLICATION OF NEW MEDIA CITY IMAGE COMMUNICATION STRATEGY BASED ON DATA MINING TECHNOLOGY

As the soft power of a city, the reasonable development and utilization of city image can create a good public opinion environment and development space for the city, and help to improve the comprehensive competitiveness of the city. Looking at the current development of urban image communication, most cities' image communication mostly follow the following steps: urban image positioning - formulating urban image communication strategies - integrating media means - evaluating the effect of urban image communication - optimizing urban image communication. Among them, the positioning of city image depends on the historical image of the city, existing resources and expectations for the future; It is particularly important to formulate effective city image communication strategies and integrate communication media means. How to comprehensively use the existing new and old media means, realize the optimal combination of media, expand the arrival rate of media, and increase the number of media coverage is a huge challenge [5].

2.1 City Image Communication Theory

With the acceleration of the market economy and globalization process, cities have long been the centers of human production activities, and the competition between cities has become increasingly fierce[9-10]. In order to enhance its own competitiveness and promote the process of urban modernization, the concept of city image has been widely spread. With the continuous development of new media, the dissemination of city image has also attracted more and more attention.

2.1.1 City image factor

The image of a city is the perception of the formation of the city by the masses' comprehensive mass media and their own experience. The main factors of city image are shown in Table 1.

Table 1. City image content division

The image of a city	Tangible city image	Natural geographical environment	Geographic location
			Natural resources and scenery
		Urban planning and construction	Construction and transportation
			Sanitation and City Appearance
	Invisible city image	Politics and economy	The level of economic development
			Government Action
		Spiritual Culture	Civic morality
			History, culture and customs
			Technology and Education Development

2.1.2 changes in the communication path of urban image in the era of big data

The audience's cognition of the city floats on the surface and fails to penetrate the audience's heart. The visual content form of big data, through the mining and collection of massive data, will show the concepts that are difficult for ordinary people in an easy to understand visual data chart and multidimensional all-round way, and finally present a three-dimensional and all-round intuitive city image in front of the audience. Moreover, when big data is combined with content, it will bring a disruptive revolution in content dissemination [6-7]. The content dissemination of big data can help solve two key problems in the dissemination of urban image, that is, to find the target urban audience, and then show the audience the optimized content that conforms to the audience's preferences[8].

2.2 New Media Communication Analysis Based on Data Mining

In the field of data mining, due to the complex data structure of the new media network, which contains a large amount of multimedia data such as text, image, and video, the mining models are also diverse. Therefore, this article uses multiple classifiers to improve accuracy.

2.2.1 Deep learning to build an interactive network

1) Initialize the label array F and set all elements to 0, i=0.

Proc. of SPIE Vol. 12604 1260407-6

2) Access starts from P_0. First, add P_0 to the label array F, and query the points connected to P_0 through the edge set E to form a subset S(P_0).

3) Sort all edges in the subset by weight, and traverse all points P_i in the subset in turn. If P_i is not visited, the weighted depth of the new initial point of P_i is preferentially traversed until all points reachable by P_0 are traversed or the number of points in the marker array f exceeds the preset threshold, the recursion ends.

2.2.2 Analysis of maximum confidence fusion algorithm

For a true random variable r with a value range on R, suppose we make n independent observations on the variable r, get n different values of r, and calculate their average value. For the variable r In other words, its Hoeffding constraint is that in the confidence interval, the true value of the variable r is at least, among which.

$$\varepsilon = \sqrt{\frac{R^2 - \ln(1/\delta)}{2n}} \tag{1}$$

In the formula (1), r represents information gain, the value range of R is lb (Classes), and Classes is the number of categories.

Information gain is used to measure the ability of a given attribute to distinguish training examples. The formula is as follows:

$$Entropy(S) = \sum_{i=1}^{c} -p_i \log(p_i) \tag{2}$$

$$Gain(S,A) = Entroy(S) - \sum_{v=Values(A)} \frac{|Sv|}{S} Entropy(Sr) \tag{3}$$

Therefore, $\vec{u}_i . \geq \vec{0}$, the first term in equation (1) guarantees that if the target x_i is not assigned to the group g_i by a specific algorithm, then these conditional probabilities end. If j=1,...,s, the group g_i is taken from the classifier, so k_j =1. The second term of formula (1) limits the difference between the prediction category of the group g_i public knowledge and the first prediction category. α is the cost factor that violates the constraint. For j=s+1,..,v, the group g_i comes from the unsupervised model, at this time k_j =0, and there are no restrictions. In the same way, if Q is fixed, the global minimum cost of $\vec{u}_i$ can be obtained.

$$\vec{q}_{j.}^{(t)} = \frac{\sum_{i=1}^{n} a_{ij} \vec{u}_i^{(t-1)} + \alpha k_j y_j .}{\sum_{i=1}^{n} a_{ij} + \alpha k_j} \tag{4}$$

$$\vec{u}_{i.}^{(t)} = \frac{\sum_{j=1}^{v} a_{ij} \vec{q}j_{.}^{(t)}}{\sum_{j=1}^{v} a_{ij}} \quad \vec{u}_{i.}^{(t)} = \frac{\sum_{j=1}^{v} a_{ij} \vec{q}j_{.}^{(t)}}{\sum_{j=1}^{v} a_{ij}} \tag{5}$$

In each iteration, each set of vertices obtains information from neighboring target nodes, thereby updating the initial value Y. At the same time, the vertex feeds back information to neighboring vertices. This directly proves that ($Q^{(t)}, U^{(t)}$) converges to the stable optimal solution.

2.3 Optimization Path of New Media City Image Communication Strategy based on Data Mining Technology

Basic information: name, gender, age, education background, income, occupation, address, contact information, etc; Behavior information: browsing records, browsing duration, browsing content, browsing frequency, browsing keywords, consumption time, consumption content, consumption frequency, consumption requirements, etc; Psychological information: personality characteristics, original consumption intention, consumption expectation, post consumption evaluation, consumption psychology, etc. [9]. The precise positioning process of city image based on big data is shown in Figure 1:

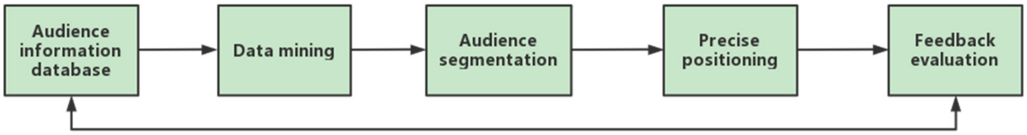


Figure 1. Big data precise positioning process

In this process, audience segmentation is the foundation. When we subdivide the urban audience, we can start from the audience data information base, from three dimensions, according to the different attributes of the audience [10].

2.3.1 City image communication strategy

First, city managers produce and disseminate the content of city image communication activities to produce a large amount of data; Secondly, data mining technology is used to analyze and summarize the existing data, so as to provide decision support for subsequent content survival and dissemination activities. This includes two links: first, in the content production link, collect the data of the urban audience, gain insight into the characteristics and needs of the audience, and produce content (theme, method, content, derived content) according to the needs of the audience; Second, in the communication link, precise information push, optimize the media mix, and personalized communication; Third, data collection to form a big data platform [11-12].

3.EXPERIMENTAL RESEARCH ON OPTIMAL PATH OF NEW MEDIA CITY IMAGE COMMUNICATION STRATEGY BASED ON DATA MINING TECHNOLOGY

3.1 Experimental program

In order to make this experiment more scientific and effective, this experiment mainly uses the questionnaire survey method for experimental research. This experiment investigates the image of local cities from the perspective of the audience. The experimental workers distributed 200 questionnaires, of which 80 valid questionnaires in the city and 100 from outside the city were collected, a total of 180 questionnaires. The populations surveyed this time are distributed under the age of 40 to ensure the validity of the experimental data. On the basis of this experiment, in order to further research and analyze this experiment, this article investigates the communication efficiency of this city in new media such as Weibo, WeChat and Toutiao, and uses mathematical statistics to calculate the results obtained. ,analyze.

3.2 Research methods

3.2.1 Questionnaire survey method

In this experiment, a targeted questionnaire was set up, and a fully enclosed method was used to investigate the effect of local city image dissemination. Its purpose is to promote the correct entry of the respondents.

3.2.2 Field research method

In this experiment, by going deep into the local city and conducting field research and collecting data on the way of city image communication, the collected data will be sorted and analyzed. These data provide a reliable reference for the topic selection of this article.

3.2.3 Mathematical Statistics

Use the relevant software to carry on the statistical analysis to the research result of this article.

3.2.4 Interview method

This experiment conducted in-depth interviews with relevant personnel on the image of local cities and recorded data. These data provide a reliable reference for the topic selection of this article.

4.EXPERIMENTAL ANALYSIS OF OPTIMAL PATH OF NEW MEDIA CITY IMAGE COMMUNICATION STRATEGY BASED ON DATA MINING TECHNOLOGY

4.1 Public awareness of city image

In order to make this experiment more scientific and effective, this experiment conducted a field questionnaire survey on the image dissemination of local cities in the form of a questionnaire survey, and used the percentile system for statistical analysis of the results. The final results are shown in Table 2.

Table 2. Public awareness of city image

	Very familiar	Familiar	General understanding	Only heard of
Familiarity	6.7	41.7	36.7	15
Internal familiarity	5.0	35.0	3.3	0
External familiarity	1.7	6.7	33.3	15

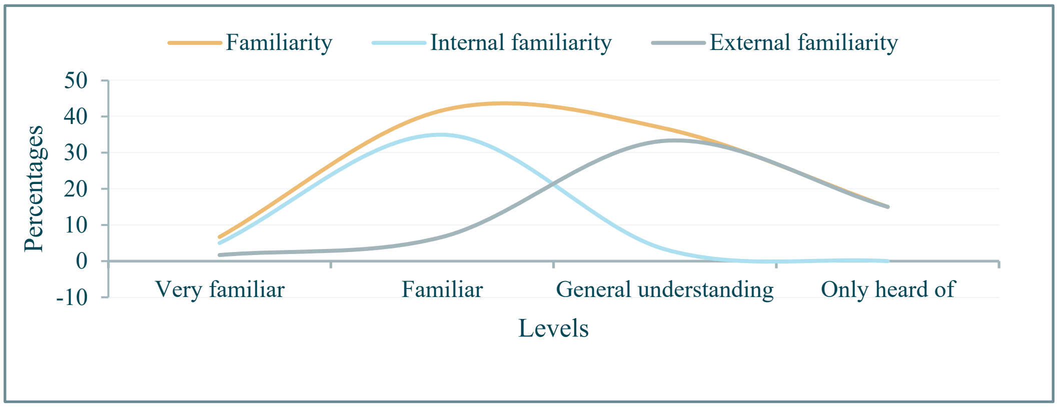


Figure 2. Public awareness of city image

It can be seen from Figure 2 that from the analysis of the image positioning of the local city, the public's perception of it is relatively vague and not clear. Only 6.7% of the respondents are very familiar with the local city, and only 41.7% are familiar with the city, of which 35 % Are long-term residents of the city. From this, it can be seen that the overall effect of image dissemination in local cities is poor, and there is still much room for improvement.

4.2 Analysis of the percentage of urban new media communication

In order to further research and analyze this experiment, this experiment was verified by researching new media channels such as Weibo, WeChat and Toutiao. The results are shown in Table 3.

Table 3. Analysis on the Percentage of New Media Communication in the City

	Politics	Economy	Humanities	Environment	Citizen	Convenience services	Urban construction	Others
Weibo	15.3	6.1	6.5	1.6	3.6	9.7	3.8	53.4
WeChat	27.7	9.9	7.2	2.4	4.8	7.9	1.4	38.7
Today's headlines	25	16.7	9.5	4.8	27.4	14.3	2.4	0

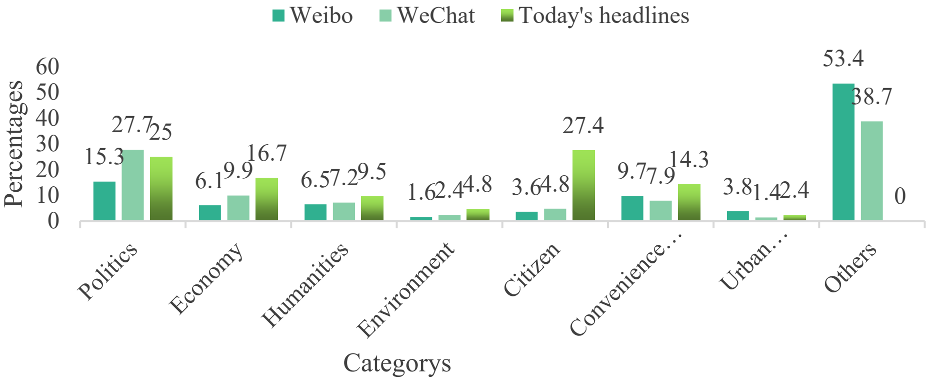


Figure 3. Analysis on the Percentage of New Media Communication in the City

It can be seen from Figure 3 that the political image accounts for the highest proportion. In addition, economic, humanities, and convenient services are more displayed. This reflects that the local city is in a period of economic development. It can be seen from this that new media has become an important channel for cities to spread their influence, strength and image.

5.CONCLUSIONS

The purpose of this paper is to analyze the optimization path of new media city image communication strategy based on data mining technology. Through an overview of urban communication theory and organic integration with new media, the application of new media and city image communication are based on this. The advantages are discussed in detail. In the end, this paper conducts an in-depth analysis of the application of data mining technology in new media communication, and uses questionnaire surveys to conduct an experimental investigation on the image communication of local cities. The final conclusion is that new media has become incompatible with city image communication. The power of neglect.

ACKNOWLEDGEMENTS

This work was supported by 2021 school level scientific research fund project of Xi'an Eurasia University：Research on the Relationship between the Construction of Xi'an New Media Matrix and the Construction of City Image (Fund number:2021XJSK16).

REFERENCES

[1] Walters T, Insch A. How community event narratives contribute to place branding[J]. Journal of place management and development, 2018, 11(1):130-144.

[2] Justine, Ferey, Florent, et al. A new optimization strategy for MALDI FTICR MS tissue analysis for untargeted metabolomics using experimental design and data modeling[J]. Analytical & Bioanalytical Chemistry, 2019, 411(17):3891–3903.

[3] Sima Shuanglong, Li Peiju. A study on international communication of city image in public diplomacy[J]. Journal of Zhenjiang College, 2018, 031(001):82-87.

[4] University H C. Image processing system, and image processing method[J]. And Then Becomes A Constant Value after Time, 2017, 1(1):65-70.

[5] Li Feng. Effectiveness of a communication model in city branding using events: The case of the Taiwan Lantern festival[J]. International Journal of Event & Festival Management, 2016, 7(2):137-148.

[6] Sun Xunqiang. Application of EPON in City Image Monitoring and Communication System[J]. Enterprise Technology Development (Academic Edition), 2017, 036(009):34-36.

[7] Yang Xue. Construction and Communication of the City Image of Changsha by Foreign Media[J]. Journal of Changsha University, 2017, 031(003):1-6.

[8] Mundaca E. Habitat Structure, Resources and Natural Enemies: Their Influence on Population Fluctuations of the Kowhai Moth Uresiphita Polygonalis Maorialis (Felder)[J]. Image Communication, 2012, 27(1):39-53.

[9] Casero-Ripolls, Andreu, Feenstra, et al. Old and New Media Logics in an Electoral Campaign. [J]. International Journal of Press/Politics, 2016, 21(3):378-397.

[10] Girard M, Stark D. Heterarchies of Value in Manhattan-Based New Media Firms[J]. Theory, Culture & Society, 2016, 20(3):77-105.

[11] Trudeau J. The role of new media on teen sexual behaviors and fertility outcomes—the case of 16 and Pregnant[J]. Southern Economic Journal, 2016, 82(3):975-1003.

[12] Aguirre E, Roggeveen A L, Grewal D, et al. The personalization-privacy paradox: implications for new media[J]. Journal of Consumer Marketing, 2016, 33(2):98-110.

Research on target grid investment optimization technology of medium and low voltage distribution network based on improved genetic algorithm

Na Yu[1,a], Ming Chen[2,b], Yan Wu[2,c*]
[1]Guangdong Power Grid Co., Ltd., Guangzhou City, Guangdong, 510080, China
[2]Grid Planning and Research Center, Guangdong Power Grid Co., Ltd., Guangzhou City, Guangdong, 510080, China
[a]317255621@qq.com, [b]13751853697@139.com, [c]*89687771@qq.com

ABSTRACT

With the continuous improvement of the electricity demand of the whole society in China and the complexity of power grid network planning and construction, how to further optimize the structure of low voltage distribution network and improve the investment efficiency and efficiency of distribution network planning has become one of the important topics facing the operation and development of power grid enterprises. This paper analyzes the current situation and problems of medium and low voltage distribution network construction in a region of GD province, combines with the relevant principles of improved genetic algorithm, optimizes the investment and construction of medium and low voltage distribution network construction in this region, and verifies the effectiveness of the model method.

Key words: improved genetic algorithm, medium and low voltage distribution network, target network frame, investment optimization

1 INTRODUCTION

In general, at the present stage, China is in the development level of distribution network optimization work is still in the process of exploration, compared with western countries, there is still a certain gap.

Document [1] combined with actual investigation, has carried out in-depth exploration of some problems in urban network planning, hoping to play a positive reference role. Literature [2] for the current distribution network network planning of line long, low line power supply rate, distribution automation coverage is not high, power supply path through forest and wetland reserve, gives the grid index data, from strong grid structure, power supply capacity, power supply quality and equipment level advanced grid planning principles and matters needing attention, put forward distribution network network planning ideas, comprehensive guarantee the rationality of distribution network line network planning, provide reference for the sustainable development of distribution network line. Literature [3] mainly analyzes the problems existing in urban distribution network in China, reduces the network loss for optimizing network structure, introduces the network structure, adopts genetic algorithm for distribution network reconstruction, and verifies the 33 nodes based on IEEE standard, provides theoretical and experimental basis for optimizing the structure of distribution network. Finally, the grid structure is applied in an urban distribution network system to verify its feasibility and practicability. Document [4] puts forward the specific requirements of distribution network structure optimization from the four aspects of economy, flexibility, stability and bearing capacity, and puts forward the idea of distribution network structure optimization. Literature [5] combined with the current situation of distribution network construction in China, analyzes the existing problems of distribution network planning, and proposes specific measures to improve the level of network network planning.

To sum up, relevant scholars have taken in-depth measures for the current situation and problems of network distribution and network frame planning in China, and put forward relevant optimization ideas, while the research on the specific methods of network frame optimization is relatively weak. Therefore, the research carried out in this paper is necessary.

International Conference on Computer Graphics, Artificial Intelligence, and Data Processing (ICCAID 2022)
edited by Ruishi Liang, Jing Wang, Proc. of SPIE Vol. 12604, 1260408
© 2023 SPIE · 0277-786X · doi: 10.1117/12.2674646

2 ANALYSIS OF THE CURRENT SITUATION OF MEDIUM AND LOW VOLTAGE DISTRIBUTION NETWORK FRAME CONSTRUCTION

This paper takes the current situation of medium and low voltage distribution network construction in a local company in GD province as an example. Through the analysis of the network structure, equipment, equipment and power supply capacity of the distribution network, the current problems of medium voltage distribution network in Chancheng District are shown in the following table:

Table 1. Diagnostic Analysis of medium and low voltage distribution

network construction in G D province.

Diagnostic content	evaluating indicator	Current situation of power grid construction
The superior power supply	capacity-load ratio	The capacity ratio of 110kV substation is 3.15, and the transformer capacity is relatively sufficient, which can meet the recent electricity demand
	Multi-main change rate	The multi-main transformer rate of the substation is 100%, which meets the reliability requirements
	10kV outgoing rate	The interval rate of 10kV outlet line in the substation is 74.4%
network topology	Main variant "N-1" pass rate	The pass rate of the main variant "N-1" was 83.6%
	Line contact rate	The connection rate of the 10kV line is 85.2%
	Typical wiring rate	Typical wiring rate is 20%
Equipment level	Pass rate of line trunk section	The qualified rate of the trunk section is 57%
	Qualified rate of line trunk length	The qualified rate of the trunk length was 78%
	Line insulation rate	The line insulation rate is 84%
Running metrics	Line average load rate	The average load rate of the line is 44%, and some lines have overloading and overload problems
	Variable load rate	The load rate of the transformer is 43%, with overloading and overload problems
	Line availability rate	Line transfer rate is 78%

3 RESEARCH ON THE OPTIMIZATION TECHNOLOGY OF THE MEDIUM AND LOW VOLTAGE DISTRIBUTION NETWORK FRAME BASED ON THE IMPROVED GENETIC ALGORITHM

3.1 Basic principle and problem analysis of simple single-parent genetic algorithm

All the genetic operations of the single-parent genetic algorithm are carried out on an individual, through a single parent. There is no cross operator of the traditional genetic algorithm, which makes the genetic operation simple and can improve the computational efficiency.

Genetic algorithm agrees with the basic idea of conventional genetic algorithm and belongs to the category of genetic algorithm. The idea is: the genetic operator acts on the current population to produce new individuals, controls the evolutionary direction of the population through selection operation, and conducts selection and genetic operation generation by generation, so as to realize the evolution of the population and reproduce the global optimal individual. According to the definition of the single-parent genetic algorithm, the only difference from the conventional genetic algorithm is that there is no cross operator. In conventional genetic algorithms, cross operations can produce new individuals and thus control the evolutionary direction of the population. Recombination operation is used to produce new individuals, and the recombination operators include shift operators, transposition operators, and inversion operators.

However, single-parent genetic algorithm has its advantages, which can solve the "precocious convergence" problem caused by conventional genetic algorithm. Conventional genetic algorithms mainly produce new individuals through the cross operation. When the two individuals acted on by the cross operator are exactly the same, the genetic operation cannot produce new individuals, and the cross operator fails. Individuals lose their diversity, the genetic iteration is difficult to go on, and the phenomenon of "precocious convergence" is easy to occur. However, the genetic operation of the single-parent genetic algorithm is all carried out on an individual, regardless of the diversity of the population, and has no requirement for the diversity of the population. Therefore, it has a good application value in the distribution network reconstruction.

3.2 Improvement of the genetic algorithm

According to the function of the fitness function, the selection of the fitness function in the genetic algorithm is very critical. Since the genetic algorithm basically does not consider the external information when selecting the operation, the only basis of the selection operation is the fitness function value, so it controls the evolutionary direction of the population. Improper selection of the fitness function may lead to slow calculation convergence speed or even failure to find the optimal individual. The fitness function is generally transformed from the objective function, which ensures that the high-quality solution has a great chance of survival.

There are two common methods for converting from the objective function value $f(x)$ at a point in the solution space to the fitness function value $Fit(f(x))$ of the corresponding individual in the search space.

1)Conversion directly to the fitness function by the objective function to be solved: if the objective function is the maximum optimization problem, then:

$$Fit(f(x)) = f(x) \qquad (1)$$

If the objective function is the minimum optimization problem, then:

$$Fit(f(x)) = \frac{1}{f(x)} \qquad (2)$$

This fitness function is simple and intuitive, but because the genetic algorithm requires non-negative values, this method is only applicable if the objective function is greater than zero. Some other functions to be solved vary greatly in the distribution of function values, so the obtained average fitness can be beneficial to reflect the average performance of the population and affect the performance of the algorithm.

2) transforms the desired objective function into a fitness function, i. e., if the objective function is a minimum optimization problem, then:

$$Fit(f(x)) = \begin{cases} C_{\max} - f(x) & f(x) > C_{\max} \\ 0 & \text{other} \end{cases} \qquad (3)$$

where $C_{\max}$ is a relatively large number. If the objective function is a maximum optimization problem, then:

$$Fit(f(x)) = \begin{cases} f(x)\text{-}C_{\max} & f(x) > C_{\max} \\ 0 & \text{other} \end{cases} \qquad (4)$$

where $C_{\min}$ is a relatively large number.

3.3 Implementation steps of grid frame optimization

According to the above distribution network frame optimization method of improved single-parent genetic algorithm, the steps are:

1) Combined with the characteristics and reconstruction characteristics of the distribution network, the switching status problem is transformed into chromosome coding binary coding, and the coding is simplified combined with the characteristics of the distribution network;

2) Determine the population size and generate chromosome groups, the chromosome length is equal to the effective number of switches of the simplified network, and the number of "" in the chromosome is the number of contact switches;

3) The reliability index of the distribution network is calculated by the failure rate influence method to obtain the chromosomal fitness value;

4) The selection method of optimal preservation and competition is used for selection operation to string the duplicated genes to pairing

storeroom;

5) Transfer and change the operation until the resulting offspring population reaches the set scale;

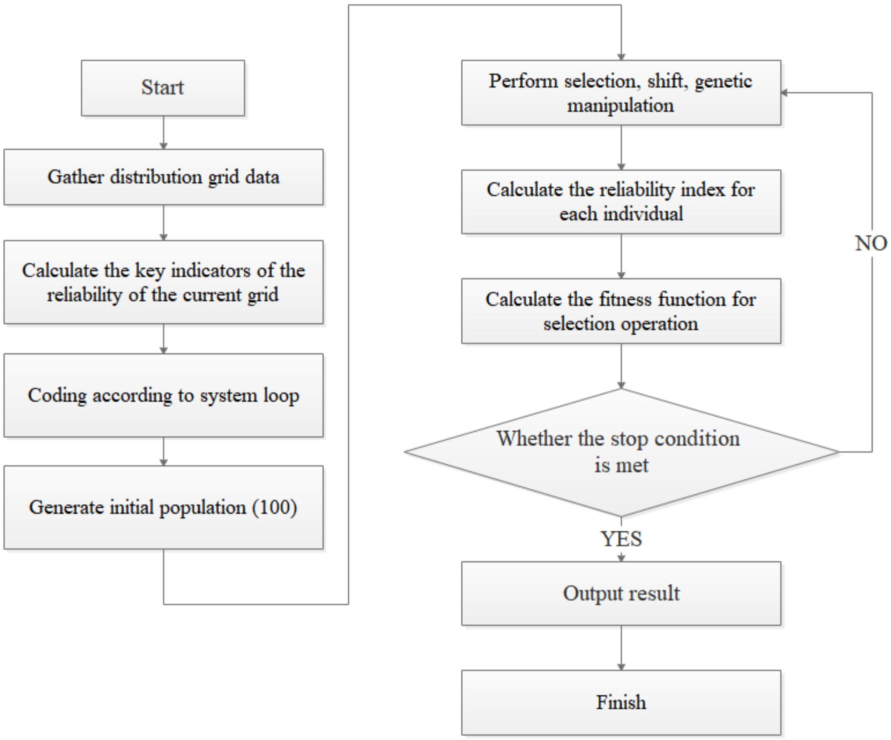

Start
Gather distribution grid data
Calculate the key indicators of the reliability of the current grid
Coding according to system loop
Generate initial population (100)
Perform selection, shift, genetic manipulation
Calculate the reliability index for each individual
Calculate the fitness function for selection operation
Whether the stop condition is met
NO
YES
Output result
Finish

Figure 1. Implementation process of medium and low voltage network frame optimization method based on improved genetic algorithm.

4 EMPIRICAL ANALYSIS

This paper takes the regional distribution network system in Chapter 2 as an example to carry out empirical analysis and research. Selected rack systems are known to include:

Two voltage levels of 11kV and 11kV, 11 kV is a radial network, can be reversed. There are 38 load points and 26 switches in the system, out of which 4 are contact switches. In the original system, switches 23,24,25, and 26 are disconnected, and the remaining switches are closed. The system network structure diagram is shown below in Figure 2.

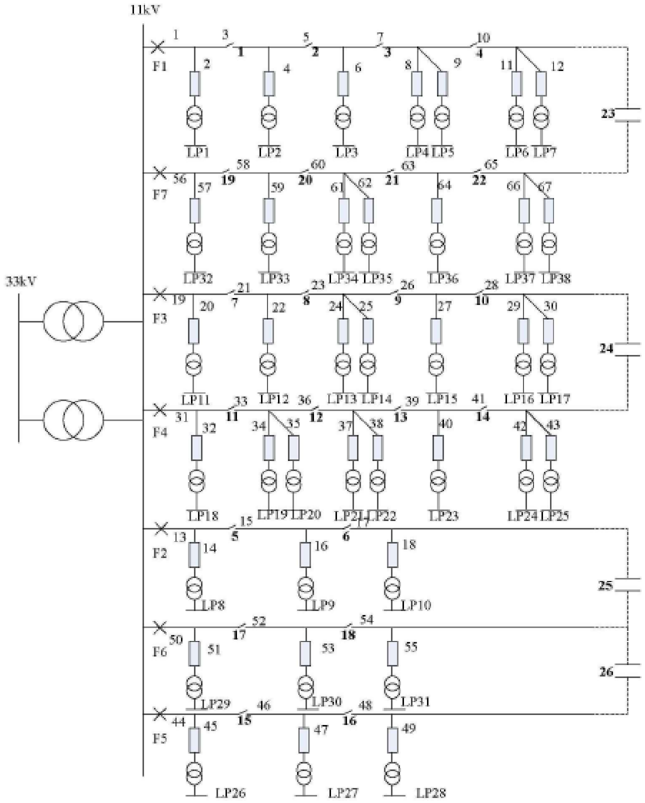


Figure 2. Structure diagram of the system grid frame.

Using the genetic algorithm reconstruction theory, optimize the distribution network structure through programming, and get the following results:

Table 1. Comparative analysis table before and after the structure optimization of medium and low voltage distribution network.

scheme	Disconnect the switch	AITC (Average power outage times of the user)	AIHV (Average power outage time of users)	SA (Power supply reliability rate%)
Original distribution network	23 24 25 26	0.4300	4.406	99.9497
Optimized network	14 23 25 26	0.4257	4.252	99.9557

As can be seen from the above table, the optimal network structure is selected through the network structure optimization of the distribution networkThe average number of users and the average power outage time of users are reduced compared with the original network, providing power supplyThe reliability rate has been increased from 99.9497% to 99.9557%. It has some effect to improve the reliability of the distribution network system, and also proves the effectiveness of the optimization algorithm.

5 CONCLUSION

Under the premise of the current situation and problems of LV distribution network construction in some area of GD province, this paper proposes the network network investment optimization model based on improved genetic algorithm, and verifies its effectiveness combined with practical cases. The method model can further improve the structure of medium and low voltage distribution network and improve the investment efficiency of enterprise grid.

REFERENCES

[1] Xing Guozhong. Analysis of urban Distribution Network Planning Problems [J]. Chinese New Technology and New Products, 2018(11):55-56.DOI:10.13612/j.cnki.cntp.2018.11.033.

[2] Dong Fude.10 kV distribution network [J]. Electrotechnical technology, 2020(24):141-142. DOI:10.19768/j.cnki.dgjs.2020.24.055.

[3] The beam bridge is new. Practical measures of distribution network transformation based on optimized network frame structure [J]. Electrotechnical technology, 2019(22):60-62.DOI:10.19768/j.cnki.dgjs.2019.22.023.

[4] to Joe. Exploration on Structure optimization strategy of transmission and distribution network frame [J]. Enterprise Technology Development, 2019,38(05):130-131+134.DOI:10.14165/j.cnki.hunansci.2019.05.039.

[5] Yin Fubin, Ren Zhongwu. Several aspects that should be paid attention to in urban network distribution and network frame planning [J]. Electronic production, 2014(20): 246. DOI: 10.16589/j.cnki.cn11-3571/tn.2014.20.067.

A DeepFake Compressed Video Detection Method Based on Dense Dynamic CNN

Xiuqing Mao[1*], Lei Sun[1], Hongmeng Zhang[1], Shuai Zhang[1]
[1] Information Engineering University, Zhengzhou/450001, China.
*Correspondence should be addressed to Xiuqing Mao; 21166813@qq.com

Abstract

The emergence of DeepFake poses serious risks to data privacy and social stability. We propose an end-to-end DeepFake video detection method based on a dense dynamic convolutional neural network (CNN) to address the poor performance of DeepFake video detection on complex compression formats and datasets of different forgery methods. In this method, extracted face images are clustered and cleaned by cosine similarity, and face images are expanded through data augmentation to improve data diversity. Dynamic dense blocks are incorporated in a CNN to address optimization difficulties in deep neural networks, and an attention mechanism further improves generalization power. Convolution kernel pruning increases processing speed by effectively reducing the computational needs due to dynamic convolution. Experiments demonstrate that this method has better results on DeepFake video detection across compression rates and datasets compared to other network models.

Keywords-component; DeepFake video detection, Dense Dynamic CNN.

1.Introduction

Advances in algorithms, computing power, and data collection have supported the development of artificial intelligence (AI) typified by deep learning, which has ushered in yet another wave of technological development[1]. As the focus of a new round of industrial transformation, deep learning has led us from the era of big data, which focuses on content and digital standards, to that of AI, which focuses on generating social value and extensive machine intervention. At the same time, AI poses a series of risks and challenges due to the cross-fertilization of innovative technologies. Unauthorized exploitation at the data level, black-box algorithms at the technical level, and blurred and hard-to-control application boundaries have led to AI misuse and risks to ethics and human rights[2].

At the end of 2017, Reddit user “deepfakes” posted a fake pornographic video of a celebrity using FakeApp, garnering significant public attention[3]. Since then, DeepFake, as a misuse of AI technology, have come to the forefront. The name combines the terms “deep learning” and “fake,” and refers to AI-facilitated, deep learning-based human image synthesis technology. The number of DeepFake projects in the GitHub open-source community is increasing, with projects like FaceApp[4], FaceSwap, DeepFaceLab[5], and other open-source code. Some of these projects allow complete face-swapping with a single click after configuring certain parameters, which greatly reduces the use threshold. DeepFake technology requires a large amount of data for model training, so public and political figures with numerous online images have become the initial victims. When combined with social media networks, DeepFakes can publicly present false information in a highly credible way, manipulating public opinion, triggering crises of social trust, provoking societal conflict, and posing serious risks to data privacy and social stability.

We conducted experiments on DeepFake public datasets such as FaceForensics++, DeepFake Detection (DFD), and DeepFake Detection Challenge (DFDC). Our results show that the accuracy of the dense dynamic CNN method incorporating an attention mechanism is higher than that of current methods. We propose a clustering algorithm based on cosine similarity to clean noisy data in the face-extraction stage, and data augmentation to increase data diversity and ameliorate an imbalanced classification problem. These methods improve the model’s accuracy when tested on a lightly compressed dataset by about 1.5%. We use DenseNet to construct a dense CNN and incorporate a dynamic convolutional module with an additional attention mechanism to improve the generalization power of the network and the expression power of its structure. Sensitivity-based pruning improves model convergence speed and reduces the number of parameters. The model’s inherent learning ability automatically extracts feature information of genuine versus forged images for training, and generates optimal model parameters for detection. These results were validated on a variety of datasets.

International Conference on Computer Graphics, Artificial Intelligence, and Data Processing (ICCAID 2022)
edited by Ruishi Liang, Jing Wang, Proc. of SPIE Vol. 12604, 1260409
© 2023 SPIE · 0277-786X · doi: 10.1117/12.2674838

The rest of this paper is organized as follows: In Sect. 2, DeepFake generation and detection methods are presented and discussed. The proposed method is presented in Sect. 3. Experimental results, comparisons, and discussions are illustrated in Sect. 4. Finally, we give out the conclusion in Sect. 5.

2.Related Work

2.1DeepFake Generation Methods

Face swapping is the replacement of a face in a target video using a face from a photo gallery. It is a technique used by cyber-attackers to penetrate identification or authentication systems and gain illegal access. Face-swapping images using deep network models such as CNN and generative adversarial networks (GANs) pose more difficulty for detection forensics because they preserve the lighting, character pose, and facial expressions in photos[6]. Faceswap-GAN[7] is a GAN-based DeepFake method that can generate images with resolutions of 64x64 (default), 128x128, and 256x256. The core techniques of VGG-Face[8] and FaceNet[9] have been incorporated in the encoder-decoder architecture of Faceswap-GAN. FaceNet's introduction of the multi-task convolutional neural network (MTCNN[10])allows for more stable face detection and alignment. VGG-Face's perceptual loss functions allow for the alignment of eye movements with facial actions, resulting in higher quality output videos.

2.2Video Sequence-based Detection Methods

Li et al.[11] proposed a blink-based method to identify DeepFake videos. The face and eye regions are extracted at the video-frame level, and a new frame sequence is created by extracting and scaling the bounding box of eye-region markers according to facial alignment. This sequence is assigned to a long-term recurrent convolutional network to achieve dynamic prediction of the open- and closed-eye states. This approach uses only blinking as a criterion to evaluate videos, and does not sufficiently consider the rationale behind the use of blink frequency for prediction. It is easily bypassed by post-processing or training of a more advanced model with blink capability.

2.3Video Frame-based Detection Methods

Detection methods based on inter-frame temporal correlation are generally constructed by deep RNNs. Creating DeepFake face-swap videos requires affine face transformation, such as scaling, rotating, and cropping, to match features of various facial regions from the source video. Artifacts left by this process can be captured by CNNs such as VGG16, ResNet50, ResNet101, and ResNet152, due to inconsistent resolutions between the distorted transformed facial regions and the surrounding environment.

3.Methodology

3.1Overview

The results and discussion may be presented separately, or in one combined section, and may optionally be divided into headed subsections. We used the general face detector Dlib for face extraction. However, we found that Dlib also located and extracted non-face regions. A clustering algorithm based on cosine similarity was subsequently used to clean the data and improve the model's training accuracy. Data augmentation was used to expand the data and improve data diversity. DenseNet[12] was used to build a dense CNN to address optimization issues of deep neural networks. An attention mechanism was fused with dynamic convolution so that the network focuses on DeepFake forgery regions such as face edges and nose and lip regions, and to increase the model's detection ability across datasets. Sensitivity-based pruning increased convergence speed and further reduced the number of model parameters. Authentic and forged images were directly input to the built network, which automatically extracted feature information for training and generated the optimal model parameters to detect image authenticity. This model of end-to-end DeepFake video detection based on a dense dynamic CNN is shown in Figure 1.

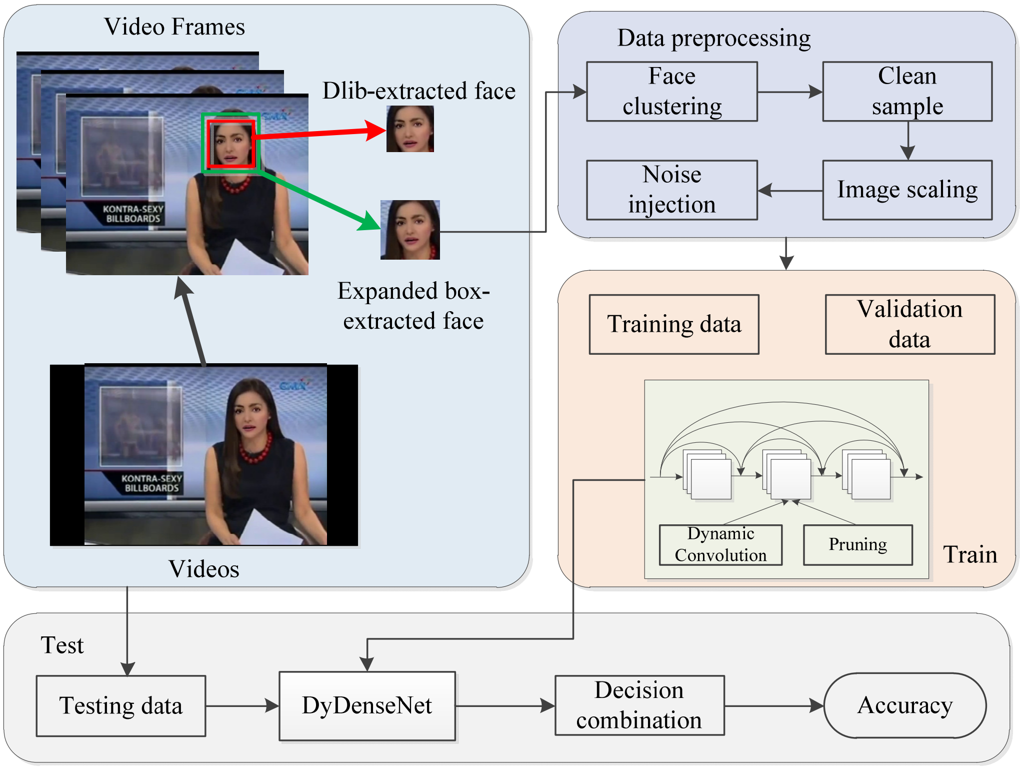


Figure 1: Overall detection process.

3.2Data Preprocessing

Data preprocessing had three parts. First, videos were split into frame images, and the dlib tool library located and extracted faces in video frames by 68 feature points. Facial extraction removes environmental interference detrimental to DeepFake detection and focuses the network on extracting subtle facial features, thus improving validation and testing accuracy. However, the rectangular box around the original dlib localized face in certain cases did not wrap the full face. Thus the original rectangular box was enlarged 1.5 times before extraction, and the cropped face was reduced to 224×224 resolution to facilitate network training.

Second, during dlib face extraction, there were interfering images that could affect the training and validation accuracy of our model. So, facial feature points were transformed to a 68×2 feature matrix, and extracted face images were clustered based on cosine similarity to clean interfering data. Given two images with feature matrices A and B, S_{AB} is the normalization of cosine similarity of A and B, and the cosine similarity θ is given by the dot product and vector length, where A_i and B_i represent the components of A and B, respectively:

$$S_{AB} = \cos(\theta) = \frac{A \cdot B}{\|A\|\|B\|} = \frac{\sum_{i=1}^{n} A_i \times B_i}{\sqrt{\sum_{i=1}^{n}(A_i)^2} \times \sqrt{\sum_{i=1}^{n}(B_i)^2}} \quad (1)$$

A correlation matrix was produced by calculating the cosine similarities of the two cropped face images, setting the relevant threshold for clustering, and removing the below-threshold images for the purpose of data cleaning. The effect is shown in Figure .

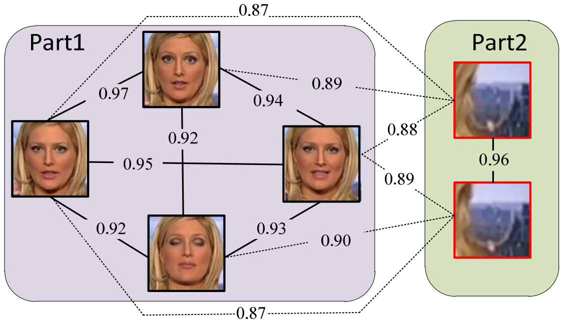
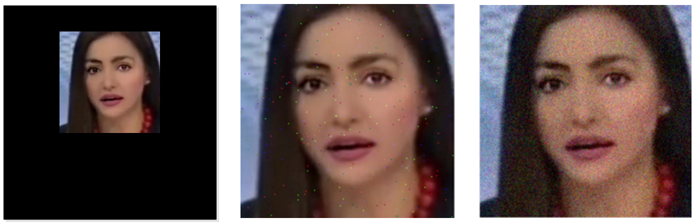


Figure 2: Clustering of noisy images and face images Video frame face cropping.

Finally, the cleaned training data were augmented with image scaling and noise injection to increase data diversity, addressing the problem of imbalanced classification and improving the model's generalization power. Inward image scaling to a lower resolution was performed while retaining the standard network size, as shown in Figure (a). Salt-and-pepper noise and Gaussian noise were applied, as shown in Figure (b) and (c). After this preprocessing, the training dataset was expanded four times from its original size.



(a) Inward image scaling (b) Salt-and-pepper noise (c) Gaussian noise

Figure 3 Data augmentation.

3.3 Constructing a Dense Dynamic CNN

At the point of submission, authors may provide all figures embedded within the manuscript at a convenient break near to where they are first referenced or, alternatively, they may be provided as separate files. All figures should be cited in the paper in a consecutive order. Where possible, figures should be displayed on a white background. When preparing figures, consider that they can occupy either a single column (half page width) or two columns (full page width), and should be sized accordingly. All figures must have an accompanying caption which includes a title and, preferably, a brief description (see Figure 1).

In this paper, instead of using a single convolution kernel on each layer of the model, multiple parallel convolution kernels are dynamically aggregated according to an attention mechanism, which dynamically adjusts the weight of each convolution kernel according to the input to generate adaptive dynamic convolution kernels. Since the attention value is a function of the input, dynamic convolution is no longer a linear function. To nonlinearly superimpose convolution kernels by attention values has stronger representation power. The traditional static perceptron is represented as $y = g(W^T x + b)$, where W and b are the weight matrix and bias vector, respectively, and g is the activation function. In this paper, the dynamic perceptron is defined by aggregating K linear functions $\{\widetilde{W}_k^T x + \tilde{b}_k\}$, as follows:

$$y = g(\widetilde{W}_k^T(x)x + \tilde{b}_k(x)) \tag{2}$$

$$\widetilde{W}(x) = \sum_{k=1}^{K} \pi_k(x)\widetilde{W}_k\,, \tilde{b}(x) = \sum_{k=1}^{K} \pi_k(x)\tilde{b}_k$$

$$\text{s.t. } 0 \leq \pi_k(x) \leq 1, \textstyle\sum_{k=1}^{K} \pi_k(x) = 1,$$

where π_k is the attention weight of the kth linear function, $\widetilde{W}_k^T x + \tilde{b}_k$. The aggregation weights $\widetilde{W}_k^T(x)$ and biases $\tilde{b}_k(x)$ are functions of the inputs, and have identical weights. The weight $\{\pi_k(x)\}$ is not fixed, but rather varies with the input x. It is a nonlinear function that represents the optimal aggregation of the linear model $\{\widetilde{W}_k^T x + \tilde{b}_k\}$ for a given input, such that the dynamic perceptron has better representation power than the static perceptron.

Just as a squeeze-and-excitation network (SENet)[13] calculates the output channel attention after a "squeeze" reinforcement learning transformation, we calculate the attention of the convolution kernel accordingly, and use it as the basis for pruning below. The attention model, as shown in Figure , has low computational complexity and consists of an average pooling layer and two full convolutional layers. Global spatial information is compressed by average pooling. Two fully connected layers, separated by a ReLU activation function, and a softmax function are used to generate normalized attention weights for the K convolution kernels.

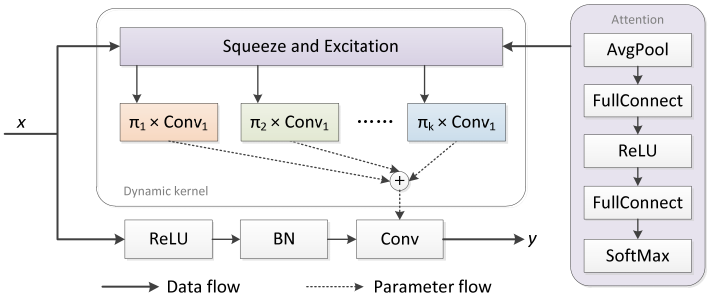


Figure 4: Schematic of a dynamic convolutional neural network incorporating an attention mechanism.

The difficulty of dynamic convolutional networks lies in the simultaneous learning of multiple convolution kernels and attention models, and this increases with the depth of the network. To address this issue, we first restrict attention to a near uniform distribution, which aids in the simultaneous learning of multiple convolution kernels during early-stage training. Second, to restrict attention values to simplify attention model learning will reduce the value space of multiple superimposed convolution kernels, so we restrict attention values between 0 and 1, with all attention summing to 1.

The main portion of a dense CNN uses the structure of dynamic dense block + transition layers. The dynamic dense block uses dense connections between layers such that the features of each layer can be passed between input and output layers at high speeds, ensuring maximum information flow between layers in the CNN. Since the input of later layers is very large, bottleneck layers are used within the dynamic dense block to reduce computation, primarily by adding 1×1 Conv to the original structure. Figure shows a schematic diagram of a three-layer densely connected dynamic dense block. The Nth layer receives the feature maps of all its predecessor layers, $X_0, X_1, \cdots, X_{N-1}$, as input, resulting in

$$X_l = H_l([X_0, X_1, \cdots, X_{N-1}]) \tag{3}$$

where $H_l(\cdot)$ is a nonlinear transformation function, which is a combinatorial operation, i.e., BN + ReLU + 1×1 Conv + BN + ReLU + 3×3 DyConv, where 1×1 Conv serves to reduce the number of features, thus improving computational efficiency, and the activation function increases the network's nonlinearity, allowing it to express more complex features.

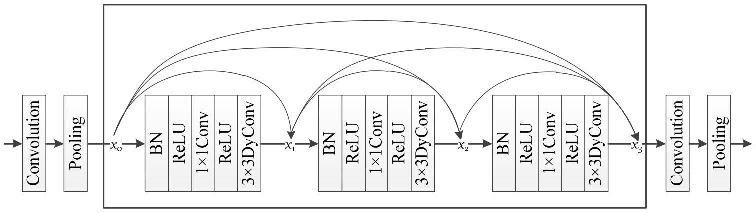


Figure 5: Schematic of dynamic dense blocks.

To merge the feature maps from the dynamic dense blocks for feature reuse, their sizes must be consistent. To accomplish this, the dense CNN reduces the size of feature maps through transition layers, which are placed between two dense blocks and adjust the number of feature map channels through a 1×1 convolutional layer. The size of feature maps is also reduced through a pooling layer. The transition layers include a 1×1 convolution layer and 2×2 average pooling layers with the structure of BN+ReLU+1×1 Conv+2×2 AvgPool, and simultaneously facilitate compression of the model. Figure shows the structure of this paper's dense CNN, containing a total of four dynamic dense blocks, connected by transition layers.

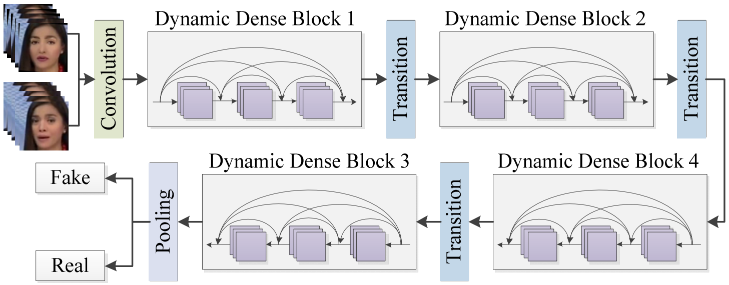


Figure 6: Framework of overall model.

Computational efficiency is improved by pruning insufficiently contributing convolution kernels, which can significantly compensate for the shortcomings of dynamic convolution. On average, fewer than 15% of an image's pixels change in DeepFake face-swapping videos. Since the video is a continuous image, facial changes are concentrated in the stitched and affine transformation regions, such as face and lip edges. The rest of the background regions can be regarded as similarity features, which are relatively constant, but require much computation in the dynamic CNN's deeper layers due to diffusion effects. Convolution kernel pruning can increase processing speed by reducing the number of below-threshold convolution kernels. It can also solve the problem of increased computational needs in dynamic CNN with deeper convolutional layers.

Inspired by the convolution pruning method proposed by Li[14], we firstly remove the convolution kernels with smaller contributions from the trained model, and then retrain the model. After calculating the sum of absolute weights W for each convolution kernel, we rank and prune the convolution kernels based on their sensitivity K_{se} and $\pi[\cdot]$. We create new convolution kernel matrices for all layers, and migrate the remaining convolution kernel weights to the new model. The weights are converged by retraining the new CNN. The loss function is based on cross-entropy, which measures the degree of difference between two probability distributions of some random variable, which is expressed in machine learning as the difference between the true and predicted probability distributions. The smaller the cross-entropy the better the model prediction. The pseudo-code is shown in Table 1.

Table 1: Pseudo-code of pruning algorithm

```
Input:  $M$: dense dynamic convolutional neural network;
        $\pi[\cdot]$: corresponding weight of convolution kernel
        $K_{se}$: percentage of convolution kernels pruned according to sensitivity
        $D_t$: real and forged image training sets for model training
Output: $M_{new}$: pruned model

if $K_{se}$ >0 then
  $F[\cdot] \leftarrow$ GetConvFilters ($M$)
  $\pi[\cdot] \leftarrow$ GetConvAttention ($M$)
  $W[\cdot] \leftarrow F[\cdot] \times \pi[\cdot]$
  for each $\omega$ in $W$ do
    $\pi[\cdot]$ [Order($F$)] += Abs($\omega$)
  end for
  $\pi[\cdot] \leftarrow$ Sort($\pi[\cdot]$)
  for all $\pi$ in $\pi[\cdot]$
    if $\pi < y$
      query the corresponding of $F$
      Calculate the proportion of kernels $K_{se}$
    end if
  end for
  $M_{new} \leftarrow$ FilterPruning ($M$;$K_{se}$)
  $M_{new} \leftarrow$ FineTuning ($M_{new}$;$D_t$)
end if
until loop all convolutional layers
```

The traditional quadratic loss function suffers from smaller parameter adjustments and slower training if errors are relatively large, which is influenced by the gradient. Cross-entropy eliminates the gradient's influence on parameter adjustment through logarithmic function derivation, achieving larger parameter adjustment amplitudes when the error is large, and increasing the network's convergence speed. The loss function under batch training is

$$L = \sum_{j=1}^{M}\sum_{i=1}^{N} y^{(i)} log\hat{y}^{(ij)} + (1 - y^{(ij)}\log(1-\hat{y}^{(ji)}) \quad (4)$$

where y and ŷ are respectively the actual and expected output of the network; M is the sample size in the current batch training; and N is the number of categories, which is 2 in this paper, for the categories of true and false.

Finally, to aggregate detection accuracy of individual frames into that of the whole video, four methods are considered as the final prediction of video detection accuracy: averaging individual frame prediction, taking the median prediction, taking the minimum prediction, and taking the maximum prediction.

4.Experimental Results

4.1Dataset and Evaluating Indicator

The FaceForensics++ dataset is an expansion of the FaceForensics dataset[15] produced by Rössler et al. It was included in the large DeepFake dataset DFDC, jointly produced by Google and Jigsaw, to add original DeepFakes and authentic videos. FF++ has been widely used as a standard dataset for training and testing DeepFake detection models. Rössler et al. divided it into four sub-datasets: DeepFake, Face2Face, FaceSwap, and NeuralTextures, each containing 1000 videos, according to the forgery methods. Comprising more than 1.8 million forged videos, they have three compression formats based on H.264 encoding: lossless, light, and strong, with compression parameters of 0, 23, and 40, respectively. The DFDC dataset[16] includes 119,197 videos, each 10 seconds long, with frame rates ranging from 15 to 30 fps, and resolution from 320×240 to 3840×2160. A variety of mainstream algorithms, such as DeepFakes and Face2Face, are used for synthetic face generation.

We used the FF++ dataset to train and validate the model, and other datasets were tested. The test sets comprised the four fake sub-datasets of FF++ and real datasets numbered 0–99. Videos numbered 100–199 were designated as validation sets, and those numbered 200–999 as training sets. When producing frame images from videos, this division of test, validation, and training sets was unchanged to ensure no data crossover. This eliminated the issue of "stealing" the test and validation set data distributions for training, thus ensuring the reliability of the model. During video frame extraction, considering feature redundancy caused by too low of a frame number, the method was set to intercept every 25 frames. The number of intercepted frames and their expanded images are shown in Table 2.

Table 2: Number of images in each dataset category

	Training set	Validation set
Before expansion	39537	2958
After expansion	158148	11832

The evaluation parameters in our experiments were the receiver operating characteristic (ROC) curves of the subjects. The following concepts are used in this study. Positive classification denotes real images, and negative classification denotes synthetic images. A true positive (TP) is actually positive and predicted to be positive. A false negative (FN) is actually positive but predicted to be negative. A false positive (FP) is actually negative but predicted to be positive. A true negative (TN) is actually negative and predicted to be negative. The false positive rate (FPR) is the proportion of all forged videos among the real videos predicted by the detection model; the true positive rate (TPR) is the proportion of all real videos among those predicted by the detection model. TPR and FPR are calculated as follows:

$$TPR = \frac{TP}{TP+FN} \quad (5)$$

$$FPR = \frac{FP}{FP+TN} \quad (6)$$

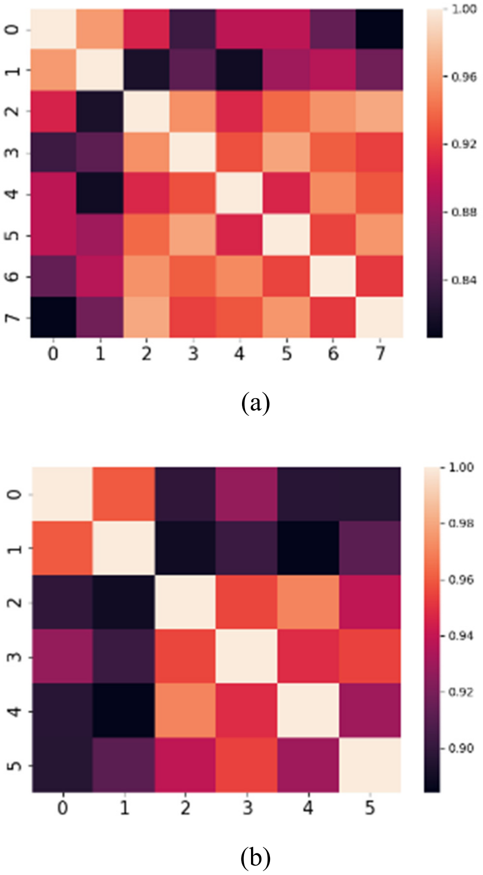
The vertical and horizontal coordinates of ROC curve are composed of TPR and FPR, respectively. The area under the ROC curve (AUC) is used to measure the classification model's efficacy. When a positive and negative sample are randomly selected, the probability that the current classification algorithm ranks the positive sample in front of the negative sample according to the calculated score value is the AUC. The higher the AUC the more likely a classification algorithm will rank positive samples in front of negative samples, thus enabling better classification.

4.2Experimental Setup

In this paper, the batch training size was set to 16, consistent with hardware conditions. Choosing a suitable batch size is important to improve the convergence speed and accuracy of the network model. The learning rate started at 0.001, decayed by 95% every 1000 steps, and used stochastic gradient descent optimization. Training was terminated at the 30th iteration cycle, and the model was fine-tuned using hard sample mining. The trained parametric model was used to test the data.

4.3Data Cleaning Experiments

First, we evaluated the face image cleaning method. Two videos, numbers 245 and 803, were randomly selected. There were eight images in video 245, and two noisy images, corresponding to the 0 and 1 coordinates in Figure (a). Six images in video 803, included two noisy images, corresponding to the 0 and 1 coordinates in Figure (b). These images were transformed to a color confusion matrix, as shown in Figure , after calculating the correlation matrix by cosine similarity. There was a large difference in color between the noisy and clean images; the noisy image was darker. After setting the threshold value to 0.92, the noisy and clean face images could be clustered accurately, and the dataset was effectively cleaned after the below-threshold classifications were removed.

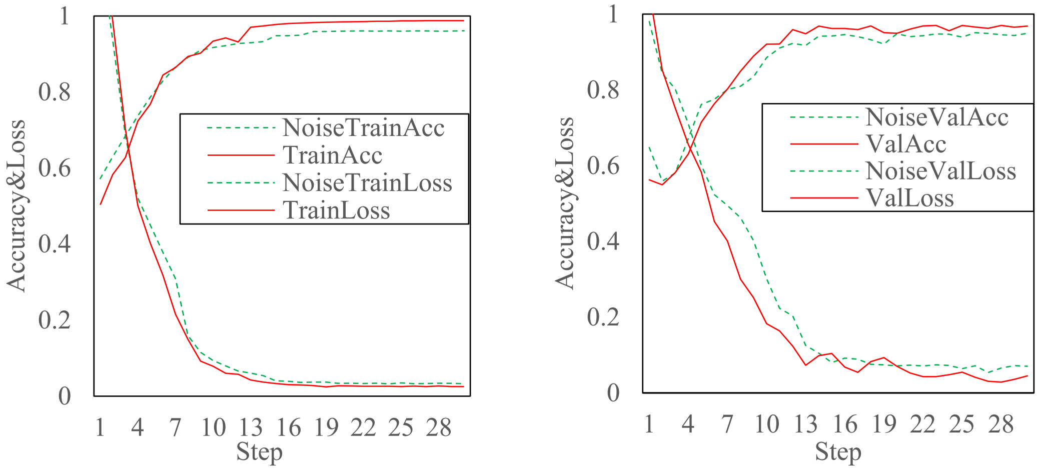


(a)

(b)

Figure 7: Partial image correlation matrix of videos 245 (a) and 803 (b).

In this paper, the clustering method was applied in training as well as testing. The proposed model was trained for 30 iterations based on the same training and validation sets, with accuracy and loss curves as shown in Figure . As the number of iterations increased, the model became increasingly powerful, but deep learning models are highly prone to overfitting. To avoid this, we adopted an early stopping strategy, terminating training when the loss function's value fluctuated very little in a certain period of time. From Figure (a), it can be seen that the final loss value of the noisy training model converged to about 0.032, and accuracy converged to about 0.96, while the loss value of the denoised training model converged to about 0.025, and the accuracy to about 0.99. From Figure (b), we can see that the final noisy model converged to a loss value of about 0.07 and accuracy of about 0.94 in the validation set, and the denoised training model converged to a loss value of about 0.04 and accuracy of about 0.97. The positive effect of denoising on model learning is clear.



(a) (b)

Figure 8: Comparison of learning validation curves with and without noise respectively is (a) and (b).

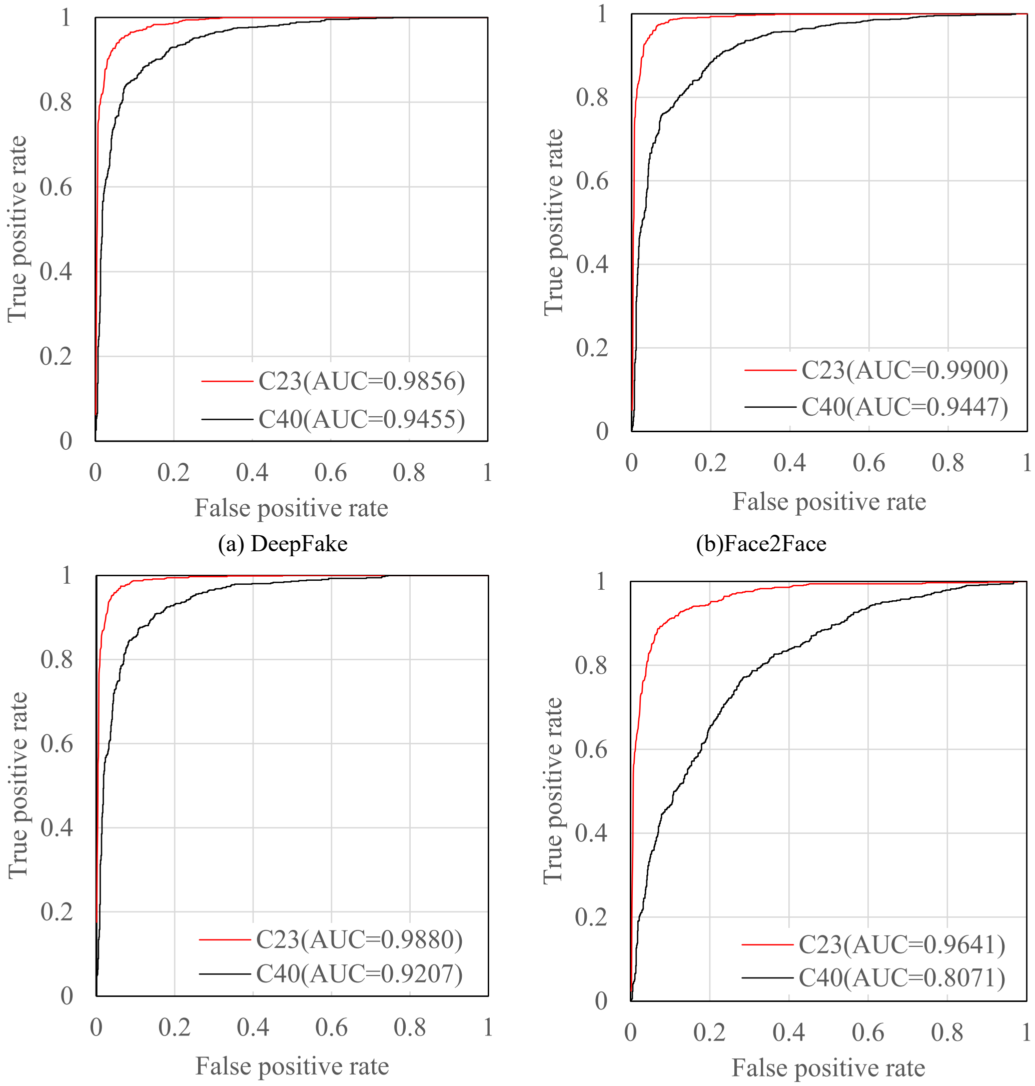
We tested on a lightly compressed dataset from FF++ and obtained the results shown in Table 3. It is evident that the proposed data cleaning method improves the model's accuracy. On lightly compressed data, data cleaning improved accuracy by 1.5%, and on heavily compressed data by 7%, demonstrating the efficacy and necessity of this method in DeepFake video detection.

Table 3: Area under ROC curve (AUC) values of different network methods on lightly compressed FaceForensics++ dataset

	Type	DeepFake	FaceSwap	Face-2Face	Neural Textures
VGG19	With noise	89.21	91.87	86.28	76.45
	Denoised	90.61	92.99	89.79	78.24
ResNet101	With noise	90.55	94.09	93.16	80.25
	Denoised	92.83	95.33	94.28	81.23
Inception_v3	With noise	91.73	96.25	94.05	87.11
	Denoised	94.70	97.72	95.57	88.95
Ours	With noise	97.32	97.68	98.01	96.55
	Denoised	98.56	98.80	99.00	97.41

4.4DeepFake Detection Experiments

Tests were conducted on the standard dataset FF++ dataset and the public DFD and DFDC datasets, and lossless compression, lightly compressed, and heavily compressed dataset versions were used. To effectively evaluate the learning ability and generalizability of the algorithm, experiments were divided into those of in-library and cross-library detection. In-library detection was tested on DeepFake, FaceSwap, Face2Face, and Neural Textures on the FF++ dataset, with ROC curves as shown in Figure . The figure demonstrates that the model's AUC took high values, and the model performed well in terms of prediction and generalization. Videos with high compression rates lose many features, making detection more difficult. The model demonstrated an ability to combat compression issues, with an average of over 90% AUC.



(a) DeepFake

(b)Face2Face

(c)FaceSwap

(d)NeuralTextures

Figure 9: Comparison of receiver operating characteristic curves for in-library DeepFake detection.

Cross-library detection was still relatively difficult, as a model trained on one tampering method can show significantly degraded performance when tested on another tampering method. Notably, the combined training of the model on multiple datasets did not degrade its performance on each tampered dataset. The experimental results also show that combined training with different tampering methods can to some extent improve model performance on highly compressed data. Therefore, to expand the training set with new tampering types is an effective solution to cross-library detection. We used the DFD and DFDC datasets, with results as shown in Table 4, where DFD had a slightly lower AUC value compared to FF++, but the model still demonstrated better generalization performance across datasets.

Table 4: ACC values of networks on different compression parameter datasets

	DeepFake		FaceSwap		Face2Face		NeuralTextures	
	C23	C40	C23	C40	C23	C40	C23	C40
Fridrich et al. [17]	77.12	69.58	79.51	60.58	74.68	57.55	76.94	60.69
Cozzolino et al. [18]	81.78	68.26	85.69	62.08	79.8	55.77	80.6	62.42
Bayar & Stamm[[19]	90.18	80.95	93.14	76.83	86.1	73.63	86.04	72.38
Raghavendra et al. [20]	-	-	-	-	93.5	82.13	-	-
Ous	**97.80**	**92.24**	**98.39**	**91.69**	**98.27**	**90.67**	92.86	78.78

We also compared different depths of neural networks to highlight the model's exceptional performance. The proposed detection method performed well on lossless compressed datasets, an uncommon format on social networks. We evaluated model performance on varying neural network depths between lightly compressed and heavily compressed datasets.

Table 5: AUC values of different networks on DFD and DFDC datasets

Model	DFD (C23)	DFD (C40)	DFDC
This method	**0.9680**	**0.9007**	**0.7989**

As shown in Tables 5 and 6, the dense dynamic CNN used in this paper achieved a high detection accuracy compared to methods in the literature. These results show that the proposed DeepFake video detection method with dense dynamic CNN incorporating an attention mechanism is effective, guarantees the accuracy of cross-library detection, and performs well in a highly compressed video environment.

Table 6: ACC values of networks on DFD and DFDC datasets

Model	DFD (C23)	DFD (C40)	DFDC
This method	**94.79**	**83.36**	**67.80**

4.5Ablation Experiments

We conducted targeted ablation experiments to verify the efficacy of the proposed method, using the image feature extraction module as the base network, and systematically adding the preprocessing, dynamic convolution, and pruning modules to calculate their accuracy values after training and validation. Five comparison models were examined.

Table 7: Ablation experiments comparing different model modules

Model	Training accuracy	Verification accuracy	Training time
DenseNet	93.14	91.43	20h31min
DenseNet+DyKernel	97.24	96.18	20h52min
DenseNet+DyKernel+ Prune	96.91	96.13	19h41min
Preprocessing + DenseNet+DyKernel	99.57	98.84	20h26min
Preprocessing + DenseNet+DyKernel+ Prune	99.35	98.56	18h53min

As shown in Table 7, following the addition of DenseNet and DyDenseNet, we found that to add dynamic convolution and an attention mechanism significantly improved the detection accuracy of DeepFake face-swap videos. Pruning slightly increased the accuracy of training and validation, and definitively improved the convergence time of model training. The preprocessing module improved training and validation accuracy and reduced convergence time. Ablation experiments showed that the proposed method yields excellent results for DeepFake video detection.

5.Conclusions

We proposed a DeepFake video detection model with a dense dynamic CNN incorporating an attention mechanism to address generalizability issues across various compression formats. To resolve the issue of insufficient samples and noisy images in the training phase, we used inward image scaling and noise injection, expanding the data volume by a factor of four. Modifying the face-extraction algorithm to enlarge the frame selection by 1.5 times, and cleaning data based on cosine similarity further refined the dataset and allowed the neural network to better capture deep semantic information of the face. A dense network incorporating dynamic convolution kernels was used to improve nonlinear fitting model performance and increase detection accuracy, while increasing generalizability to complex image synthesis methods and multiple compression formats. Sensitivity-based pruning of the model was found to increase model convergence speed and further reduce the number of model parameters, thus reducing the computational needs of a dynamic CNN. Experiments on various compression formats across different datasets showed that the proposed method predicted DeepFakes and performed better than other networks on heavily compressed datasets, strongly supporting robustness and generalizability.

References

[1] Tianchen Z,Xiang X,et al . “Learning Self-Consistency for Deepfake Detection,” from arxiv, ICCV 2021 Oral.

[2] Aayushi Agarwal,Akshay Agarwal,et al. “MD-CSDNetwork: Multi-Domain Cross Stitched Network for Deepfake Detection,” Calif. L. Rev., 107: 1753. 2021.

[3] FakeApp, https://www.malavida.com/en/soft/fakeapp/, 2022.

[4] FaceApp. https://faceapp.com/app. 2022.

[5] DeepFaceLab github. https://github.com/iperov/DeepFaceLab, 2022.

[6] Korshunova I, Shi W, Dambre J, et al. “Fast face-swap using convolutional neural networks,” *IEEE International Conference on Computer Vision*. 2017: 3677-3685.

[7] Faceswap-GAN. https://github.com/shaoanlu/faceswap-GAN 2022.

[8] Keras-VGGFace. “VGGFace implementation with Keras framework,” https://github.com/rcmalli/keras-vggface. 2022.

[9] FaceNet. https://github.com/davidsandberg/facenet. 2022.

[10] Zhang K, Zhang Z, Li Z, et al. “Joint face detection and alignment using multitask cascaded convolutional networks,” *IEEE Signal Processing Letters*, 2016, 23(10): 1499-1503.

[11] Li Y, Chang M C, Lyu S. “In ictu oculi: Exposing ai created fake videos by detecting eye blinking,” *IEEE International Workshop on Information Forensics and Security*. 2018: 1-7.

[12] Huang G, Liu Z, Laurens VDM, et al. “Densely Connected Convolutional Networks,” *IEEE Conference on Computer Vision and Pattern Recognition*. Honolulu, HI, 2017: 2261-2269.

[13] Jie H, Li S, Gang S, et al. “Squeeze-and-excitation net-works,” *IEEE Conference on Computer Vision and Pattern Recognition*. 2018.

[14] Li H, Kadav A, Durdanovic, I., et al. “Pruning filters for efficient convnet,” *International Conference on Learning Representations*. 2017.

[15] Rössler A, Cozzolino D, Verdoliva L, et al. “FaceForensics: A large-scale video dataset for forgery detection in human faces,” https://arxiv.org/abs/1803.09179. 2021.

[16] Brian D, Joanna B, Ben P, et al. “The DeepFake Detection Challenge. DFDC Dataset,” arXiv preprint:2006.07397. 2021

[17] Fridrich J, Kodovsky J. “Rich Models for Steganalysis of Digital Images,” *IEEE Transactions on Information Forensics and Security*, 7(3): 868-882. 2012.

[18] Cozzolino D, Poggi G, Verdoliva L. “Recasting Residual-based Local Descriptors as Convolutional Neural Networks,” *5th ACM Workshop on Information Hiding and Multimedia Security*. Philadelphia, Pennsylvania, USA. New York, pp. 159-164. 2017.

[19] Bayar B, Stamm MC. “A Deep Learning Approach to Universal Image Manipulation Detection Using a New Convolutional Layer,” *4th ACM Workshop on Information Hiding and Multimedia Security*. Vigo, Galicia, Spain. New York, pp. 5-10. 2016.

[20] Raghavendra R, Raja KB, Venkatesh S, et al. “Transferable Deep-CNN Features for Detecting Digital and Print-Scanned Morphed Face Images,” *IEEE Conference on Computer Vision and Pattern Recognition Workshops*. Honolulu, HI, USA. Piscataway, pp.1822-1830. 2017.

Construction of Multi-variety Combination Trading Model for Electricity Market

Dunnan Liu, Zhixin Dong
School of economic and management, North China Electric Power University
Beijing, China
*dongzhixin10@163.com

Abstract

At present, the actual situation of spot transactions in China's power market is relatively complicated, and it is necessary to expand diversified power products to meet different market participants. Therefore, we can introduce the Nordic power market block trading products as a reference to continuously improve our own construction. Combination transactions based on block transactions and their derivatives are aimed at enabling power market participants to choose transaction types that are more in line with their own characteristics and making transactions more flexible and diverse. Based on the current situation of China's high proportion of renewable energy, combined with the Nordic day ahead power trading model, this paper first analyzes the diversified trading needs, and designs flexible and diverse power market trading varieties. Then, we constructs a power spot market clearing model which considers the flexible combination of multiple trading varieties. The calculation example shows that the flexible transaction mechanism has greatly stimulated market vitality, achieving a high transaction rate and low abandonment rate on both the supply side and the demand side. While meeting the power demand of users, it also guarantees the interests of power generation entities with different economic characteristics.

Keywords-renewable energy; diversified power product; Combination transactions;

1. INTRODUCTION

At present, China has included Guangdong and other 8 provinces in the first batch of spot market pilot construction provinces, and it is required to start trial operation before the end of 2018. In order to do better in the construction of the spot market pilot, it is necessary to strengthen the learning and reference of the practical experience of foreign spot markets. Especially in recent years, in the Nordic day-ahead power spot market, flexible block transactions and bidding mechanisms have been introduced, which allows power generation companies and power users to select suitable flexible blocks according to their power generation technical characteristics and actual demand for electricity. Bidding for transactions reflects the willingness of transactions, participates in market competition, and realizes integrated and centralized clearing of various types of transactions, making the power transaction process more autonomous and fair, and achieving the effect of efficient matching of power resource supply and demand.

In view of the above problems, it is particularly necessary to seek innovation in the bidding mechanism and transaction mode, and propose a market transaction mode that adapts to the characteristics of China's new power system. The bidding mechanism is the core issue of the power market, which has been deeply studied and practiced at home and abroad. At present, it mainly includes time-sharing bidding model, segment bidding model, hybrid bidding model and Nordic energy block bidding model.

Literature [1] introduced the trading varieties and bidding mechanism of flexible block trading in the northern European spot power market, allowing power generation enterprises and power users to select suitable flexible block trading for bidding according to their power generation technical characteristics and actual power consumption needs, reflecting the trading willingness, participating in market competition, and realizing integrated and centralized clearing of various trading varieties, so as to make the power trading process more independent and fair, and achieve the effect of efficient matching of power resource supply and demand. On the basis of clarifying the theory and disadvantages of time-sharing bidding, literature [2] proposed the theory and market mechanism of segment bidding for the first time, and compared with time-sharing and segment bidding, proposed that the segment bidding mode was more in line with the operating characteristics of the power market. Literature [3] establishes the corresponding market operation mode and bidding model for the segmented bidding mode. Analysis and demonstration show that the segmented bidding mode greatly

International Conference on Computer Graphics, Artificial Intelligence, and Data Processing (ICCAID 2022)
edited by Ruishi Liang, Jing Wang, Proc. of SPIE Vol. 12604, 126040A
© 2023 SPIE · 0277-786X · doi: 10.1117/12.2674860

reduces the power purchase cost. Literature [4] divides the load curve into three sections based on the segment bidding theory: base load, waist load and peak load, and designs the corresponding bidding mechanism for the three sections. In view of the difficult problem of dividing the waist load and peak load boundaries, the boundary between them is determined through system clearing and optimization. Literature [5] studies the impact of the number of load segments on the total power purchase cost and operation efficiency of the market based on the segment bidding mechanism. The results show that the more load segments, the lower the total power purchase cost of the market, and the higher the market efficiency. Document [6] introduces electricity segment trading into bilateral contracts, which divides electricity into several consecutive segments according to the duration and carries out transactions.

However, there is still a lack of research on the trading mechanism design of multi energy bidding on the same platform. As the actual situation of spot trading in China's power market is relatively complex, it is necessary to expand diversified power products to meet different market participants, it is appropriate to introduce Nordic power market block trading products as a reference to continuously improve our own construction. Therefore, based on the current situation of new power system construction in China and combined with the recent power trading mode in northern Europe, this paper first analyzes the different flexible trading needs of multi market players, and then designs the corresponding power market trading varieties, and constructs the combination mode of multi trading varieties; Finally, the clearing model of multi variety combinations was constructed.

2. LESSONS FROM THE NORDIC DAY-AHEAD TRADING MODEL

The Nordic electricity spot market has designed three different types of trading varieties: hourly trading (horlyorder), block trading (blockorder), and flexible hourly trading (flexible hourly trading). Market members can choose any one of them according to their actual electricity generation and consumption needs. The combination of the three is used for bidding. The transaction allows power generation companies and power users to select suitable flexible block transactions for bidding according to their power generation technical characteristics and actual electricity demand, reflecting the willingness of transactions, participating in market competition; realizing integration of various types of transactions Centralized clearing, making the power transaction process more autonomous and fair, and achieving the effect of efficient matching of power resource supply and demand; flexible block trading allows market members to declare a certain amount of power supply or demand, win bids for 3 hours or more, and provide market members with Multiple choices such as mobile block trading, linked block trading, and extended link block trading, give market members full independent choice.

3. CHINA'S CLEAN ENERGY USER CHARACTERISTICS AND DEMAND ANALYSIS

In the gradually opening electricity market, the types of market entities continue to increase. In addition to traditional power plants, power grids, and large users, today's power market also contains many emerging entities, such as new energy generator sets, power sales companies, load aggregators, virtual power plants, energy storage resources, and electric vehicles. The emergence of emerging entities not only reflects technological progress, but also reflects the perfection of market mechanisms. Therefore, the multiple entities in the market also exhibit the following three characteristics:

(1) More flexibility in power generation output. According to the "China Renewable Energy Development Report 2019" statistics, China's clean energy consumption accounted for 23.4% in 2019. During the "13th Five-Year Plan" period, my country's renewable energy annual growth rate was about 12%, and the average annual share of renewable energy power generation installed capacity has exceeded 50%. More traditional thermal power, mature hydropower, and the application scale of renewable energy continue to expand. The market needs to provide adequate guarantees to deal with the intermittent, volatility, and instability of renewable energy power generation.

(2) Load resources are more controllable. On the load side, with the development of demand response, more and more load-side resources participate in market transactions to achieve peak-shaving and valley-filling of the power system and increase the consumption of new energy. The establishment of the spot market provides conditions for the marketization of demand response. Whether it is emergency demand response or economical demand response, it is a collection of dispersed adjustable loads.

(3) The contrast of cost is obvious. The participation of multiple entities highlights the economic characteristics of different types of resources. Compared with thermal power and hydropower, the cost of new energy power generation is limited by the level of technology and policy support. In an environment where the country vigorously promotes new

Proc. of SPIE Vol. 12604 126040I-81

energy power generation to subsidize and grid-connected at a fair price, photovoltaic and wind turbines need clear price signals to maximize profits with unstable efforts.

In order to adapt to the uncertainty of multiple entities participating in market transactions, give full play to the advantages of different entities, and seize price signals, multiple entities have created new demands for market mechanisms. Different types of entities have different economic characteristics. Under a fixed market environment, it is impossible for multiple entities to effectively choose the optimal operation strategy. Traditional trading varieties limit the scope for multiple entities to play. Take Shanghai demand response as an example. In 2020, the State Grid Shanghai Electric Power Company will conduct demand response transactions. In addition to basic peak-shaving and valley-filling responses, new medium- and long-term peak-shaving, valley-filling, intra-day peak-shaving, valley filling and fast-tracking are added. There are 6 types of peak cutting and valley filling. For the first time in the bidding, three types of call methods, namely switch type, step type, and curve type, which are closer to the actual demand of the power grid, are set up for the first time. Customers can independently choose the method to be called according to their own energy use characteristics, so that the actual response load is more accurate. It can be seen that diversified trading varieties can help market entities make more rational choices.

4. DESIGN OF ELECTRICITY MARKET TRADING VARIETIES

In order to deal with the uncertainty of multiple entities, combined with the experience of the Nordic power market, the trading varieties are designed from the following four perspectives: (1) It is related to time. Link trading varieties to time, with 1h as the unit, perform hourly transactions and flexible hourly transactions, and market entities declare electricity and prices. (2) It is related to the output. The market entity declares the volume-price combination, and determines the time of the entity's output when the market clears. (3) It is related to time and output. The market entity declares the quantity-price combination within a specific time period to form an "energy block", and the output period for clearing out coincides with or is completely the same as the declared period. The following table summarizes the design of trading varieties that meet the needs of multiple entities. Market entities can choose one or more combinations to participate in the market according to their own needs.

Among them, block transactions, regular block transactions, and flexible block transactions need to be marked with supply and demand attributes. The supply block needs to provide a minimum price, and the demand block needs to provide a maximum price. The transaction time is more than 3h (including 3h). Block transactions indicate the volume of the transaction, the period of the transaction (which can be a discontinuous period), and the minimum transaction rate of the transaction. The standard format is shown in Table 1, and the values in the table are examples. Market entities can refer to this template when declaring block transaction orders.

TABLE 1 THE DESIGN OF TRADING VARIETIES ADAPTED TO THE NEEDS OF MULTIPLE ENTITIES

Design direction	Transaction type	Transaction type	Bidding method	Applicability
Related to time	Hourly transaction	In hours, market entities declare 24h electricity and price	Independent quotation for each period, no hourly coupling constraints	Independent quotation for each period, no hourly coupling constraints
	Flexible hourly transaction	In hours, market entities declare acceptable output time, electricity and price,	bidding at most once per hour	Market entities cannot predict the output curve for a day, but they can predict the output curve for certain periods of time
Related to output	block transaction	The market entity declares the electricity and price without a specific time scale (the default time length is 3h)	All the declared electricity has won the bid or failed the bid, and the output time will be determined after the cleared out	Applicable to market entities that do not have requirements for the time of output
Related to time and output	Regular block transaction	Market entities declare electricity and price with time scale (≥3h)	All the declared electricity quantities have won the bid, or all of them have not won the bid	Applicable to market entities who are confident about the power of 3h and above

Proc. of SPIE Vol. 12604 126040A-3

Flexible block transaction	Market entities combine multiple (≥3) conventional blocks and set a minimum acceptance rate	When the clearing result is greater than the minimum acceptance rate, the flexible block transaction is completed	Suitable for market entities whose output cannot be accurate to the hour

5. CONSTRUCTION OF AN ELECTRICITY SPOT MARKET CLEARING MODEL CONSIDERING THE FLEXIBLE COMBINATION OF MULTIPLE TRADING VARIETIES

5.1. Market transaction mechanism design

The time-length transaction mechanism constructed in this paper to adapt to the interaction of source, network, load and storage adopts the organizational method of "two-way quotation, centralized bidding, unified clearing, and marginal pricing". The specific bidding mechanism is: power generators, e-commerce sellers and large users can choose hourly transactions, various types of energy block transactions, etc. for bidding declarations. Based on the consideration of grid security constraints and the goal of maximizing social welfare, the Electric Power Exchange conducts unified market optimization and clearing, and determines the hourly bidding power of each market member, the bidding status of block transactions, and the hourly market marginal output. Clear price. When the transmission section of each price range is blocked, the zone price can be formed.

5.2. Clearing model building

Market clearing takes the maximization of social welfare as the trading goal, including hourly trading and block trading. The objective function is shown in formula (1).

$$\begin{aligned} \max = &\sum (I^{h}_{g,t} \cdot Q^{h}_{g,t} \cdot P^{h}_{g,t} + I^{lh}_{g,t} \cdot Q^{lh}_{g,t} \cdot P^{lh}_{g,t} + I^{b}_{g,t} Q^{b}_{g,t} \cdot P^{b}_{g,t} + \\ &I^{cb}_{g,t} \cdot Q^{cb}_{g,t} \cdot P^{cb}_{g,t} + I^{lb}_{g,t} \cdot Q^{lb}_{g,t} \cdot P^{lb}_{g,t}) + \\ &\sum (I^{h}_{b,t} \cdot Q^{h}_{b,t} \cdot P^{h}_{b,t} + I^{lh}_{b,t} \cdot Q^{lh}_{b,t} \cdot P^{lh}_{b,t} + I^{b}_{b,t} \cdot Q^{b}_{b,t} \cdot P^{b}_{b,t} + \\ &I^{cb}_{b,t} \cdot Q^{cb}_{b,t} \cdot P^{cb}_{b,t} + I^{lb}_{b,t} \cdot Q^{lb}_{b,t} \cdot P^{lb}_{b,t}) \end{aligned} \tag{1}$$

The constraints of this model mainly include power and electricity balance constraints, and the number of trade clearing constraints. The constraint condition of power and electricity balance is shown in formula (2).

$$\begin{aligned} &\sum (Q^{h}_{g,t} + Q^{lh}_{g,t} + Q^{b}_{g,t} + Q^{cb}_{g,t} + Q^{lb}_{g,t} + Q^{u}_{g,t} \\ &+ Q^{h}_{b,t} + Q^{lh}_{b,t} + Q^{b}_{b,t} + Q^{cb}_{b,t} + Q^{lb}_{b,t} + Q^{u}_{b,t}) = 0 \end{aligned} \tag{2}$$

Define the set of winning bids as, then the restriction on the number of clearing transactions is shown in formula (3).

$$0 \le I_i \le 1, \forall i \tag{3}$$

The flexible hourly transaction clearing schedule is shown in formula (4), where it is the set of hours acceptable to the subject.

$$\sum I^{lh}_{i} = 1, i \in \{H \mid h_1, h_2, h_3, \ldots\} \tag{4}$$

After the trading center obtains the flexible quotation from the market members, it classifies the orders according to the supply side and the demand side, and sorts the orders within the same time period according to the sorting rules. Then the trading center simulates market transactions with the goal of maximizing social welfare, and transactions are constrained by the number of clearings, the overall bid-winning rate, the balance of electricity and electricity, and prices. Finally, the trading center releases the trading results that meet the 24 time periods. When the clearing result does not meet the constraints, the calculation time of the clearing result is limited, and the last calculation result shall prevail.

6. CASE ANALYSIS

Assuming that the market quotation is a single-segment quotation, there are 5 power generation entities and 5 household electricity entities in the market.The specific clearance results are shown in figure 1-2.

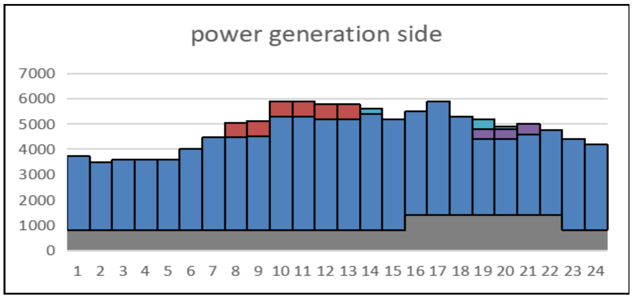


Figure 1. The result of Power generation side

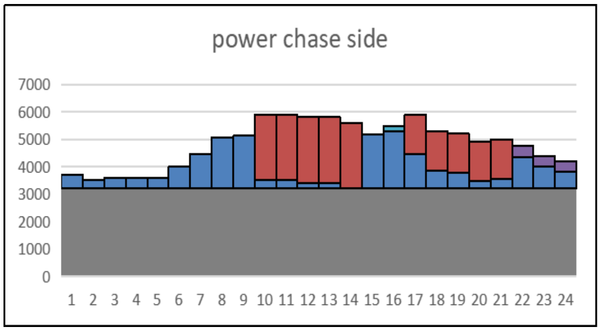


Figure 2. The result of Power chase side

7. CONCLUSION

Based on the current situation of China's high proportion of renewable energy, and combined with the recent power trading mode in northern Europe, this paper firstly analyzes the diversified trading needs. Secondly, it designs flexible and diverse trading varieties in the power market, and then constructs a clearing model of the power spot market considering the flexible combination of multiple trading varieties. Finally, through the analysis of examples, it is shown that the flexible trading mechanism has greatly stimulated the market vitality, achieved a high transaction rate and a low abandonment rate on the supply side and the demand side, and guaranteed the interests of power generation entities with different economic characteristics while meeting the electricity demand of users. At the same time, in view of the current situation of China's dual track system and the fact that some electricity can not be traded when the current model is cleared, the following aspects can be studied in depth.

(1)According to different trading varieties, combined with the current situation of China's dual track system and the contract priority in the actual trading process, the transaction sequence of different trading varieties is designed to achieve accurate positioning of different varieties and reflect the value of different trading varieties;

(2)Aiming at the flexible energy block clearing mode, a matching power and electricity balance mechanism should be designed to ensure the real-time balance of power on both sides of the power generation and consumption, and the safe and stable operation of the system.

Acknowledgment

The Guangzhou Trading Center Science and Technology Project "Research on flexible configuration mode of power mid and long-term trading business functions".

References

[1] Z. Xinyu, C. Qixin, G. Rui, et al, “Clearing model of electricity spot market considering flexible block orders,” J. Automation of Electric Power Systems, 2017, 41(24) : 35-41

[2] W. Xifan, “Block bidding model based power market,” J．Proceedings of the CSEE，2001(12)：2-7

[3] G. Jian，W. Xifan，D. Xiaoying，et al, “Models of block bidding in power market and comparisons with hourly bidding,” J．Proceedings of the CSEE，2003(9)：22-27

[4] X. Ming，L. Linchuan，S. Wei，et al, “Realization of clearing algorithm for segmented bidding power market,” J．Automation of Electric Power Systems，2003(23)：12-16

[5] G. Jian，W. Xifan，C. Haoyong，et al, “Simulation, analysis and comparison of segmented bidding power market,” J．Journal of Xi'an Jiaotong University，2003(10)：1043-1047

[6] Z. Xian，W. Xifan，W. Jianxue，et al, “The application of segmented transaction in the electricity bilateral contract market,” J．Automation of Electric Power Systems，2004(11)：13-16

Research on Evaluation of Equipment Support Capability of Land Aviation Brigade Based on FAHP

Huining Gu*, Yaolong Zhang, Hongli Jia and Xianmin Shi
Department of Equipment Command and Management, Army Engineering University, Shijiazhuang 050003, China
Email: *441751312@qq.com

Abstract

To test to evaluate the aviation brigade combat equipment support ability, based on the combat equipment safeguard drill results in recent years, with the relevant regulations standard, through consulting literature, investigation and other methods, constructed the aviation brigade combat equipment support ability evaluation index system, and through the AHP to determine evaluation index weight of every layer, The fuzzy comprehensive evaluation method is used to calculate the evaluation results, and finally a more scientific and comprehensive evaluation is made for the actual combat equipment support capability of the army aviation brigade.

Keywords: Land navigation brigade; Equipment support capability; Fuzzy comprehensive evaluation; AHP

1. Introduction

With the rapid development of the Army aviation force of the PLA, the number of helicopters has increased sharply, models have been constantly updated. The generation and development of equipment support capability of the Army aviation force is facing more and more challenges. For the army aviation brigade, equipment support capacity is an important factor to improve combat effectiveness, is a key supporting element. How to improve the equipment support capability of the Army aviation force is a realistic problem urgently to be studied and solved.

As a basic work, the assessment of actual combat equipment support capability plays a directional and guiding role in the process of demonstration of equipment support system construction requirements and scientific decision-making. Carrying out the assessment of actual combat equipment support capability plays an important role in promoting the equipment support capacity construction of the army aviation force [1].

In this paper, based on the combat equipment support throughout the years practice activities, combined with the relevant rules and regulations standard, making use of the analytic hierarchy process (AHP) construct the aviation brigade combat equipment support ability evaluation index system [2], and the fuzzy analytic hierarchy process(FAHP) related quantitative study of the combat equipment support ability evaluation for the aviation units to provide the reference.

2. Overview of Assessment Methods

2.1. Analytic Hierarchy Process (AHP)

AHP is a systematic analysis method combining qualitative analysis and quantitative analysis proposed by Professor T.L.Saaty, a famous American operational research scientist. It transforms a complex problem into a hierarchy of target layer, criterion layer and scheme layer, and then compares factors at the same level in pairs by assigning values, and then conducts quantitative and qualitative analysis [3]. Yaahp software is generally used for weight calculation.

2.2. Fuzzy comprehensive evaluation (Fuzzy)

Fuzzy is a kind of based on the theory of Fuzzy comprehensive evaluation method [4], is used to solve the problem of vague, difficult to quantitative analysis, this method can take the related evaluation index according to the Fuzzy evaluation standard to quantitatively, then Fuzzy mathematics model is set up, on the basis of the quantitative value, calculate the relevant index factors It is mainly used for comprehensive evaluation of evaluation objects with complex factors and multiple levels.

Considering the multi-level and complexity of the evaluation factors of the actual combat equipment support capability of the army aviation brigade, this paper adopts the analytic hierarchy process to determine the evaluation index system and

International Conference on Computer Graphics, Artificial Intelligence, and Data Processing (ICCAID 2022)
edited by Ruishi Liang, Jing Wang, Proc. of SPIE Vol. 12604, 126040B
© 2023 SPIE · 0277-786X · doi: 10.1117/12.2674713

weight vector, calculates the evaluation results of each layer of the index system by fuzzy comprehensive evaluation, and finally obtains the evaluation results of actual combat equipment support capability [5].

3. Establishment of the evaluation model for the actual combat equipment support capability of the Army Aviation Brigade

3.1. Establish an evaluation index system

It is a very complex system engineering to construct the evaluation index system of the equipment support capability of the army aviation brigade for actual combat. It needs to analyze the components and main contents of the equipment support of the army aviation brigade in multiple channels and aspects [6], and grasp the internal structure and operation process of the equipment support system of the army aviation brigade under actual combat conditions. Through the investigation of the Army Aviation force, this paper analyzes the problems faced by the actual combat equipment support work of the Army Aviation force at the present stage. In view of the performance characteristics of the army aviation helicopter, based on the references, The evaluation index system of actual combat equipment support capability of the Army Aviation Force is constructed from four aspects of command and control capability, field four stations and maintenance support capability, supply support capability and data support capability [7], as shown in Figure 1.

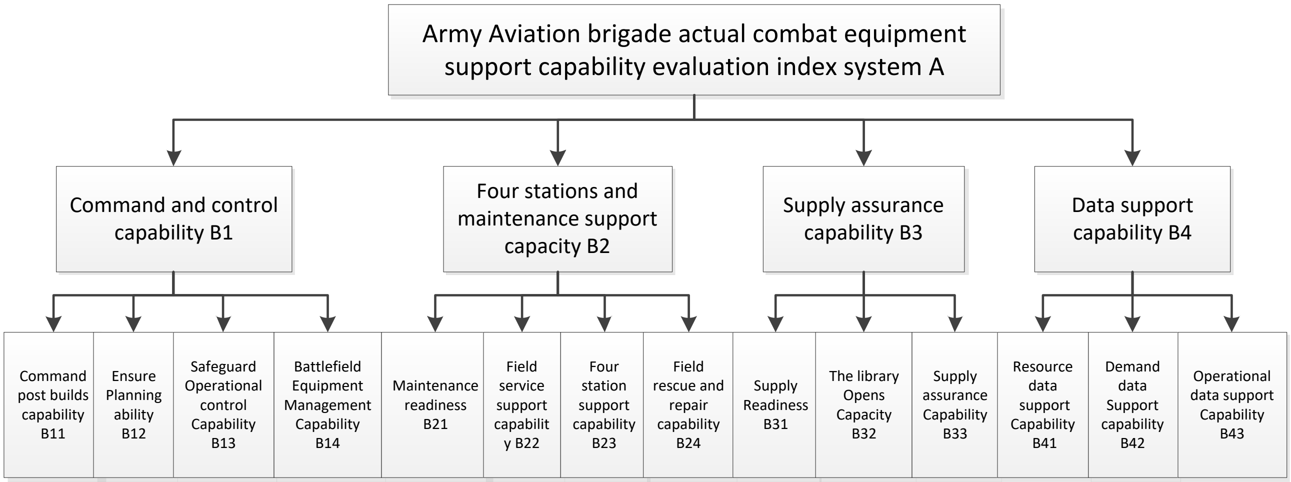

Army Aviation brigade actual combat equipment support capability evaluation index system A
Command and control capability B1
Four stations and maintenance support capacity B2
Supply assurance capability B3
Data support capability B4
Command post builds capability B11
Ensure Planning ability B12
Safeguard Operational control Capability B13
Battlefield Equipment Management Capability B14
Maintenance readiness B21
Field service support capability B22
Four station support capability B23
Field rescue and repair capability B24
Supply Readiness B31
The library Opens Capacity B32
Supply assurance Capability B33
Resource data support Capability B41
Demand data Support capability B42
Operational data support Capability B43

Figure 1: Actual combat equipment support capability evaluation index system of Army Aviation Brigade

3.2. Establishment of factor set and comment set [8]

Establish a set of assessment factors U={U_1,U_2,…,U_m}, On the premise of taking all factors into consideration, AHP is used to divide attribute relations and realize multi-level fuzzy comprehensive evaluation. The determined factor set is the criterion layer and index layer of the evaluation index system.

The establishment of comment set is to delimit the evaluation results of the evaluation object into a certain level. A collection of these grades V={V_1, V_2,…,V_4}.This paper determines that the evaluation level of the actual combat equipment support capability evaluation of the Army Aviation Brigade is level 4:V_1 to Excellent,V_2 to Benign,V_3 to Marginal,V_4 to Flunk. It's called a set of fuzzy comments V={V_1,V_2,V_3,V_4}.

3.3. Determination of the weight of evaluation indicators

In the evaluation indexes, due to the different degree of influence on equipment support capability, each factor index occupies different proportions in the overall evaluation. We give corresponding weight coefficient W_i to each factor U_i to feedback its importance. In this paper, analytic hierarchy process is adopted to determine the weight of indicators [9].

Step 1: Building the judgment matrix

The judgment matrix represents the relative importance of an element at the upper level determined by pairwise comparison between the relevant elements at the same level. If the judgment matrix is $[a_{ij}]_{n*n}$, a_{ij} represents the numerical representation of the relative importance of a_i at this level compared with a_j for upper element usually 1-9 and their reciprocal [10].

Blements B1,B2... ,Bn is compared with the element B of the upper layer, and the judgment matrix can be obtained as follows:

$$A=\begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ a_{31} & a_{32} & \cdots & a_{3n} \\ a_{41} & a_{42} & \cdots & a_{4n} \end{bmatrix}, a_{ij} > 0, a_{ij} = 1/a_{ji}, i, j = 1, 2, \cdots, n$$

Step 2: Weight calculation and consistency test of judgment matrix

First, the maximum eigenroot and corresponding eigenvector of the judgment matrix are calculated (yaahp software calculation). Then, consistency index CI, random consistency index RI and consistency ratio CR were used for consistency test. If it passes, the feature vector (after normalization) is the weight vector; If not, the pairwise comparison matrix is reconstructed.

1) The maximum eigenroots and their eigenvectors are calculated

Here we use the normalized summation. The calculation steps are as follows:

(1) Formalize each column of the judgment matrix, i.e:

$$\bar{a}_{ij} = a_{ij} / \sum_{(k=1)}^{n} \bar{a}_{ij}, i, j = 1, 2, \cdots, n \tag{1}$$

(2) Add the processed judgment matrix by row, i.e:

$$\bar{w}_i = \sum_{(j=1)}^{n} \bar{a}_{ij}, j = 1, 2, \cdots, n \tag{2}$$

(3) Normalize the vector$\bar{w}_i = (\bar{w}_1, \bar{w}_2, \cdots, \bar{w}_n)^T$, i.e:

$$w_i = \bar{w}_i / \sum_{(j=1)}^{n} \bar{w}_j, i = 1, 2, \cdots n \tag{3}$$

The vector$w = (w_1, w_2, \cdots, w_n)^T$ is the eigenvector.

(4) The maximum characteristic root is：

$$\lambda_{\max} = \sum \frac{(Bw)_i}{nw_i} \tag{4}$$

In the Green band $(B_w)_i$ is the ith element of the vector Bw.

2) Consistency test of judgment matrix

(1) Consistency index CI, defined as

$$CI = \frac{\lambda_{\max} - n}{n-1} \tag{5}$$

When CI = 0, $\lambda_{\max} = n$, The judgment matrix has complete consistency.

(2) Average random consistency index RI, whose value is given in Table 1.

Table 1 Average random consistency index values

dimension	3	4	5	6	7	8	9
RI	0.58	0.90	1.12	1.24	1.32	1.41	1.45

(3) Random consistency index CR, defined as：$CR = CI / RI$

General requirements for consistency testing. If the consistency test results do not meet the requirements, it is necessary to check whether the relationship values of each element in the judgment matrix are set reasonably, and then adjust accordingly. If the judgment matrix passes the consistency test, the eigenvector W is determined to be the weight of this level.

3.4. Calculation of fuzzy comprehensive Evaluation Results [11]

Step 1: Establish the fuzzy evaluation matrix

First of all, evaluators evaluate the rating standard of the evaluation object factors to determine the degree of membership of a single factor to each element in the evaluation set and form a fuzzy judgment matrix, namely:

$$R = \begin{bmatrix} r_{11} & r_{12} & \cdots & r_{1j} \\ r_{21} & r_{22} & \cdots & r_{2j} \\ \vdots & \vdots & \ddots & \vdots \\ r_{i1} & r_{i2} & \cdots & r_{ij} \end{bmatrix}$$

Among them, R_{ij} is the membership degree of FACTOR U_i rated as V_j, and check whether it meets the requirement $\sum_{j=1}^{n} R_{ij} = 1$. If not, normalization is required. R_{ij} was determined by expert evaluation method.

Step 2: Calculate the comprehensive evaluation results

1) Establishment of fuzzy judgment set:

$$B = A \circ R = (B_1, B_2, \cdots, B_n) \tag{6}$$

Thereinto $B_i = \sum_{i=1}^{n}(w_i r_{ij}), j = 1,2,\cdots,n$。 If $\sum_{j=1}^{n} B_j \neq 1$，let's do normalization.

2) Determine the set of fractions$C = (C_1, C_2, \cdots, C_n)^T$， Thereinto $C_j(j \in [1, n])$ represents the score of grade J. The evaluation criteria were determined as follows: If Cn>Cn critical，Is considered to have reached the corresponding level. If the score is 85, A good score of Cn critical is 80, the evaluation result of this capability index is considered to be good.

3) Calculate the evaluation result：S=B・C

S value is the final evaluation score of the evaluation object, and then it is placed in the evaluation set to obtain the corresponding evaluation grade, which is the evaluation result.

4. Case Analysis of actual Combat Equipment support Capability of The Army Aviation Brigade [12]

An army aviation brigade participated in a real-combat equipment support drill. It was evaluated by FAHP method.

4.1. Determine the weight set

The hierarchical structure model of army aviation brigade's actual combat equipment support capability evaluation briefly describes the factors related to the actual combat equipment support capability of army aviation brigade and their relationship with each other. However, due to the different degree of influence on equipment support capability, each factor index occupies different proportion in the overall evaluation, so we construct the judgment matrix through pairwise comparison.

According to the 1-9 scale criterion of analytic hierarchy process, the pair-wise comparative judgment matrices of relevant factors at each level are given. Firstly, the comparative judgment matrix of the first-level index to the target layer is

$$A = \begin{bmatrix} 1 & \frac{1}{2} & \frac{1}{2} & 3 \\ 2 & 1 & 1 & 2 \\ 2 & 1 & 1 & 2 \\ \frac{1}{3} & \frac{1}{2} & \frac{1}{2} & 1 \end{bmatrix}$$

According to the formula, its maximum eigenvalue and corresponding eigenvector are respectively

$$\lambda_{\max} = 4.0104, w^{(1)} = (w_1^{(1)}, w_2^{(1)}, w_3^{(1)}, w_4^{(1)})^T = (0.2269, 0.3063, 0.3220, 0.1447)^T$$

The corresponding consistency indicator is $CI = 0.0522, RI = 0.90, CR = 0.0580$

That is, pass the consistency test. Therefore, the matrix meets the consistency, and the weight of evaluation index of equipment support capability of criterion layer is: $W = (0.2269,0.3063,0.3220,0.1447)$

Secondly, the judgment matrix of the secondary index, the calculation of consistency test and the determination of combined weight vector are determined. The pair comparison judgment matrices of the four indexes of the sub-criterion layer to the criterion layer are as follows:

$$B_1=\begin{bmatrix}1 & \frac{1}{2} & \frac{1}{3} & \frac{1}{2}\\ 2 & 1 & 2 & 3\\ 3 & \frac{1}{2} & 1 & 2\\ 2 & \frac{1}{3} & \frac{1}{2} & 1\end{bmatrix},\quad B_2=\begin{bmatrix}1 & 2 & 3 & \frac{1}{5}\\ \frac{1}{2} & 1 & \frac{1}{2} & \frac{1}{4}\\ \frac{1}{3} & 2 & 1 & \frac{1}{3}\\ 5 & 4 & 3 & 1\end{bmatrix},\quad B_3=\begin{bmatrix}1 & 5 & \frac{1}{2}\\ \frac{1}{5} & 1 & \frac{1}{6}\\ 2 & 6 & 1\end{bmatrix},\quad B_4=\begin{bmatrix}1 & \frac{1}{2} & \frac{1}{3}\\ 2 & 1 & 2\\ 3 & \frac{1}{2} & 1\end{bmatrix}$$

Then, the maximum eigenvalue and the corresponding eigenvector are calculated respectively, and the consistency test is performed, then:

$$\lambda_{1\max}=4.0200, w_1^{(2)}=(w_{11}^{(2)},w_{12}^{(2)},w_{13}^{(2)},w_{14}^{(2)})^T=(0.2021,0.3016,0.2729,0.2234)^T$$
$$CI(1)=0.0075, RI(1)=0.90, CR(1)=0.0833$$

That is, matrix B1 passes the consistency test.

$$\lambda_{2\max}=4.0200, w_2^{(3)}=(w_{21}^{(3)},w_{22}^{(3)},w_{23}^{(3)},w_{24}^{(3)})^T=(0.1533,0.3412,0.2528,0.2528)^T$$
$$CI(2)=0.0075, RI(2)=0.90, CR(2)=0.0833$$

That is, matrix B2 passes the consistency test.

$$\lambda_{3\max}=3.0000, w_3^{(4)}=(w_{31}^{(4)},w_{32}^{(4)},w_{33}^{(4)})^T=(0.3744,0.1682,0.4573)^T$$
$$CI(3)=0.0000, RI(3)=0.58, CR(3)=0.0000$$

That is, matrix B_3 passes the consistency test.

$$\lambda_{4\max}=3.0178, w_4^{(5)}=(w_{41}^{(5)},w_{42}^{(5)},w_{43}^{(5)})^T=(0.2703,0.3771,0.3528)^T$$
$$CI(4)=0.0171, RI(4)=0.58, CR(4)=0.0295$$

That is, matrix B_4 passes the consistency test.

A questionnaire was sent to the experts, who were invited to directly give the weight of each factor by referring to the evaluation set of the evaluation index system of the army Aviation Brigade's actual combat equipment support capability, and then calculate the arithmetic average of multiple weights of a factor and determine the corresponding membership vector. See Table 2.

Table 2 Evaluation factor system of actual combat equipment support capability

Of Army Aviation Brigade

criterion layer		index level		fuzzy evaluation matrix			
Risk Evaluation Factors	Weightiness	Risk Evaluation Factors	Weightiness	Excellent	Benign	Marginal	Flunk
Command and control capability	0.2269	Command post builds capability	0.2021	0.7	0.3	0	0
		Support planning capability	0.3016	0.6	0.2	0.2	0
		Ensure operational control capability	0.2729	0.5	0.3	0.1	0.1
		Battlefield equipment management capability	0.2234	0.6	0.2	0.1	0.1

Four stations and maintenance support capacity	0.3063	Maintenance readiness	0.1533	0.7	0.2	0.1	0
		Site support capability	0.3412	0.4	.03	0.2	0.1
		Four station support capacity	0.2528	0.5	0.2	0.2	0.1
		Field rescue and repair ability	0.2528	0.4	0.3	0.2	0.1
Supply assurance capability	0.3220	Supply readiness	0.3744	0.6	0.3	0.1	0
		Ability to open a library	0.1682	0.7	0.3	0	0
		Supply assurance capability	0.4573	0.4	0.4	0.2	0
Data support capability	0.1447	Resource data support ability	0.2703	0.8	0.1	0.1	0
		Ability to support demand data	0.3771	0.7	0.2	0.1	0
		Ability to support action data	0.3528	0.6	0.3	0.1	0

4.2. Calculation of fuzzy evaluation results

1) Calculate the fuzzy evaluation matrix R

According to Formula (1), the membership vector of each criterion is calculated and normalized:

$$B_1 = W_1 \circ R_1 = (0.2021, 0.3016, 0.2729, 0.2234) \circ \begin{pmatrix} 0.7 & 0.3 & 0 & 0 \\ 0.6 & 0.2 & 0.2 & 0 \\ 0.5 & 0.3 & 0.1 & 0.1 \\ 0.6 & 0.2 & 0.1 & 0.1 \end{pmatrix} = (0.59, 0.25, 0.11, 0.05)$$

And that's the same thing B_2、B_3、B_4， Thus, the membership matrix of criterion layer and target layer can be obtained:

$$R = \begin{pmatrix} B_1 \\ B_2 \\ B_3 \\ B_4 \end{pmatrix} = \begin{pmatrix} 0.59 & 0.25 & 0.11 & 0.05 \\ 0.36 & 0.20 & 0.14 & 0.30 \\ 0.53 & 0.34 & 0.13 & 0 \\ 0.69 & 0.21 & 0.10 & 0 \end{pmatrix}$$

2) Work out the membership vector of the target layer and normalize it:

$$B = W \circ R = (0.2269, 0.3063, 0.3220, 0.1447) \circ \begin{pmatrix} 0.59 & 0.25 & 0.11 & 0.05 \\ 0.36 & 0.20 & 0.14 & 0.30 \\ 0.53 & 0.34 & 0.13 & 0 \\ 0.69 & 0.21 & 0.10 & 0 \end{pmatrix} = (0.52, 0.26, 0.12, 0.10)$$

Calculation of evaluation results

Compare with the scoring table in Table 3 to obtain the evaluation score S of actual combat equipment support capability of the army Aviation brigade.

Table 3 Evaluation table of actual combat equipment support capability of

The Army Aviation Brigade

grade	excellent	benign	marginal	flunk

score C	90	80	70	60

$$S = B \circ C = 0.52 \times 90 + 0.26 \times 80 + 0.12 \times 60 + 0.10 \times 40 = 78.8$$

4) Determine the assessment level

According to the evaluation score obtained in the previous step and the classification of the evaluation grade, it is determined that the evaluation result of the actual combat equipment support ability of an Army Aviation brigade is above passing grade, close to good.

5. Conclusion

In this paper, based on FAHP evaluation method, the evaluation index system of the actual combat equipment support ability of the army aviation brigade is constructed. The weight and score of the index are determined by the expert evaluation method, and the model is used for quantitative analysis and calculation, so as to reduce the subjectivity of the expert scoring and improve the credibility and objectivity. Through the example analysis, the evaluation results of this method can accurately feedback the actual combat equipment support capability of the army aviation brigade, and the results have certain reference significance.

References

[1] Wang Peng, Song Huawen, Chen Xiangbin, Li Shu et al. Main practice, development trend and enlightenment of foreign military equipment support actual combat training [J]. Journal of equipment academy, 2015,26 (4) : 43-47.

[2] Qiu Wei, Zhang Zenglei, Tian Wenxiang, Fan Jun. Ability evaluation of equipment support personnel based on analytic hierarchy process and fuzzy comprehensive evaluation [J], journal of ordnance equipment engineering, 2018,39 (4): 108-113.

[3] Hou Bao 'e, Tian Hengdou, Gao Yang. Fuzzy-ahp based safety evaluation method for shipborne anti-torpedo weapon system [J]. Fire control & command control, 2019,44 (11) : 102-106.

[4] LIU Tao. Risk assessment of goaf based on AHP-Fuzzy method [J]. Nonferrous Metals (Mine), 2016 (5) : 49-52.

[5] ZHAO Wuzhou, Hu Hongwei. Research on index evaluation of energy-saving housing based on AHP method [M], Building Materials World, 2009.30 (5), 86-92.

[6] Yan Zhiteng, Dai Rong, Ma Haodong. Computer measurement and control, 2019,27 (9) : 267-271,287.

[7] Xiao Fei, TIAN Yanni, YU Kaimin. Command control & simulation, 2015,37 (1) : 85-88.

[8] Miao Qiguang, Liu Juan, Ning Shuting. Evaluation of Airfield Strike Effect based on Fuzzy comprehensive Evaluation [J], Systems Engineering and Electronics, 2012 34 (7) :1395-1399.

[9] Shi Quan, Wang Lixin, Shi Xianming, Zhao Mei. System Decision and Modeling [M], Beijing: National Defense Industry Press, 2016.7, 108-114.

[10] Xie Jingci. Evaluation model of Shandong port logistics comprehensive capacity based on Fuzzy-AHP [J]. Logistics Science and Technology, 2014 (7) : 13-16,31.

[11] Zhu Yongsong. Application of analytic Hierarchy Process in multi-objective investment decision [J]. Science and Technology Entrepreneurship Monthly, 2004 (12) : 144-146,148.

[12] Su Xujun, Chen Jiansi. Fuzzy Comprehensive Evaluation of Equipment Development Risk Assessment [J], Fire Control and Command Control, 2013 38 (4), 118-120,124.

Prediction of highway pavement performance based on combined model

Guangli Ren *[a], Peng Zhang [b], Yixin Cui [c]
[a]China Highway Engineering Consulting Group Co., Ltd., CHECC DATA CO., LTD.,
[b]China Highway Engineering Consulting Group Co., Ltd., CHECC DATA CO., LTD.,
[c]China Highway Engineering Consulting Group Co., Ltd., CHECC DATA CO., LTD.,
*renguangli123@163.com;

ABSTRACT

This article focuses on the analysis of the performance of asphalt pavement, and establishes a prediction model for the performance of asphalt pavement in Guidu based on the data of traffic volume, climate, and road surface smoothness. After fully understanding the performance evaluation indicators of various asphalt pavements, the International Roughness Index is selected as the performance evaluation indicators of asphalt pavements. According to the technical performance of the asphalt mixture (high temperature stability, low temperature crack resistance, water stability, anti-fatigue performance, anti-aging performance, etc.), the influencing factors (temperature, rainfall days, traffic volume, etc.) of the asphalt pavement performance are derived. Collect traffic flow and asphalt pavement performance data on the spot, and process and analyze the data. Establish a gray forecast model, a moving average forecast model, as well as a multiple regression forecast model and a VAR model that consider the four variables of traffic volume, truck ratio, temperature, and rainfall days. Predict the performance of the asphalt pavement through the above model, and get the prediction result.

Keywords: Highway transportation, Asphalt pavement performance, RQI, VAR model, Combined model

1. INTRODUCTION

The operation of a highway after construction is very important. The high volume of vehicles passing through the asphalt pavement every day will inevitably have an impact on its performance, predicting pavement performance and making timely maintenance to maintain the safety and smoothness of highway use, so pavement performance prediction is very important and will receive more and more attention. Accurate prediction of asphalt pavement performance can help road management and maintenance departments to carry out short and medium-term planning, promote the scientific determination of pavement maintenance timing, so that maintenance work is more regular and targeted. It can maintain the high level of asphalt pavement performance as much as possible and minimize the impact on traffic operation, so as to use the limited funds more rationally and reduce the waste of human and material resources, and at the same time make the asphalt pavement performance greatly improved and increase the service life of the pavement.

In 1992, Johnson and Cation classified pavements into structural and nonstructural deterioration, analyzed the predictability of structural deterioration, smoothness, and other indicators, and based on this, proposed a pavement life cycle prediction model to predict pavement condition by virtue of rutting and cracking of asphalt pavements, which was successfully applied in South Dakota, USA [1]. In 2016, Kong studied the pavement inspection data of eight typical highways in Beijing. The prediction models of each individual index in the form of indices as well as the composite index were established using statistical methods. Among them, for the pavement exercise quality index and the comprehensive pavement index, separate prediction models were also developed based on whether normal maintenance was performed or not. These prediction models have shown good accuracy and practicality in their applications [2]. In addition, there are researchers who have developed polynomial models [3] and logistic regression models [4] for pavement usage performance prediction. With the development of mathematics, computers, and other related disciplines, methods such as gray theory [5-6], mixed-effects models [7], and various time-series models [8-9] have also been used in the prediction of pavement use performance. In 1991, Liu Boying analyzed the Beijing road network data and decided to use a probabilistic prediction model with pavement damage (measured by PCI), ride quality (measured by RQI), and major damage types as predictor variables for short- and medium-term pavement performance. However, each single prediction model has defects, and in order to reduce the influence of defects on the prediction results, a combined prediction model can be established.In 2011, Jianke Luo established a combined prediction model of gray prediction model and BP neural

International Conference on Computer Graphics, Artificial Intelligence, and Data Processing (ICCAID 2022)
edited by Ruishi Liang, Jing Wang, Proc. of SPIE Vol. 12604, 126040C

network prediction model [10]; in 2015, Lan Zhou established a combined prediction model of gray prediction model and multiple regression prediction model [11]; in 2015, Suna Hu established a combined prediction model of gray prediction and Markov prediction model [12]; in 2019, Yuan Zufeng used gray theory and linear regression to analyze the asphalt pavement usage performance of highways in Anhui Province, and established a combined prediction model by determining their respective weights with the entropy value method. Among them, the accumulated traffic load ESAL was used as the independent variable in the regression prediction model analysis, and the asphalt pavement service performance indexes PCI, RQI and RD were used as the dependent variables, respectively. According to the accuracy test results, the combined prediction model was found to have better prediction results [13].

Selecting a suitable single prediction model to establish a combined model can build on the strengths and avoid the weaknesses, adapt to the prediction needs in different situations such as lack of data and many influencing factors, and provide ideas for applications under similar conditions. Considering that most asphalt pavement performance predictions are made in years or months, and the asphalt pavement performance of roads changes rapidly and may reach a high degree of damage in a short period of time, a combination of prediction models with more accurate short-term predictions is selected to strengthen the short-term prediction capability of asphalt pavement performance and make the prediction more flexible and accurate.

1 Asphalt pavement performance factors and evaluation indicators

1.1 Evaluation index

According to the Technical Specification for Maintenance of Asphalt Pavement (JTJ 073.2-2001), the common diseases of asphalt pavement include Cracks, Crowding, Subsidence, Rutting, Rubbing and Waves, Frost heave and slurry, Potholes, Sagging and Loose, Oiling, Peeling, Edge Gnawing, Polishing.

The road surface ride quality index RQI is used to evaluate the comfortable performance of vehicle driving. There is a quantitative relationship between road surface smoothness and driving comfort, so RQI can be calculated from IRI as follows formula 1.

$$RQI = \frac{100}{1 + a_0 e^{a_1 IRI}} \tag{1}$$

IRI is the cumulative vertical displacement value of a quarter car at a speed of 80km/h as the IRI value, IRI stands for International Flatness Index. a_0 . a_1 represents the model parameters. a_0 =0.026, a_1 = 0.65.

1.2 Influencing Factors

The impact on the performance of asphalt pavements can be divided into the following eight categories of macro factors: road surface type, Climatic conditions. Road age factor, Road grade, Traffic volume, Engineering Factors, Road surface material.

There are many kinds of influencing factors, but if all the influencing factors are included in the subsequent modeling, it will make the modeling more difficult and increase the chance of overfitting; Considering that the reliability of the prediction model is inevitably reduced if factors with minimal relevance to asphalt pavement performance are introduced into the model. The factors affecting asphalt pavement performance are uncertain, complex and diverse, and previous studies lacked rationality in considering the influencing factors, so this paper attempts to select the influencing factors through the technical performance of asphalt mixtures.

The factors considered in this paper include temperature, number of days of rainfall, traffic volume, proportion of large vehicles, and time of day. In order to prevent the model estimation from being distorted or difficult to estimate accurately due to the existence of exact correlation or high correlation among the influencing factors, it is also necessary to do the multicollinearity test, and the results obtained are shown in the following table.

Table1. Multicollinearity test results

Percentage of Trucks	Percentage of Trucks	week average peak hourly volume	Number of Rainfall Days	Weekly average
Weekly average peak hour traffic volume	1.000	0.303	0.15	-0.153
Number of weekly rainfall days	0.303	1.000	0.055	0.354
Weekly average temperature	0.15	0.055	1.000	-0.392
Percentage of Trucks	-0.153	0.354	-0.392	1.000

The contents of Table 1 show the results of the multicollinearity test results between several pavement technical performance influencing factors. The numbers in the table are proportional variances, and a proportional variance close to 1 means that there is multicollinearity between the two variables. From the table, it can be seen that there is no number close to 1 within the proportional variance, and there is no multicollinearity among these four influencing factors.

1.3 Evaluation Indicators

This chapter describes three widely used error analysis criteria[14-15]:

(1) the average absolute prediction error of the forecast results (MAE)

$$MAE = \frac{1}{N}\sum_{i=1}^{N} | \hat{y}(t) - y(t) | \tag{2}$$

(2) Mean of squared prediction error（RMSE）

$$RMSE = \sqrt{\frac{1}{N}\sum_{t=1}^{N}\left(\hat{y}(t) - y(t)\right)^{2}} \tag{3}$$

(3) Mean absolute percentage error（MAPE）

$$MAPE = \frac{1}{N}\sum_{t=1}^{N} \left| \frac{\hat{y}(t) - y(t)}{y(t)} \right| \times 100\% \tag{4}$$

$\hat{y}(t)$ is the predicted value and $y(t)$is the measured value.， If the MAE、RMSE、MAPE of the model are lower, the better the prediction of the model.

2. COMBINED PREDICTIVE MODELS

2.1 Single prediction model

Commonly used single forecasting models include three categories: (1) time series models (TSMs), including gray models (GMs), integrated moving average autoregressive models (ARIMAs), and exponential smoothing (ESs); (2) causal analysis models (CAMs), such as regression analysis and elasticity coefficient analysis; and (3) nonlinear dynamic forecasting models (NDFMs) including support vector regression (SVR), genetic programming (GP), and artificial neural networks (ANNs). These methods have not entirely consistent strengths and inherent weaknesses, The following table gives a description.

Table2. Commonly used single prediction models

Predictive Model		Advantages	Disadvantages
Time Series Model (TSMS)	Integrated Moving Average Autoregressive Model (ARIMA)	Model simplicity. Only endogenous variables are required, no exogenous variables are needed.	Requires stable time-series data or stable after differencing. Essentially can only capture linear relationships.
	Grey model (GM)	Suitable for forecasting complex systems containing uncertainties. High short-term prediction accuracy.	Strong dependence on historical data. Low accuracy in medium- and long-term forecasting.
Causal Analysis Models (CAMs)	Regression analysis method	Easy to analyze multi-factor models. Accurate measurement of correlation between different factors and model fit	Large amount of data and complicated data processing.
	Elasticity coefficient analysis	Requires less data. Flexible in application.	Only the relationship between two variables can be considered. The results may be inaccurate when the elasticity coefficient changes over time.
Nonlinear Dynamic Prognostic Models (NDFMs)	Support vector regression (SVR)	High computational efficiency. High global optimality and generalization capability.	The problems caused by ambiguity cannot be solved well. It is not easy to solve multi-classification problems.
	Genetic Programming (GP)	Ultra-high speed, fault tolerance and error tolerance.	The calculation is cumbersome.
	Artificial Neural Networks (ANNs)	It can realize functions such as autonomous learning and rapid finding of optimal solutions.	The grid structure is difficult to determine. Slow convergence.

Currently, quantitative forecasting of road traffic volume is usually performed using causal analysis forecasting method with time series analysis forecasting method. If we continue to subdivide, causal analysis forecasting covers methods such as regression analysis, support vector machine, neural network, etc. Time series analysis forecasting covers methods such as exponential smoothing and gray theory. Since time series analysis uncovers its change pattern directly from data, its model is more time-sensitive and less subjective, so time series analysis method is more suitable for traffic volume forecasting than causal analysis method. This paper focuses on time series analysis to select the best model.

If non-negative raw data series are available, GM (1,1) model can be built:

$$x^{(1)}(t+1)=\left(x^{(0)}(1)-\frac{b}{a}\right)e^{-at}+\frac{b}{a} \tag{5}$$

The exponential smoothing method can be understood as a special moving average method. Unlike the general moving average method, data from different periods are not equally weighted and recent values are heavily weighted.

The basic formula is shown below.

$$S_t = ay_t + (1-a)S_{t-1} \tag{6}$$

S-- The smoothed value at time t； *yt* -- The actual value at time t； *St*–1-- The smoothed value at time *t* –1； *a*-- Smoothing constant, Numerical size between 0 and 1.

The calculation formula is as follows:

$$S_t^2 = aS_t^{(1)} + (1-a)S_{t-1}^{(2)} \tag{7}$$

S_t^2 −−the quadratic exponential smoothing value at time t; $S_t^{(1)}$ --primary exponential smoothing value at time *t*; $S_{t-1}^{(2)}$ -- the quadratic exponential smoothing value at time *t* –1;*a* -- weighting factors (i.e., smoothing constants).

2.2 Multiple regression forecasting model and VAR forecasting model

Multiple regression prediction model, all the remaining variables of interest are considered as independent variables, Only one of these variables is left as the dependent variable. Build the mathematical model. Based on the model relationship equation, the required independent variables are entered and the dependent variable is predicted. The equation is as follows.

$$Y = \partial_0 + \partial_1 X_1 + \partial_2 X_2 + \ldots + \partial_n X_n + \varepsilon \tag{8}$$

The p-order vector autoregressive model VAR(p) is to turn the variables and error terms in the p-order autoregressive model AR(p) into vector form and the coefficients into coefficient vector or matrix form. The presentation is shown below:

$$Y_t = c + \sum_{k=11}^{p} \Pi_k Y_{t-k} + \varepsilon_t, t = 1, 2, \ldots, T \tag{9}$$

2.3 Portfolio Modeling

First, we need to define the rules for model formulation. The division is to approximate the performance index and determine the weight coefficients.

Let the measured value of RQI be $Y(t)$. t =1,2,⋯, n, the predicted value of the combined forecasting model be $\widehat{Y}(t)$. The gray forecasting model and the VAR forecasting model predict $\widehat{Y}_1(t)$ and $\widehat{Y}_2(t)$,The closer the value of $\widehat{Y}(t)$ is to both $\widehat{Y}_1(t)$ and $\widehat{Y}_2(t)$, the better, and there are different combinatorial prediction models for choosing different approximation metrics. The following are three different approaches to combined prediction models built with common approximation performance metrics.

(1) Weighted arithmetic mean portfolio forecasting model

$$\widehat{Y}(t) = \omega_1 \widehat{Y}(t) + \omega_2 \widehat{Y}_2(t), t = 1, 2, \ldots, n \tag{10}$$

(2) Weighted square and average portfolio forecasting model

$$\widehat{Y} = \sqrt{\omega_1 \widehat{Y}_1^2 + \omega_2 \widehat{Y}_2^2}, t = 1, 2, \ldots, n \tag{11}$$

(3) Weighted Proportional Average Portfolio Forecasting Model

$$\widehat{Y}(t) = \frac{\omega_1 \left(\widehat{Y}_1(t)\right)^2 + \omega_2 \left(\widehat{Y}_2(t)\right)^2}{\omega_1 \widehat{Y}_1(t) + \omega_2 \widehat{Y}_2(t)} \tag{12}$$

Determination of the weighting coefficients by the prediction results of the combined model

$$\min J(t) = \sum_{i=1}^{n} \left(Y(t) - \widehat{Y}(t)\right)^2 \tag{13}$$

And need to meet the following two conditions:

1） $\omega 1 + \omega 2 = 1$；

2） $\omega 1$， $\omega 2 \geq 0$

2.4 Model Analysis

The model parameters were calibrated by the traffic volume, truck ratio and weather data of a highway in Guizhou from 2016 to 2020, it is about 10000 pieces of data. In this paper, we find the model of GM (1,1) as:

$$F\left(k\right)=49294.8-\frac{49203.6}{e^{0.002*(k-1)}} \tag{14}$$

The forecasting equation in the exponential smoothing forecasting model is obtained by minimizing the MAE as:

$$y_t=31S_{t-1}^{(1)}-55S_{t-1}^{(2)}+25S_{t-1}^{(3)} \tag{15}$$

The multivariate regression prediction model is obtained as:

$$Y=91.163-1.607*10^{-3}X_1-19.07X_2\text{-0.042X}_3\text{ -0.085X}_4 \tag{16}$$

The VAR model has the same forecasting properties as the combined forecasting model, so it is presented here in more detail. The obtained VAR model:

$$\begin{aligned}R_t&=0.17R_{t-1}+0.143R_{t-2}-0.03V_{t-1}-0.03V_{t-2}-15.159TP_{t-1}-1.4TP_{t-2}\\&-0.045RA_{t-1}-0.001RA_{t-2}+0.06T_{t-1}+0.028T_{t-2}+1.435\end{aligned} \tag{17}$$

Since the VAR model has the same forecasting characteristics as the combined forecasting model although it is a single forecasting model, it is not used to build the combined forecasting model. Based on the principle that the forecast value is as close to the actual value as possible, a quadratic programming model is established using Lingo12. to determine the weight coefficients of the GM (1,1) model, the exponential smoothing forecasting model and the multiple regression forecasting modeω2.

By combining the models, the combined forecasting model can be obtained as a combined exponential smoothing and multiple regression model as follows.

$$\widehat{Y}\left(t\right)=0.803\widehat{Y}_1\left(t\right)+0.197\widehat{Y}_2\left(t\right) \tag{18}$$

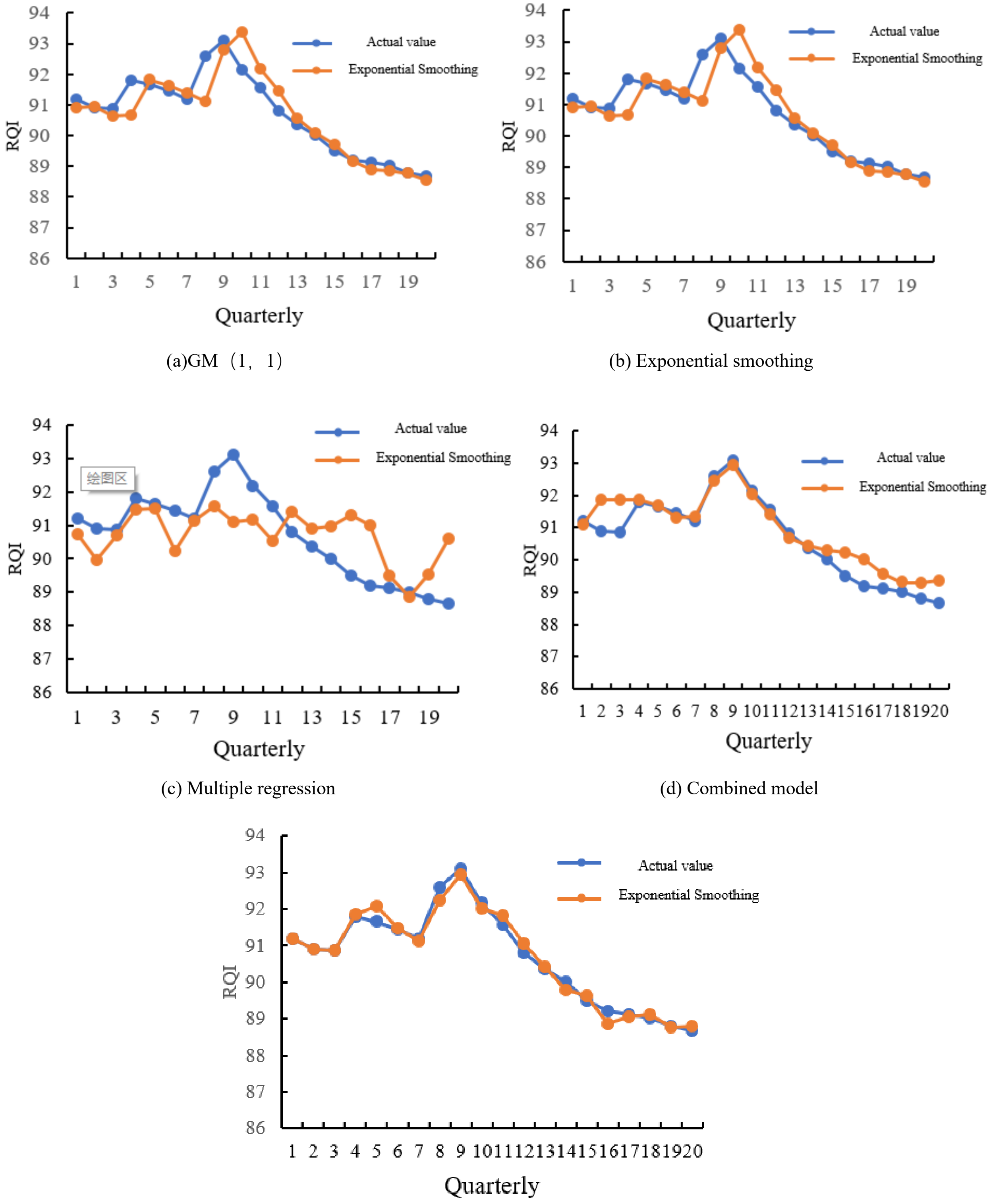
Table 3. Comparison of prediction error analysis results

Prediction Model	MAE	RMSE	MAPE
GM (1, 1)	0.606	0.830	0.007
Exponential smoothing	0.383	0.562	0.004
Multiple regression	0.865	1.059	0.01
Combined model	0.356	0.494	0.004
VAR	0.139	0.188	0.002

The above table can provide a clear comparison of the prediction effects of single prediction models and combined prediction models. It can be found that the combined prediction model obtained by reasonable selection of combined prediction model types tends to have a better prediction effect on asphalt pavement service performance than the single prediction model.

As can be seen from the table, the MAE, RMSE, and MAPE values of the VAR model are the smallest, indicating that the prediction effect of the VAR prediction model, compared to the rest of the above prediction models, is better. The MAE, RMSE, and MAPE values of the combined exponential smoothing and multiple regression forecasting model are only greater than those of the VAR model.

The combined performance proves that the combined exponential smoothing and multiple regression prediction model has a better prediction effect on asphalt pavement performance than the VAR(1) model.



(a)GM（1，1）

(b) Exponential smoothing

(c) Multiple regression

(d) Combined model

(e)VAR

Note: The horizontal coordinate is in units of one quarter and the vertical coordinate is RQI

Figure 1. Comparison of asphalt pavement performance prediction results

It is not difficult to find that, among the single models of forecasting, the gray forecasting model is not suitable for long-term forecasting with a quadratic curve trend, but for short-term forecasting or forecasting of variables with a trend close to linearity; the exponential smoothing forecasting model can have a better forecasting effect when the trend is smoother, but not for long-term forecasting with an inflection point, otherwise it will produce a large forecasting bias; the combined forecasting can correct the The combined forecast can correct the shortcomings of a single model to a certain extent and make its forecast value more consistent with the actual trend. Compared with the rest of the prediction models, the VAR model not only has better prediction accuracy, but also can better represent the change trend of pavement performance, which is helpful to guide the further prediction of its future changes.

3. CONCLUSION

This paper mainly collects data on the service performance and road conditions of a highway asphalt pavement in Guizhou, climate data, etc, The above results were used to carry out a study on the prediction of the service performance of asphalt pavements of highways. The following conclusions were drawn.

In this paper, the technical performance of asphalt mixture was used to select the influencing factors, the average weekly temperature, weekly rainfall days, average weekly peak hour traffic volume, truck ratio, and time were selected as the influencing factors of asphalt pavement service performance. When the prediction of asphalt pavement service performance was carried out the above factors were found to have a significant influence on it. By collecting a large amount of data, we successively established GM(1,1) model, exponential smoothing method prediction model, multiple regression prediction model, VAR(1) model, and combined them on this basis, and tried to establish the combined GM(1,1) and multiple regression prediction model of asphalt pavement usage performance, combined exponential smoothing method and multiple regression prediction model, combined ARIMA(0,1,0) and multiple regression combined prediction model. The final result is that the combined prediction models are basically better than the single prediction models in predicting the asphalt pavement service performance, among which the VAR model has the best prediction effect. Therefore, the VAR model can be used to predict the asphalt pavement performance of highways in practical production applications.

Due to the time and conditions of the study, the original data in this paper are small, the selected road sections are also small, and the material and structure are the same, so the influence of road material and structure differences are not considered. In the subsequent study, We can select the road with great differences in road materials and structures as the research objects.

REFERENCES

[1] Johnson K D, Cation K A. Performance prediction development using three indexes for North Dakota pavement management system[J]. Transportation Research Record, 1992 (1344).

[2] Durango P L. Adaptive optimization models for infrastructure management[J]. 2004.

[3] Kong, Xiangjie. Research on asphalt pavement service performance prediction and maintenance and repair decision method [D]. Beijing University of Technology, 2015.

[4] Xu Ran. Research on Pavement Performance Prediction and Maintenance Countermeasures of Ning-Hang Expressway[D]. Southeast University, 2016.

[5] Wu M, Wang Danyi, Lei Chaoxu. Research on the prediction model of asphalt pavement performance[J]. Guangdong Highway Traffic,2009(01):5-9.

[6] Zhao Fei, Shi Faken, Chen Li. Prediction of asphalt pavement leveling model based on BP neural network and Logistic regression analysis[J]. Road Construction Machinery and Construction Mechanization,2019,36(07):110-114.

[7] Chen L. Research on asphalt pavement levelness prediction method based on LTPP[D]. Chang'an University,2018.

[8] Yang G.F., Wang H.Y., Pan Y.L.. Prediction of asphalt pavement service performance based on mixed-effects model[J]. Highway Traffic Science and Technology,2018,35(08):19-27.

[9] Ni FJ, Fang Y, Xue ZM. Application of time series in pavement leveling prediction[J]. Journal of Southeast University (Natural Science Edition),2006,36(4):634-637.

[10] Liu BY, Yao ZK. Performance prediction of asphalt pavements [J]. Chinese Journal of Highways,1991(02):5-15+33.

[11] Luo JK. Application of neural network and combined prediction in highway pavement management system[D]. Southwest Jiaotong University, 2011.

[12] Zhou Lan. Evaluation and prediction of highway asphalt pavement performance[D]. Southeast University, 2015.

[13] Hu, Suna. Research on the prediction of the service performance of highway asphalt pavement [D]. Chongqing Jiaotong University,2015

[14] P. Du, J. Wang, W. Yang, T. Niu, A novel hybrid model for short-term wind power forecasting, Appl. Soft Comput. J. (2019).

[15] Y. Xu, W. Yang, J. Wang, Air quality early-warning system for cities in China, Atmos. Environ. (2017).

Research on denoising of skinned point cloud based on multi-feature point parameter weight optimization

Binpeng Li, Jian Mao*, Jie Yang, Hang Cai
School of Mechanical and Automotive Engineering, Shanghai University of Engineering Science, Shanghai 2162, China
* Corresponding author: jmao@sues.edu.cn

ABSTRACT

The effect of point cloud denoising is very important for the subsequent surface fitting and modeling design of the 3D scanning process. How to extract feature points quickly and accurately has become a research hotspot. However, the key to point cloud denoising lies in singular values and outliers. Therefore, this paper proposes a denoising model coupled with multi-feature parameters, discusses the influence degree of each feature point parameter separately, and uses the swarm intelligence algorithm to solve a set of optimal parameter weights to determine the point cloud denoising model, and to achieve the optimal denoising effect of 3D scattered point cloud. The simulation results show that the swarm intelligence algorithm used is faster and less time-consuming than the existing differential evolution algorithm. At the same time, the point cloud denoising model proposed in this paper has better performance than radius filtering and statistical filtering. denoising effect.

Keywords: 3D point cloud; point cloud denoising; swarm intelligence algorithm; point cloud feature points;

1.INTRODUCTION

In the field of 3D scanning and imaging, line laser scanning has been widely used in the digital inspection of parts. However, due to the influence of measurement equipment and environment factor, the collected point cloud data often contains a lot of noise, which hinders the subsequent reconstruction of 3D models. In order to make the point cloud data meet high-quality requirements of subsequent surface fitting and modeling design for point cloud data, it is necessary to decrease noise and smooth the point cloud data obtained by scanning first.

The point cloud collection includes spatial features and color features. Spatial features are used in the field of point cloud processing due to their obvious feature information. The feature points contain important information of model, such as: edge, sharp corner and ridge, eat. These feature points can reflect the basic geometry of model and play a key role in judging whether the appearance of model is correct or not. The feature points are also called geometric feature points [1-2], which have the great stability and are widely used in 3D point clouds [3-4]. Feature point detection can be divided into two types: grid-based model and point cloud-based model. For feature detection of grid models, Shin [5] et al. used polynomial fitting to estimate normal vector of point cloud. Yutaka [6] et al. reconstructed the implicit function to solve the point cloud curvature information. Charlie et al. [7] used Bilateral filtering is used for feature detection, but this method has a large error in detecting the boundary of the point cloud, and the topological relationship of the grid will also change accordingly. For feature detection of point cloud models, Yang [8] used principal component analysis (PCA) to calculate the curvature of the point cloud. T. Gatzke [9] used multiple fittings to calculate the curvature map. Huang [1] et al. After triangulation, the normal vector and curvature of the point cloud are estimated, but this model cannot represent the original scattered point cloud, and can only approximate the original scattered point cloud infinitely, so the accuracy cannot be guaranteed. Kris [11] designed a new feature detection method, which does not need to detect the curvature of the point cloud and the grid model, but calculates the point cloud with a large normal change through the region growing algorithm, and obtains the initial classification of the point cloud., and then divide different types of data to construct a minimum spanning tree to obtain the feature line of the point cloud model. This algorithm is suitable for uniformly distributed point cloud data. Wang [12] proposed a feature detection method that comprehensively considers the curvature of the point cloud, the normal angle between the point cloud and the neighborhood point, and the average distance from the point cloud to the neighborhood point, to distinguish the feature points and non-feature points of the point cloud model. However, the degree of influence of each feature point parameter on the point cloud denoising algorithm is not discussed. Chen [13] used the principal component analysis method to solve the curvature of the point cloud, and construct a denoising model considering multiple feature parameters, but also did not analyze the influence of different feature parameter weights on the denoising algorithm.

International Conference on Computer Graphics, Artificial Intelligence, and Data Processing (ICCAID 2022)
edited by Ruishi Liang, Jing Wang, Proc. of SPIE Vol. 12604, 126040D
© 2023 SPIE · 0277-786X · doi: 10.1117/12.2674547

In this paper, a point cloud denoising algorithm is proposed, which comprehensively considers the point cloud curvature, the angle between the point cloud normal and the neighborhood point normal, the distance from the point to the center of gravity of the neighborhood point, and the average distance from the point to the neighborhood point. The weight of each feature parameter is defined, and the influence of different feature parameters on the point cloud denoising algorithm is discussed. At the same time, the swarm intelligence algorithm is used, and the peak signal-to-noise ratio is used as the objective function to solve the optimal feature point parameter weight, and then determine the mathematical model of point cloud denoising and the discrimination threshold between noise points and feature points. This paper will use Stanford University's Bunny model and a skinned point cloud model to verify the algorithm.

2. POINT CLOUD MULTI-FEATURE POINT DETECTION

2.1 Point cloud curvature and angle between point cloud normal and neighbor point normal

The curvature of the point cloud is an important judgment basis for the recognition of surface features, and the curvature of the point cloud reflects the degree of concavity and convexity of the surface. The point cloud curvature is generally estimated by gridding the data and calculating the point cloud curvature. In this paper, Principal Component Analysis (PCA) will be used to construct the covariance matrix of the k neighborhood points of the point, and estimate the normal and curvature of the point cloud. This method will greatly simplify the calculation process and shorten the calculation time. The specific mathematical process is as follows:

Point cloud set:

$$P = \{p_i(x_i, y_i, z_i) | i \quad = \quad 1,2, \dots, N\}$$

(1) Among them, P is the input point cloud data (only contains position information); p_i represents a certain point cloud data; N represents the scale of point cloud data.

Neighborhood point set:

$$\{p_{ij}(x_{ij}, y_{ij}, z_{ij}) | j \quad = \quad 1,2, \dots, k\}$$

(2) Among them, k represents the number of neighbor points. The barycentric coordinates of k neighbor points can be solved by neighborhood points coordinates.

$$O_i = \frac{\sum_{j=1}^{k} p_{ij}}{k} \tag{3}$$

According to the above formula, the covariance matrix of the data points can be established

$$A_i = \begin{bmatrix} p_{i1} - O_i \\ p_{i2} - O_i \\ \dots \\ p_{ik} - O_i \end{bmatrix}^T \begin{bmatrix} p_{i1} - O_i \\ p_{i2} - O_i \\ \dots \\ p_{ik} - O_i \end{bmatrix} \tag{4}$$

The eigenvalues $\lambda_1, \lambda_2, \lambda_3$ and corresponding eigenvectors e_1, e_2, e_3 are solved by numerical analysis. Then determine the smallest eigenvalue.

$$\lambda_m = min\{\lambda_1, \lambda_2, \lambda_3\} \tag{5}$$

Then the normal vector of data point is the eigenvector e_m corresponding to the minimum eigenvalue λ_m.

$$n_i = e_m \tag{6}$$

Curvature C_i

$$C_i = \frac{\lambda_m}{\lambda_1 + \lambda_2 + \lambda_3} \tag{7}$$

The angle between the normal vector of the point cloud and the normal vector of the neighboring points is also an important indicator for judging whether the surface is flat or not. The feature point and the noise point are distinguished by the angle value. The solution process is as follows:

$$\theta = arcos\left(\frac{p_{gi} \cdot p_{gj}}{|p_{gi}| \times |p_{gj}|}\right), \ \ \theta \in [0, \pi] \tag{8}$$

Among them, p_{gi} is the normal direction of point cloud; p_{gj} is the normal direction of neighboring points of the data point; θ is the angle between the two vectors.

2.2 The distance from the point to the center of gravity of neighbor point

Compared with the standard test cloud point model, the general point cloud models are mostly unevenly distributed, and the complexity is greater. Therefore, the distance from the point to the center of gravity of the neighborhood point is considered as a judgment size. According to the known point cloud set and the center of gravity, the distance can be calculated [12].

$$d_{1i} = |p_i - O_i| \tag{9}$$

Among them, p_i is any point cloud data point; O_i is the center of gravity of neighborhood point.

2.3 Average distance from point to neighbor point

The size of average distance from point to neighbor point reflects the density of point clouds in a certain area. The larger the average distance, the greater number of point clouds in the area, and vice versa. Therefore, the distance is also used as the evaluation scale of feature points, and the calculation process is as follows [13].

$$d_{2i} = \frac{\sum_{j=1}^{k} |p_i - p_{ij}|}{k} \tag{10}$$

Among them, d_{2i} is the average distance; k is the number of neighborhood points; p_i and p_{ij} have the same meanings as before.

2.4 Point cloud denoising with integrated multi-feature parameters

In order to consider the influence of multiple feature point parameters on the denoising effect of point cloud at the same time, this paper constructs a mathematical model of point cloud denoising. The model also includes the curvature of the point cloud and the angle between the normal vector of the point cloud and the normal vector of the neighbor points, the distance from the point to the center of gravity of the neighbor point, and the average distance from the point to the neighbor point, and analyzes the influence of the parameters of each feature point. At the same time, the swarm intelligence algorithm is used, and the peak signal-to-noise ratio is used as the objective function to solve the optimal solution of the parameter weights of each feature point, to obtain the mathematical model of point cloud denoising, which is used to denoise the point cloud model.

According to the feature point parameters calculated above, construct a mathematical model of point cloud denoising.

$$A_i = \frac{\lambda_1 C_i + \lambda_2 \theta_i + \lambda_3 D_{1i}}{\lambda_4 D_{2i}} \tag{11}$$

Among them, λ_i is the corresponding weight of each feature. And discriminant threshold is constructed.

$$Th = \varphi \times \frac{1}{N} \sum_{i=1}^{N} A_i \tag{12}$$

$$\begin{cases} Feature\ point & if\ A_i > Th \\ Noise & else \end{cases} \tag{13}$$

Through the above mathematical model, the feature points and noise points can be judged. In order to determine the weight value of each feature, this paper adopts the sparrow search algorithm (SSA) [14], and takes the peak signal-to-noise ratio as the objective function to solve a set of optimal feature weights. The SSA algorithm has good global search ability, so it can solve the optimal value under the constraint of the objective function. The objective function is constructed as follows:

The target point cloud set:

$$P(x_{pi}, y_{pi}, z_{pi}), pi = 1,2, \dots, m_1 \tag{14}$$

The point cloud set after denoising:

$$Q(x_{qi}, y_{qi}, z_{qi}), qi = 1,2, \dots, m2 \tag{15}$$

where m_i is the size of the point cloud data, respectively.

The peak signal-to-noise (PSNR) mathematical model is as follows:

$$\begin{cases} PSNR = 10 \times log_{10}\frac{\max\{P\}\times\max\{Q\}}{\sum_{i=1}^{N}[(xp_i-xq_i)^2+(yp_i-yq_i)^2+(zp_i-zq_i)^2]} \\ \\ N = \min(m1, m2) \end{cases} \tag{16}$$

Through the objective function constructed above, the SSA algorithm is used for iterative optimization, and a set of optimal solutions are solved to determine the mathematical model of point cloud denoising, then point cloud denoising mathematical model for point cloud denoising can be determined. The specific process of algorithm is displayed in Fig.1.

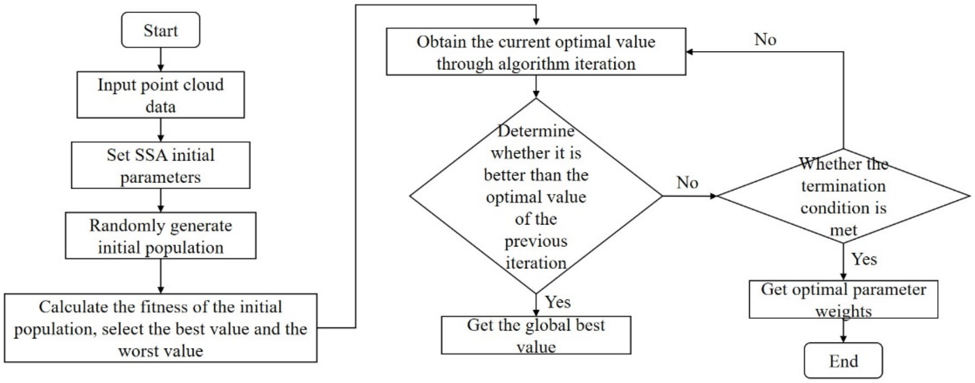

Start
Input point cloud data
Set SSA initial parameters
Randomly generate initial population
Calculate the fitness of the initial population, select the best value and the worst value
Obtain the current optimal value through algorithm iteration
No
Determine whether it is better than the optimal value of the previous iteration
No
Whether the termination condition is met
Yes
Get the global best value
Yes
Get optimal parameter weights
End

Figure 1. The flow chart of algorithm

3. SIMULATION AND EXPERIMENT

This article will use MATLAB R2020a to conduct simulation experiments, and the simulation uses the Bunny model of Stanford University. First, the feature information parameters of the point cloud are obtained, then the optimal feature weight is determined by the sparrow search algorithm, and finally the point cloud denoising mathematical model is used to denoise the point cloud to verify the feasibility of the algorithm in this paper.

3.1 Random noise experiment

In order to verify the theory proposed above, this paper will conduct random interference experiments, and compare with the algorithm based on the radius filtering principle and the statistical filtering principle, reflecting the superiority of the algorithm in this paper. The principle of radius filtering [15] is to assume that each data point in the initial point cloud contains a certain number of neighborhood points in the specified radius neighborhood, and the data points that do not meet the assumption conditions are regarded as noise points to be eliminated. The principle of statistical filtering [16-17] is to perform a statistical analysis on the neighborhood of each point to eliminate point clouds that do not meet certain conditions.

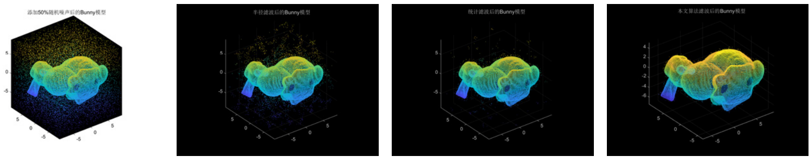


Figure 2a. Comparison diagram of Bunny model denoising algorithm with 50% noise

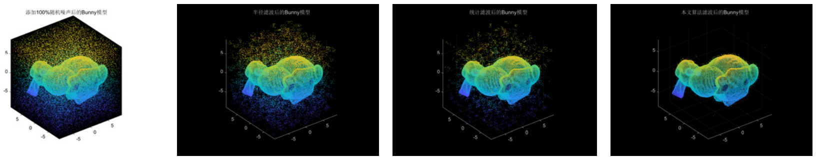


Figure 2b. Comparison diagram of Bunny model denoising algorithm with 100% noise

After simulation, Figure 2 shows the denoising comparison chart under different algorithms. After the Bunny model test, the point cloud denoising comparison effect can be found compared to radius filtering and statistical filtering. The 3D scattered point cloud denoising algorithm proposed in this paper has better denoising effect under the premise of ensuring the model feature points. In order to further reflect the effect of the algorithm in this paper in actual processing, a skin point cloud model is selected for simulation experiments, and the skin is scanned by a laser scanning device to form a skin point cloud model. Due to the influence of the experimental equipment, the scanned skin point cloud itself has noise points, so there is no need to add random noise points, and denoise it directly. The experimental equipment is shown in 3. Figure 4 shows the comparison chart of different denoising algorithms. Through comparison, it is found that the algorithm in this paper is better than the compared algorithms in terms of denoising quality while retaining the skin point cloud features.

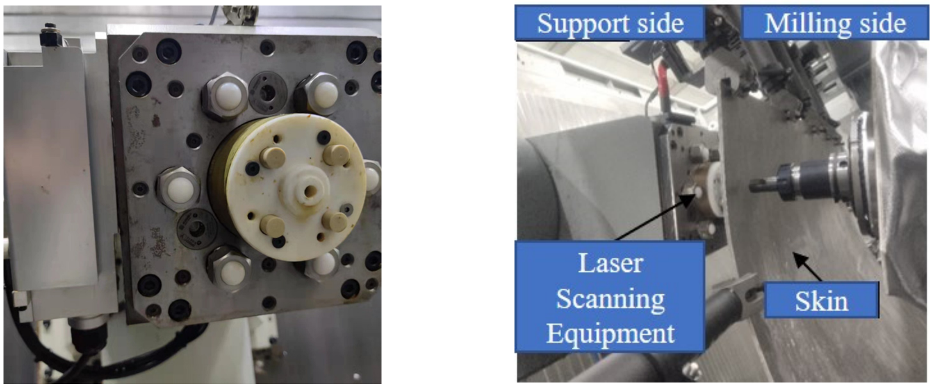
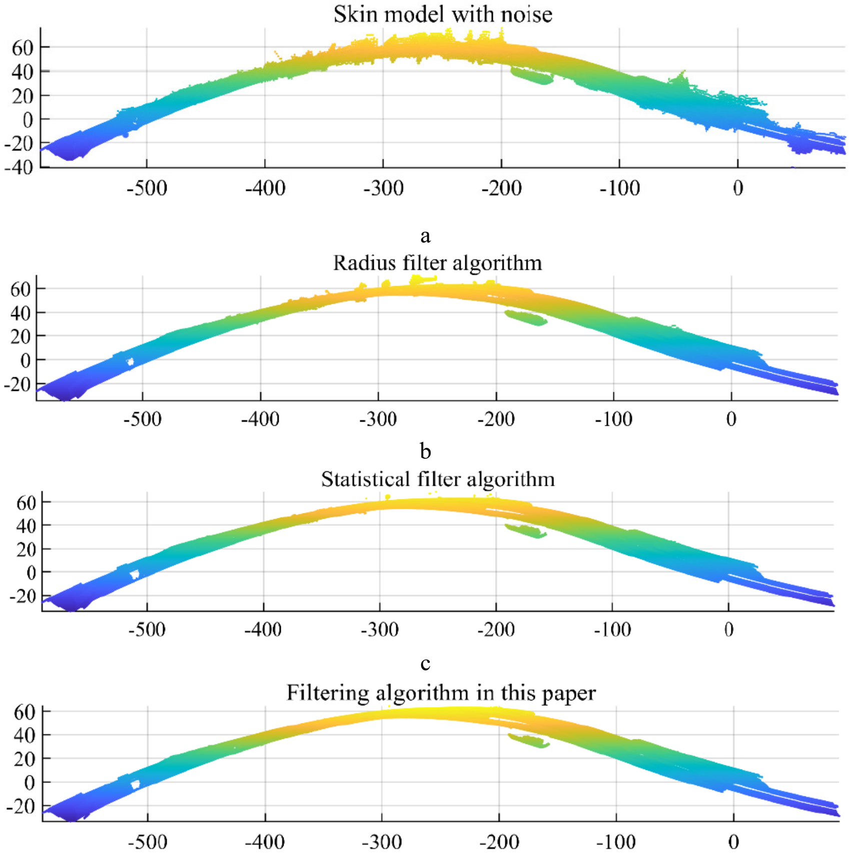


Figure 3. Experiment Device



a

b

c

d

Figure 4. Comparison of different denoising algorithms

Table 1 shows the weight of the feature information parameter of the Bunny model, and Table 2 shows the weight of the feature information parameter of a skin model. By observing the weight of the feature information parameter of the Bunny model, the weight of the curvature weight is λ_1, the weight of the normal angle is λ_2, and the weight of the distance from the point to the neighborhood point is the center of gravity λ_3. The change of the average distance λ_4 from the point to the neighborhood point is disordered and has nothing to do with the added noise, but the threshold coefficient φ is positively related to the noise.

Table 1. The feature information parameter weights of Bunny model

Weight name	λ_1	λ_2	λ_3	λ_4	φ
25% noise	11.320	31.098	82.506	91.416	0.40326
50% noise	81.174	59.602	36.502	76.886	0.37202
75% noise	10.701	54.926	67.107	94.771	0.16322
100% noise	53.935	16.364	23.014	84.840	0.10599

Table 2. The parameter weight of feature information of a skin model

Weight name	λ_1	λ_2	λ_3	λ_4	φ
Skin model with noise	28.385	54.134	44.655	73.869	0.18312

4. CONCLUSION

This paper proposes a point cloud denoising algorithm, which considers the point cloud curvature and the angle between the point cloud normal and the neighbor point normal, the distance from the point to the center of gravity of the neighbor point, and the average distance from the point to the neighbor point. The influence degree of different feature parameters on the point cloud denoising algorithm is discussed. At the same time, the swarm intelligence algorithm is used, and the peak signal-to-noise ratio is used as the objective function to solve a set of optimal feature point parameter weights, so as to determine the mathematics model of point cloud denoising. And then perform point cloud denoising. The feasibility of the proposed 3D scattered point cloud denoising algorithm is verified by simulation experiments, which provides a good foundation for the subsequent fitting and surface reconstruction of the 3D point cloud model. In addition, it also has certain guiding significance for practical application.

REFERENCES

[1] Yongtao Yang, Kun Zhang, Guoyan Huang, Peiliang Wu. Outliers Detection Method Based on Dynamic Standard Deviation Threshold Using Neighborhood Density Constraints for Three Dimensional Point Cloud[J]. Journal of Computer-Aided Design & Computer Graphics,2018,30(6):1034-1045.

[2] Hui Zeng, Huijuan Wang, Jiyuan Dong. Robust 3D keypoint detection method based on double Gaussian weighted dissimilarity measure[J]. Multimedia Tools and Applications,2017,76(24):26377-26389.

[3] Sai Manoj Prakhya, Bingbing Liu, Weisi Lin, Vinit Jakhetiya, Sharath Chandra Guntuku. B-SHOT: a binary 3D feature descriptor for fast Keypoint matching on 3D point clouds[J]. Autonomous Robots,2017,41(7):1501-1520.

[4] D. N. Brito, C. F. G. Nunes, F. L. C. Padua and A. Lacerda, "Evaluation of Interest Point Matching Methods for Projective Reconstruction of 3D Scenes," in IEEE Latin America Transactions,2016, 14(3): 1393-1400.

[5] Shin Yoshizawa, Alexander Belyaev, Hans-Peter Seidel. Fast and robust detection of crest lines on meshes[P]. Solid and physical modeling,2005:227-232.

[6] Yutaka Ohtake, Alexander Belyaev, Hans-Peter Seidel. Ridge-valley lines on meshes via implicit surface fitting[J]. ACM Transactions on Graphics (TOG),2004,23(3):609-612.

[7] Wang Charlie C L. Bilateral recovering of sharp edges on feature-insensitive sampled meshes. [J]. IEEE transactions on visualization and computer graphics,2006,12(4):629-639.

[8] Yong-Liang Yang, Yu-Kun Lai,Shi-Min Hu, Helmut Pottmann. Robust principal curvatures on multiple scales[P]. Geometry processing,2006:223-226.

[9] Gatzke T, Grimm C. Feature detection using curvature maps and the min-cut/max-flow algorithm[C]. In Proc. Geometric Modeling and Processing,2006,578-584.

[10] Huang, Jian-bing Menq Chai-Hsiang H. Automatic data segmentation for geometric feature extraction from unorganized 3-D coordinate points[J]. IEEE Robotics and Automation Society, 2001,17(3):268-279.

[11] Kris Demarsin, Denis Vanderstraeten, Tim Volodine, Dirk Roose. Detection of closed sharp edges in point clouds using normal estimation and graph theory[J]. Computer-Aided Design,2006,39(4):276-283.

[12] Lihui Wang, Baozong Yuan. Feature Point Detection for 3D Scattered Point Cloud Model[J]. Signal Processing, 2011,27(6):932-938.

[13] Long Chen, Yong Cai, Jiansheng Zhang, Beiping Xiang. Feature point extraction of scattered point cloud based on multiple parameters hybridization method[J]. Application Research of Computers,2017,34(9):2867-2870.

[14] Xue, J., & Shen, B. (2020). A novel swarm intelligence optimization approach: sparrow search algorithm. Systems Science & Control Engineering (1), 22-34.

[15] Song Bi, Yuhao Wang. LiDAR Point Cloud Denoising Method Based on Adaptive Radius Filter[J]. Transactions of the Chinese Society for Agricultural Machinery,2021,52(11):234-243.

[16] Liuyi Li, Yufeng Zhu. Research on Denoising Algorithm of Point Cloud Data Based on Hybrid Filtering[J]. Jiangxi Science,2021,39(3):525-529,533.

[17] Shuo Wei, Nanxiang Zhao, Minle Li, Yihua Hu. Single photon denoising algorithm combined with improved DBSCAN and statistical filtering[J]. Laser Technology,2021,45(5):601-606.

Proc. of SPIE Vol. 12604 126040I-7

Electricity retail market package decision model based on electricity consumption characteristics

HongPeng Qi[1a*], Rong Zheng[1b*], Zhe Lin[1c], YunWang Hu[1d], FeiFei Zhou[1e,] JianZhao Li[1f], Hui Zhao[1g]

[1] Guangxi Power Exchange Center Co., Ltd., Nanning, China

[a*]qi_hp@gx.csg.cn, [b*]zheng_r@gx.csg.cn, [c]lin_z@gx.csg.cn, [d]hu_yw@gx.csg.cn,
[e]zhou_ff@gx.csg.cn, [f]li_jz@gx.csg.cn, [g]zhao_h@gx.csg.cn.

ABSTRACT

This paper starts from the purpose of assisting electricity market entities to make better decisions, providing guidance for the design and selection of retail packages, improving market efficiency, and guiding the healthy development of the market, and considering the future development trend of the electricity retail market, designs an optimal package decision model under the background of competitive electricity retail market. The model can be based on retail customers' electricity consumption information and risk preference analysis and select the optimal retail package that meets their needs from the retail package library. Finally, an example is used to illustrate the package decision-making process that considers the consistency of the customers' power consumption curve and the time-of-use electricity price and considers the market transaction needs of different customers in the electricity retail market.

Keywords: Retail packages; decision models; risk preference

1 INTRODUCTION

With the continuous deepening of the reform and further development of the electricity retail market, the influence of the retail side on the organization and operation of the electric power system and the electric power market will also increase. Under the development trend that market players are faced with numerous choices, how to assist market players to make better two-way decisions, provide guidance for the design and selection of retail packages, improve market efficiency and guide the healthy development of the market is a problem that needs to be studied and solved in depth at this stage. For this reason, this paper carries out the research on the optimal package decision model for the electricity retail market.

Foreign related studies mainly focus on recommending packages with high ratings for the same type of customers based on package ratings and customer classification. Literature [1] collects electricity consumption data from customers through smart meters, analyzes individual customers' preferences for various packages, and uses collaborative filtering algorithms to recommend optimal packages. The [2] proposes an evaluation system for power customer, discusses the basis of customer category classification and evaluates 10 power customer value benefit and designs four types of packages. In [3] a method of electricity consumption behavior clustering and pricing packages based on data mining is proposed, and a distributed clustering framework combining DTW k-medoids algorithm is designed, the segmentation of electricity consumption behavior can realize effective personalized electricity package recommendation service for customers.

In summary, domestic and international research analyses have recommended retail packages for customers from the end-customer electricity consumption characteristics but have not considered the risk preferences of customers for different retail packages. By studying the optimal package decision-making model in the electricity retail market, this paper analyzes customers' subjective risk preferences and different electricity consumption characteristics in the spot market, aiming to put forward suggestions for market construction.

2 RETAIL MARKET PACKAGE DECISION MODEL

2.1 Retail market package decision model

The structure of the optimal decision model for the retail menu is shown in Figure 1. Firstly, the information of customers' electricity consumption characteristics curve is input in the model. The model will evaluate and analyze the input curve parameters, assume the main decision based on the customers' electricity consumption characteristics, classify the electricity customers' electricity consumption characteristics, match the electricity retail packages suitable for the

International Conference on Computer Graphics, Artificial Intelligence, and Data Processing (ICCAID 2022)
edited by Ruishi Liang, Jing Wang, Proc. of SPIE Vol. 12604, 126040E
© 2023 SPIE · 0277-786X · doi: 10.1117/12.2674540

electricity customers, and make the retail customers' electricity load distribution consistent with the electricity time-of-use price.

Risk preference analysis undertakes to assist in decision-making. The prices of some retail packages in the current market need to be formed using the transaction prices in the future wholesale electricity market as the Base Price, so in this pricing model, it is necessary to consider the customers' risk preference for future market prices and incorporate the customers' gaming and risk preference for future prices into the package decision.

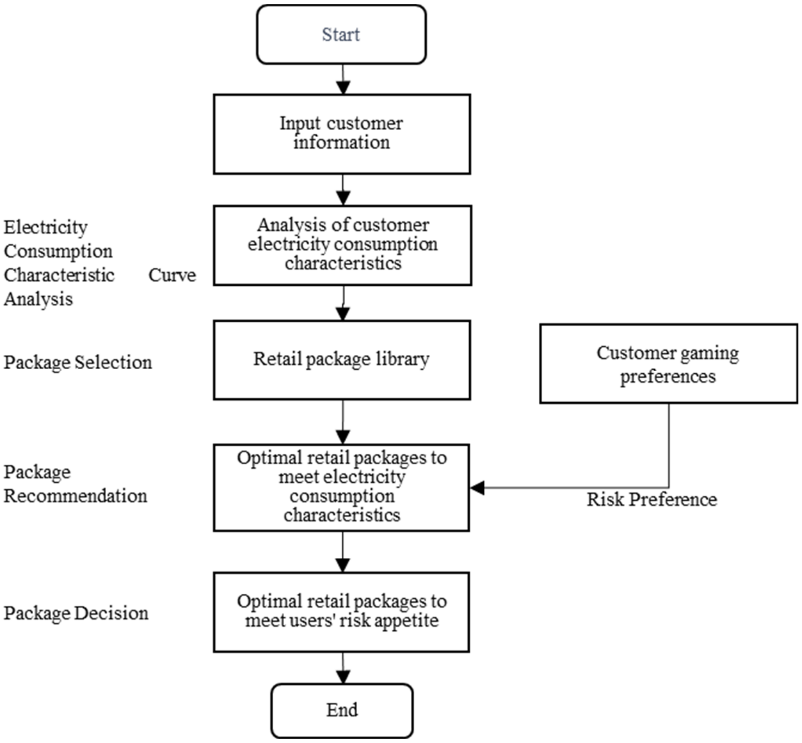

Start
Input customer information
Electricity Consumption Characteristic Curve Analysis
Analysis of customer electricity consumption characteristics
Package Selection
Retail package library
Customer gaming preferences
Package Recommendation
Optimal retail packages to meet electricity consumption characteristics
Risk Preference
Package Decision
Optimal retail packages to meet users' risk appetite
End

Fig.1 Construction and geometrical dimensions of specimens

2.2 Information on electricity consumption characteristics

The information of electricity consumption characteristics includes selecting the customer type, entering the annual typical date load curve, and entering the annual estimated electricity consumption of the customer.

(1) Select the type of customer into 3 categories: large industrial customer, general industrial customer, and commercial customer.

(2) Enter the estimated annual electricity consumption of the customer. The unit is kWh/year.

(3) Enter the ratio of the customer's electricity consumption by month.

(4) Enter the typical time-of-day load curve for power consumers.

2.3 Customer risk preferences

There are 3 descriptions of trading risk preferences in economics: Risk averse, Risk neutral, and Risk-seeking[4].

(1) Risk aversion is the tendency to prefer transactions that are safer but may also have lower expected returns when a person is faced with uncertain returns.

(2) Risk preference means that when faced with a transaction with uncertain returns, a person prefers a transaction with a greater degree of risk and higher expected returns.

(3) The risk-neutral person's decision is not influenced by the uncertainty of the resulting from the decision. Risk-neutrals have the same preference for both decisions with the same expected returns but different levels of risk.

The base price is a factor in the retail price package that changes for both buyers and sellers, and the base price may increase or decrease in the future. Market operators can provide market participants with historical base price as a reference for choosing a retail package for the current period.

2.4 Retail Package Library

The retail package library is formed by the standard retail packages issued by the electricity sales company. The electricity sales company designs the retail packages in advance and the issued retail packages are open to all customers, who can freely choose from the issued retail packages. The electricity sales company will also issue packages that are tailored to the electricity consumption characteristics of the customer, and the customer can choose such packages in conjunction with their own electricity consumption characteristics and choose the retail package with the lowest rate that meets their risk preference.

3 RETAIL PACKAGE TARIFF MODEL

At present, the types of retail packages in the retail market are fixed price packages, linked price packages, mixed rate packages and proportional share packages, and the price calculation methods and the gaming level of the four types of packages are as follows.

Table 1 Retail package content and degree of gaming

Package Type	Content	Calculation method
Fixed Price Packages	Fixed price packages for both peak and valley times	$y \ =a$ $y_p =f_1 * a$ $y_v =f_2 * a$
Proportional share package	According to the "base price + share ratio" to form the plain price, then in accordance with the peak and valley price ratio to form the peak and valley tariffs	$y \ =a-(a-b)*m\%$ $y_p =f_1 * \left[a-(a-b)*m\%\right]$ $y_v =f_2 * \left[a-(a-b)*m\%\right]$
Mixed Rate Packages	The electricity package price is formed by two parts: "fixed price + linkage price".	$y \ =n_1 * b * x + n_2 * a$ $y_p =f_1 * (n_1 * b * x + n_2 * a)$ $y_v =f_2 * (n_1 * b * x + n_2 * a)$
Linking Price Package	The package price is formed in accordance with the "base price + linking ratio" to form a plain price, and in accordance with the peak and valley price ratio to form a peak and valley tariffs	$y \ =b * x$ $y_p =f_1 * b * x$ $y_v =f_2 * b * x$

In the above equation, y、y_p、y_v is the plain, peak and valley retail transaction prices respectively; f_1、f_2 is the peak and valley price ratio, the peak and valley prices are formed by multiplying the plain price by the peak and valley price ratio; a is the fixed price component of the retail price; b is the linking price component of the retail price; $m\%$ is the share ratio; n_1 and n_2 are the proportion of fixed price and linking price components in the mixed rate package, the sum of which is 1. x is the package price linking ratio of the linked price package.

4 ELECTRICITY CONSUMPTION CHARACTERISTICS ASSESSMENT

As early as 2001, G. Chicco started using load characteristic metrics to represent the electricity load curve[5], aiming to represent the characteristics of daily or weekly customer electricity consumption behaviours. The factors associated with the load curve include peak load factor (f_1), valley load factor (f_2), night factor (f_3), and noon factor (f_4).

$$f_1 = \frac{P_{av,day}}{P_{max,day}}, f_2 = \frac{P_{min,day}}{P_{av,day}}, f_3 = \frac{P_{av,night}}{P_{av,day}}, f_4 = \frac{P_{av,lunch}}{P_{av,day}}$$

Each type of factor is used to quantitatively evaluate the proportion of the customer's peak and valley loads on the electricity consumption curve throughout the day.

(1) Peak load ratio (f_1) indicates the ratio of the average load to the maximum load throughout the day.

(2) Valley load ratio (f_2) indicates the ratio of the lowest load to the average load for the whole day.

(3) The night factor (f_3) indicates the ratio of the average evening electricity load to the average load of the whole day.

(4) The noon factor (f_4) indicates the proportion of the average midday electricity load to the average load of the whole day.

The influence of the above values of factors on the electricity consumption characteristics of large industrial customers can be divided according to the following intervals.

Table 2 Large industrial customers' electricity load factor interval

Factors	f1	f2	f3	f4
Low	[0.2,0.3)	[0,0.2)	[0,0.2)	[0,0.5)
Medium	[0.2,0.4)	[0.2,0.4)	[0.2,0.3)	[0.5,0.7)
High	[0.4,0.5)	[0.4,0.5)	[0.3,0.4)	[0.7,0.9)
Very high	[0.5,0.6)	[0.5,0.6)	[0.4,0.6)	[0.9,1.2]
Extremely high	[0.6,1]	[0.6,1]	>0.6	>1.2

For example, for large industrial customers, the four basic classification rules only need to use f_1 and f_3 load factors, because the difference between different large industrial customers is mainly whether the electricity consumption is peak at night and whether the electricity consumption is stable all day.

Table 3 Classification rules for large industrial customers

Conditions		Customer Categories
f_1 is a very high level and	f_3 is an extremely high level	Stable consumption
f_1 is above high level and	f_3 is low or medium level	Daily consumption
f_1 is above high level and	f_3 high, very high	Night consumption
Other		Unstable consumption

5 ELECTRICITY CONSUMPTION CHARACTERISTICS ASSESSMENT

5.1 Electricity consumption characteristics assessment

Taking a large industrial customer's package decision process in a single month as an example, setting the planned power consumption of the large industrial customer in a certain month as 1000MWh, and the typical curve of the customer's power consumption in that month is in Fig.2.

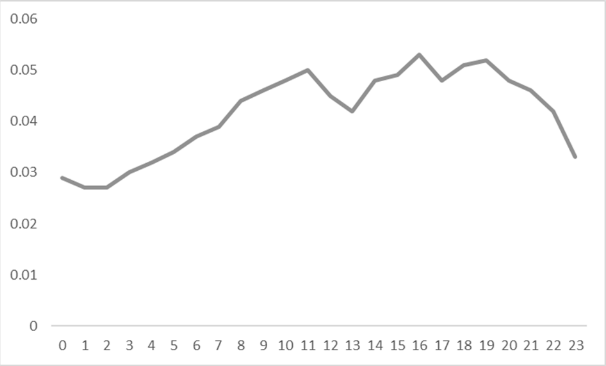


Fig.2 Typical electricity consumption curve of a large industrial customer

According to the calculation method of load factor and the typical power curve distribution of this customer, the load factor of this customer is calculated, the results are shown in the Table 4.

Table 4 Load factor of a typical curve for a large industrial customer

Factors	**f_1**	**f_2**	**f_3**	**f_4**
Calculation results	0.786	0.648	0.768	1.08
Grade	Extremely high	Extremely high	Extremely high	Very high

According to this customer's load factor results, $f_1$and f_2 are extremely high, the customer should belong to the stable electricity consumption type then will select the stable electricity consumption packages in the retail package library.

5.2 Package tariffs Comparison

Assuming that the customer is a risk- preference customer, the market manager provides the historical "base price" for the last three years as a reference for the market participants. The retail package selection system or retail trading platform will provide customers with the lowest price package in each category that matches the customer's electricity consumption characteristics, so that each customer only needs to choose the retail package that matches his or her preference among the four types of packages.

The 4 types of package tariffs in the retail package library for stable electricity consumption are listed in the Table 5.

Table 5 Package rates for each retail package

Historical "Base Price"(¥/kWh)	Fixed Price Packages(¥/kWh)	Proportional Share Package(¥/kWh)	Mixed Rate Packages(¥/kWh)	Linking Price Package(¥/kWh)
0.360	0.400	00.370	0.390	0.396
0.400	0.400	0.390	0.400	0.44
0.440	0.400	0.410	0.409	0.484

The market "base price" in the above parameter is selected as the comprehensive monthly price in the Day-Ahead market before.

5.3 Customer risk preference decision

Risk- preference customers will choose the package with the potential maximum revenue as their retail price package based on the goal of maximizing revenue, i.e., choose the "base price" of 36 cents/kWh as the final decision, the peak and valley price ratio $f_1 = 1.5$, $f_2 = 0.5$, the peak-valley price between different packages is shown in the Table 6.

Table 6 Final retail package peak and valley tariffs rates

Time period	Fixed Price Packages(¥/kWh)	Proportional share package(¥/kWh)	Mixed Rate Packages(¥/kWh)	Linking Price Package(¥/kWh)
Plain	0.400	0.370	0.390	0.432
Peak	0.600	0.555	0.586	0.648
Valley	0.200	0.185	0.195	0.216

5.4 Monthly costing

5.4.1 Peak and valley power calculation. According to the typical curve of this customer, the total electricity consumption of the customer is decomposed in different time periods in the following table, and the distribution of the monthly electricity consumption of the customer is 382MWh in peak period, 369MWh in the plain period, and 249MWh in the valley period.

Table 7 Peak and valley time division

Segmentation of electricity consumption	Time period
Peak period	9:00-12:00, 18:00-23:00
plain period	7:00-9:00, 12:00-18:00
Valley period	0:00-7:00, 23:00-24:00

5.4.2 Monthly electricity tariffs. Based on the customer's electricity consumption on different time periods with different package tariff, the monthly electricity cost of the customer is calculated. The monthly electricity cost under the above four categories with different package tariffs are shown in the Table 8.

Table 8 Monthly tariffs of different electricity retail packages

Price	Fixed Price Packages	Proportional share package	Mixed Rate Packages	Linking Price Package
Plain Fee (¥)	147600	136530	144057.6	159408
Peak fee (¥)	229200	212010	223699.2	247536
Valley Fee (¥)	49800	46065	48604.8	53784
Monthly Fee (¥)	426600	394605	416361.6	460728
Difference with fixed price (¥)	0%	-8%	-2%	8%

The following conclusions can be drawn from the above tariffs results.

(1) From the point of view of the total cost of electricity, the overall cost of electricity is the lowest when the customer chooses the proportional share package, but customers may not get this price in the end if the choice of the retail package associated with the market base price, the final price will have a certain risk.

(2) At the price gaming level, the proportional share package can obtain an objective expected return of 8% lower than the fixed price compared to the fixed price package in the market, and the full return can only be obtained when the final "base price" is equal to the historical lowest price.

For other customers with different risk preference, still taking the above package prices as an example, if the customer's risk preference is risk-averse, then faced with potential market risks, such customers will choose a fixed-price package that fits their electricity consumption characteristics in the whole market; if the customer is risk-neutral, the customer will make their own decisions according to the probable expected price of each type of package.

In the above example, the model is illustrated for one month only. In practice, retail customers can estimate the cost of different packages for multiple months in the future.

6 CONCLUSION

With the further expansion of the electricity retail market the two-way decisions of retail market participants have become more complex. The electricity retail package decision model proposed in this paper can analyse retail customers' electricity consumption information and risk preferences to select the type of retail package that meets their needs from the retail package library. It can provide qualitative analysis suggestions for electricity retail market participants to design and select electricity retail packages.

ACKNOWLEDGMENTS

This work was financially supported by the Planning Projects of Guangxi Power Grid Co., Ltd. (GXDW2021-17).

REFERENCES

[1] Energy; New Findings on Energy from University of Salamanca Summarized (Multi-agent Recommendation System for Electrical Energy Optimization and Cost Saving in Smart Homes) [J]. Energy Weekly News, 2019

[2] G. Chicco, R. Napoli, P. Postolache, M. Scutariu and C. Toader, “Electric Energy Customer Characterisation for Developing Dedicated Market Strategies”, in Proc. 2001 IEEE PowerTech, 10-13 September 2001, Porto, Portogul, pp. 1-6

[3] Zhang S, Fan Y, Wang B. A method of electricity consumption behaviour clustering and pricing packages based on data mining[J]. Information Systems and Signal Processing Journal, 2020, 5(1): 18-23.

[4] Ritchken, P.H., and Tapiero, C.S., "Warranty design under buyer and seller risk aversion", Naval Research Logistics Quarterly 33 (1986) 657-671

[5] G. Chicco, “Customer Behaviour and Data Analytics,” 2016 Int. Conf. Expo. Electr. Power Eng.,no. Epe, pp. 771–779, 2016.

Implementation and visualization of Weighted A-Star algorithm and Bidirectional Weighted A-Star algorithm under large-scale road network

Dexin Yu*[a], Luchen Wang[a], Xincheng Wu[a,b], Zhuorui Wang[c,d], Jianyu Mao[e], Xiyang Zhou[c,d]
[a]Navigation College, Jimei University, Xiamen, 361021, China;
[b]Navigation College, Xiamen Ocean Vocational College, Xiamen, 361012, China;
[c]School of Transportation, Jilin University, Changchun, 130022, China;
[d]Jilin Engineering Research Center for Intelligent Transportation, Changchun, 130022, China;
[e]School of Architecture and Transportation Engineering, Guilin University of Electronic Technology, Guilin, 541004, China;
[*] Corresponding author: yudx@jmu.edu.cn

ABSTRACT

In the real world, traffic scenes are complex and contain intricate road networks, which makes the shortest path computation on large-scale road networks a challenging task. Existing research has concentrated on small-scale urban road networks or grid maps. In practical scenarios, however, we are often faced with seeking the shortest paths on large-scale road networks. For this reason, it is imperative to develop efficient shortest path searching algorithms, as it offers significant savings in time and resources. To tackle this issue, this paper proposes two improved A* algorithms, namely the Weighted A* algorithm and the Bidirectional Weighted A* algorithm. To verify the effectiveness of our proposed algorithms, we validated the performance of our proposed algorithms against the conventional Dijkstra and A* algorithms on urban road networks of different sizes. Our results significantly demonstrate the effectiveness of our solution, as both algorithms significantly outperform Dijkstra's and A* algorithms, with little loss of accuracy.

Keywords: OSM, Shortest Path, Weighted A*, Bidirectional Search

1. INTRODUCTION

In the real world, traffic scenes are complex with the rapid urbanization and increased mobility. Research on shortest path computation aims to improve algorithm efficiency and reduce customer's travel costs. Peter E. Hart et al.[1] proposed the A* algorithm in 1968, which reduced the number of search nodes by introducing heuristic information to accelerate the algorithm. On top of this, Pohl et al. further proposed to set the expansion factor on the heuristic function as the original version of the Weighted A* algorithm, which is called the WA* algorithm[2]. In practice, computing the shortest path is usually converted into another problem, i.e., how to search a relatively short path and reduce double-counting time. The Weighted A* algorithm is a computing strategy applicable for this problem. However, there are few studies on Weighted A* algorithms under large-scale urban road networks. Previous studies based on small-scale networks, including fewer points and edges graphs, can hardly characterize the complexity of road networks and verify the efficiency of the algorithm. As such, we propose a Weighted A* algorithm and extend it to different sizes of large-scale urban road for validation. This paper supports that the Weighted A* algorithm has a good performance in balancing accuracy and efficiency under large-scale road networks and has significant improvement over the Dijkstra algorithm and A* algorithm with almost no loss of accuracy.

On the other hand, based on the original Dijkstra algorithm, Luby and Raged proposed the Bidirectional Dijkstra algorithm[3]. The algorithm speed is two times faster than the previous version in the ideal case. Hence, Some investigators proposed a Bidirectional A* algorithm[4]-[6], which inherits the efficiency of the Bidirectional Dijkstra algorithm. In most cases, the A* algorithm produces a "tree" scattered in a fan shape on the map. It leads to an increase in the number of traversal points in the search process and will eventually slow it down. Based on this problem, this paper proposes a Bidirectional Weighted A* algorithm for large-scale road networks based on Weighted A* algorithm, which combines the advantages of the above algorithms and shows better performance in most cases.

International Conference on Computer Graphics, Artificial Intelligence, and Data Processing (ICCAID 2022)
edited by Ruishi Liang, Jing Wang, Proc. of SPIE Vol. 12604, 126040F
© 2023 SPIE · 0277-786X · doi: 10.1117/12.2674530

The remainder of this paper is organized as follows: Section 2 introduces the data and two improved algorithms. Section 3 describes the validation results of our design. Section 4 summarizes our conclusion, and we give some research recommendations.

2. ALGORITHM DESIGN

2.1 Data Structure

In this paper, the data structure stores the directed graph in the form of edges and points (obtaining data sources based on OSMnx parsing open street map (OSM)). All algorithms in this paper are based on this data structure to ensure the consistency of the experimental environment. The data structure is shown in Table 1 and Table 2.

Table 1. Data structure of the road segment.

u	v	length	all length	distance	highway
0	589	257.319	inf	…	tertiary
0	18	786.54	inf	…	trunk
…	…	…	…	…	…

Table 2. Data structure of nodes.

osmid	y	x	lon	lat
0	2719200.998	491673.039	110.9178	24.58682
1	2719858.12	491214.2378	110.9132	24.59276
…	…	…	…	…

In Table 1, the "u" and "v" denote the osmid (id of the point), and the direction is from "u" to "v". "length" refers to the actual length of the road segment from "u" to "v". "all length" represents the total cost (actual cost + estimated cost) in the forward direction and is initially set to "inf", which means infinity. “distance” indicates the predicted distance from point v to the endpoint (in the one-way case), which is calculated from the Euclidean distance, and the algorithm degenerates to the Dijkstra algorithm when the distance is constant to zero. “highway” represents the grade of the road segment. "lon" and "lat" denotes the latitude and longitude in the WGS84 coordinate system respectively. "y" and "x" refer to the calculated Mercator projection coordinates. With this data structure, the benefits are 1) the simplicity of the structure, which is easy to view; 2) the ability to readily extend the set of edges and points; and 3) in combination with other data labels, such as "traffic condition" or "highway" mentioned in the table above, “length” can be easily assigned with different weights.

2.2 Data Acquisition and Visualization

This subsection introduces urban road network acquisition methods with pervasive applicability and road network visualization methods with interactive nature.

There are often several problems in the shortest theoretical studies in the field of transportation: in the theoretical studies where GIS is not introduced, geographic information is ignored, resulting in heuristic information being almost unavailable, the classical A* algorithm is limited, and most of the studies are restricted to Dijkstra's algorithm which does not contain heuristic information, and the algorithm is less efficient from the overall point of view; in the theoretical studies where GIS is introduced, geographic information is included The distance between nodes may be regarded as Euclidean distance, which is not consistent with the reality despite satisfying the conditions of A* algorithm; meanwhile, the introduced GIS data may have a non-uniform format (directed and undirected graphs) and non-disclosure, which makes it difficult for subsequent studies to reproduce or compare in the same scenario.

In order to solve the above problems, this paper obtains urban road network geographic information data based on OSMnx parsing open street map, and simplifies the number of nodes, "road section" is represented by the real line type between two points, and the calculation process abstracts the middle "road section" into The calculation process abstracts the intermediate "road segments" into actual trip lengths for calculation, and finally Folium performs web presentation. The steps are as follows.

Step1: Based on OSMnx download and parse OSM data (simplifying the road network method please refer to this literature[7]) into GeoDataFrame data structure, and store it into graphML[8] format data (to facilitate subsequent calls again without re-downloading the road network data).

Step2: Screen the GeoDataFrame data to form the key information of edges and points and store them in DataFrame format.

Step3: If it is A* algorithm then use NumPy broadcast mechanism to calculate each point inspired information. Otherwise, skip to Step4.

Step4: Shortest path calculation, return distance and path nodes.

Step5: Call GeoDataFrame in Folium and add the base map for visualization.

2.3 Weighted A* Algorithm

The traditional A* algorithm adds a heuristic function with constraints to Dijkstra's algorithm (which can also be thought of as an algorithm that guarantees an optimal solution by adding certain restrictions to the estimation function of A's algorithm), which is expressed as follows.

$$f(n) = g(n) + h(n) \tag{1}$$

Where "n" is the current node, $f(n)$ is the total estimated cost, $g(n)$ is the cost from the starting point to the current node, and $h(n)$ is the predicted cost from the current node to the endpoint and satisfies $h(n) \leq h^*(n)$, where $h^*(n)$ denotes the actual cost from node n to the endpoint. It is pointed out in the literature[9] that the use of the A* algorithm satisfies the admissibility and consistency when the Euclidean distance is used as the heuristic value and the actual length of the road segment between two points is used to represent the cost in a vector map, so it can find the optimal solution using the A* algorithm in a vector map. The A* algorithm satisfies $h(n) \leq h^*(n)$ if the predicted value of the heuristic function can approximate the actual cost distance as close as possible. The A* algorithm can maintain the fastest speed to find the optimal solution. In practice, due to the expansion of the number of nodes, the shortest-path problem usually converts to how to find a more optimal solution in less time, and the classical A* algorithm does not dominate.

For the above proposed situation, if the presentation of the heuristic function can be changed to expand the heuristic information appropriately so that more predicted values are closer to the true values, a faster computational process can be exchanged for a smaller cost. The presentation is as follows.

$$f(n) = g(n) + \varepsilon * h(n) \tag{2}$$

The ε can take a value slightly greater than 1. $h(n)$ is considered as the two-dimensional Euclidean distance from point n to the end point. The formula is as follows.

$$h(n) = \sqrt{(x_t - x_n)^2 + (y_t - y_n)^2} \tag{3}$$

where x_t and y_t denote the projection coordinates of the target node, and x_n and y_n denote the projection coordinates of the current node. With the above expression, Dijkstra and A* algorithms are transformed into a special case of Weighted A* algorithm, i.e., Dijkstra's algorithm when $\varepsilon = 0$, $f(n) = g(n)$. When $\varepsilon = 1$, $f(n) = g(n) + h(n)$, i.e., the A* algorithm. By appropriately expanding the weight coefficients ε, the estimated cost of more nodes is made to approximate the true value, thus obtaining faster computation speed. The shortest paths of certain road sections themselves to the end point present straight lines (approximating the Euclidean distance), which, after multiplying by the weight coefficient ($\varepsilon > 1$), make $\varepsilon * h(n) > h^*(n)$, violating the requirement that the estimated cost is less than or equal to the actual cost, which is the cause of the failure to guarantee the optimal solution.

2.4 Bidirectional Weighted A* Algorithm

Based on the original Dijkstra algorithm, some scholars have proposed the Bidirectional Dijkstra algorithm, which performs the Dijkstra algorithm from both the starting point and the endpoint and is twice as fast as the original Dijkstra algorithm in the ideal case. In this paper, we propose the Bidirectional Weighted A* algorithm after combining the advantages of the above algorithms. The algorithm has the following properties: (1) satisfies $f(n) = g(n) + \varepsilon * h(n)$ ($\varepsilon > 0$), and $h(n)$ is the Euclidean distance from the point to the endpoint. (2) The Weighted A* algorithm is executed only once in a cycle, and the execution is determined by $g(n)$. (3) The algorithm stops when the one encounters each other. See Figure 1 for details.

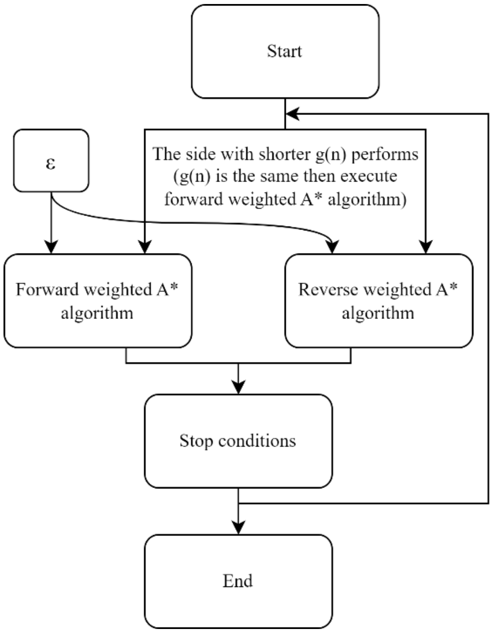

Start
ε
The side with shorter g(n) performs (g(n) is the same then execute forward weighted A* algorithm)
Forward weighted A* algorithm
Reverse weighted A* algorithm
Stop conditions
End

Figure 1. Flowchart of the Bidirectional Weighted A* algorithm.

We use the forward-weighted A* algorithm as an example to show the performed process. As shown in Figure 2, three DataFrames are used to store the data separately. The first DataFrame contains the total road segments. The second one stores the sections that are popped from the total road segment and are not yet included in the optimal section from the source to any point *v*. The third one stores the optimal sections from the source to any point v. The latter two DataFrames correspond to the OPEN sets and CLOSED sets in the general shortest-path calculation process, respectively.

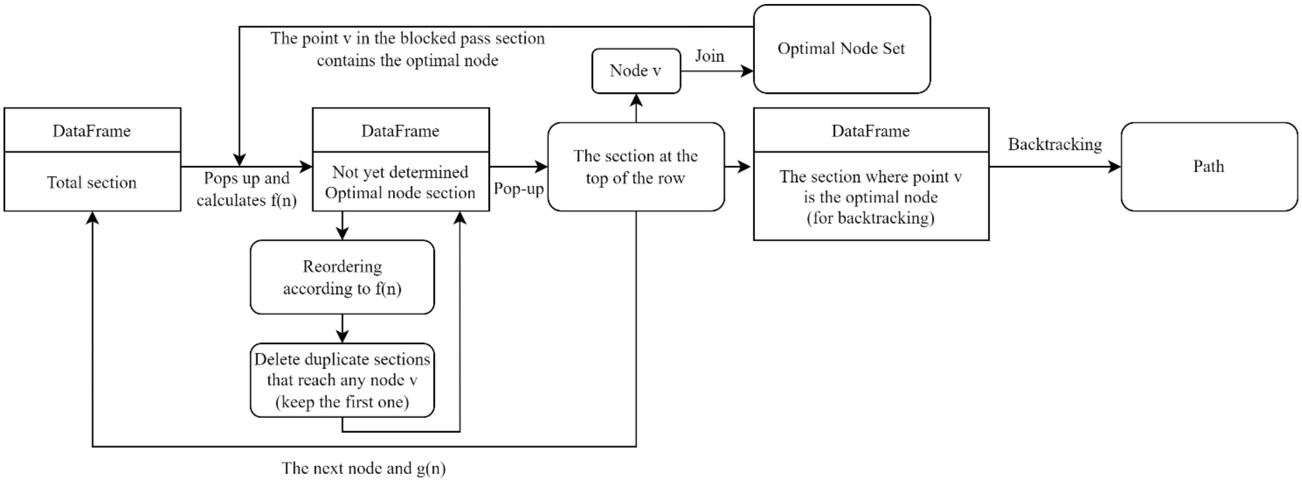

The point v in the blocked pass section contains the optimal node
Node v
Join
Optimal Node Set
DataFrame
Total section
Pops up and calculates f(n)
DataFrame
Not yet determined Optimal node section
Pop-up
The section at the top of the row
DataFrame
The section where point v is the optimal node (for backtracking)
Backtracking
Path
Reordering according to f(n)
Delete duplicate sections that reach any node v (keep the first one)
The next node and g(n)

Figure 2. The specific process of the Bidirectional Weighted A* algorithm (one of the directions, the other similar).

In the above process, the two nodes in the road segment, *u* and *v*, can be understood as node *v* and its previous node *u*. As for sorting, it corresponds to building a priority queue. When sorted, removing duplicate nodes *v* ensures the minimum cost of currently reaching node *v* (when ε is not zero, the $f(n)$ includes the predicted cost of node v to the endpoint).

2.5 Evaluation of the Considered Algorithms

For evaluation, we use the following metrics to estimate and compare different methods:

$MPEL$**: Mean percentage error** on shortest paths **length computation**, which is the computed average of percentage errors by which the calculated path length l_ε of the considered algorithm a differs from the actual optimal distance l_{best} under different parameter ε. We apply it to measure the average accuracy loss.

$$MPEL = \frac{1}{n}\sum_{\varepsilon=j}^{k}\frac{l_{a_\varepsilon}-l_{best}}{l_{best}} \quad (\varepsilon \in [j,k]) \tag{4}$$

$MaxPEL$: **Maximum percentage error** on shortest paths **length computation**, which is used to quantify the maximum accuracy loss between the max calculated path length and the actual optimal distance l_{best} under different parameter ε.

$$MaxPEL = \frac{max\,(l_j,\ldots,l_k)}{l_{best}} - 1 \quad (\varepsilon \in [j,k]) \tag{5}$$

$MPEE_{a\_b}$: **Mean percentage error** on **efficiency computation,** which is obtained to measure the efficiency improvement of algorithm *a* relative to algorithm *b* under different parameters *ε*. The running time of the algorithm is denoted by *t*.

$$MPEE_{a\_b} = \frac{1}{n}\sum_{\varepsilon=j}^{k}\left(\frac{t_{b_\varepsilon}}{t_{a_\varepsilon}} - 1\right) \quad (\varepsilon \in [j,k]) \tag{6}$$

3. ALGORITHM VALIDATION

3.1 Environment

The computer configuration for the verification of the algorithm in this paper is as follows.

CPU：12th Gen Intel(R) Core (TM) i7-12700F 2.10 GHz

RAM：16.0 GB DDR4 2133 MHz

Disk：ST2000DM005-2U9102 2000 GB, 5400 r/min, 256 MB

The computer system uses Windows 11, the development environment is Pycharm 2022.2.2, the development language is Python, and the version is Python 3.11.

3.2 Results

Data validation was conducted in four cities of different sizes (Chongqing, Guilin, Shanghai, and Beijing) with the following data sets.

Table 3. City size and path characteristics.

Paths	**City**	**Number of nodes**	**Number of edges**	**Features**	**Start id**	**End id**	**Selection of road sections**
Path 1	Chongqing	70322	170963	It shows a dense road network in the central and western part and a sparse road network in the north.	2137	56228	From the sparse area and through the dense area. From northeast to southwest.
Path 2	Guilin	9316	23891	The road network is dense at the east and west ends of the core and sparse in the middle.	4694	7292	From the edge of the core, traverse the dense zone - sparse zone - dense zone. From east to west.
Path 3	Shanghai	50265	139420	Dense road network.	4694	43120	The path is located inside the dense area. It runs from southeast to northwest.
Path 4	Beijing	87452	239670	Typical square grid and circular radial urban structure with dense road network.	16997	65221	The path is located inside the dense area. It runs from southwest to northeast.

Compare the efficiency and accuracy of Weighted A* algorithm and Bidirectional Weighted A* algorithm with different *ε* (when *ε*=0 the results are calculated for Dijkstra and Bidirectional Dijkstra respectively). The results are as follows.

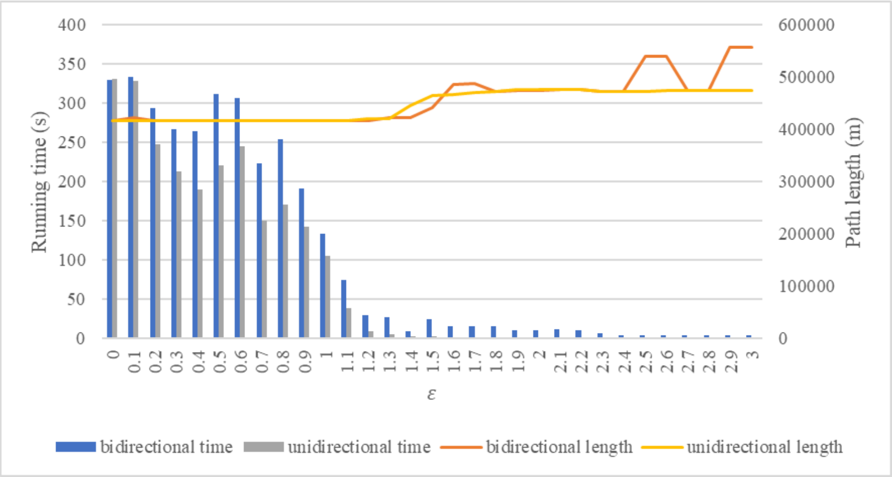


Figure 3. Running time and path length of path 1 under different *ε*.

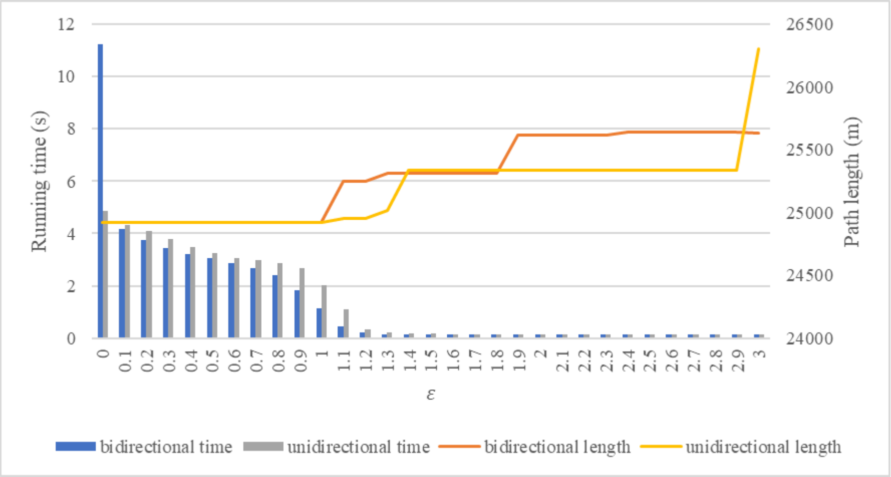


Figure 4. Running time and path length of path 2 under different *ε*.

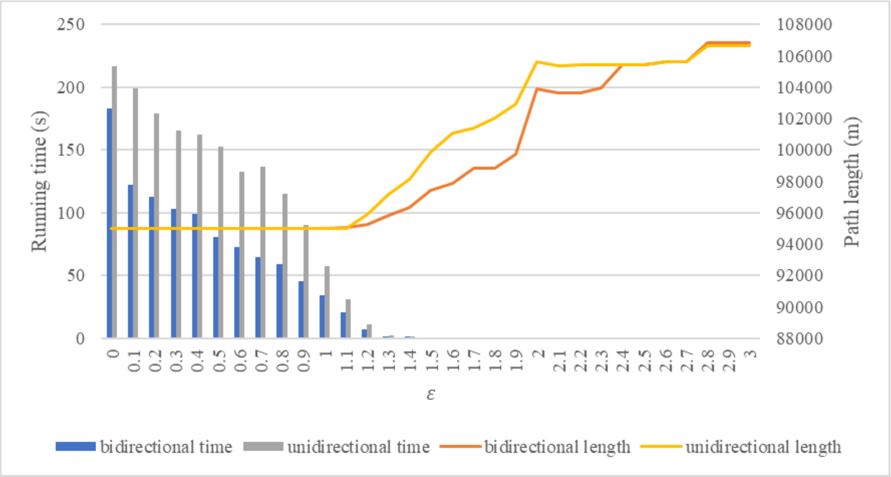


Figure 5. Running time and path length of path 3 under different *ε*.

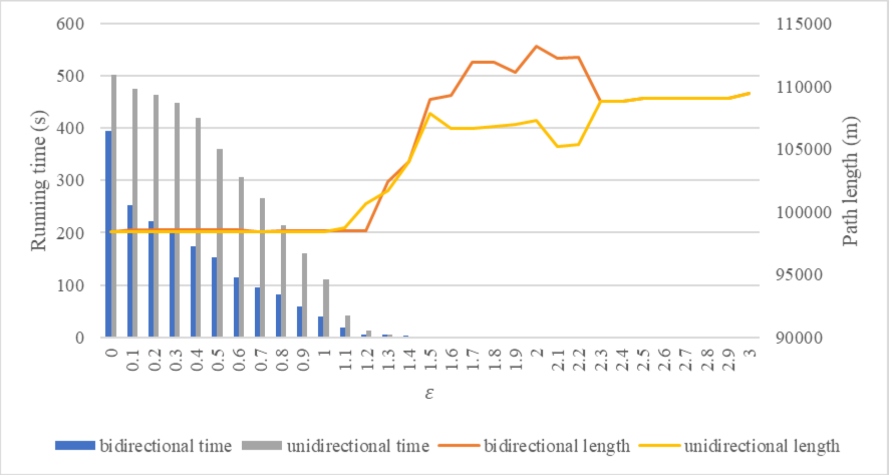


Figure 6. Running time and path length of path 4 under different ε.

The Dijkstra algorithm and Bidirectional Dijkstra algorithm search for the best result when $\varepsilon = 0$ while the Weighted A* algorithm satisfies the optimum when $\varepsilon \leq 1$.

The Bidirectional Dijkstra algorithm does not consistently outperform the Dijkstra algorithm with the guarantee that the Bidirectional Dijkstra can obtain the optimal solution, and sometimes the running time even reaches twice that of the one-way Dijkstra. This case occurs in path 2. Please refer to Figure 4.

Figure 7 shows there exist some special cases when the single-ended road network is dense and encounters obstacles; It may lead to Bidirectional Weighted A* efficiency lower than Weighted A*, which is due to the obstacles encountered in the inverse Weighted A* algorithm, resulting in the algorithm traversing more points. This case occurs in path 1. Please refer to Figure 3.

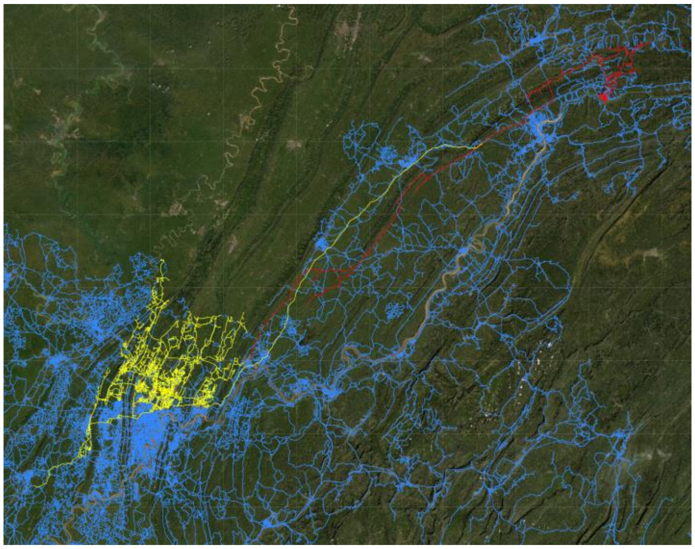


Figure 7. Path 1 is incorporated into the optimal roadway in both directions at $\varepsilon = 1.5$ (red for the forward direction and yellow for the reverse direction).

As ε ($\varepsilon \leq 1.5$) increases, the speed of the improved algorithm is significantly improved compared with Dijkstra's algorithm and Bidirectional Dijkstra's algorithm ($\varepsilon = 0$). In Path 4, the average efficiency of the two improved algorithms over Dijkstra's algorithm is improved by 9463.19% and 11736.98% ($\varepsilon \in (1.0, 1.4]$), respectively. Please refer to Table 5 below. The results show that the weighted A* and Bidirectional Weighted A* algorithms have strong applicability in real road networks.

When $\varepsilon \in (0, 1.0]$, Bidirectional Weighted A* does not necessarily satisfy the optimum, but the algorithm almost does not lose accuracy (the maximum average loss is 0.13% and the maximum loss is 1.35% in this paper), and the speed is usually faster than that of Weighted A* algorithm with the same ε.

In a larger scale road network, when $\varepsilon \in (1.0, 1.4]$, the Bidirectional Weighted A* algorithm is faster than the Weighted A* algorithm under the same ε, the accuracy loss is slightly less than or approximately equal to the Weighted A* algorithm.

The two improved algorithms ($\varepsilon > 1.0$) are faster than the A* algorithm and the Bidirectional A* algorithm (ε = 1.0), respectively. Please refer to Figure 3 - Figure 6.

In this paper, we suggest setting ε around $(1.0, 1.4]$ when using the Bidirectional Weighted A* algorithm under the real road network and setting the size of ε to balance accuracy and efficiency according to the demand. The experimental results show that the maximum accuracy loss in this range is 2.47% and the speed improvement is obvious (62.52%, 30.53%, and 56.89% in the three paths located in Guilin, Shanghai, and Beijing, respectively, compared with Weighted A*).

The results of the segmentation statistics are provided in Table 4 and Table 5.

Table 4. Result Statistics of Path 1 and Path 2

Method	**Paths**	**Path 1**				**Path 2**			
	Metric	**0.1-1.0**	**1.1-1.4**	**1.5-3.0**	**0.1-3.0**	**0.1-1.0**	**1.1-1.4**	**1.5-3.0**	**0.1-3.0**
BWA^*	$MPEL\%$	0.13	0.64	17.92	9.69	0.00	1.43	2.52	1.53
	$MaxPEL\%$	1.35	1.34	33.37	33.37	0.00	1.56	2.89	2.89
$BWA^*\ vs\ Dijkstra$	$MPEE_{BWA^*\_Dij}\%$	37.51	1438.19	5048.63	2896.86	94.98	2254.03	3238.67	2059.49
WA^*	$MPEL\%$	0.00	2.01	13.28	7.35	0.00	0.57	1.90	1.09
	$MaxPEL\%$	0.00	6.80	14.10	14.10	0.00	1.66	5.56	5.56
$WA^*\ vs\ Dijkstra$	$MPEE_{WA^*\_Dij}\%$	81.20	5432.63	26281.52	14768.23	56.37	1565.20	3414.89	2048.76
$BWA^*\ vs\ WA^*$	$MPEE_{BWA^*\_WA^*}\%$	-22.76	-66.84	-81.19	-59.80	20.01	62.52	-4.47	12.62

Table 5. Result Statistics of Path 3 and Path 4

Method	**Paths**	**Path 3**				**Path 4**			
	Metric	**0.1-1.0**	**1.1-1.4**	**1.5-3.0**	**0.1-3.0**	**0.1-1.0**	**1.1-1.4**	**1.5-3.0**	**0.1-3.0**
BWA^*	$MPEL\%$	0.00	0.67	8.59	4.67	0.08	2.47	11.96	6.73
	$MaxPEL\%$	0.00	1.42	12.48	12.48	0.13	5.68	15.03	15.03
$BWA^*\ vs\ Dijkstra$	$MPEE_{BWA^*\_Dij}\%$	218.46	7574.16	28899.14	16495.58	400.87	9463.19	32911.73	18948.30
WA^*	$MPEL\%$	0.00	1.64	9.95	5.53	0.00	2.88	9.52	5.46
	$MaxPEL\%$	0.00	3.31	12.28	12.28	0.00	5.63	11.16	11.16
$WA^*\ vs\ Dijkstra$	$MPEE_{WA^*\_Dij}\%$	76.27	7415.47	40237.36	22474.08	94.22	11736.98	56909.31	31947.97
$BWA^*\ vs\ WA^*$	$MPEE_{BWA^*\_WA^*}\%$	79.44	30.53	-24.76	17.34	145.71	56.89	-41.92	33.80

4. CONCLUSIONS

The Weighted A* algorithm and the Bidirectional Weighted A* algorithm show superior performance under large-scale road networks. On the one hand, although the efficiency improvement of the two algorithms relative to Dijkstra's algorithm is not consistent across different road networks, the overall situation shows that the two algorithms still have a 10-500 times speedup relative to Dijkstra's algorithm while maintaining a small loss of accuracy in large-scale road networks. They are also faster than the A* algorithm. On the other hand, the adjustable parameter ε allows researchers to determine either higher accuracy or faster efficiency is needed depending on the actual needs. We suggest that in subsequent studies research scholars could express the parameter ε in the form of a function and explore the relationship between parameter ε and prediction costs in large-scale road networks to achieve higher efficiency and less loss of accuracy.

REFERENCES

[1] Hart, Peter E., Nils J. Nilsson, and Bertram Raphael. "A formal basis for the heuristic determination of minimum cost paths." IEEE transactions on Systems Science and Cybernetics 4.2 (1968): 100-107.

[2] Ebendt, Rüdiger, and Rolf Drechsler. "Weighted A∗ search–unifying view and application." Artificial Intelligence 173.14 (2009): 1310-1342.

[3] Luby, Michael, and Prabhakar Ragde. "A bidirectional shortest-path algorithm with good average-case behavior." Algorithmica 4.1 (1989): 551-567.

[4] Nannicini, Giacomo, et al. "Bidirectional A∗ search for time-dependent fast paths." International Workshop on Experimental and Efficient Algorithms. Springer, Berlin, Heidelberg, 2008.

[5] Nannicini, Giacomo, et al. "Bidirectional A* search on time‐dependent road networks." Networks 59.2 (2012): 240-251.

[6] Rice, Michael, and Vassilis Tsotras. "Bidirectional A* search with additive approximation bounds." International Symposium on Combinatorial Search. Vol. 3. No. 1. 2012.

[7] Boeing, Geoff. "OSMnx: New methods for acquiring, constructing, analyzing, and visualizing complex street networks." Computers, Environment and Urban Systems 65 (2017): 126-139.

[8] Brandes, Ulrik, et al. "GraphML progress report structural layer proposal." International Symposium on Graph Drawing. Springer, Berlin, Heidelberg, 2001.

[9] Liu, Hao, and Bao, Yuan-Lu. "Application of A* algorithm in vector map optimal path search." Computer Simulation 25.4 (2008): 253-257.

A Study on the Distribution Characteristics of the Arrival Time Interval of Different Ship Types

LinHao Wu, DaNing Xing*
Navigation College, Jimei University
* Corresponding author: dnxing@jmu.edu.cn

Abstract

Based on the ship data of a certain port in China within one year, this paper analyzes the distribution characteristics of ship arrival time interval of the port within one year by using mathematical statistics method, and the Matlab tool is used to fit the three probability distributions of arrival probability of each ship type to obtain the goodness of fit, sum variance and root mean square difference. Through the analysis of fitting data results of negative exponential distribution, Weibull distribution and Erlang distribution, it can be compared and concluded that: when ship types are not distinguished, the arrival laws of all ships obey the Weibull distribution; When the ship types are distinguishes, container ships, bulk carriers and general cargo ships obey the Weibull distribution, while oil tanker arrival law obeys both the Weibull distribution and the low-order Erlang distribution.

Keywords: arrival law; Matlab fitting; Negative exponential distribution; Weibull distribution; Erlang distribution

1.Introduction

Ships are affected by natural factors such as typhoons, fog and unforeseeable events when sailing at sea, which makes the arrival time of ships uncertain. Liu Jingxian et al. used the mathematical statistics method to fit the samples with the normal distribution density curve and Poisson distribution density function curve, and found that the number of daily arrivals of ships was more subject to normal distribution[1]. Yu Jin et al. compared the measured data points with the theoretical values of the probability distribution by means of time guarantee rate, and showed that the number of daily ship arrivals in inland waterways follows a normal distribution, and the daily bow spacing is approximately an Erlang distribution[2]. Wang Nuo et al. conducted a global and local analysis of the arrival time of container liner, and concluded that the ship arrival time interval follows the Erlang first-order and Erlang higher-order distribution[3]. Wu Di et al. analyzed the deviation between the actual arrival time of the container liner and the shipping schedule, and proved that the probability of the container liner's arrival at night follows the gamma distribution[4]. Song Yunting et al. concluded that the deviation probability of container liner arrival time conforms to Erlang distribution under the conditions of scheduling constraints and disturbance uncertainty[5].

The above research results have significant limitations. Some scholars only conduct research based on the overall operation state, without considering the differences between different ship types, and have not discovered the inherent law of ship arrivals. Some scholars analyzed only container liners, but the ship type studied is relatively single, so it is not representative and lacks comprehensive research. The law of ship arrival is of great significance to the optimization of berth scheduling and the scientific management of wharf operation. This paper makes an in-depth analysis of the overall law of the arrival of all types of ships and the arrival law of different ship types, and verifies the rationality of the established probability distribution function model by statistically analyzing the arrival time of nearly 2,000 ships in a port in China for one year.

2.Probability distribution function model

Based on the data obtained by relevant departments, the ship traffic situation can be quantitatively analyzed to understand the spatio-temporal characteristics of ship traffic. Probability theory and mathematical statistics are commonly used to analyze the actual situation of ship traffic. Probability theory and statistical distribution are commonly used to analyze the actual situation of ship traffic, and the basic form to reflect the characteristics and laws of ship traffic is statistical distribution. Due to the characteristics of uncertainty in the arrival law of ships, In this paper, the probability density function curves of negative exponential distribution, Weibull distribution and Erlang distribution are mainly used for fitting and comparison, and the probability density function that can most accurately describe the law of arrival of different types of ships is determined. The Pearson test, which is most commonly used in statistics, is used for testing.

International Conference on Computer Graphics, Artificial Intelligence, and Data Processing (ICCAID 2022)
edited by Ruishi Liang, Jing Wang, Proc. of SPIE Vol. 12604, 126040G
© 2023 SPIE · 0277-786X · doi: 10.1117/12.2674522

2.1 Negative exponential distribution

The probability density function is:

$$f(x;\lambda) = \begin{cases} \lambda e^{-\lambda x}, x \geq 0 \\ 0, x < 0 \end{cases} \tag{1}$$

In Equation (1), x is a random variable and λ is a rate parameter.

The distribution function is:

$$F(x;\lambda) = \begin{cases} 1 - e^{-\lambda x}, x \geq 0 \\ 0, x < 0 \end{cases} \tag{2}$$

Mathematical expectation:

$$E(x) = \frac{1}{\lambda}, \tag{3}$$

Variance:

$$D(x) = \frac{1}{\lambda^2} \tag{4}$$

2.2 Weibull distribution

The probability density function is:

$$f(x;\lambda,k) = \begin{cases} \frac{k}{\lambda}(\frac{x}{\lambda})^{k-1} e^{-(\frac{x}{\lambda})^k} & x \geq 0 \\ 0 & x < 0 \end{cases} \tag{5}$$

In Equation (5), x is a random variable, $\lambda > 0$ is a scale parameter, and $k > 0$ is a shape parameter. Its cumulative distribution function is an extended exponential distribution function. When $k = 1$, it's exponential; When $k = 2$, it's a Rayleigh distribution.

Mathematical expectation:

$$E = \lambda \Gamma\left(1 + \frac{1}{k}\right) \tag{6}$$

Variance:

$$Var = \lambda^2 \left[\Gamma\left(1 + \frac{2}{k}\right) - \Gamma(1 + \frac{1}{k})^2\right] \tag{7}$$

In Equations (6) and (7), γ is a gamma function

2.3 Erlang distribution

The probability density function is:

$$f(t) = \frac{k\mu(k\mu t)^{k-1}}{(k-1)!} e^{-k\mu t} (t \geq 0) \tag{8}$$

In Equation (8), k is the order and μ is the mean. When $k = 1$, it is a negative exponential distribution function. When k increases, the image of Erlang distribution function gradually forms symmetry. When $k \geq 30$, the distribution is approximately normal. When $k \to \infty$, Erlang becomes a deterministic distribution function.

The distribution function is:

$$F(t) = 1 - e^{-k\mu t}(1 + k\mu t + \cdots + \frac{(k\mu t)^{k-1}}{(k-1)!} (t \geq 0) \tag{9}$$

Mathematical expectation:

$$E(t) = \frac{1}{\mu} \tag{10}$$

Variance:

$$Var(t) = \frac{1}{k\mu^2} \tag{11}$$

2.4 Pearson's test

Statistic of test:

$$\chi^2 = \sum_{i=1}^{r} \frac{(n_i - np_i)^2}{np_i} \tag{12}$$

In Equation (12), n_i is the measured frequency; np_i is the theoretical frequency; n is the total number of samples.

In the χ^2 distribution, the parameters depend only on the degrees of freedom R.

$$R = r - s - 1 \tag{13}$$

In Equation (13), s is the number of estimated parameters of theoretical distribution calculated by statistical distribution. s is 1 for Poisson distribution and 2 for normal distribution. r is the number of groups. According to Pearson's theorem, if the significance level α is given, the critical quantity χ_α^2 of the χ^2 distribution can be calculated. If the statistic χ_n^2 is less than $\chi_{\alpha,R}^2$, the hypothesis test is accepted; Otherwise, the assumed theoretical distribution is rejected. The critical values of partial chi-square test needed in this paper are shown in Table 1.

Table 1. Critical values of chi-square test

Degree of Freedom	Significance Level（α）					
	0.50	0.25	0.10	0.05	0.03	0.01
5	4.351	6.626	9.236	11.070	12.833	15.086
6	5.348	7.841	10.645	12.592	14.449	16.812
7	6.346	9.037	12.017	14.067	16.013	18.475
8	7.344	10.219	13.362	15.507	17.535	18.475
9	8.343	11.389	14.684	16.919	19.023	21.666
10	9.342	12.549	15.987	18.307	20.483	23.209
11	10.341	13.701	17.275	19.675	21.920	24.725

2.5 Evaluation indicator

R-Square:

$$R^2 = 1 - \frac{\sum (y_i - \hat{y}_i)^2}{\sum (y_i - \overline{y}_i)^2} \tag{14}$$

In Equation (14), y_i is the measured value, $\hat{y}_i$ is theoretical value, $\overline{y}_i$ is desired value, $R^2 \in [0,1]$, The closer the value of R^2 is to 1, the better the fitting degree of the curve to the measured value. On the contrary, the closer the value of R^2 is to 0, the worse the fitting degree of the curve to the measured value is.

SSE:

$$SSE = \sum_{i=1}^{n} (y_i - \hat{y}_i)^2 \tag{15}$$

In Equation (15), y_i is the measured value, $\hat{y}_i$ is theoretical value, n is The total number of samples. In the case of the same dataset, the smaller the SSE, the smaller the error and the better the model effect.

RMSE:

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n} (y_i - \hat{y}_i)^2}{n}} \tag{16}$$

In Equation (16), y_i is the measured value, $\hat{y}_i$ is theoretical value, n is The total number of samples. In the case of the same dataset, the smaller the RMSE, the smaller the error and the better the model effect.

3.Instance analysis

In 2020, 1060 ships entered and left a certain port in China, with an annual throughput of 90 million tons. Statistics on the arrival time of nearly 2,000 ships in 2020 were recorded. Firstly, the interval time rule of all ships in the port is analyzed, and then the frequency of interval time distribution of each ship type is analyzed according to different ship types. Matlab programming is used to calculate the theoretical frequencies of negative exponential distribution, Weibull distribution and Erlang distribution under the two conditions of whether ship types are distinguished. Then, the goodness of fit, sum variance and root mean square difference of the three probability distributions were obtained, and the most suitable probability distribution model to describe the ship arrival interval was found out. Finally, Pearson card method was used to further test.

3.1 Overall analysis of all ship arrival times

Without distinguishing ship types, the data of all ships arriving at the port are statistically analyzed. The fitting effect of the three probability distribution models is shown in Figure 1. The detailed goodness of fit and error are shown in Table 2. It can be seen from the relevant charts that the fitting effect of Weibull distribution curve is the best. Therefore, it is assumed that the arrival time interval of all ships follows Weibull distribution, and the chi-square test is shown in Table 3.

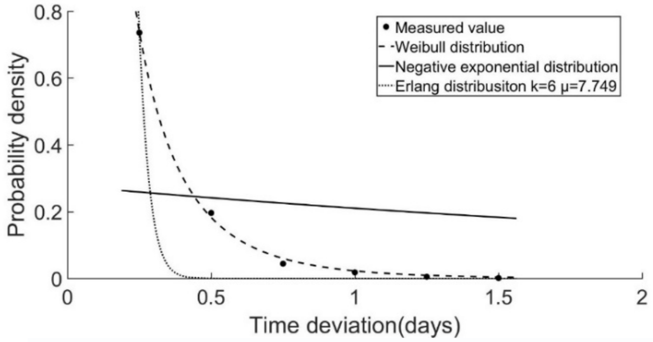


Figure 1. Distribution function of arrival time interval of all ships

Table 2. Fitting results of the distribution of all ships' arrival time intervals

Fitting result Fitting function	R-Square	SSE	RMSE
Weibull	0.9988	0.00048	0.01098
Negative exponent	0.1119	0.36920	0.27170
Erlang	0.9023	0.04062	0.09013

Table 3. Arrival time intervals of all ships subject to Weibull distribution Chi-square test analysis table

A_i	n_i	p_i	np_i	$\frac{(n_i - np_i)^2}{np_i}$
$A_1: t = 0.25$	1461	0.738	1464.192	0.007
$A_2: t = 0.5$	388	0.180	357.120	2.670
$A_3: t = 0.75$	88	0.052	103.168	2.230
$A_4: t = 1$	36	0.020	39.680	0.341
$A_5: t = 1.25$	9	0.007	13.888	1.720
$A_6: t = 1.5$	2	0.003	5.952	2.624
$\sum$	1984	1.0000	1984.000	9.593

When $\alpha = 0.05$, look-up table to $\chi^2_{0.05}(r - s - 1) = \chi^2_{0.05}(6 - 0 - 1) = \chi^2_{0.05}(5) = 11.070$, the Weibull distribution $\chi^2 = 9.593 < 11.070$, so under the significance level of 0.05, The hypothesis that the arrival time interval of all ships follows Weibull distribution ($k = 0.6725, \lambda = 10.9779$) is valid, that is, the theoretical value is consistent with the measured value.

3.2 Analysis of port arrival rules of different ship types

The arrival time data of four typical ship types at the port in 2020 were extracted for analysis, including 650 container ships, 1230 bulk carriers, 64 general cargo ships and 30 oil tankers.

3.2.1 Analysis of the regularity of container ship arrival time interval

The law of container ship arrival time interval is analyzed. The fitting effect of the three probability distribution models is shown in Figure 2, and the detailed goodness of fit and error are shown in Table 4. As can be seen from the relevant charts, when the Erlang order is 1, it is equivalent to a negative exponential distribution, and its curve coincides with the negative exponential. By comprehensive comparison, it can be concluded that the fitting effect of Weibull distribution curve is the best. Therefore, it is assumed that the arrival time interval of all ships follows Weibull distribution, and its chi-square test is shown in Table 5.

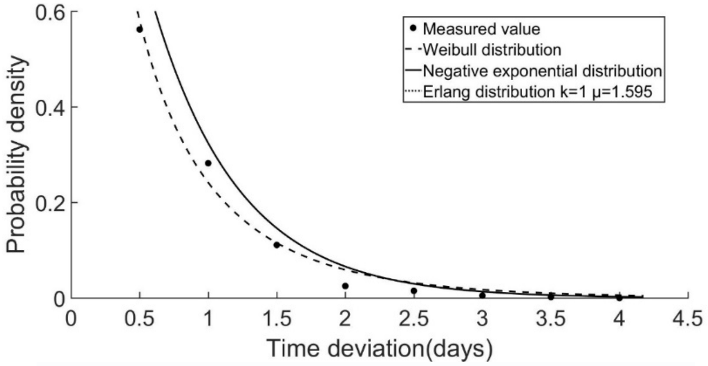


Figure 2. Distribution function of container ship arrival time interval

Table 4. Fitting results of container ship arrival time interval distribution

Fitting result Fitting function	R-Square	SSE	RMSE
Weibull	0.9870	0.00368	0.02479
Negative exponent	0.8952	0.02962	0.06505
Erlang	0.8952	0.02962	0.06505

Table 5. Analysis Table of Chi-square test with Weibull distribution for the arrival time of container ships

A_i	n_i	p_i	np_i	$\frac{(n_i - np_i)^2}{np_i}$
$\mathrm{A_1: t = 0.5}$	365	0.568	369.200	0.048
$\mathrm{A_2: t = 1}$	183	0.243	157.950	3.973
$\mathrm{A_3: t = 1.5}$	72	0.115	74.750	0.101
$\mathrm{A_4: t = 2}$	16	0.040	26.000	3.846
$\mathrm{A_5: t = 2.5}$	10	0.020	13.000	0.692
$\mathrm{A_6: t = 3}$	3	0.009	5.850	1.388
$\mathrm{A_7: t = 3.5}$	1	0.005	3.250	1.558
$\sum$	650	1.000	650.000	11.606

When $\alpha = 0.05$, look-up table to $\chi^2_{0.05}(\mathrm{r - s - 1}) = \chi^2_{0.05}(7 - 0 - 1) = \chi^2_{0.05}(6) = 12.592$, the Weibull distribution $\chi^2 = 11.606 < 12.592$, so under the significance level of 0.05, The hypothesis that all ship arrival time intervals follow Weibull distribution ($k = 0.7891, \lambda = 2.0103$) is valid, that is, the theoretical value is consistent with the measured value.

3.2.2 Analysis of time interval of arrival of bulk carrier

The law of arrival time interval of bulk carrier is analyzed. The fitting effect of the three probability distribution models is shown in Figure 3, and the goodness of fit and error in detail are shown in Table 6. As can be seen from the relevant charts, when the Erlang order is 1, it is equivalent to a negative exponential distribution, and its curve coincides with the negative exponential. By comprehensive comparison, it can be concluded that the fitting effect of Weibull distribution curve is the best. Therefore, it is assumed that the arrival time interval of bulk carrier complies with Weibull distribution, and its chi-square test is shown in Table 7.

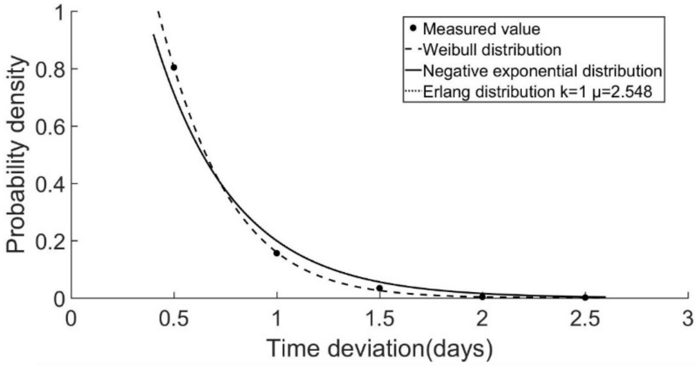


Figure 3. Distribution function of arrival time interval of bulk carrier

Table 6. Fitting results of distribution of arrival time interval of bulk carrier

Fitting result Fitting function	R-Square	SSE	RMSE
Weibull	0.9998	0.00008	0.005188
Negative exponent	0.9771	0.01083	0.05202
Erlang	0.9771	0.01083	0.05202

Table 7. Analysis Table of the arrival time interval of bulk container ships subject to Weibull distribution Chi-square test

A_i	n_i	p_i	np_i	$\frac{(n_i - np_i)^2}{np_i}$
$A_1: t = 0.5$	989	0.803	987.690	0.002
$A_2: t = 1$	192	0.159	195.570	0.065
$A_3: t = 1.5$	42	0.031	38.130	0.393
$A_4: t = 2$	5	0.005	6.150	0.215
$A_5: t = 2.5$	2	0.002	2.460	0.086
$\sum$	1230	1.000	1230.000	0.761

When $\alpha = 0.05$, look-up table to $\chi^2_{0.05}(r - s - 1) = \chi^2_{0.05}(5 - 0 - 1) = \chi^2_{0.05}(4) = 9.488$， Weibull distribution $\chi^2 = 0.761$＜9.488. Therefore, at the significance level of 0.05, the hypothesis that the arrival time interval of all ships follows Weibull distribution $k = 1.1630, \lambda = 2.6659$) is valid, that is, the theoretical value is consistent with the measured value.

3.2.3 Analysis of the time interval of arrival of general cargo ships

By analyzing the law of the arrival time interval of general cargo ships, the fitting effect of the three probability distribution models is shown in Figure 4, and the goodness of fit and error in detail are shown in Table 8. As can be seen from the relevant charts, when the Erlang order is 1, it is equivalent to a negative exponential distribution, and its curve coincides with the negative exponential. By comprehensive comparison, it can be concluded that the fitting effect of Weibull distribution curve is the best. Therefore, it is assumed that the arrival time interval of general cargo ships complies with Weibull distribution, and its chi-square test is shown in Table 9.

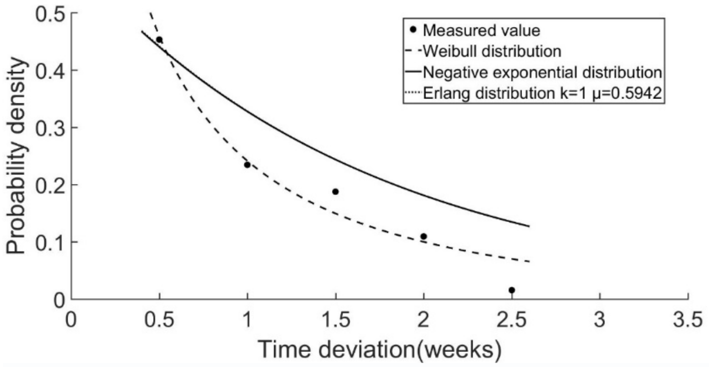


Figure 4. Distribution function of arrival time interval of general cargo ship

Table 8. Fitting results of distribution of arrival time interval of general cargo ship

Fitting result Fitting function	R-Square	SSE	RMSE
Weibull	0.9573	0.00459	0.03915
Negative exponent	0.7088	0.03133	0.08851
Erlang	0.7088	0.03133	0.08851

Table 9. Arrival time interval of cargo container ship subject to Weibull distribution Chi-square test analysis Table

A_i	n_i	p_i	np_i	$\frac{(n_i-np_i)^2}{np_i}$
$\mathrm{A}_1\mathrm{:t}=0.5$	29	0.460	29.440	0.007
$\mathrm{A}_2\mathrm{:t}=1$	15	0.241	15.424	0.012
$\mathrm{A}_3\mathrm{:t}=1.5$	12	0.149	9.536	0.637
$\mathrm{A}_4\mathrm{:t}=2$	7	0.100	6.400	0.056
$\mathrm{A}_5\mathrm{:t}=2.5$	1	0.050	3.200	1.513
$\sum$	64	1.000	64.000	2.224

When $\alpha = 0.05$, look-up table to $\chi^2_{0.05}(\mathrm{r}-\mathrm{s}-1)=\chi^2_{0.05}(5-0-1)=\chi^2_{0.05}(4)=9.488$, Weibull distribution $\chi^2 = 2.224 < 9.488$. Therefore, at the significance level of 0.05, the hypothesis that all ship arrival time intervals follow Weibull distribution ($k = 0.6597, \lambda = 1.1746$) is valid, that is, the theoretical value is consistent with the measured value.

3.2.4 Analysis of time interval of tanker arrival at port

The law of arrival time interval of oil tanker is analyzed. The fitting effect of the three probability distribution models is shown in Figure 5, and the detailed goodness of fit and error are shown in Table 10. It can be seen from the relevant charts that the fitting effect of Weibull distribution curve and Erlang distribution is the best. Therefore, it is assumed that the arrival time interval of oil tanker conforms to both Weibull distribution and Erlang distribution, and the chi-square test is shown in Table 11-12.

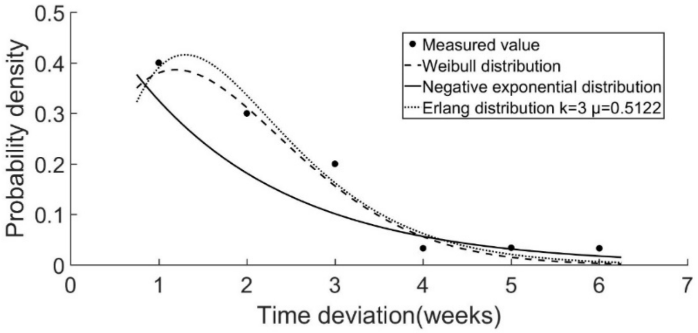


Figure 5. Distribution function of tanker arrival time interval

Table 10. Fitting results of tanker arrival time interval distribution

Fitting result Fitting function	R-Square	SSE	RMSE
Weibull	0.9667	0.00421	0.03246
Negative exponent	0.7625	0.03008	0.07756
Erlang	0.9644	0.00450	0.03002

Table 11. Chi-square test analysis Table with Weibull distribution for the arrival time interval of oil tanker

A_i	n_i	p_i	np_i	$\frac{(n_i - np_i)^2}{np_i}$
$\mathrm{A_1: t = 1}$	12	0.392	11.760	0.017
$\mathrm{A_2: t = 2}$	9	0.316	9.480	0.219
$\mathrm{A_3: t = 3}$	6	0.185	5.550	0.314
$\mathrm{A_4: t = 4}$	1	0.057	1.710	1.874
$\mathrm{A_5: t = 5}$	1	0.028	0.840	0.263
$\mathrm{A_5: t = 6}$	1	0.022	0.660	0.004
$\sum$	30	1.000	30.000	2.690

When $\alpha = 0.05$, look-up table to $\chi^2_{0.05}(\mathrm{r-s-1}) = \chi^2_{0.05}(6-0-1) = \chi^2_{0.05}(5) = 11.070$, the Weibull distribution $\chi^2 = 2.690 < 11.070$. Therefore, at the significance level of 0.05, the hypothesis that all ship arrival time intervals follow Weibull distribution ($k = 1.6940, \lambda = 0.4950$) is valid, that is, the theoretical value is consistent with the measured value.

Table 12. Chi-square test Analysis Table of the arrival time interval of oil tanker subject to third-order Erlang distribution

A_i	n_i	p_i	np_i	$\frac{(n_i - np_i)^2}{np_i}$
$\mathrm{A_1: t = 1}$	12	0.390	11.700	0.008
$\mathrm{A_2: t = 2}$	9	0.336	10.080	0.116
$\mathrm{A_3: t = 3}$	6	0.163	4.890	0.252
$\mathrm{A_4: t = 4}$	1	0.062	1.860	0.398
$\mathrm{A_5: t = 5}$	1	0.021	0.630	0.217
$\mathrm{A_5: t = 6}$	1	0.028	0.840	0.030
$\sum$	30	1.000	30.000	1.021

When $\alpha = 0.05$, look-up table to $\chi^2_{0.05}(\mathrm{r-s-1}) = \chi^2_{0.05}(6-0-1) = \chi^2_{0.05}(5) = 11.070$ Erlang distribution $\chi^2 = 1.021 < 11.070$. Therefore, at the significance level of 0.05, the hypothesis that all ship arrival time intervals follow the Erlang distribution ($k = 3, \mu = 0.5122$) is valid, that is, the theoretical value is consistent with the measured value.

4.Conclusion

Through the statistical analysis of the arrival time of ships in a port, the law of the arrival time interval of ships can be fitted by Weibull distribution density function curve, negative exponential distribution density function curve and Erlang distribution density function curve, and the fitting degree is relatively ideal. The optimal probability distribution model was obtained by comparing the goodness of fit (R-Square), sum of squared errors (SSE) and root mean square error (RMSE). The chi-square test method was used to test the probability distribution. The regularity of arrival time interval of all ships and distinguished ship types was more consistent with Weibull distribution.

When the law of arrival of all ships is studied, due to the diversity of ship types, the arrival of ships is still random on the whole, so the time interval of successive arrival of ships is Weibull distributed. When studying the law of arrival of different types of ships, because the goods carried by different types of ships have their own attributes, different ship types also have diversity in the distribution of arrival time intervals. However, the successive arrival time intervals of container ships, bulk carriers and general cargo ships also follow Weibull distribution. However, the interval of successive arrival time of oil tanker conforms to Weibull distribution and third-order Erlang distribution. This conclusion has certain reference value for the rational allocation of port resources and the optimization of port facilities in the future, and also provides an effective analysis method for the arrival law of ships in other ports.

REFERENCES

[1] Liu Jingxian, Li Yunbin. Statistical Analysis of Ship Arrival Law in Main Channel of Tianjin Port _ Liu Jingxian [J]. Journal of Wuhan University of Technology (Traffic Science and Engineering Edition), 2008(2): 351-353, 357.

[2] Yu Jin, Zhang Wei, Jiang Jihong, et al. Probability Distribution Characteristics of Ship Flow in Xijiang Waterway _ Yu Jin [J]. Journal of Traffic and Transportation Engineering, 2006(2): 88-93.

[3] Wang Nuo, Xu Lingjie, Song Nanqi, et al. Distribution Function and Empirical Study of Container Liner Arrival Law _ Wang Nuo [J]. Journal of Dalian Maritime University, 2013, 39(4): 107-110.

[4] Wu Di, Wang Nuo, Wu Nuan, et al. Study on the Law of Late Shift Arrival of Container Liner and Its Application _ Wu Di. Journal of Dalian Maritime University, 2015, 41(1): 77-82.

[5] Song Y T, Wang N U. On probability distributions of the time deviation law of container liner ships under interference uncertainty[J]. Journal of The Royal Statistical Society Series A-statistics in Society, 2021,184(1):354-367

IM2DP: An Intensity-Based Approach to Loop Closure Detection and Optimization for LiDAR Mapping

Lu Qiang*, Jiahang Liu
Department of Astronautics, Nanjing University of Aeronautics and Astronautics, Nanjing 210016, China
* Corresponding author: qianglu@nuaa.edu.cn

ABSTRACT

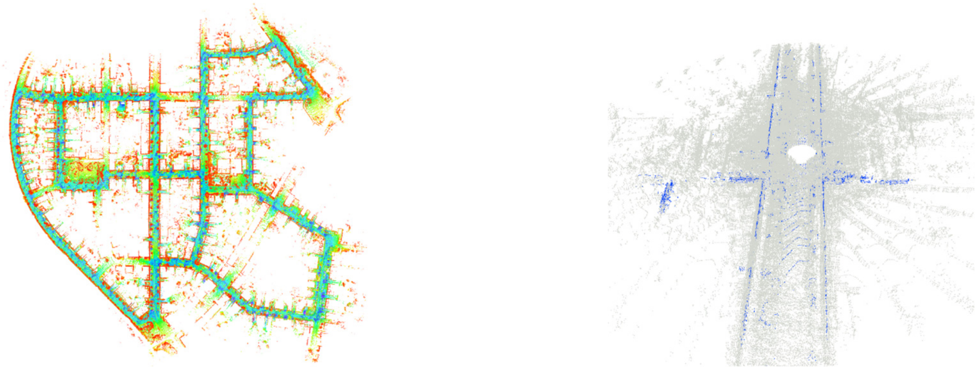
We propose a brand-new global point cloud descriptor called IM2DP that combines shape and intensity data to conduct loop closure detection. By incorporating intensity data, our method expands the multiview 2D projection (M2DP) descriptor. We calculate intensity characteristics from multiple 2D projections of the point cloud and apply them to the M2DP shape features. Then, singular value decomposition (SVD) is used to compute the compressed descriptors for reducing the dimensionality. We combine this algorithm with the LOAM and conduct experiments on the KITTI public dataset. The results show that our approach is effective in reducing trajectory error and shows competitiveness when compared to the M2DP.

Keywords: intensity data, loop closure detection, global optimization

1. INTRODUCTION

Simultaneous localization and mapping (SLAM) has been actively explored and implemented in the field of robotics and autonomous driving in recent years. A large amount of related-work to SLAM has been published in literature. Solutions to SLAM problems have evolved from early statistical-based approaches[1] to current graph optimization approaches[2]. Over the past decade, typical LiDAR SLAM solutions have evolved into a framework with four main parts, including front-end odometry, loop closure detection, mapping, and global optimization[3-5].

The ability to recognize a visited place is referred to loop closure detection, also known as place identification. In SLAM problems, the estimation of states and trajectories is often accompanied by unavoidable drifts. The robot can decrease drift errors by identifying revisited places. In addition, it can prevent multiple registrations of the same landmarks, thus creating a globally consistent map.



(a)

(b)

Figure 1. A KITTI dataset example using the suggested strategy. (a) shows the mapping result of the proposed method on KITTI dataset sequence 00. (b) shows a loop closure detection example, where the colorful portion represents the point cloud in the current frame and the gray portion represents the historical map.

Typically, vision-based or LiDAR-based techniques can be used to perform loop closure detection. Vision-based techniques have the advantage of being able to extract a wealth of semantic data from the images. The images are compactly structured, low-dimensional (2D), and simple to process using computer vision methods. Its drawbacks, however, include being hypersensitive to changes in lighting, moving objects, seasonal variations, and other elements. In ORB-SLAM[7-8],

International Conference on Computer Graphics, Artificial Intelligence, and Data Processing (ICCAID 2022)
edited by Ruishi Liang, Jing Wang, Proc. of SPIE Vol. 12604, 126040H
© 2023 SPIE · 0277-786X · doi: 10.1117/12.2674664

the bag-of-words model (BoW)[6] has been utilized. In comparison to VLAD descriptor[10], the learning-based technique NetVLAD[9] achieves superior performance. Contrarily, point clouds are three-dimensional (3D), and the added complexity makes computation more difficult. Point clouds, on the other hand, are impervious to the effects of changing lighting and seasons and can directly represent the geometrical characteristics of the surroundings.

Inspired by the study of location identification using point cloud intensity[11-12], in this paper, we present an innovative descriptor called intensity multiview 2D projection (IM2DP), which extends the M2DP descriptor by merging intensity information to increase its accuracy in loop closure detection. Intensity distributions are calculated using several 2D projections of the point cloud along with geometric data. Geometrical characteristics are connected to these intensity distributions. Then, using the same procedures as M2DP, we minimize the dimension by applying SVD to produce a compact signature, which we then utilize as our final descriptor.

The algorithm is integrated into LOAM and the results of the loop closure detection are used to compute the transformation relationships between the corresponding keyframes and add them to the factor graph as loop closure factors for global optimization to reduce errors. For global optimization, the most common approach is pose graph optimization[13]. Several popular open-source C++ libraries[14-16] can be used to solve global optimization problems, among which GTSAM[15] archives good optimization result with low computational cost[17]. For this reason, GTSAM is used in the present study for global optimization.

The following is the structure of this paper. Section 2 of this paper reviews previous research on place recognition. Section 3 provides information on how our descriptor IM2DP is implemented. We discuss the experimental results of the algorithm used on the KITTI dataset in Section 4. Finally, we conclude the paper in Section 5.

2. RELATED WORK ON POINT CLOUD LOOP CLOSURE DETECTION

Numerous approaches for mobile robots to recognize their global locations have been put forward. The BoW model, which compares the current scene to the global map using a previously trained visual vocabulary, is frequently used in vision-based approaches[6,18]. Visual receptors, however, are extremely sensitive to variations in lighting and perspective. Numerous studies have attempted to solve this issue, yet vision-based solutions are frequently still insufficient. The robustness of LIDAR to light and variations in perspective has made LIDAR-based approaches quite popular. Local descriptor-based method, global descriptor-based method, and learning-based method can be used to categorize LIDAR-based location identification techniques.

2.1 Local descriptor-based method

The foundation of local descriptor-based approaches is the extraction and matching of local point cloud descriptors. FPFH[19] utilized local surface normal vectors to produce local descriptors, which decreased the computational cost of the prior method. To create descriptors, SHOT[20] separated the sphere's center-of-mass zone into regions. Each region's normal angle histograms were then compiled. A key point voting mechanism was proposed by Bosse et al[21]. To find potential location matches, a local descriptor database was queried for a fixed number of nearest neighbor votes for each key point, then the results were pooled. A technique using 3D fragment matching was suggested by Dubé et al[22]. The process pulled pieces from the point cloud, compared them to fragments from previously visited locations, and then used a geometric verification step to identify candidates for place identification. Guo et al[12] proposed a new probabilistic keypoint voting approach and included intensity information to SHOT descriptors. The recognition of distinct critical locations with great reproducibility is still a difficult task, though.

2.2 Global descriptor-based method

The extraction of significant points and laborious local geometry computations are typically necessary for the recognition of local descriptors. Global descriptor matching is more effective than local descriptor matching. A voxel grid was used by Wohlkinger et al[23] to approximate the real surface of a point cloud using ESF, which did so without computing normals by employing shape characteristics instead. A global description based on a point cloud height distribution histogram was proposed by Rö Kling et al[24]. M2DP was suggested by He et al[25], which projected the point cloud into numerous 2D planes and defined the point cloud using the density distribution in the 2D planes. A point cloud description dubbed DELIGHT[11], which alluded to the SHOT feature extraction technique, was proposed by Cop et al. Kim et al[26] presented Scan Context (SC), a self-centered spatial descriptor. SC transformed the whole point cloud from a 3D LiDAR scan into a matrix using the height information contained in the point cloud. It has demonstrated that the strategy of simply extracting the highest points of the visible point cloud was superior to other global descriptors already in use.

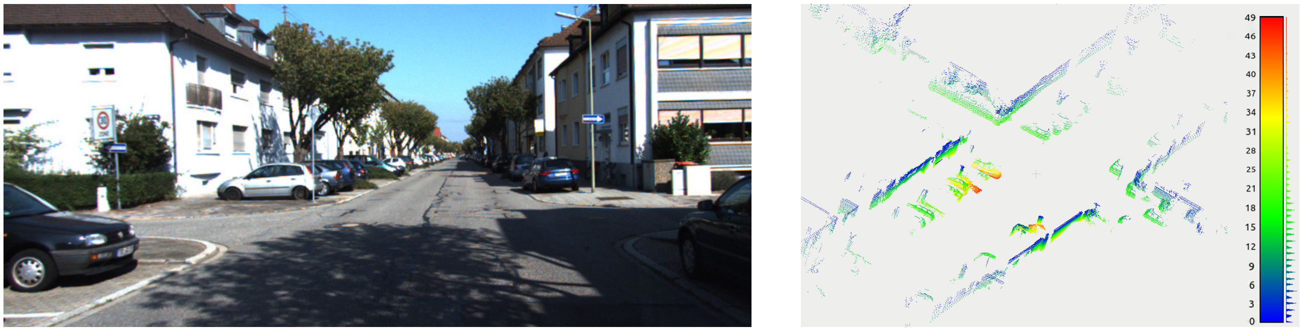
2.3 Learning-based method

Several approaches based on learning have also been proposed. LocNet was a semi-manual deep learning system presented by Yin et al[27] that addresses the place recognition problem as a comparable modeling problem. Deep learning, which incorporated the networks of PointNet[29] and NetVLAD[9], was used by Uy et al to extract point cloud descriptors. CNN[30] was utilized by Kim et al to train SC pictures for long-term position recognition. These approaches' shortcomings were their high data requirements and lengthy training times.

3. METHODOLOGY

The M2DP descriptor, which is generated using shape data obtained from numerous 2D projected perspectives of the point cloud, indicates a point cloud with compact features. It is advised to read the original[25] for additional information on its original construction. By fusing the calculated intensity histogram with the extracted point cloud's shape properties, we expand its design. By utilizing the intensity data in the point cloud, we hope to increase descriptiveness and detection accuracy.

First, the input point cloud's shape features are created using the M2DP algorithm. After that, we add the process of gathering intensity data and incorporating it into the descriptors. Multiple 2D planes are generated using distinct p azimuth angles $[0,\frac{\pi}{p},\frac{2\pi}{p},\ldots,\pi]$, and q elevation angles $[0,\frac{\pi}{2q},\frac{2\pi}{2q},\ldots,\frac{\pi}{2}]$, up to a total of $p\times q$. The input point cloud P is projected onto each 2D plane X, and the shape signature matrix and intensity signature matrix are generated from each projection P_X respectively. Then, we use singular value decomposition (SVD) to reduce the dimensionality and use the first left and right singular vectors of the signature matrix as descriptors, and finally, the descriptors computed from the two matrices are combined to obtain the final IM2DP descriptors.



(a) Camera view of a crossroad.

(b) Intensity scan reading of the crossroad.

Figure 2. An example of intensity reading from KITTI sequence 00.

3.1 Point cloud pre-processing

LiDAR senses its surroundings by sending and receiving laser beams. In general, distance is calculated by trip time, whereas surface reflectivity can be estimated using the returned energy level (i.e., intensity). The intensity values reveal the structure of the surrounding surface reflections. Existing LiDAR research demonstrates that various objects produce varied intensity readings[31]. Retroreflective material, such as metal plate, typically returns high value and concrete returns low value, as seen in Figure 2, where we present an example of the KITTI dataset. The same location is shown in both the point cloud and the image. The intensity return for 3D LiDAR typically consists of an 8-byte integer (0-255).

However, the intensity channel is noisy because it is influenced by the acquisition shape (e.g., distance), instrument effects (e.g., transmitted energy), and target surface properties (e.g., roughness, surface reflectance). Calibration is therefore required to lessen interference from these other factors. We use the method[12] to calibrate the intensity reading.

In applications, it is necessary to perform some pre-processing on the LiDAR scans to remove redundant information. It has been noted that LiDAR noise rises with distance. As a result, untrustworthy sites are initially removed from LiDAR data by defining a distance threshold, *Lmax*. In addition, ground points are often not important for describing the surrounding spatial information, so they are optimized beforehand using a relatively effective method[32].

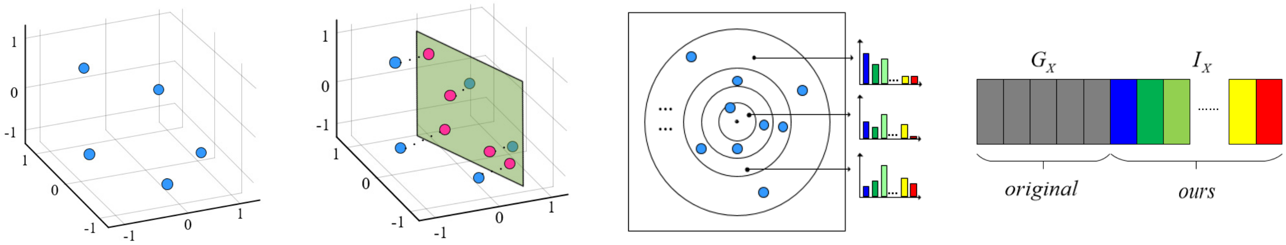
3.2 IM2DP descriptors

To guarantee the shift and rotation invariance of the IM2DP descriptor, our initial step is to adhere to the original M2PD algorithm. We use the point cloud's centroids (i.e., mean points) as the descriptors' reference frame origin and move the point cloud with zero mean. After that, we run these points using principal component analysis (PCA). We assume that each point cloud has two dominating orientations and utilize the first and second principal components of the descriptor's reference frame as the X and Y axes.

In Figure 3, we describe how an IM2DP descriptor from each 2D projection is computed. Figure 3-(b) shows the projection of an input point cloud *P* onto a 2D plane *X*. *X* indicates a perspective that is determined by the azimuth angle θ and elevation angle φ of the normal vector *m*, with respect to the origin. The 2D plane is split into *l* concentric circles, with radii [r, 2^2r , . . . , l^2r] centered at the centroid. The maximum radius is determined by our distance threshold *Lmax*. Following that, a shape feature matrix is calculated for each projection P_X by first splitting each concentric circle into *t* bins, each of which is indexed by the *x*-axis, and then counting the points that fall inside each bin.

To fuse the intensity information into the shape features, we first divide the intensity values in the range of 256 into *n* intervals, then compute the intensity histogram for each concentric circle, and connect the intensity histograms of concentric circles on all projection planes to generate the intensity feature matrix. Finally, the obtained shape feature matrix and intensity feature matrix are decomposed by SVD separately to obtain the final descriptors, as shown in Figure 3-(c).

Our method aims to calculate the intensity histogram for each concentric circle of the two-dimensional projection instead of the bin of each shape, dividing the intensity range into *n* intervals instead of each intensity value, thus, we prevent a si gnificant rise in dimensionality.



(a) Example of a point cloud.

(b) Project a point cloud on a 2D plane.

(c) Compute intensity histograms for each concentric circle.

(d) Concatenate shape and intensity signatures.

Figure 3. Additional IM2DP procedures to generate an intensity signature for each 2D point cloud projection.

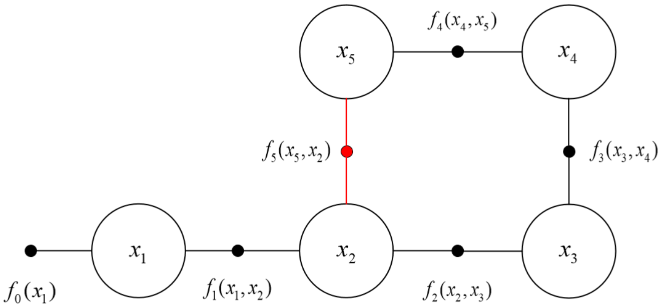


Figure 4. A factor graph example.

3.3 Loop closure matching and back-end optimization

In the sequence, IM2DP descriptors are calculated for each point cloud. The *L2* norm matching algorithm is then used to compare the most similar descriptors amongst point clouds in order to identify the loop closure. We follow the setup[25] and use a window size of ±50 frames to exclude neighbors of the current frame during the matching process. If a match is below the *L2* distance threshold, it is determined to be a cyclic closure.

We use a factor graph to optimize the back-end pose.The factor graph is one of the probabilistic graphical models. As seen in Figure 4, it has factor nodes, variable nodes, and edges. The black ones are the odometry measurements; the red ones are the loop closure factors added by our algorithm. The circles represent the variable nodes. Each edge connects a factor node and a variable node.

In the LiDAR SLAM problem, the variable nodes represent the pose to be optimized and the factor nodes represent the LiDAR ranging measurements. The pose estimation problem can be transformed into MAP estimation of the variable nodes as shown in Equation (1).

$$\{x\}^{*}=\mathrm{argmax}(x_0)\prod P(x_k|,x_{k-1}|,u_k) \tag{1}$$

The issue is a nonlinear least squares problem when Gaussian noise is assumed, and it can be resolved iteratively using the Gauss-Newton method or the Levenberg-Marquardt method. As the system runs, the size of the factor graph expands and the sparsity of the Jacobian matrix in the least-squares problem gradually decreases, which causes inconvenience to the solution. Therefore, it is necessary to efficiently update the pose while maintaining the sparsity of the problem. Kaess et al. address this problem by incremental smoothing and mapping (iSAM)[33] and iSAM2[34]. In factor graphs without closed-loop, only the latest node is optimized; while in factor graphs with closed-loop, only the nodes in the closed-loop chain are optimized. In this paper, we use the Levenberg-Marquardt method from the GTSAM library to solve the nonlinear least squares problem. We use iSAM2 to incrementally construct the map and optimize the pose with closed-loop constraints. When adding closed-loop constraints, we use a Cauchy robustness noise model combined with ICP matching adaptation scores to improve the robustness of the system to false-positive loop closure detection results.

4. THE EXPERIMENTAL TEST

To validate the effectiveness of our method in real scenes, we adopted sequences 00,05,06 and 09 from the KITTI dataset[35] for evaluation. All tests were done based on the Robot Operating System (ROS) that was installed on a laptop with an Intel i5-7300HQ processor, a 16 GB RAM and the Ubuntu platform.

4.1 Experiments settings

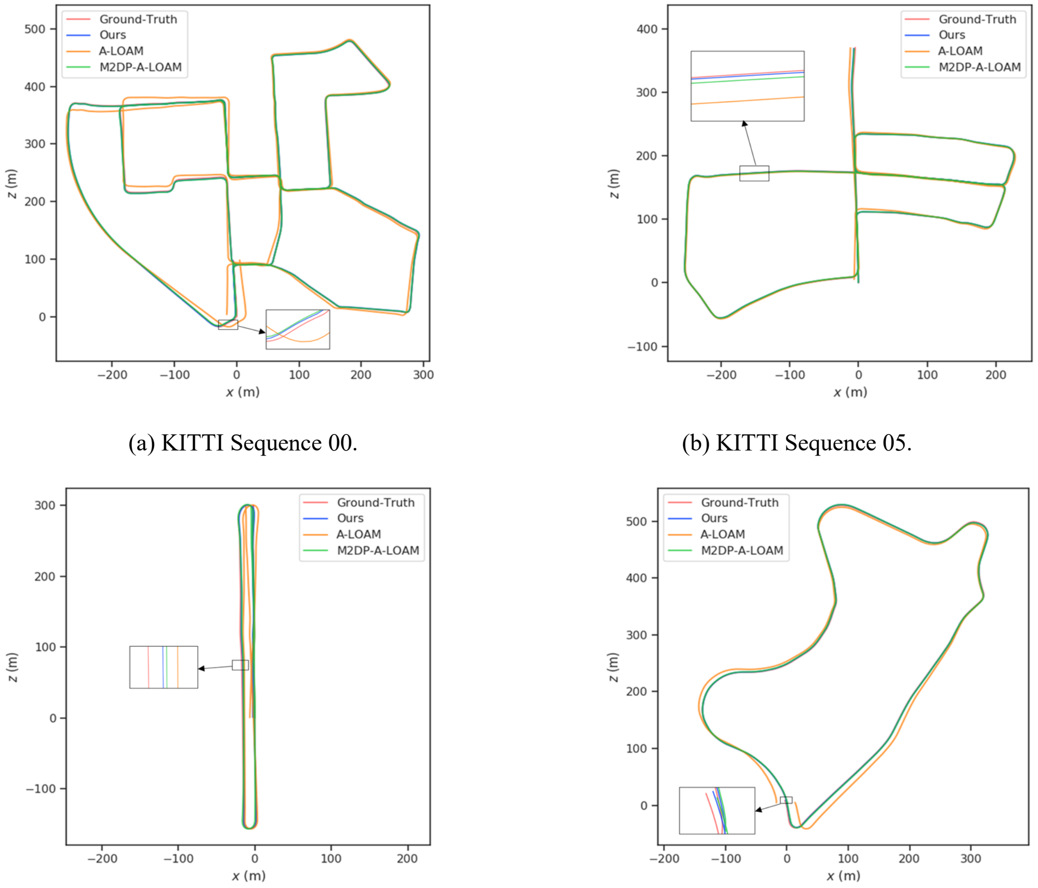
As indicated in Table 1. we use the parameter values[25] for both M2DP and IM2DP descriptors. We set the number of intensity histogram intervals equal to the number of bins per concentric circle *n* = *t*. The IM2DP descriptor ends up being a vector of size 384, compared to the original size of 192 for the M2DP descriptor.

Table 1. M2DP and IM2DP parameters.

Parameters	M2DP	IM2DP
Azi*m*uth angles (*p*)	4	4
Elevation angles (*q*)	16	16
Concentric circles (*l*)	8	8
Shape bins (*t*)	16	16
Intensity bins (*n*)	—	16

4.2 Evaluation on KITTI dataset

Figure 5 shows a comparison of the trajectories of the method in this paper, the LOAM method under the original M2DP algorithm, and the community-maintained version of the A-LOAM algorithm without loop closure detection module compared to the ground truth, and we can see that for the four sequences in the experiment, the method proposed in this paper outperforms the competing methods by providing the trajectories closest to the ground truth.



(a) KITTI Sequence 00.

(b) KITTI Sequence 05.

(c) KITTI Sequence 06.

(d) KITTI Sequence 09.

Figure 5. Comparison of trajectories on KITTI dataset.

We also calculate the Absolute Pose Error (APE) and the Relative Pose Error (RPE) for each method under four sequences, i.e., root mean square error (RMSE), median, mean and standard deviation (Std). Table 2 shows the results. Where the A-LOAM algorithm uses the baseline results of the community-maintained version[36], without the loop closure mechanism.

Table 2. Results on KITTI sequence 00,05,06 and 09.

Sequence	Methods	APE(m)				RPE(m)			
		RMSE	Median	Mean	Std	RMSE	Median	Mean	Std
00	A-LOAM	8.02	5.84	6.88	4.11	3.77	3.54	3.67	0.87
	M2DP-A-LOAM	1.38	1.10	1.25	0.58	3.75	3.53	3.65	0.87
	Ours	**1.29**	**1.04**	**1.14**	**0.52**	**3.74**	**3.52**	**3.64**	**0.87**
05	A-LOAM	3.61	2.54	3.15	1.76	3.76	3.36	3.66	0.89
	M2DP-A-LOAM	0.98	**0.76**	**0.87**	0.45	3.76	**3.36**	3.65	0.89
	Ours	**0.95**	0.83	0.88	**0.39**	**3.74**	3.52	**3.64**	**0.87**
06	A-LOAM	4.46	3.44	3.87	2.2	4.22	3.81	4.09	1.03
	M2DP-A-LOAM	1.60	1.42	1.45	0.67	4.33	3.93	4.21	**1.02**
	Ours	**1.28**	**1.18**	**1.21**	**0.43**	**4.22**	**3.81**	**4.09**	1.03
09	A-LOAM	8.52	4.80	7.01	4.84	**3.92**	**3.74**	**3.80**	**0.93**
	M2DP-A-LOAM	1.52	1.43	1.45	0.46	3.99	3.77	3.83	0.94
	Ours	**1.36**	**1.29**	**1.29**	**0.43**	3.94	3.77	3.83	0.93

Proc. of SPIE Vol. 12604 126040I-6

From the statistics, we can see that the error metric of the proposed method in this paper decreases to different degrees in all scenarios. Specifically, in terms of RMSE and std of APE, the proposed method reduces 6.5% and 10.3% over M2DP-A-LOAM for 00 sequences, 3.1% and 13% for 05 sequences, performs best for 06 sequences with 20% and 35.8% reduction, and 10.5% and 6.5% for 09 sequences. The results show that the method performs well and can improve the map quality.

5. CONCLUSION

In this paper, we develop a point cloud descriptor for loop closure detection that integrates point cloud shape features with intensity data. It is a development of M2DP, a global point cloud descriptor. In addition to maintaining the useful form properties of M2DP, the IM2DP constructs intensity features by computing and concatenating intensity histograms from several 2D projections of the point cloud. This method is combined into the LOAM algorithm to correct structural errors and reduce cumulative errors. Our method performs better than original M2DP algorithm on the KITTI Odometry dataset, showing its competitiveness.

REFERENCES

[1] H. Durrant-Whyte and T. Bailey, "Simultaneous localization and mapping: part I," in IEEE Robotics & Automation Magazine, vol. 13, no. 2, pp. 99-110, June 2006.

[2] D. Koller and N. Friedman, Probabilistic graphical models: principles and techniques. MIT press, 2009.

[3] H. Wang, C. Wang and L. Xie, "Intensity-SLAM: Intensity Assisted Localization and Mapping for Large Scale Environment," in IEEE Robotics and Automation Letters, vol. 6, no. 2, pp. 1715-1721, April 2021.

[4] L. Li et al., "SA-LOAM: Semantic-aided LiDAR SLAM with Loop Closure," 2021 IEEE International Conference on Robotics and Automation (ICRA), 2021, pp. 7627-7634.

[5] J. Jiang, J. Wang, P. Wang, P. Bao and Z. Chen, "LiPMatch: LiDAR Point Cloud Plane Based Loop-Closure," in IEEE Robotics and Automation Letters, vol. 5, no. 4, pp. 6861-6868, Oct. 2020.

[6] D. Galvez-López and J. D. Tardos, "Bags of Binary Words for Fast Place Recognition in Image Sequences," in IEEE Transactions on Robotics, vol. 28, no. 5, pp. 1188-1197, Oct. 2012.

[7] R. Mur-Artal, J. M. M. Montiel and J. D. Tardós, "ORB-SLAM: A Versatile and Accurate Monocular SLAM System," in IEEE Transactions on Robotics, vol. 31, no. 5, pp. 1147-1163, Oct. 2015.

[8] R. Mur-Artal and J. D. Tardós, "ORB-SLAM2: An Open-Source SLAM System for Monocular, Stereo, and RGB-D Cameras," in IEEE Transactions on Robotics, vol. 33, no. 5, pp. 1255-1262, Oct. 2017.

[9] R. Arandjelovic, P. Gronat, A. Torii, T. Pajdla and J. Sivic, "NetVLAD: CNN Architecture for Weakly Supervised Place Recognition," 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2016, pp. 5297-5307.

[10] R. Arandjelovic and A. Zisserman, "All About VLAD," 2013 IEEE Conference on Computer Vision and Pattern Recognition, 2013, pp. 1578-1585.

[11] K. P. Cop, P. V. K. Borges and R. Dubé, "Delight: An Efficient Descriptor for Global Localisation Using LiDAR Intensities," 2018 IEEE International Conference on Robotics and Automation (ICRA), 2018, pp. 3653-3660.

[12] J. Guo, P. V. K. Borges, C. Park and A. Gawel, "Local Descriptor for Robust Place Recognition Using LiDAR Intensity," in IEEE Robotics and Automation Letters, vol. 4, no. 2, pp. 1470-1477, April 2019.

[13] L. Feng, M. E. Evangelos, “Globally Consistent Range Scan Alignment for Environment Mapping.” Autonomous Robots, 4, 333–349 (1997).

[14] S. Agarwal and K. Mierle, "Ceres solver: Tutorial & reference," Google Inc, vol. 2, no. 72, p. 8, 2012.

[15] F. Dellaert, “Factor graphs and GTSAM: A hands-on introduction,” Georgia Institute of Technology, 2012.

[16] G. Grisetti, et al, "g2o: A general framework for (hyper) graph optimization." Proceedings of the IEEE international conference on robotics and automation (ICRA), Shanghai, China. 2011.

[17] A. Jurić, F. Kendeš, I. Marković and I. Petrović, "A Comparison of Graph Optimization Approaches for Pose Estimation in SLAM," 2021 44th International Convention on Information, Communication and Electronic Technology (MIPRO), 2021, pp. 1113-1118.

[18] A. Glover, W. Maddern, M. Warren, S. Reid, M. Milford and G. Wyeth, "OpenFABMAP: An open source toolbox for appearance-based loop closure detection," 2012 IEEE International Conference on Robotics and Automation, 2012, pp. 4730-4735.

[19] R. B. Rusu, N. Blodow and M. Beetz, "Fast Point Feature Histograms (FPFH) for 3D registration," 2009 IEEE International Conference on Robotics and Automation, 2009, pp. 3212-3217.

[20] S. Salti, F. Tombari, and L. D. Stefano, "SHOT: Unique signatures of histograms for surface and texture description." Computer Vision and Image Understanding 125 (2014): 251-264.

[21] M. Bosse and R. Zlot, "Place recognition using keypoint voting in large 3D lidar datasets," 2013 IEEE International Conference on Robotics and Automation, 2013, pp. 2677-2684.

[22] R. Dubé, D. Dugas, E. Stumm, J. Nieto, R. Siegwart and C. Cadena, "SegMatch: Segment based place recognition in 3D point clouds," 2017 IEEE International Conference on Robotics and Automation (ICRA), 2017, pp. 5266-5272.

[23] W. Wohlkinger and M. Vincze, "Ensemble of shape functions for 3D object classification," 2011 IEEE International Conference on Robotics and Biomimetics, 2011, pp. 2987-2992.

[24] T. Röhling, J. Mack and D. Schulz, "A fast histogram-based similarity measure for detecting loop closures in 3-D LIDAR data," 2015 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), 2015, pp. 736-741.

[25] L. He, X. Wang and H. Zhang, "M2DP: A novel 3D point cloud descriptor and its application in loop closure detection," 2016 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), 2016, pp. 231-237.

[26] G. Kim and A. Kim, "Scan Context: Egocentric Spatial Descriptor for Place Recognition Within 3D Point Cloud Map," 2018 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), 2018, pp. 4802-4809.

[27] H. Yin, L. Tang, X. Ding, Y. Wang and R. Xiong, "LocNet: Global Localization in 3D Point Clouds for Mobile Vehicles," 2018 IEEE Intelligent Vehicles Symposium (IV), 2018, pp. 728-733.

[28] M. A. Uy and G. H. Lee, "PointNetVLAD: Deep Point Cloud Based Retrieval for Large-Scale Place Recognition," 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2018, pp. 4470-4479.

[29] R. Q. Charles, H. Su, M. Kaichun and L. J. Guibas, "PointNet: Deep Learning on Point Sets for 3D Classification and Segmentation," 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2017, pp. 77-85.

[30] G. Kim, B. Park and A. Kim, "1-Day Learning, 1-Year Localization: Long-Term LiDAR Localization Using Scan Context Image," in IEEE Robotics and Automation Letters, vol. 4, no. 2, pp. 1948-1955, April 2019.

[31] A. G. Kashani, et al, "A review of LiDAR radiometric processing: From ad hoc intensity correction to rigorous radiometric calibration." Sensors 15.11 (2015): 28099-28128.

[32] D. Zermas, I. Izzat and N. Papanikolopoulos, "Fast segmentation of 3D point clouds: A paradigm on LiDAR data for autonomous vehicle applications," 2017 IEEE International Conference on Robotics and Automation (ICRA), 2017, pp. 5067-5073.

[33] M. Kaess, A. Ranganathan and F. Dellaert, "iSAM: Incremental Smoothing and Mapping," in IEEE Transactions on Robotics, vol. 24, no. 6, pp. 1365-1378, Dec. 2008.

[34] M. Kaess, et al, "iSAM2: Incremental smoothing and mapping using the Bayes tree." The International Journal of Robotics Research 31.2 (2012): 216-235.

[35] A. Geiger, P. Lenz and R. Urtasun, "Are we ready for autonomous driving? The KITTI vision benchmark suite," 2012 IEEE Conference on Computer Vision and Pattern Recognition, 2012, pp. 3354-3361.

[36] HKUST Aerial Robotics Group, “A-LOAM: Advanced Implementation of LOAM,” 2018, https://github.com/HKUST-Aerial-Robotics/A-LOAM.

Multi-label Feature Selection Algorithm Based on HSIC-Lasso

Chengwen Li, Jianhui Li*, Jiadong Zhu
Foshan Polytechnic, Foshan, Guangdong, 528137, China
*Corresponding author: joe863@163.com

ABSTRACT

The Hilbert-Schmidt Independence Criterion (HSIC) was originally design-ed to measure the statistical dependence of distribution-based Hilbert spaces embedding in statistical inference. In recent years, due to the validity and efficiency of this standard, it has been witnessed that this criterion can tackle a large number of learning problems owing to its effectiveness and high efficiency[1]. Unlike traditional binary classifications or multi-class single label problems, one goal of multi-label problems may be related to multiple labels. The rich relationship between labels makes the analysis of multi-label problems more complex. To solve the problem of how to make use of the relationship between features and labels, a multi-label feature selection algorithm based on HSIC is presented. This method uses Lasso (The least absolute shrinkage and selection operator) to solve a non-convex HSIC problem, and converts it to solve a lasso optimization problem, which can effectively calculate the global optimal solution. Finally, experiments show that our algorithm can improve the performance of multi-label classification.

Keywords: Multi-label; Feature Selection; Relevance; HSIC; Lasso

1. INTRODUCTION

The purpose of feature selection is to identify the feature subset that is most useful and discriminative for data analysis. In practical applications, the larger the data set, the higher the data collection cost and the more difficult the model interpretation are[2]. Moreover, the size of the data set is closely related to the prediction efficiency, computing cost and generalization ability[3]. Therefore, it is necessary to select features before training the model. Traditionally, feature selection technology has two classification methods from different perspectives[4,5]. First, feature selection can be divided into supervised, semi supervised and unsupervised methods according to whether the supervised information such as output tags is used in the classification problem. The supervision method works when there is enough label information, while the unsupervised method does not need any supervision information. Semi supervised feature selection is a trade-off between supervised and unsupervised methods. When the available labeled data is limited, this method can use both unlabeled data and labeled data. And then, for different selection strategies, feature selection algorithms can generally be divided into Wrapper based method, Filter based method, and Embedded based method. Wrapper iteratively evaluates the importance of candidate feature subsets according to the classification performance of a specific multi label learning algorithm until a stop criterion is met or a feature subset corresponding to the required learning performance is obtained. Filter method focuses on the inherent attributes of multi label data, usually uses predefined evaluation criteria to sort the features, and selects the subset that makes the evaluation criteria optimally correspond, independent of any multi label learning algorithm. Embedded method integrates feature selection into the learning process of a specific multi label learning algorithm.

In real life, data mostly exists in the form of multiple labels, which makes multi label feature selection, classification and recognition become one of the important research directions in machine learning, and has a very wide range of application scenarios. Compared with traditional single label data, multi label data becomes very challenging due to the existence of complex and changeable target objects and huge label combination space[6]. The most important feature of multi label data is the correlation between multiple labels of data. Exploring the semantic information and interrelationship of tags is one of the important means to improve the performance of multi tag learning methods[4]. For example, in the classic INRIA Person Dataset, pictures are divided into four categories: only cars, only people, with cars and people, and no cars and no people. Among them, pictures with cars and people are typical multi label images. How to accurately identify pedestrians in pictures with cars and people is a common problem in multi label research[7].

In this paper, we propose a new multi label classification algorithm framework, called the multi label learning algorithm based on Hilbert-Schmidt independence criterion. We use Hilbert-Schmidt independence criterion to evaluate the correlation between features and their labels, and improve it. We use polynomial kernel function instead of linear kernel

International Conference on Computer Graphics, Artificial Intelligence, and Data Processing (ICCAID 2022)
edited by Ruishi Liang, Jing Wang, Proc. of SPIE Vol. 12604, 126040I
© 2023 SPIE · 0277-786X · doi: 10.1117/12.2674561

function to measure the correlation between features and labels, and add label weight matrix considering the different contributions of different labels to classification.

2. HILBERT-SCHMIDT INDEPENDENCE CRITERION

Hilbert-Schmidt independence criterion is a variable correlation evaluation method based on kernel function. This method first calculates the cross covariance of two variables in the reproducing kernel Hilbert space (RKHS), and then selects the features suitable for multi label classification from these variables[8].

Suppose F is the reproducing kernel Hilbert space of the function $f: X \longrightarrow R, X$ is a separable metric space, and R is a collection of real number. For point $x, x' \in X$, exists feature mapping $\phi: X \longrightarrow F$ making $\phi(x), \phi(x') \in F$, thus the kernel $k: X \times X \longrightarrow R$ can be defined as:

$$k(x, x') = \langle \phi(x), \phi(x') \rangle_F \tag{1}$$

The second RKHS G is also defined with respect to the separable metric space Z , and for the point $z, z' \in Z$, the feature map is defined as $\varphi: Z \longrightarrow G$, and the corresponding kernel is defined as:

$$l(z, z') = \langle \varphi(z), \varphi(z') \rangle_F \tag{2}$$

Assuming that P_{xz} is the joint probability distribution of (x, z)over $X \times Z$, the mutual covariance operator $G_{xz}: G \rightarrow F$ of ϕ and φ is defined as:

$$C_{xz} = E_{x,z}\Big[\big[\phi(x) - E_x[\phi(x)]\big] \otimes \big[\varphi(x) - E_z[\varphi(z)]\big]\Big] \tag{3}$$

Where ⊗ defined as the tensor product, $E_{x,z}$, E_x and E_z are the expectations of the joint probability distribution P_{xz} and the edge probability distributions P_x and P_z, respectively. The square of the Frobenius or Hilbert–Schmidt norm(3) of this cross-covariance operator is the so-called HSIC:

$$HSIC(F, G, P_{xz}) = \|C_{xz}\|_{HS}^2 = E_{x,x',z,z'}[k(x, x')l(z, z')]$$

$$-2E_{x,z}\big[E_{x'}[k(x, x')]E_{z'}[l(z, z')]\big] + E_{x,x'}[k(x, x')]E_{z,z'}[l(z, z')] \tag{4}$$

Where $E_{x,x',z,z'}$ is the joint expectation of $(x, z) \backsim P_{xz}$ and $(x', z') \backsim P_{xz}$.

Given the data set $D = \{(x_i, z_i)\}_{i=1}^n$, the empirical estimate of HSIC can be given as:

$$HSIC(F, G, D) = \frac{1}{(n-1)^2} tr[HKHL] \tag{5}$$

Where $H, K, L \in R^{n \times n}$, $K_{ij} = k(x_i, x_j)$ and $L_{ij} = l(z_i, z_j)$ are the kernel and Gram matrix about the observation value respectively, $H = I_n - e_n e_n^T / n \in R^{n \times n}$ is the central matrix, $I_n \in R^{n \times n}$is the unit matrix, $e_n \in R^n$ is the column vector whose element values are all 1, e_n^T is the transpose of e_n, and $tr(\cdot)$ is the trace operator of the matrix. The empirical estimate of HSIC has been proved theoretically to have the advantages of fast convergence speed and simple calculation. The larger the empirical estimate is, the stronger the correlation between X and Z is. When it is equal to 0, X and Z are independent.

3. MULTI-LABEL FEATURE SELECTION ALGORITHM

Although HSIC can be used to evaluate the correlation of two variables in the kernel space, for multi-label data, the importance of different labels for the same feature is different. Moreover, there is some correlation between multiple tags of multi tag data. Therefore, we improved the HSIC criterion, used polynomial kernel to measure the correlation between tags, and added tag weights.

Given the labeled data set $D = \{(x_i, y_i) \in X \times Z\} | i = 1, 2, \cdots, v$, where X and Z are the spaces of the sample feature set and label set respectively. Assuming that the total number of possible categories of the sample is m, the label y_i of the labeled sample $x_i (i = 1, 2, \cdots, l)$ is a m dimensional column vector, and:

$$y_{ij} = \begin{cases} 1, x_i \ belongs\ to\ categoty\ j \\ 0, \qquad\quad otherwise \end{cases} \tag{6}$$

Given that the kernel functions on X and Z are $k(x,x'),(x,x' \in X)$ and $l(y,y'),(y,y' \in Z)$ respectively, their Gram matrices K and L with respect to X and Y can be obtained, thus:

$$HSIC(F,G,X,Y) = \frac{1}{(n-1)^2} tr[HKHL] \tag{7}$$

Where, F and G are respectively reproducing kernel Hilbert spaces of X and Y, H is defined in the same way as equation (5), and $n = v$ represents the total number of samples. The objective of this method is to solve the maximum value of equation (7), thus:

$$\max HSIC(F,G,X,Y) = \frac{1}{(n-1)^2} \max tr[HKHL] \tag{8}$$

However, because K is a positive semi definite Gram matrix and H is a symmetric matrix, HKH is also a positive semi definite matrix. If there are no other restrictions, equation (8) has no maximum in fact. This method introduces Lasso[9] and modifies (8) as the optimization objective:

$$\begin{gathered} \min_{w} \frac{1}{2} \left\| \bar{L} - \sum_{i=1}^{d} w_i \overline{K_\iota} \right\|_{Fro}^{2} + \lambda \|w\|_1 \\ s.t\ w_i \geq 0, i = 1, \cdots, d \end{gathered} \tag{9}$$

Where $\bar{K} = HKH, \bar{L} = HLH$, K_i is the kernel matrices of the i-th feature of all samples, and $\|\cdot\|_{Fro}$ is Frobenius norm. The first term indicates the correlation between the linear combination of the kernel matrix $\{\overline{K_\iota}\}_{i=1}^{d}$ (a feature is associated with a kernel) of the input data and the kernel matrix $\bar{L}$ of the output tag. The second term indicates that the weight value (combination coefficient) of the irrelevant feature (kernel) tends to be zero, because the l_1 norm tends to produce a sparse solution. In addition, the first term of equation (9) can be in the following form:

$$\begin{aligned} &\frac{1}{2} \left\| \bar{L} - \sum_{i=1}^{d} w_i \overline{K_\iota} \right\|_{Fro}^{2} \\ &= \frac{1}{2} \langle \bar{L}, \bar{L} \rangle_{Fro} - \sum_{i=1}^{d} w_i \langle \bar{L}, \overline{K_\iota} \rangle_{Fro} + \frac{1}{2} \sum\nolimits_{i=1}^{d} \sum\nolimits_{j=1}^{d} w_i w_j \langle \overline{K_\iota}, \overline{K_J} \rangle_{Fro} \\ &= \frac{(n-1)^2}{2} HSIC(L,L) - (n-1)^2 \sum_{i=1}^{d} w_i HSIC(L, K_i) \\ &\qquad + \frac{(n-1)^2}{2} \sum\nolimits_{i=1}^{d} \sum\nolimits_{j=1}^{d} w_i w_j\, HSIC(K_i, K_j) \end{aligned} \tag{10}$$

The calculation of $HSIC(K,L)$ is shown in equation (5). It can be seen that (10) is a convex optimization problem, so the global optimal solution can be found. For the input data x, we first normalize the input x to the unit standard deviation, and then use the Gaussian kernel:

$$k(x,x') = exp\left(-\frac{(x-x')^2}{2\sigma^2}\right), x, x' \in X \tag{11}$$

Where, σ parameter is taken as the average of the Euclidean distance of any two points on the sample feature set. The kernel for y shown in equation (6) is taken as a linear kernel:

$$l(y,y') = y'^{T} y, y, y' \in Z \tag{12}$$

4. EXPERIMENT

4.1. Data collections

In this paper, the proposed algorithm is validated by using two datasets: the prediction of anticancer activity （NCI[10]） and the prediction of toxicity （PTC[11]） . A brief description of the datasets is shown in Table 1, where "AvgL" represents the average number of labels assigned to each map.

(i) Anticancer activity prediction (NCI). This data set selects 10 groups of NCI cancer bioassay compounds from PubChem database. Each compound is represented as a graph, with atoms as nodes and chemical bonds as edges. After preprocessing such as balancing the size of positive and negative sample sets and connecting 10 sets of data sets, a data set containing 812 graph structures is finally obtained for the graph data classification task. NCI anti-cancer activity data set is widely used for graph data classification, and each data set belongs to a bioassay task anti-cancer activity prediction. If a compound in the data set is positive for the corresponding cancer, its label is positive, represented by 1; If a compound in the data set is inhibitory to the corresponding cancer, its label is negative, represented by 0. Because the labels of the original data are unbalanced (about 5% of the positive samples), the data needs to be preprocessed. We randomly select a certain size of negative data set from each data set, and remove some incomplete records. Finally, 812 graphs with 10 labels are obtained. Table 1 briefly describes 10 groups of data, where "Pos (%)" represents the average percentage of active compounds in each experiment.

Table1. Details of the anti-cancer activity prediction task (NCI1 dataset)

Bioassay-ID	Class Name	Active (Pos %)	Cancer Type
1	NCI-H23	35.6	Lung cancer
33	UACC-257	47.7	Melanoma
41	PC-3	38.5	Prostate cancer
47	SF-295	34.1	Nervous sys. tumor
81	SW-620	17.5	Colon cancer
83	MCF-7	59.2	Breast cancer
109	OVCAR-8	42.2	Ovarian tumor
123	MOLT-4	73.5	Leukemia
145	SN12C	54.8	Renal cancer
330	P388	33.4	Leukemia

(ii) Compound toxicity prediction (PTC). The second data set uses the standard data set PTC, which contains the carcinogenic information of 417 different compounds acting on four different mice. The four different kinds of mice are: small male mice are represented by MM, large male mice are represented by MR, small female mice are represented by FM, and large female mice are represented by FR. Every kind of mice injected with any compound will have a medical reaction whether it is carcinogenic. The medical reaction whether it is carcinogenic can be divided into the following categories: CE, SE, P, E, EE, IS, NE, N. Among them, E, EE and IS belong to ambiguous labels, that is, if such labels are detected on mice, it cannot indicate whether the compounds have carcinogenic effects after being injected into mice. Therefore, we removed such labels in the experiment; P. SE and CE belong to the positive label category, that is, if such labels are detected on mice, it indicates that the compounds have no carcinogenic effect after being injected into mice; N and NE belong to the negative label category, that is, if such labels are detected in mice, it indicates that the compounds have carcinogenic effects after being injected into mice. We removed the defective data sets (i.e. data with E, EE and IS tags) in four different mouse models, and finally obtained the experimental data of 253 compounds. Then, these data were represented in the form of graphs, and four types of tags (i.e. MR, FR, MM, FM) were assigned. The attribute of each type of label can be marked as+1, - 1 or 0, that is,+1 means no carcinogenic effect, 0 means not marked, and 1 means carcinogenic effect. Specific data are shown in Table 2, where "Pos (%)" represents the average percentage of active compounds in each experiment.

Table 2. Details of toxicology prediction task (PTC dataset)

Class Name	Active (Pos %)	Animal Model
MR	41.9	Male Rats
FR	36.0	Female Rats
MM	38.7	Male Mice
FM	43.1	Female Mice

4.2 Comparing methods

To demonstrate the practicability and effectiveness of this method, the following comparative experiments will be set up:

(i) Single label feature selection Binary IG+SVM algorithm. This method first decomposes multi label graph data into multiple single label graph data through one to many binary decomposition. For each binary task, we use Information Gain as an entropy to select the most recognizable feature subset from frequent sub-graphs. SVM is used as a binary classifier to classify the graph into multiple binary categories. Here we use the libSVM software package to train SVM, and the parameters are set by default. This algorithm is mainly used to compare the impact of combining multiple labels to classify data on classification performance for multi label data.

(ii) HSICFS+SVM. First, use our method to find the optimal sub graph feature set, and then use SVM to train each class one to many for multi label classification. Here we use the libSVM software package to train SVM, and the parameters are set by default. This comparison is mainly used to illustrate that even though multiple tags are used to select data. However, the classifier uses two classifiers to classify it, which affects the classification performance

(iii) HSICFS +BoosTexter. We first use our method to find the optimal sub-graph feature set, and then classify it with the multi-label classifier BoosTexter.

4.3 Evaluation Metrics

The evaluation of multi-label learning questions is much more complicated than that of single-label learning. Literature[12] defines five commonly used evaluation indexes in multi-label learning. This paper, Average Precision and Hamming Loss are selected to evaluate the multi-label classification performance. Specific formulas can be found in the original text.

(i) Average Precision: the average precision reflects the possibility that the category whose confidence value is greater than the confidence value of the real category is the real category of the sample.

(ii)Hamming Loss: after the threshold value is specified, the category of any unlabeled sample can be predicted by the sample category confidence value. If y_{ij} is greater than the threshold value, the i-th sample is considered to belong to category j. HammingLoss can measure the degree of inconsistency between the predicted results and the actual category of the sample, that is, the sample belongs to a certain category but is not recognized, or the sample does not belong to a certain category but is wrongly judged.

4.4 Experimental Result

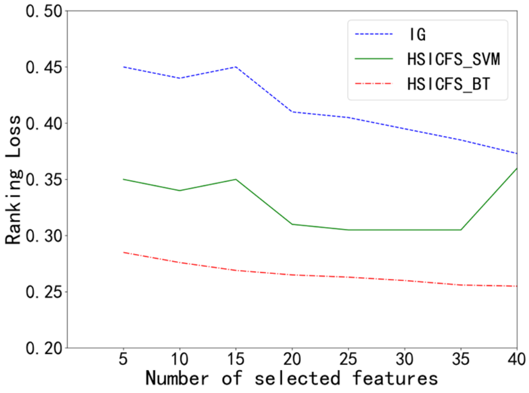


Figure 1. HammingLoss

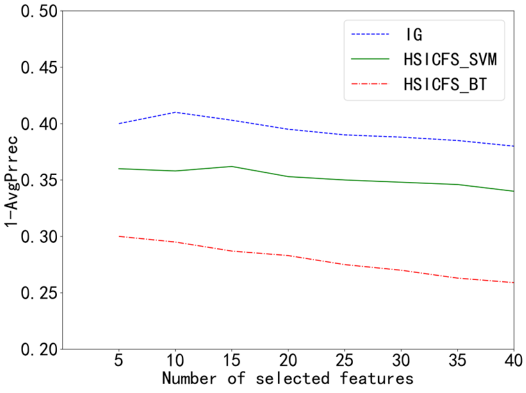


Figure 2. Average Precision

In the experiment, the whole dataset is divided into 10 equal sized parts. One of them is used as a verification set, the other nine are used as a training set, and then repeated 10 times until each one has been used as a verification set, and the rest is folded as a training set. The experimental results are shown in graph 1 and graph 2. The method proposed in this paper (HSICFS_BT) has better performance than the algorithm directly applying the binary classification algorithm to multi label classification (IG). At the same time, the classifier also plays a decisive role in the classification performance of multi label data. The classification performance of multi label classifier (HSICFS_BT) is better than that of single label classifier (HSICFS_SVM).

5. CONCLUSION

In this paper, HSIC is used to evaluate the correlation between features and label, and Lasso is introduced. Considering the different contributions of different tags to classification, tag weight matrix is added, and Gaussian kernel and linear kernel are selected as the data set and tag set kernel functions respectively. By solving a lasso optimization problem, we can effectively calculate the global optimal solution, and then select the features that play a greater role in classification. In the future research, we can consider extending the algorithm to the field of multi label image recognition.

Acknowledgement

The research work is supported by Foundation for Disinguished Young Talents in Higher Education of Guangdong, China(2020KQNCX213), Foundation for Creation and Innovation of Higher Education of China - Innovation Project of Next Generation Information Technology (2020ITA03037),Results of Constructing A Youth Developed Urban Research Topic in Foshan City(5), High-level Talent Research Project of Foshan Polytechnic, China (Grant no.KY2020G01), and Special Projects in Key Areas of Colleges and Universities in Guangdong Province (Grant no.2020ZDZX3104), High level talent research project of Foshan Vocational and Technical College.(KY2021G02)

REFERENCES

[1] L. Song, A. Smola, A.Gretton, J. Bedo, K. Borgwardt, Feature selection via dependence maximizat-ion, J. Mach. Learn. Res.13 (2012) 1393-1434.

[2] Wang T, Dai X, Liu Y. Learning with Hilbert–Schmidt independence criterion: A review and new perspectives[J]. Knowledge-based systems, 2021(234-Dec.25).

[3] Cole E, Aodha O M, Lorieul T, et al. Multi-Label Learning from Single Positive Labels:, 10.48550/arXiv.2106.09708[P]. 2021.

[4] I. Guyon, A. Elisseef, An introduction to variable and feature selection, J.Mach.Learn.Res.3(2003)1157-1182.

[5] J. Li, K.Cheng, S. Wang, F. Morstatter, R.P. Trevino, J.T ang, H. Liu, Feature selection: A data perspect-ive, ACM Comput.Surv.50 (6)(2018)94.

[6] Li Wei. Multi label image recognition based on graph neural network [D]. Harbin University of Technology, 2021. DOI: 10.27061/d.cnki.ghgdu.2021.000861.

[7] Ben-Baruch E, Ridnik T, Zamir N, et al. Asymmetric Loss For Multi-Label Classification[C]// 2020.

[8] Zhang Jujie. Research on key issues in multi-label learning [D]. Xidian University,2016.

[9] Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. Journal of the Royal Statistical Society, Series B, 58(1), 267-288.

[10] X. Kong, P. S. Yu. Multi-Label Feature Selection for Graph Classification[C]. In IEEE, pages 274-283, 2010.

[11] Helma C, King R, Kramer S, Srinivasan A (2001) The predictive toxicology challenge 2000-2001.Bioinformatics 17(1): 107-108

[12] Grigorios T, Eleftherios S, Jozef V, et al. MULAN: a java library for multilabel learning J.J Mach Learn Res, 2011,12(7): 2411-2414.

RS-Fusion: A novel virtual reality localization method based on RTK and visual SLAM

Zhitian Li[1], Weimin Zhang[1,2,3,*], Ye Tian[1,2,3] and Fangxing Li[1,2,3]

[1] School of Mechatronical Engineering, Beijing Institute of Technology, Beijing 100811, China

[2] Key Laboratory of Biomimetic Robots and Systems, Beijing Institute of Technology, Ministry of Education, China.

[3] Beijing Advanced Innovation Center for Intelligent Robots and Systems, China.

[*] Corresponding author email: zhwm@bit.edu.cn

Abstract

Virtual reality technology is a new multidisciplinary integrated technology, in which localization technology is the key technology to determine the user experience, and is the core of VR technology. RTK and visual slam are two common localization technologies, however they are limited by satellite conditions and rely heavily on feature point extraction and matching effects, which will affect the accuracy of localization process. So as to achieve high-precision virtual reality spatial localization, this paper proposed a novel virtual reality localization method called RS-fusion， which combined the RTK carrier phase difference technology and the visual SLAM technology by Kalman fusion. In addition, the gain matrix is used to calculate the spatial displacement state to meet the accuracy requirements of virtual reality and augmented reality localization in indoor and outdoor environment, and to obtain a more matching virtual reality fusion effect and simulation mapping. Experiment results show the robustness of the method and RS-fusion can realize the simultaneous localization of cameras in real space and virtual space.

Keywords: virtual reality localization; visual SLAM; RTK technology; Kalman filter

1.Introduction

As a new way of human interaction, virtual reality aims to improve people's cognitive ability. Users will manipulate objects in the virtual environment. From this perspective, the virtual world is a special real world[1-2]. Nowadays, virtual reality technology has practical applications in all walks of life. With the aid of virtual reality technology, the working mode and interaction mode in these fields are gradually changing[3].

According to the localization method of the head-mounted display (HMD), the existing virtual reality system can be divided into a passive localization scheme and an active localization scheme. As one of the best-selling virtual reality devices, HTC VIVE adopts a passive localization scheme[4].The localization and tracking of the HTC VIVE head-mounted display and hand controller requires at least 2 base stations (called Lighthouse), the base station is equipped with an infrared laser transmitter, and the head-mounted display and the hand controller are installed with photosensitive sensors. The yaw angle and pitch angle of each photosensitive sensor in the base station coordinate system can determine the position and motion trajectory of the head-mounted display and the hand controller[5].According to the research of Niehorster et al. [6], this passive tracking method performs well in terms of accuracy and delay. Many motion capture systems (such as Optitrack and Vicon) use this passive tracking localization. However, this localization solution needs to install the base station in advance, which increases the time cost and space cost of use, and also brings inconvenience to the expansion and transfer of the venue. Therefore, some active localization virtual reality devices have been proposed, such as HTC VIVE Focus, Oculus Rift S, etc, which rely on the camera and inertial sensors in the head-mounted display and based on simultaneous localization and mapping (SLAM) technology to locate and track head-mounted displays in continuously updated digital maps[7]. This active localization scheme does not need to establish a base station in advance, which makes the expansion and transfer of the venue more convenient.

However, in virtual reality (VR) and augmented reality (AR) applications, high-precision localization is required to drive simulation agents that observe subjects. In indoor applications, optical localization and laser localization can be used, but in outdoor applications, optical localization and laser localization cannot meet the needs of use due to their erection methods and sunlight exposure. Accurate localization of virtual reality in outdoor scenes has become one of the research hotspots.

International Conference on Computer Graphics, Artificial Intelligence, and Data Processing (ICCAID 2022)
edited by Ruishi Liang, Jing Wang, Proc. of SPIE Vol. 12604, 126040J
© 2023 SPIE · 0277-786X · doi: 10.1117/12.2674691

With the development of SLAM technology, its localization accuracy and robustness have been continuously improved, so its application has become more and more extensive. Visual SLAM technology based on visual sensors has undergone significant changes and breakthroughs in both theory and practice, and is rapidly moving from laboratory research to market application[8]. On the other hand, Beidou differential localization selects Real Time Kinematic (RTK) technology that combines satellite localization and inertial measurement, which has the advantages of all-weather, global coverage, and high efficiency. It can provide real-time sensor information such as the position, attitude and speed of the carrier.

Therefore, this paper proposes the RS-Fusion algorithm that combines RTK and visual SLAM localization data with Kalman fusion, so as to achieve stable and high-precision outdoor localization, and provide a new idea for accurate localization of virtual reality in outdoor scenes.

2. Method

2.1 Real-time localization based on visual SLAM

The visual SLAM localization technology extracts the feature points of the environment in its field of view through the binocular fisheye camera, and estimates the pose of the camera during the camera movement process to obtain a global unified movement trajectory and corresponding map. The real-time position of the camera is calculated through feature point matching, and an environment model is established at the same time[9]. The schematic diagram of the extracted feature points is shown in Figure 1.

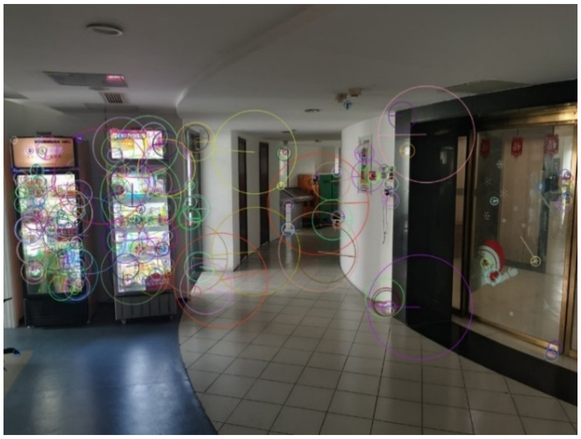


Figure 1. Feature points extraction results based on SLAM technology

Taking the displacement increment of the visual localization data as the state quantity, the state equation as shown below is established:

$$\begin{bmatrix}\bar{x}_{k,k-1}\\ \bar{y}_{k,k-1}\end{bmatrix}=A\cdot\begin{bmatrix}\bar{x}_{k-1,k-1}\\ \bar{y}_{k-1,k-1}\end{bmatrix}+\begin{bmatrix}\sigma x_k\\ \sigma y_k\end{bmatrix}+w_k \tag{1}$$

In the above formula, vector $\begin{bmatrix}\bar{x}_{k,k-1}\\ \bar{y}_{k,k-1}\end{bmatrix}$ is the prior estimation of the visual localization data at time k. A denotes the identity matrix $\begin{bmatrix}1 & 0\\ 0 & 1\end{bmatrix}$. $\begin{bmatrix}\bar{x}_{k-1,k-1}\\ \bar{y}_{k-1,k-1}\end{bmatrix}$ is the posterior estimation of the visual localization data at time k-1. $\begin{bmatrix}\sigma x_k\\ \sigma y_k\end{bmatrix}$ is the displacement increment of the visual localization data. w_k is the covariance matrix of the process noise, which is measured by experiments and is an adjustable parameter. The preferred matrix is $\begin{bmatrix}w_k & 0\\ 0 & w_k\end{bmatrix}$.

The accuracy of visual SLAM localization technology can reach the millimeter level, which can meet the accuracy requirements of virtual reality and augmented reality, but it is slightly insufficient in other aspects[10]. First of all, the position of the visual localization output is in the local coordinate system, and the position of each startup is the origin of the coordinate system, which needs to be mapped to the world coordinate system for use. Secondly, visual localization depends on the number and quality of feature points in the surrounding environment. It rely heavily on feature point extraction and matching effects, which will affect the accuracy of localization process .When there are few feature points in the environment or the features are not obvious, the localization accuracy will decrease or even diverge. Thirdly, with the increase of usage time, visual localization will produce a certain cumulative error. Although it can be corrected through feature points, it will also generate large data jumps during the correction.

2.2 Beidou differential localization based on RTK technology

RTK technology is a mathematical model for high-precision relative localization based on the carrier phase of satellite signals. The basic principle is to use the spatial correlation between two adjacent observation stations to eliminate or weaken the errors generated in the measurement process. Use the carrier phase after the integer ambiguity has been solved for the localization solution. By this method, high-precision localization results can be obtained[11].

Assuming that the Beidou differential localization output based on RTK technology is represented by latitude and longitude, we need to convert it to UTM coordinates first[12]. Let the longitude be φ, the latitude be λ, and the UTM coordinates be (E,N). Calculate parameters $v(\varphi)$ and $s(\varphi)$.

$$V=\frac{1}{\sqrt{1-e^2\sin^2\varphi}} \tag{2}$$

$$s=\left(1-\frac{e^2}{4}-\frac{3e^2}{64}-\frac{5e^6}{256}\right)\varphi-\left(\frac{3e^2}{8}+\frac{3e^4}{32}+\frac{45e^6}{1024}\right)\sin 2\varphi+\left(\frac{15e^4}{256}+\frac{45e^6}{1024}\right)\sin 4\varphi-\frac{35e^6}{3072}\sin 6\varphi \tag{3}$$

where $Z=\left\lfloor\frac{\lambda}{6}\right\rfloor+31$, $\lambda_0=(Z-1)\times 6-183$, $A=(\lambda-\lambda_0)\cos\varphi$. Then calculate the UTM coordinates:

$$E=E_0+k_0av\left(A+(1-T+C)\frac{A^3}{6}+\left(5-18T+T^2\right)\frac{A^5}{120}\right) \tag{4}$$

$$N=k_0a\left(s+v\tan\varphi\left(\frac{A^2}{2}+\left(5-T+9C+4C^2\right)\frac{A^4}{24}+\left(61-58T+T^2\right)\frac{A^6}{720}\right)\right) \tag{5}$$

where $T=\tan^2\varphi, C=\frac{e^2}{1-e^2}\cos^2\varphi, k_0=0.9996$，$E_0=500$，$e=0.00818192$.

After the UTM coordinates are obtained, they are used as observations to establish the observation equation as shown below:

$$\begin{bmatrix}\widehat{X}_k\\ \hat{y}_k\end{bmatrix}=C\cdot\begin{bmatrix}x_k\\ y_k\end{bmatrix}+r_k \tag{6}$$

In the above formula, the vector $\begin{bmatrix}\hat{x}_k\\ \hat{y}_k\end{bmatrix}$ is the posterior estimate of the Beidou differential localization data at time k, C is the observation matrix, the unit matrix $\begin{bmatrix}1 & 0\\ 0 & 1\end{bmatrix}$ is taken, $\begin{bmatrix}x_k\\ y_k\end{bmatrix}$ is the output data of the Beidou differential localization system, and r_k is the observation noise matrix. Measured by experiments, it is an adjustable parameter, and the preferred matrix is $\begin{bmatrix}r_k & 0\\ 0 & r_k\end{bmatrix}$.

The longitude, latitude and elevation data of the Beidou differential localization real-time output carrier in the world coordinate system have no cumulative error, and the localization accuracy can reach centimeter level. Centimeter-level localization accuracy is good enough for everyday navigation applications, however, it is far from enough for virtual reality and augmented reality applications. The centimeter-level virtual-real correspondence error will cause great distortion, which will affect the simulation effect and results.

2.3 Improved method

Combined with the characteristics of Beidou differential localization and visual SLAM localization, Kalman fusion processing of the two sets of pose data can obtain high-precision and stable localization data. Visual localization has the characteristics of high precision and good continuity, and its displacement increment is taken as the state transition equation. Beidou differential localization has a stable output in the world coordinate system, and its coordinate data is taken as the observation equation.

According to the state equation of equation (1) and the observation equation of equation (6), the Kalman fusion algorithm can be used to fuse Beidou differential localization data and visual SLAM localization data. Set the displacement vector $\bar{x}_{k-1,k-1}=\begin{bmatrix}\bar{x}_{k-1,k-1}\\ \bar{y}_{k-1,k-1}\end{bmatrix}$, and the initial value of the covariance matrix $P_{k-1,k-1}$ is the identity matrix.

The Kalman fusion algorithm can be divided into two stages: state prediction and state update. In the state prediction stage, the covariance matrix at time k is updated by the covariance matrix at time k-1:

$$P_{k,k-1}=P_{k-1,k-1}+Q_k \tag{7}$$

where Q_k is the covariance matrix of the system process noise.

The displacement is estimated a priori through the state transition equation:

$$\overline{X}_{k,k-1}=A\cdot\bar{x}_{k-1,k-1}+\sigma x_k+w_k \tag{8}$$

After the a priori estimate of the displacement is obtained, the state update phase is entered. First, the Kalman gain matrix K_k at time k is calculated through the covariance matrix $P_{k,k-1}$.

$$K_k=\frac{P_{k,k-1}\cdot C_k{}^T}{C_k\cdot P_{k,k-1}\cdot C_k{}^T+R_k} \tag{9}$$

where R_k is the measurement noise matrix.

Then the displacement state estimate at time k is calculated by the Kalman gain matrix K_k at time k.

$$\overline{X}_{k,k}=\bar{x}_{k,k-1}+K_k\left(\hat{x}_k-\bar{x}_{k,k-1}\right) \tag{10}$$

In the above formula, $\hat{x}_k$ is obtained by the observation equation $\begin{bmatrix}\hat{x}_k\\ \hat{y}_k\end{bmatrix}=C\cdot\begin{bmatrix}x_k\\ y_k\end{bmatrix}+r_k$.

Finally, update the covariance matrix $P_{k,k}$ to prepare for the next round of state prediction:

$$P_{k,k}=(I-K_k\cdot C_k)\cdot P_{k,k-1} \tag{11}$$

The flow chart of the Kalman data fusion algorithm is shown in Figure 2 below.

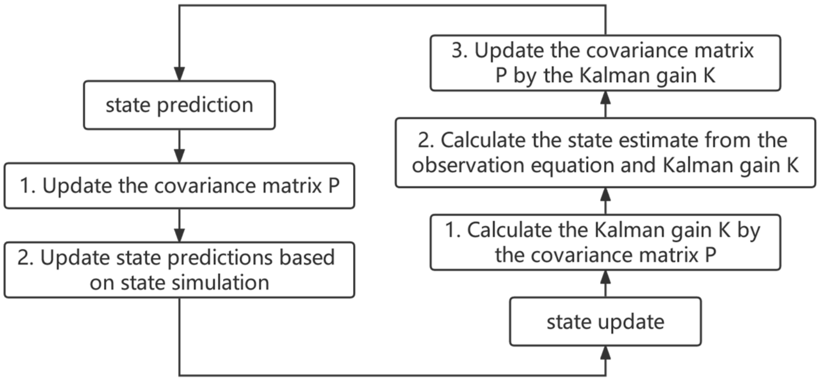

state prediction
1. Update the covariance matrix P
2. Update state predictions based on state simulation
state update
1. Calculate the Kalman gain K by the covariance matrix P
2. Calculate the state estimate from the observation equation and Kalman gain K
3. Update the covariance matrix P by the Kalman gain K

Figure 2. RS-fusion algorithm flow chart

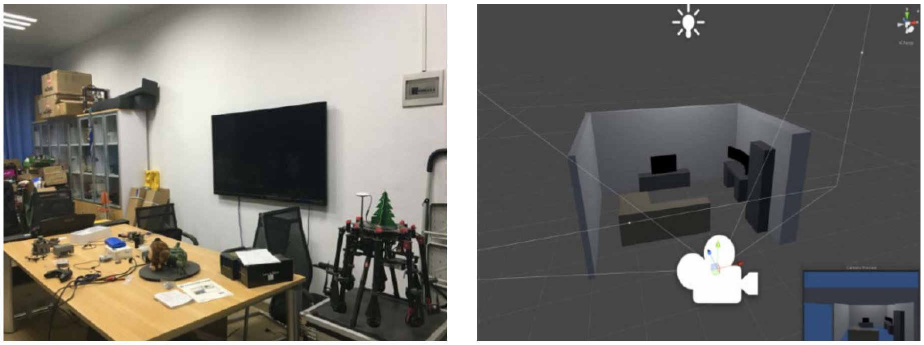
3. Experimental Results

3.1 Experimental setup

The monocular camera used in this paper is a Zenmuse X3 camera with an image resolution of 1280 × 720; the RTK device uses a NovAtel OEM-615 receiver board with a localization accuracy of 5 cm. In order to ensure the image quality and the processing performance of the algorithm, the image acquisition frequency is set to 20Hz, and the RTK acquisition frequency is set to 5Hz. In addition, Unity3D is used to create the virtual space scene required for the experiment and publish it on Linux.

3.2 Experimental results

We use Unity3D to build a simple indoor virtual space scene and publish it to Linux. In order to make users visually feel the real-time transformation in the virtual scene, some virtual objects are placed in the virtual scene as reference objects, such as TV sets, sofas, computers, wardrobes, tables and so on. Figure 3 below shows an example of the real space scene selected for the experiment and the virtual space scene created by the experiment.

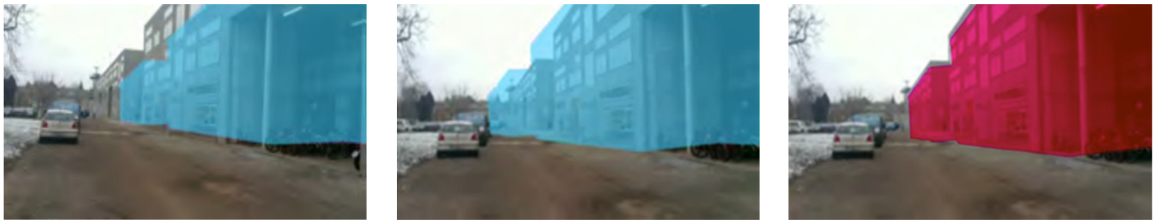


(a) Selected real-space scenarios (b) Created virtual space scene

Figure 3. Real space and virtual space scene diagram

To demonstrate the effectiveness of Kalman filter fusion, three experiments were conducted: (a) the localization effect when using RTK technology alone; (b) the localization effect when using SLAM technology alone;(c) The localization effect of Kalman fusion of RTK technology and SLAM technology.

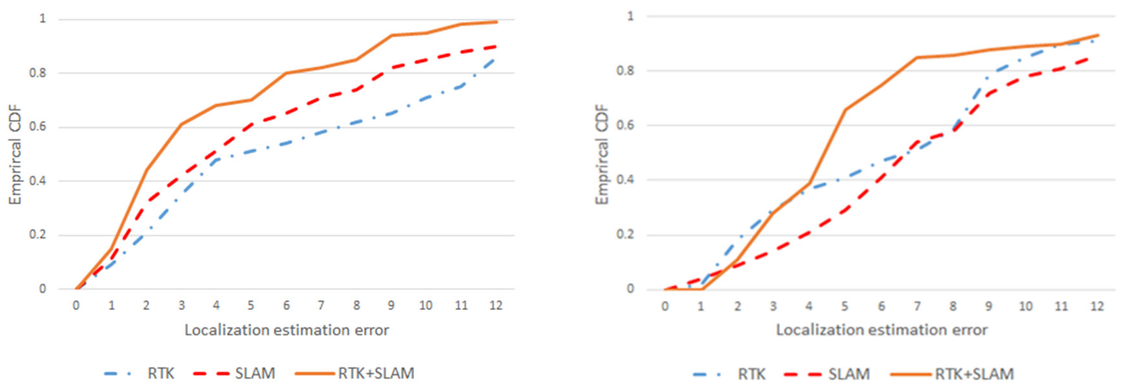
The visual results of the experiment are shown in Figure 4. The red or blue region means the localization effect of a building. It can be seen from the figure that the localization results obtained by RTK technology and SLAM technology are not accurate enough, and the localization results obtained by the Kalman fusion method that combines the two technologies are the most accurate. Therefore, the virtual reality localization method that integrates RTK technology and SLAM technology proposed in this paper is effective.



(a)RTK (b)SLAM (c)RS-Fusion

Figure 4. Visual comparisons of RTK and SLAM technologies

In order to reflect the localization effect more accurately, the scene is divided into indoor and outdoor, and the localization error comparison test is carried out for the separate RTK technology, SLAM technology, and the RS-Fusion. The test results are shown in Figure 5, where (a) shows the indoor scene results, and (b) shows the outdoor scene results. The results of Empirical CDF are given in Table1. After the RS-Fusion, results with smaller errors are obtained, and better accuracy is obtained in both indoor and outdoor scenes. It is worth noting that the error of outdoor scenes is generally higher than that of indoor scenes, because there are trees and other obstacles near the scene, which increase the localization error.



(a) Indoor scene (b) Outdoor scene

Figure 5. Localization error curves (CDF means cumulative distribution function)

Table 1. The Empirical CDF of indoor scene and outdoor scene

Scene	Method	Empirical CDF							
		1	2	3	4	5	6	7	8
Indoor	RTK	0.35	0.51	0.54	0.58	0.62	0.65	0.71	0.75
	SLAM	0.42	0.61	0.65	0.71	0.74	0.82	0.85	0.88
	RS-Fusion	0.61	0.70	0.80	0.82	0.85	0.94	0.95	0.98
Outdoor	RTK	0.29	0.37	0.41	0.47	0.59	0.79	0.85	0.91
	SLAM	0.14	0.21	0.29	0.41	0.58	0.72	0.78	0.86
	RS-Fusion	0.28	0.39	0.66	0.75	0.86	0.88	0.89	0.93

4.Conclusions

In order to obtain accurate localization of virtual reality in indoor and outdoor scenes, this paper proposed the RS-Fusion algorithm to fuse the localization data obtained by RTK carrier phase difference technology and visual SLAM technology. The stable and continuous high-precision position output in the world coordinate system is obtained by the proposed method, which meets the accuracy requirements of outdoor virtual reality and augmented reality localization, and obtains a matched virtual-real fusion effect and simulation mapping. Experimental results show that the localization error of RS-Fusion is small, so RS-Fusion can be applied to indoor and outdoor scenes and has good robustness.

Future research will focus on using machine learning based methods to achieve fusion positioning. At the same time, from the content required for positioning, more consideration will be given to centimeter level and millimeter level equipment in hardware design to improve the accuracy and reliability of positioning.

Acknowledgments

This work was supported by National Natural Science Foundation of China, Granted No. 61973031.

References

[1] Zhang F, Dai G, and Peng X. "A survey on human-computer interaction in virtual reality." *Scientia Sinica Informationis* 46.12 (2016): 1711-1736.

[2] Besançon L, et al. "Mouse, tactile, and tangible input for 3D manipulation." *Proceedings of the 2017 CHI conference on human factors in computing systems*. 2017.

[3] Zhang Q, et al. "Research progress of nuclear emergency response robot." *IOP Conference Series: Materials Science and Engineering*. Vol. 452. No. 4. IOP Publishing, 2018.

[4] Andrejevic M. "Exploiting YouTube: Contradictions of user-generated labor." *The youtube reader* 413.36 (2009): 406-423.

[5] Niehorster DC., Li L, and Markus L. "The accuracy and precision of position and orientation tracking in the HTC vive virtual reality system for scientific research." *i-Perception* 8.3 (2017): 2041669517708205.

[6] Bragdon A, et al. "Code space: touch+ air gesture hybrid interactions for supporting developer meetings." *Proceedings of the ACM International Conference on Interactive Tabletops and Surfaces*. 2011.

[7] Gao X, et al. "14 lectures on visual SLAM: from theory to practice." *Publishing House of Electronics Industry, Beijing* (2017).

[8] Quan M, Piao S, and Guo L. "An overview of visual SLAM." *Journal of Intelligent Systems* 11.6 (2016): 768-776.

[9] Jiang X, Zhu L, Song A. Research on SLAM-based virtual reality six-degree-of-freedom input system [J]. *Measurement and Control Technology*, 1-6, 2022.

[10] Campos C, et al. "Orb-slam3: An accurate open-source library for visual, visual–inertial, and multimap slam." *IEEE Transactions on Robotics* 37.6 (2021): 1874-1890.

[11] Xu G, and Yan X. "Applications of GPS theory and algorithms." *GPS. Springer*, Berlin, Heidelberg, 2016.

[12] Wu Y, Wang X, Yang L and Cao P. Monitoring and Analysis of Autonomous Integrity of GNSS/INS Compact Integrated Navigation System [J]. *Chinese Journal of Surveying and Mapping*. (2014), 43(8):786-795.

Design and Implementation of Coroutine Management Environment

Bin Cai[*1], Bingcheng Qiu[1], Jian Yao[2, 1], Jun Fu[1], Yunlan Xue[1], Xiaofei Liu[1],
Ruilong Wu[3], Xu Zhang[4]
[1]Guangdong Open University (Guangdong Polytechnic Institute), Guangzhou, Guangdong
[2]Wuhan University, Wuhan, Hubei
[3]China Hi-Tech Group Co. Ltd., Guangzhou, Guangdong
[4]Blue Dot Power Data Technology Co. Ltd., Beijing
[*]Corresponding author: bcai@gdrtvu.edu.cn

Abstract

Advances in network and storage technologies enable data centers to provide high-concurrency and low-latency processing responses. To meet user needs, a variety of low-latency technologies can be used to improve the overall concurrent response capability of the system on the premise of ensuring system stability. User mode thread pool is one of the techniques that can be used. In a multi-core processor system, the scheduling management of the user mode thread pool is much more complicated. At present, the scheduling management of the user mode thread pool mainly focuses on the resource utilization of the user mode thread pool itself and the scheduling management of many user-mode threads, such as dynamically adjusting the number and usage of user mode threads in the user mode thread pool according to the load of the system, and replacing the system thread scheduling with user mode thread scheduling, etc. The existing user mode thread scheduling methods do not distinguish the performance requirements of different data processing tasks, treat all user mode thread tasks in the same way, cannot reflect the performance requirements of different user mode thread tasks, and cannot provide more fine-grained user mode thread execution timing control; the existing user mode thread scheduling methods do not provide more effective scheduling control for the host thread, but instead use the thread scheduling mechanism of the operating system itself to control the execution of all kernel threads which is difficult to isolate the host thread and other kernel threads in the system; The existing user mode thread scheduling methods cannot give full play to the parallel execution advantages of the multi-core processor architecture, and it is difficult to implement unified resource allocation management among multiple user mode thread pools. This paper provides a two-level scheduling management method for user mode thread pools that effectively integrates the operating system scheduling mechanism and user mode thread scheduling mechanism, which can distinguish the needs of different data processing tasks and implement more targeted and fine-grained user mode threads scheduling and execution control, without affecting the scheduling mechanism of the operating system itself, realize the isolation of host threads and non-host threads, implement differentiated scheduling control. And it can realize unified resource allocation management of multi-user thread pools for multi-core processor architecture so that the overall performance of the system can be freely scaled with the change of data task processing requirements to provide a more efficient user mode thread pool management for high concurrency and low latency systems.

Keywords-Multi-core system, concurrent service, coroutine, coroutine pool, synchronization mechanisms

1. Introduction

With the advancement of network and storage technologies, users have higher requirements for high concurrency and low latency data center services. Under the premise of ensuring system stability, these systems can use coroutine technology to improve the overall concurrency of the system[1,2,3,4].

Coroutines, also known as user mode threads, correspond to kernel threads and are time-sharing technologies within kernel threads. It can have independent register contexts and execution stacks so that a single kernel thread can process multiple tasks at the same time without the need for switching between user mode and kernel mode to reduce system loss. Compared to the traditional thread running model, the coroutines are more controllable, lightweight, fast, and efficient with more obvious advantages[5,6,7].

Coroutines can significantly improve the running efficiency of the application. The traditional thread running model is scheduled and controlled by the operating system, and the application itself cannot control the execution timing of its child threads; coroutines let the application itself implement the scheduling control of coroutine task in the user mode without

International Conference on Computer Graphics, Artificial Intelligence, and Data Processing (ICCAID 2022)
edited by Ruishi Liang, Jing Wang, Proc. of SPIE Vol. 12604, 126040K
© 2023 SPIE · 0277-786X · doi: 10.1117/12.2674720

the need to switch back and forth between user mode and kernel mode. The scheduling mechanism is, therefore, more flexible, controllable, and lightweight, with higher operating efficiency and less system loss.

Coroutines can significantly reduce the synchronization overhead of an application. In the traditional thread running model, access to shared data must be performed synchronously (such as first locking, then updating, and then unlocking the data in the critical section to prevent data inconsistency), which will inevitably lead to performance loss. A coroutine usually uses a "non-preemptive" scheduling mechanism. Multiple coroutines in the same kernel thread do not compete for access to critical section data, which avoids the overhead of synchronization operations and improves coroutine concurrency[8,9,10].

Coroutines can effectively simplify the asynchronous programming of an application. The programming based on coroutine can transform the complex asynchronous callback method into a synchronous serial method, which is more in line with the processing flow of application logic and can simplify the complexity of programming when implementing a high-concurrency and low-latency system.

This paper implements an environment for coroutine management, which decouples the business processing of the application system from the management of coroutines, realizes common functions of coroutine creation, scheduling, synchronization, migration, and destruction, shields coroutines from applications the complexity of coroutine operation and management, and provides an application development interface similar to the thread model. Coroutine applications only need to use the application development interface provided by the environment to register the business task processing functions in the environment in the form of coroutines, without caring about the scheduling, execution, and management of specific task processing functions. It greatly reduces the complexity of coroutine application development and the difficulty of upgrading and maintenance; Meanwhile, this environment effectively integrates the coroutine management with the scheduling mechanism of the operating system to achieve efficient coroutine creation, scheduling, synchronization, migration, and destruction. After obtaining the requirements of the coroutine application, it processes the registration, execution, and synchronization of the task function transparently to the coroutine application to realize the isolated operation of the coroutine and the kernel thread, so that the overall performance of the system can be scaled with a load of processing tasks to provide a more efficient environment for coroutine management for high concurrency and low latency systems.

2. Related Works

Traditional high-performance server-side applications generally use thread pool technology or callback-based asynchronous I/O technology to improve server-side performance.

Multi-threading technology improves the utilization of multi-core processors, realizes the concurrent execution of logic processing, and improves the throughput and responsiveness of server applications[11,12,13,14]. However, in a high-concurrency environment, a large number of threads are scheduled and frequently switched between user mode and the kernel mode which requires a large number of instruction cycles to complete, so that the overall performance of the system cannot be effectively improved; when the business volume increases, trying to increase the processing power by increasing the number of worker threads may cause most of the system resources to be consumed in thread management and thread switching, and the system overhead will increase sharply, making the overall performance of the system to deteriorate with just the opposite effect[15,16,17].

Callback-based asynchronous I/O technology executes program logic asynchronously in an event-driven manner so that server-side applications can continue to run without waiting for I/O events to complete. The callback-based asynchronous I/O technology can reduce the I/O blocked waiting time of the server application and improve the asynchronous I/O responsiveness of the server application, but in a high concurrency environment, a large number of registered callback functions, or even the nesting of callback functions will disrupt the overall logic processing flow of the application, make the coherent business processing flow scattered, and increase the difficulty of application code maintenance. And the callback-based asynchronous I/O technology does not respond in a natural way to the application logic which is not conducive to application development[18,19,20,21].

Although the coroutine model can significantly improve the performance and efficiency of the thread model, the implementation of coroutine management functions is relatively complex, involving the creation, scheduling, synchronization, migration, and destruction of coroutines. Existing application systems using the coroutine model (hereinafter referred to as "coroutine applications") not only need to implement business-related task processing functions but also need to implement the management functions of coroutines according to their respective needs, introducing difficulties such as tight coupling of the coroutine management and business processing logic, complicated system development, difficult upgrade, and maintenance. Therefore, a general coroutine management environment is needed, which can realize functions such as creation, scheduling, synchronization, migration, and destruction of coroutines, shield the complexity of coroutine management externally, and provide an application development interface similar to the thread model.

3. Design and Implementation of Coroutine Management Environment

The coroutine management environment includes three modules, namely, the coroutine environment service module, the coroutine environment management module, and the coroutine environment interface module. Figure 1 shows the system hierarchy diagram of the coroutine management environment. The functions of the modules are described as follows:

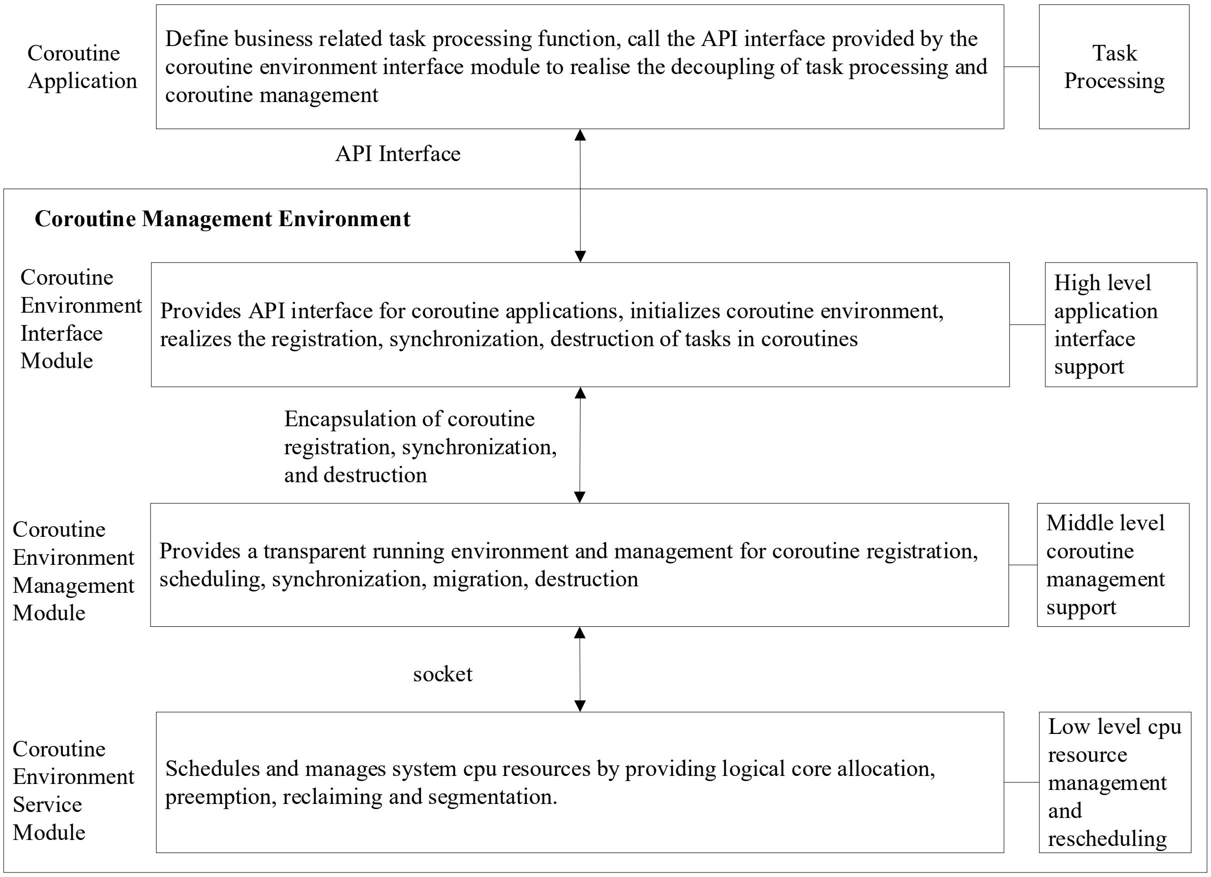

Coroutine Application
Define business related task processing function, call the API interface provided by the coroutine environment interface module to realise the decoupling of task processing and coroutine management
Task Processing
API Interface
Coroutine Management Environment
Coroutine Environment Interface Module
Provides API interface for coroutine applications, initializes coroutine environment, realizes the registration, synchronization, destruction of tasks in coroutines
High level application interface support
Encapsulation of coroutine registration, synchronization, and destruction
Coroutine Environment Management Module
Provides a transparent running environment and management for coroutine registration, scheduling, synchronization, migration, destruction
Middle level coroutine management support
socket
Coroutine Environment Service Module
Schedules and manages system cpu resources by providing logical core allocation, preemption, reclaiming and segmentation.
Low level cpu resource management and rescheduling

Figure 1. System hierarchy diagram of the coroutine management environment.

- Coroutine Environment Service Module

This module is the low-level process of the coroutine management environment. It is responsible for allocating and managing the processor logic core resources of the system, providing the core allocation, pre-emption, and reclaiming functions of the coroutine application, and providing the scheduling and isolation of threads of the coroutine (hereafter referred to as “host threads”) from traditional kernel threads. When the coroutine application starts running, it applies the required number of logical cores through the API interface provided by the environment, and the coroutine service module transparently allocates the required logical core resources according to the available logical cores and the running status of each host thread.

- Coroutine Environment Management Module

The management module is the middle layer of the coroutine management environment. It is responsible for providing the running environment and management mechanism for the creation, scheduling, synchronization, migration, and destruction of the coroutine, and provides a globally unified coroutine pool scheduling management function on the multi-core processor architecture. After the coroutine application registers the processing tasks in the environment as coroutines through the API interface provided by the environment, the coroutine management module transparently allocates coroutine context to each task processing according to the load of the coroutine pool and the task types and performs scheduling, execution, migration, synchronization.

- Coroutine Environment Interface Module

This module is the user interface layer of the coroutine management environment. It is responsible for providing API calling interfaces for coroutine applications, including “environment initialization interface”, “register shared coroutine interface”, “register exclusive coroutine interface”, “sleep interface”, “block interface”, “coroutine join interface”, “yield interface”, and “task join interface”.

3.1. Coroutine Environment Service Module

The coroutine environment service module defines four states of the host thread, namely: “initial state”, “running state”, “blocked state” and “pre-empted state”. The definitions are described as follows:

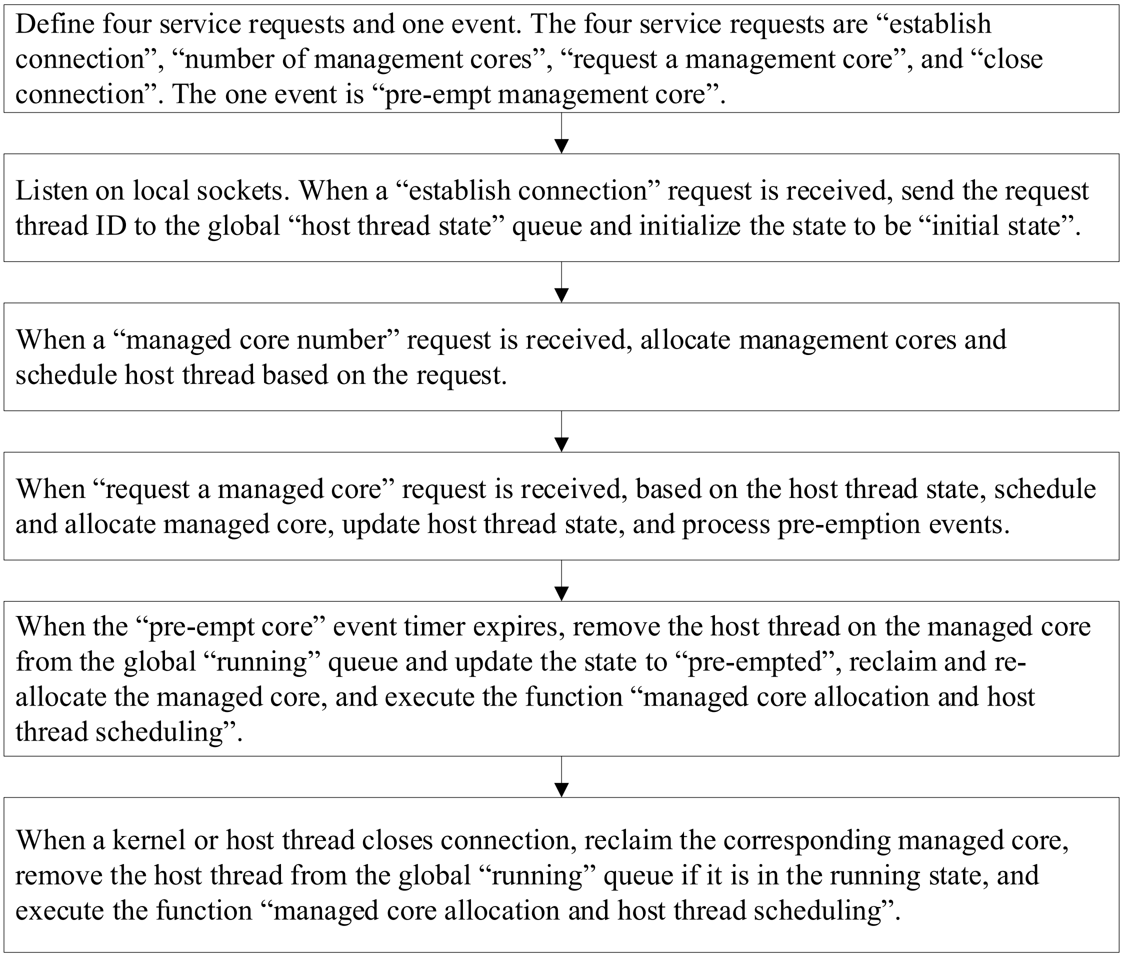
- “Initial state” means that the host thread has just established a connection with the service process and has not yet run on the managed core.

- “Running state” means that the host thread has acquired a managed core and is running on this managed core.

- “Blocked state” means that the host thread has not yet acquired a managed core and is waiting to allocate a managed core, or the host thread has successfully applied for a managed core but is currently not running on this managed core and is in a suspended state.

- “Pre-empted state” means that the host thread once ran on a managed core, but due to the shortage of managed core resources, its managed core was pre-empted by other host threads, and it was migrated to run on an unmanaged core. When the managed core on which the host thread in the “pre-empted state” last ran is reclaimed, the host thread in the “pre-empted state” will be rescheduled back to that managed core; only when that managed core is still being occupied, the host thread in the “pre-empted state” will be scheduled to run on another available managed core.

The main execution process of the coroutine environment service module is: starting the service process, looping to monitor the service request on the local socket, entering the corresponding request processing process according to the received service request type, establishing connections for each kernel thread, and host thread, allocate managed cores and schedule the execution of individual host threads in real time, handle “managed core pre-emption” events, and handle connection closures. Its execution flow is shown in Figure 2.



Define four service requests and one event. The four service requests are “establish connection”, “number of management cores”, “request a management core”, and “close connection”. The one event is “pre-empt management core”.

Listen on local sockets. When a “establish connection” request is received, send the request thread ID to the global “host thread state” queue and initialize the state to be “initial state”.

When a “managed core number” request is received, allocate management cores and schedule host thread based on the request.

When “request a managed core” request is received, based on the host thread state, schedule and allocate managed core, update host thread state, and process pre-emption events.

When the “pre-empt core” event timer expires, remove the host thread on the managed core from the global “running” queue and update the state to “pre-empted”, reclaim and re-allocate the managed core, and execute the function “managed core allocation and host thread scheduling”.

When a kernel or host thread closes connection, reclaim the corresponding managed core, remove the host thread from the global “running” queue if it is in the running state, and execute the function “managed core allocation and host thread scheduling”.

Figure 2. Process for the coroutine environment service module.

3.2. Coroutine Environment Management Module

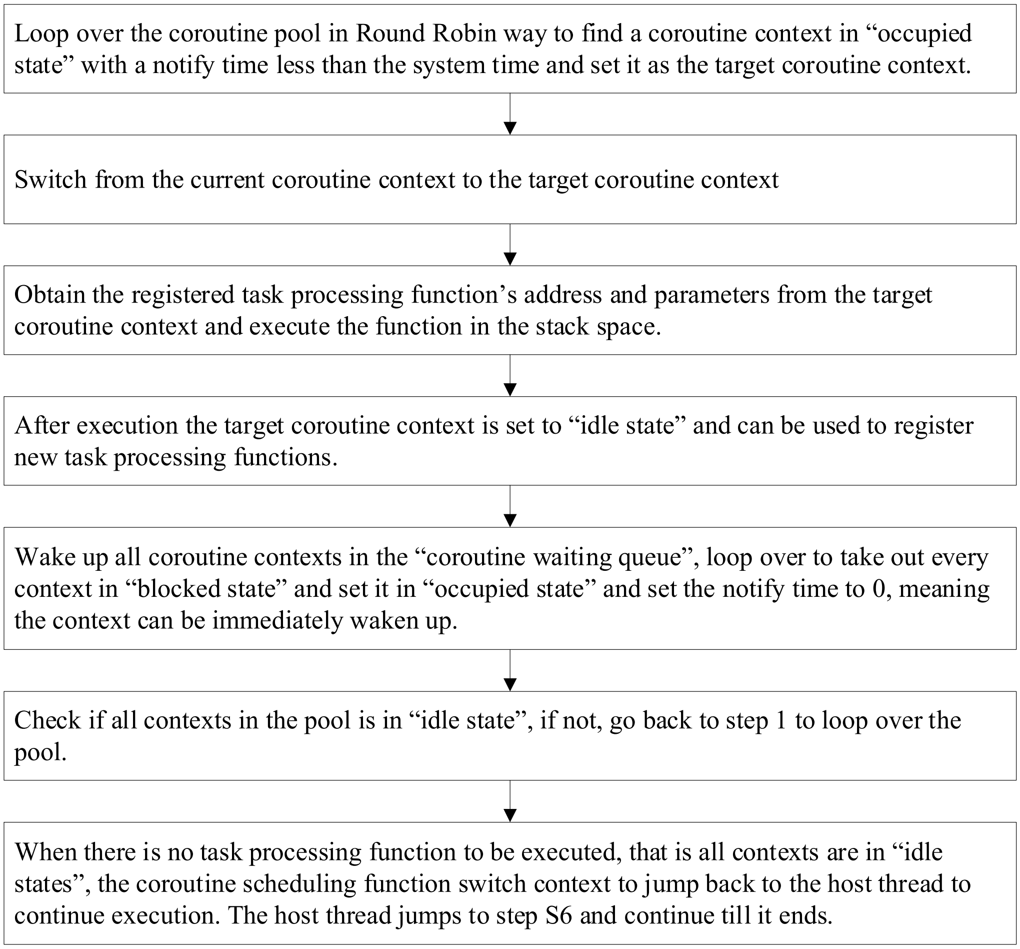
The coroutine environment management module defines three coroutine context states, namely “occupied state”, “idle state” and “blocked state”. The definitions are described as follows:

- The coroutine context of “occupied state” means that the coroutine context has been assigned to a task processing, and the coroutine task scheduling function can schedule and execute the task processing.
- The coroutine context in “idle state” means that the coroutine context is still in the initialization state and is not assigned to a task processing.
- The coroutine context in “blocked state” means that the coroutine context has been assigned to a task processing, but the task processing is in a suspended state, and the coroutine context cannot be scheduled by the coroutine task scheduling function until the coroutine context is awakened.

The coroutine environment management module also defines two processing tasks: one is “shared coroutine” and the other is “exclusive coroutine”. A coroutine pool on a managed core can accommodate multiple “shared coroutines” at the same time, which is called a “shared managed core”; an “exclusive coroutine” completely occupies a managed core, and the coroutine pool has only one coroutine, and can no longer accommodate other coroutine tasks, this kind of managed core is called “exclusively managed core”.

“Shared coroutines” can be “migrated” to other “shared managed cores” for execution, while “exclusive coroutines” will always occupy the “exclusively managed core” until execution is complete. When the coroutine application creates “exclusive coroutines” due to business needs and there is a shortage of “exclusively managed cores”, all “shared coroutines” on a “shared managed core” can be added to the coroutine pool of another “shared managed core”, and the “shared managed core” can be converted into an “exclusive managed core” to satisfy the application needs. At the same time, when the coroutines on the “exclusively managed core” have finished executing and become idle, the “exclusively managed core” can be converted to a “shared managed core”. This dynamic adjustment mechanism of logical core resources enables coroutine applications to scale with demand changes and is transparent to coroutine applications.

The main execution flow of the coroutine environment management module is: creating a host thread for the coroutine application, initializing the coroutine running environment for each host thread and the main thread, creating the coroutine pool resource, creating the coroutine context, allocating the coroutine stack space, allocating the coroutine context to the task processing function, implementing the coroutine context switch, executing the coroutine task scheduling function, and executing the task processing function. Its execution process is shown in Figure 3.



Loop over the coroutine pool in Round Robin way to find a coroutine context in “occupied state” with a notify time less than the system time and set it as the target coroutine context.

Switch from the current coroutine context to the target coroutine context

Obtain the registered task processing function’s address and parameters from the target coroutine context and execute the function in the stack space.

After execution the target coroutine context is set to “idle state” and can be used to register new task processing functions.

Wake up all coroutine contexts in the “coroutine waiting queue”, loop over to take out every context in “blocked state” and set it in “occupied state” and set the notify time to 0, meaning the context can be immediately waken up.

Check if all contexts in the pool is in “idle state”, if not, go back to step 1 to loop over the pool.

When there is no task processing function to be executed, that is all contexts are in “idle states”, the coroutine scheduling function switch context to jump back to the host thread to continue execution. The host thread jumps to step S6 and continue till it ends.

Figure 3. Coroutine Environment Management Module Process.

3.3. Coroutine Environment Interface Module

The coroutine environment interface module provides an environment initialization interface, an interface for registering a shared coroutine, an interface for registering an exclusive coroutine, a sleep interface, a block interface, a coroutine join interface, a pre-empt interface, and a task join interface. Its functions are described as follows:

- Environment Initialization Interface. Initialize the coroutine management environment of each coroutine application, create a host thread, create a coroutine pool resource, create a coroutine context, and jump to the coroutine context to realize switching from the host thread stack space to coroutine stack space.

- Register Shared Coroutine Interface. Register the task processing function defined by the business in the coroutine management environment in the form of “shared coroutine”, and transparently provide the coroutine context for the task processing function.

- Registering Exclusive Coroutine Interface. Register the task processing function defined by the business in the coroutine management environment in the way of "exclusive coroutine", and transparently provide the coroutine context for the task processing function

- Sleep Interface. Set the sleep time of the current task. Before the time expires, the task function will not be executed. Only after the sleep time expires, the task processing function can be executed.

- Block Interface. It is used to set the current task to a suspended state, and the suspended task function cannot be executed until another task function calls the "coroutine join interface".

- Coroutine Join Interface. It is used to resume the execution of the task processing function called the block interface. The task processing function that is called this interface blocks and waits until the resumed task processing function finishes execution to continue to execute.

- Yield Interface. The current task voluntarily gives up execution and is rescheduled for execution next time.

- Task Join Interface. The coroutine application blocks and waits for the end of each host thread, so that the entire application completes running.

4. Coroutine Synchronization Control

4.1. Coroutine Synchronization Control Mechanism

The existing coroutine synchronization mechanism mainly realizes the synchronization operations of "yield" and "resume". The operation of "yield" makes the executive authority jump from the coroutine that gives up execution to the target coroutine by coroutine context switching and executing the target coroutine; the operation of "resume" returns to the breaking point of the yielding coroutine to continue to execute. However, this single cooperative synchronization operation cannot accommodate complex coroutine synchronization requirements. To this end, this paper effectively integrates the scheduling mechanism of coroutines and provides six synchronous control operations, namely: "yield", "mutex", "condition", "sleep", "block", and "join", which can realize complex synchronization process, reduce the complexity and maintenance difficulty of coroutine synchronization control, and facilitate the use of coroutine applications. Figure 4 shows the interaction between the various modules when the coroutine is running. The description of each synchronization control operation is as follows:

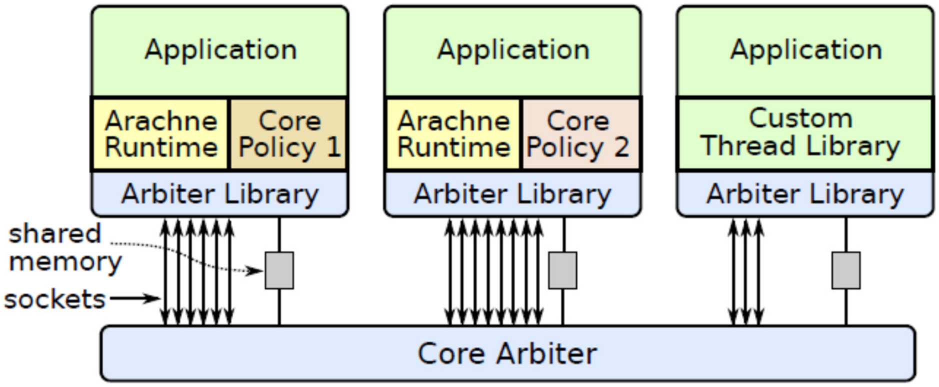


Figure 4. Coroutine Environment Interaction Process

- "Yield" control: By switching between different coroutine contexts, the transfer of executive authority is realized, and the current coroutine context is actively jumped to the target coroutine context. The current coroutine context state transitions to a "blocked state".

- "Mutex" control: The current coroutine applies for a mutex lock until it is successfully acquired. When the mutex is not successfully acquired, the current coroutine context state transitions to a "blocked state".

- "Condition Variable" control provides the waiting operation of the current coroutine, as well as the operation of waking up the coroutine context in the "blocked state" or the coroutine context in the "sleep state". When the wait operation is performed, the current coroutine context state is converted to a "blocked state"; when the wake-up operation is performed, the awakened coroutine context state is converted to an "occupied state".

- "Sleep" control: Provides the operation that the current coroutine wakes up after sleeping for a while. The current coroutine context state transitions to "sleep state", and when awakened, its state transitions to "occupied state".

- "Block" control: The current coroutine suspends execution unconditionally until it is awakened. The current coroutine context state transitions to a "blocked state" until it is woken up.

- "Join" control: The current coroutine waits for another coroutine to finish executing before being woken up. The current coroutine context state transitions to a "blocked state".

The above six synchronous control operations are supported by the underlying "lock" algorithm, "unlock" algorithm, "join" algorithm, and "notify" algorithm.

4.2. Lock and Unlock Algorithms

The algorithm flow of "lock" and "unlock" performed by the coroutine includes the following five steps. The algorithm flow of "lock" and "unlock" is shown in Figure 5 and Figure 6.

1. When the coroutine needs to perform mutually exclusive access to a shared (critical area) data, first perform the "acquire mutex lock" operation on the shared data.

2. The "Acquire Mutex" operation first defines an atomic Boolean variable and initializes it to FALSE. FALSE means that the lock acquisition was unsuccessful, and TRUE means the lock acquisition was successful.

3. Through the atomic "comparison and exchange" operation (CAS operation) and memory barrier instructions provided by the processor, the "acquire mutex lock" operation loops to perform the comparison between the Boolean variable and TRUE, if it is TRUE, it indicates that the lock is acquired successfully, and then save the current coroutine context, exit the loop, and the coroutine starts to access the shared area (critical area) data mutually exclusively.

4. If the comparison result is FALSE, indicating that the lock has not been successfully acquired and performs a "yield" operation. When it is rescheduled, enter this step again and execute it in a loop until the mutex is successfully acquired.

5. When the shared data access is completed, the coroutine performs the "release mutex lock" operation and resets the Boolean variable to FALSE through the atomic "memory store" operation (Store operation) and memory barrier instructions provided by the processor.

```
 inline void lock() {
        uint64_t startOfContention = 0;
        while (locked.exchange(true, std::memory_order_acquire) != false) {
               if (startOfContention == 0) {
                      startOfContention = Cycles::rdtsc();
                } else {
                             uint64_t now = Cycles::rdtsc();
                      if (Cycles::toSeconds(now - startOfContention) > 1.0) {
                          startOfContention = now;
                      }
               }
               if (shouldYield) {
                   yield();
               }
        }
       owner = core.loadedContext;
}
```

Figure 5. Lock Algorithm Flow.

```
 inline void unlock() {
       locked.store(false, std::memory_order_release);
}
```

Figure 6. Unlock Algorithm Flow.

4.3. “Wait” and “Notify” Algorithms

The algorithm flow of coroutine execution “wait” includes the following four steps, and the algorithm flow is shown in Figure 7.

1. The coroutine first acquires a mutex until the mutex is successfully acquired.

2. The coroutine performs the “conditional wait” operation, which adds the current coroutine to the “coroutine waiting queue”, and then releases the acquired mutex lock.

3. This operation executes the coroutine task scheduling function once, yields the execution of the current coroutine, and selects the next target coroutine context. The status of the current coroutine is set to “blocked state” by the coroutine task scheduling function, waiting for other coroutines to wake it up.

4. After the coroutine is woken up, it returns from the coroutine task scheduling function, acquires a mutex again, and continues to execute.

```
 void ConditionVariable::wait(LockType& lock) {
     blockedThreads.push_back(ThreadID);
     lock.unlock();
     dispatch();
     lock.lock();
}
```

Figure 7. Wait Algorithm Flow.

The algorithm flow of coroutine execution “notify” includes the following six steps, and the algorithm flow is shown in Figure 8.

1. The coroutine first acquires a mutex until the mutex is successfully acquired.

2. Determine whether the “coroutine waiting queue” is empty, if it is empty, indicating that no coroutine needs to be wakened up, release the acquired mutex, and then return.

3. If the queue is not empty, select the coroutine at the head of the queue as the notify object, and remove the coroutine from the “coroutine waiting queue”.

4. Determine whether the status of the coroutine context is “blocked”, if so, set its wake-up time to 0, indicating that it can be scheduled for execution immediately, release the acquired mutex, and then return.

5. If it is not “blocked”, judge whether it is in a "sleep state”, if so, set its wake-up time to 0, indicating that it can be scheduled for execution immediately, release the acquired mutex, and then return.

6. The process of the “Notify all coroutines waiting for conditions” operation looping executes the “Notify a coroutine waiting for conditions” operation on the “coroutine waiting for queue” until it is empty.

```
 void ConditionVariable::notifyOne( ) {
     if ( blockedThreads.empty() )  return;
     ThreadId awakenenThread = blockedThreads.front();
     Blockedthreads.pop_front();
     Signal(awakenedThread);
}
 void ConditionVariable::notifyAll( ) {
     while (!blockedThreads.empty() )
            notifyOne();
  }
```

Figure 8. Notify Algorithm Flow.

5. Performance evaluation

We conducted the performance evaluation on the coroutine management environment implemented in this paper, and compares the cost of basic coroutine operations with C++ std::thread, Goroutine, and uThreads. std::thread is based on kernel threads, Go implements threads at user level, and uThreads uses kernel threads to multiplex user threads. The hardware environment used was an Intel dual-processor 8-core with the memory of 64G. The operating system was Linux CentOS 7 operating system, and the kernel version was 3.10.0-957.el7.x86_64.

We designed coroutine mechanism not just to minimize latency, but also to provide high throughput. We ran two experiments to measure coroutine creation throughput and condition variable notify throughput. The results of the three sets of performance comparison tests are as follows.

5.1. Average latency test for coroutine creation

Creation costs are measured end-to-end, from initiation in one thread until the target thread wakes up and begins execution on a different core. coroutine creates all threads on a different core from the parent, Go always creates Goroutines on the parent's core, and uThreads uses a round-robin approach to assign

threads to cores. The performance test results are shown in Table 1.

Table 1. Comparison test results on the average latency of coroutine creation.

operation	Coroutine	Goroutine	uThreads	Std::thread
Thread Creation cost	350(ns)	536ns(ns)	7234(ns)	14532(ns)

This test demonstrates that a single coroutine can spawn more than 5 million new threads per second, which is 1.5x the rate of Go, at least 50x the rate of std::thread, and at least 20x the rate of uThreads. This experiment demonstrates the benefits of performing load balancing at thread creation time.

5.2. Average latency test for notifying condition variable

In "Yield" coroutine synchronization, control passes from the yielding thread to another runnable thread on the same core, "Thread Exit Turnaround" measures the time from the last instruction of one thread to the first instruction of the next thread to run on a core. The performance test results are shown in Table 2.

Table 2. Comparison test results on the average latency of notifying condition variable.

operation	Coroutine	Goroutine	uThreads	Std::thread
condition variable notify cost	274(ns)	513(ns)	5234(ns)	5244(ns)

This test demonstrates that the throughput of condition variable notify is 2x the rate of Go, at least 10x the rate of std::thread, and at least 10x the rate of uThreads. This experiment demonstrates the benefits of coroutine synchronization control mechanism.

6. Conclusion

This article provides an environment for coroutine management, which decouples the business processing of the application system from the operation management of coroutines, realizes common functions of coroutine creation, scheduling, synchronization, migration, destruction, and shields applications from coroutine operation and management complexity, and provides an application development interface similar to the thread model. Coroutine applications only need to use the application development interface provided by the environment to register the task processing functions defined by the business in the environment in the form of coroutines, without caring about the scheduling, execution, and management of specific task processing functions. It greatly reduces the complexity of coroutine application development and the difficulty of upgrading and maintenance.

At the same time, this paper also effectively integrates coroutine management with the scheduling mechanism of the operating system to achieve efficient coroutine creation, scheduling, synchronization, migration, and destruction functions. After obtaining application requests, it transparently registers, executes, and synchronizes the task processing functions, which realizes the isolated operation of the coroutine and the kernel thread, so that the overall performance of the system can be scaled with the load change of the processing task, which provides a more efficient coroutine management environment for high concurrency and low latency systems.

Acknowledgments

This study was supported by Guangdong Provincial Department of Education's University Scientific Research Platform and New Generation Information Technology Key Fields Project "Research on Panoramic Information Fusion Computing and Evaluation in an Open Smart Learning Environment" (Grant No. 2021ZDZX1076); China University Industry-University-Research Innovation Fund "Smart Campus Data Governance Methods and Practices" (Grant No. 2021BCE01008); China University Industry-University-Research Innovation Fund "Research on Distributed Storage Optimization and Concurrent Acceleration Technology for AI Computing" (Grant No. 2021LDA04002); Guangdong Open University (Guangdong Polytechnic Institute) Talent Special Project "Intelligent Database Multi-Storage Engine Fusion Technology for Domestic Processor Platform" (Grant No. RC2002); Key Platform and Research Project of Guangdong Provincial Department of Education "Recommendation Learning Algorithms Driven by Hybrid Knowledge Graph and Deep Learning and Its Application in Lifelong Education" (Grant No. 2020KTSCX401); Guangdong Open University (Guangdong Polytechnic Institute) school-level scientific research project "Research on Deep Learning Models of Semantic Context in Event Extraction and Prediction Tasks based on Temporal Events and Semantic Context" (Grant No. ZD2001).

References

[1] L. Barroso, M. Marty, D. Patterson, and P. Ranganathan. Attack of the Killer Microseconds. Communications of the ACM, 60 (4): 48–54, Mar. 2017.

[2] A. Dragojevic, D. Narayanan, M. Castro, and O. Hodson. FaRM: Fast Remote Memory. In 11th USENIX Symposium on Networked Systems Design and Implementation (NSDI 14), pp. 401–414, 2014.

[3] H. Qin. Arachne. https://github.com/ PlatformLab/Arachne.

[4] Redis. http://redis.io.

[5] J. Leverich and C. Kozyrakis. Reconciling High Server Utilization and Sub-millisecond Quality-of-Service. In Proc. Ninth European Conference on Computer Systems, EuroSys '14, pp. 4:1–4:14, 2014.

[6] H. Lim, D. Han, D. G. Andersen, and M. Kaminsky. MICA: A Holistic Approach to Fast In-Memory Key-Value Storage. In Proc. 11th USENIX Symposium on Networked Systems Design and Implementation (NSDI 14), pp. 429–444, Apr. 2014.

[7] Mutilate: high-performance Memcached load generator. https://github.com/leverich/mutilate.

[8] Nginx. https://nginx.org/en/.

[9] EISENMAN, A., GARDNER, D., ABDELRAHMAN, I., AXBOE, J., DONG, S., HAZELWOOD, K., PETERSEN, C., CIDON, A., AND KATTI, S. Reducing DRAM footprint with NVM in Facebook. In European Conference on Computer Systems (EuroSys '18) (New York, NY, USA, 2018), pp. 42:1–42:13.

[10] JEONG, S., LEE, K., LEE, S., SON, S., AND WON, Y. I/O stack optimization for smartphones. In USENIX Annual Technical Conference (USENIX ATC '13) (San Jose, CA, USA, 2013), pp. 309–320.

[11] JOSHI, K., YADAV, K., AND CHOUDHARY, P. Enabling NVMe WRR support in Linux block layer. In USENIX Workshop on Hot Topics in Storage and File Systems (HotStorage'17) (Santa Clara, CA, USA, 2017).

[12] A. Lottarini, A. Ramirez, J. Coburn, M. A. Kim, P. Ranganathan, D. Stodolsky, and M. Wachsler. vbench: Benchmarking video transcoding in the cloud. In Proceedings of the Twenty-Third International Conference on Architectural Support for Programming Languages and Operating Systems, pp. 797–809, 2018.

[13] J. Ousterhout, A. Gopalan, A. Gupta, A. Kejriwal, C. Lee, B. Montazeri, D. Ongaro, S. J. Park, H. Qin, M. Rosenblum, et al. The RAMCloud Storage System. ACM Transactions on Computer Systems (TOCS), 33 (3): 7, 2015.

[14] G. Prekas, M. Kogias, and E. Bugnion. ZygOS: Achieving Low Tail Latency for Microsecond-scale Networked Tasks. In Proc. of the 26th Symposium on Operating Systems Principles, SOSP '17, pp. 325–341, 2017.

[15] Memtier benchmark. https://github.com/RedisLabs/memtier_benchmark.

[16] L. Merritt and R. Vanam. x264: A high-performance H. 264/AVC encoder. http://neuron2.net/library/avc/overview_x264_v8_5.pdf.

[17] M. Mitzenmacher. The Power of Two Choices in Randomized Load Balancing. IEEE Transactions on Parallel and Distributed Systems, 12 (10): 1094–1104, 2001.

[18] AHN, S., LA, K., AND KIM, J. Improving I/O resource sharing of Linux cgroup for NVMe SSDs on multi-core systems. In USENIX Workshop on Hot Topics in Storage and File Systems (HotStorage ‘16) (Denver, CO, USA, 2016).

[19] CAULFIELD, A. M., DE, A., COBURN, J., MOLLOW, T. I., GUPTA, R. K., AND SWANSON, S. Moneta: A high-performance storage array architecture for next-generation, non-volatile memories. In Annual IEEE/ACM International Symposium on Microarchitecture (MICRO ‘10) (Atlanta, GA, USA, 2010), pp. 385–395.

[20] CAULFIELD, A. M., MOLLOW, T. I., EISNER, L. A., DE, A., COBURN, J., AND SWANSON, S. Providing safe, user-space access to fast, solid state disks. In International Conference on Architectural Support for Programming Languages and Operating Systems (ASPLOS XVII) (New York, NY, USA, 2012), pp. 387–400.

[21] CHIDAMBARAM, V., PILLAI, T. S., ARPACI-DUSSEAU, A. C., AND ARPACI-DUSSEAU, R. H. Optimistic crash consistency. In ACM Symposium on Operating Systems Principles (SOSP ‘13) (Farmington, PA, USA, 2013), pp. 228–243.

The transfer method on the prompt for the summarization task

Xinhao Guo[a], Da Xiao*
School of Cyberspace Security, Beijing University of Posts and Telecommunications, Beijing, China
[a]shawn_g@qq.com
*Corresponding author: xiaoda99@bupt.edu.cn

Abstract

Prompt tuning is a parameter-efficient method that even surpasses traditional fine-tuning methods in few-shot scenarios. Nowadays, pre-trained language models are getting larger and larger, with more and more parameters, which makes the traditional fine-tuning method impractical to implement and consumes a lot of computing resources. Therefore, prompt-based methods have a broad application prospect. In the experiments, it is found that prefix tuning, a prompt-based method, has the problem of non-convergence or is quite slow to converge when the training samples are small. This paper proposes a cross-task parameter transfer method, which transfers the trained parameters from prompt tuning tasks to prefix tuning to improve the training speed and alleviate the problem of non-convergence or slow convergence in prefix tuning tasks.

Keywords - transfer learning; prompt tuning; text generation; natural language processing

1. Introduction

In natural language processing tasks, implementing fine-tuning on pre-trained language models is a prevalent way to train downstream tasks and has achieved good results on many tasks. However, as today's language models are getting larger and larger, the volume of parameters is increasing, and it is difficult to continue to implement the traditional full-model fine-tuning methods to do downstream tasks, which requires a lot of resources. To solve this problem, a new training paradigm has emerged. The power of large pre-trained language models, such as GPT-3, means that for some downstream tasks, we can just use GPT-3 to do what we want, without even fine-tuning. To achieve better results, we can add task-specific prompts to guide the model. Based on this intuition, some researchers have proposed prompt-based methods, which introduce some external parameters independent of the pre-trained language model. During training, the language model is not fine-tuned, only the introduced external parameters are tuned [1], [2], which has become a popular training paradigm. And in some tasks, the effect is even better than the traditional fine-tuning. However, the existing prompt-based methods have the problem of slow convergence [3]. This paper aims to solve the problem of convergence of prefix tuning on text generation tasks. The result of the experiments shows that the parameter transfer method proposed in this paper can accelerate the convergence speed of the model, save training time, and help to improve the performance of the model.

2. Related Work

2.1. Prompt tuning

Prompt-based methods can be subdivided into two, discrete prompts and continuous prompts. Discrete prompt, also called hard prompt, can be selected manually. For example, 'TL; DR' namely too long, don't read is a hard prompt for summarization tasks. But this kind of prompt needs to be selected by the human being and it is time-consuming, so some automatic search methods have been proposed[4], which will search for the optimal prompt in the vocabulary space. The expression space of such a hard prompt is limited to the glossary, which limits the ability of this method. Therefore, continuous prompts, also called soft prompts, are proposed, and introduced prompts can be trained and expand the expressive space of prompt [5], [6]. Figure 1(a) is an example of soft prompt tuning.

2.2. Prefix tuning

Prefix tuning[2], a training method based on prompt, is different from the method described in section A, which introduces additional parameters at each layer of the model, not at the word embedding layer. The number of parameters is more than prompt tuning, but still several orders of magnitude less than the number of parameters of the model. Figure 1(b) is an example of prefix tuning.

International Conference on Computer Graphics, Artificial Intelligence, and Data Processing (ICCAID 2022)
edited by Ruishi Liang, Jing Wang, Proc. of SPIE Vol. 12604, 126040L
© 2023 SPIE · 0277-786X · doi: 10.1117/12.2674622

2.3.Transferability

Transfer learning is to use the trained task parameters to transfer to the new task to solve the problem of fewer training samples for the new task and accelerate the training process. At present, the task transfer based on prompt uses the source prompt to initialize the prompt parameters of the target task to improve the performance of the model, and cross-model transfer is also applied [7]. This paper proposes a new transfer method that transfers prompt tuning's parameters to prefix tuning and alleviates the problem of convergence and improves the performance of the model.

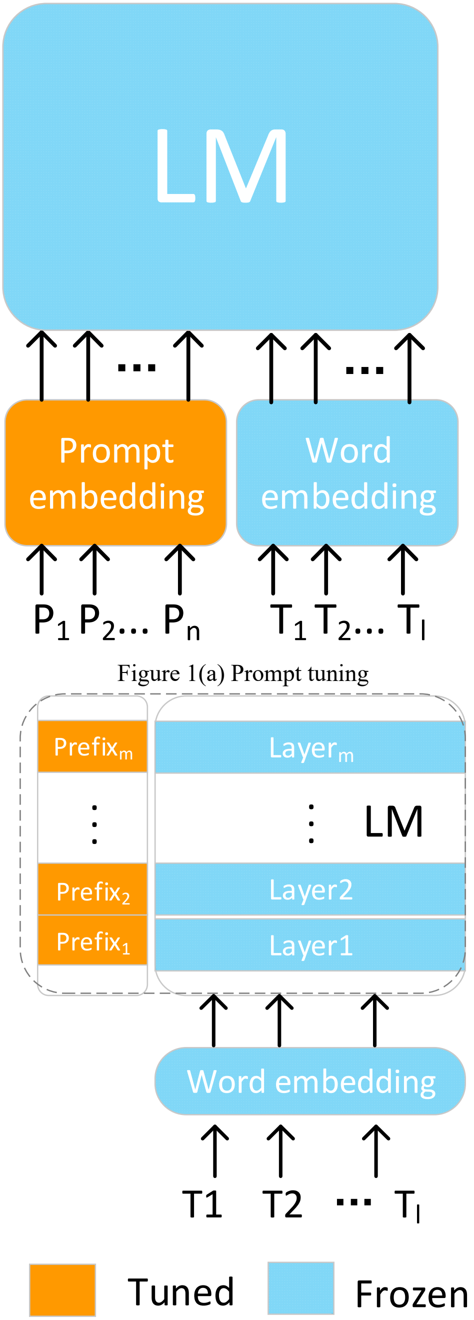


Figure 1(a) Prompt tuning

Figure 1(b) Prefix tuning

Figure 1 Illustration of prompt tuning and prefix tuning: P_i(i=1,2,…,n) denotes the id of the prompt token, n denotes the length of the prompt, T_i(i=1,2,…,l) denotes the id of the text token, and l denotes the number of text tokens. LM stands for language model. Tuned means parameters need to be trained, and Frozen means parameters needn't be trained.

3. METHOD AND EXPERIMENT SETUP

3.1.The Transfer Method

The transfer method proposed in this paper is shown in Figure 2. First, we train a prompt tuning task and then input the trained prompt embedding to the model. After passing through the model, each layer of the model will have a corresponding hidden state which refers to the output of the key and value vector of each layer. And we save these keys and values vectors and use them to initialize the prefix parameters.

Prompt tuning works by using the keys and values calculated at each layer after a forward pass to act on the text that follows, while prefix tuning uses the fixed key and value vector directly at the front of each layer. Therefore, we can use this method to initialize prefixes. The only difference between the two is that the key and value of prompt tuning need to be computed during the forward pass, while the key and value of prefix tuning are fixed in front of each layer and don't need to pass.

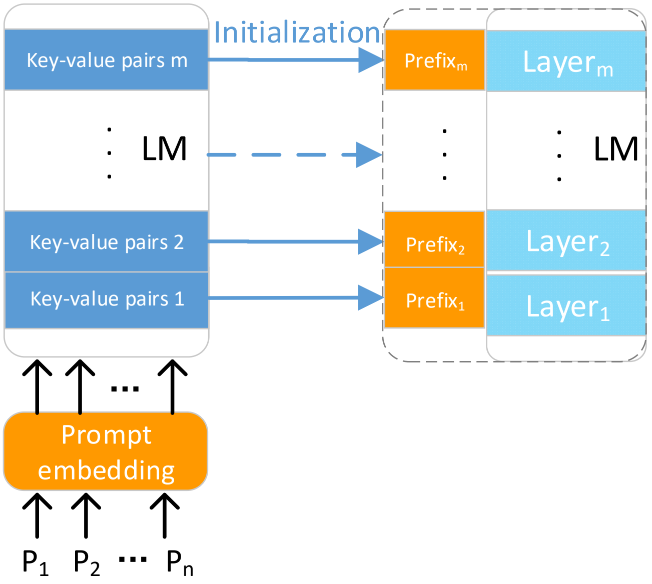


Figure 2 Illustration of transfer method: Key-value pairs i(i=1,2,…,m) represent the hidden state output of each layer of the model, and m represents the total number of layers of the model. Prefix_i(i=1,2,…,m) denotes the prefix parameters that need to be tuned for each layer.

3.2.Datasets and Metrics

The experiments are implemented on the summarization task of text generation and the dataset used in the experiment is the XSUM dataset[8], which is a news article summary dataset. An example of data is shown in Table 1. There are about 225k examples and only part of them are used in our experiment. The ratio of the training set to the validation set is 5 to 1.

The metrics we use are Rouge-1, Rouge-2, and Rouge-L. Rouge-1 refers to the overlap of unigram between the generated summaries and reference summaries. Rouge-2 refers to the overlap of bigrams between the generated summaries and reference summaries. Rouge-L is based on longest common subsequence (LCS) based statistics. The longest common subsequence problem takes into account sentence-level structure similarity naturally and identifies the longest co-occurring in sequence n-grams automatically. The perplexity (PPL) of the model is recorded during the training process, and the smaller perplexity, the better performance of the model.

Table 1 Example of XSUM dataset

Text	Summary
The firm said it intends to focus on residential customers and is looking to sell off its non-household retail division. The division is currently based in Bradford and Barnsley. A spokesman for the company said it was "early days" in the process, adding that it was important to keep staff informed. More on this and other local stories from across Yorkshire. He added that about 100 staff were directly involved in the non-household retail business at sites in Bradford and Barnsley. Yorkshire Water currently has about 138,000 business customers.	Yorkshire Water has announced plans to sell off its business supply division, with 100 jobs potentially at risk.

3.3.Training method and details

The language model we use is GPT-Neo[9], and the number of parameters is about 1.3B. The experiments are based on HuggingFace AI Community.

Firstly, a fixed epoch prompt tuning task is trained, which uses epoch=5 and does not require its convergence. After training, the prefix parameter is initialized by the forward result of tuned prompt embedding and starts to train prefix tuning. The prompt embedding is initialized with the task-relevant word summary.

In the experiment, we use the early stopping method to stop the training process when overfitting occurs. Only PPL before the lowest point is presented, and the overfitting part of PPL is not shown. The experiments compare the transfer method with random initialization which adopts the Xaiver initialization method. The model is trained on a single NVIDIA TITAN RTX 24GB. At decoding time, we use multinomial sampling to sample, the parameters such as top_p and top_k taken as default values, the temperature is set to 0.3, and the maximum length of generated summary is set to 60.

The main hyperparameters we tuned are the training data size and the length of the prefix, and the specifically tuned parameters are shown in Table 2. In low-data settings, the training data size and the prefix length is the key factor for the model performance [2], so we tune them. Because we focus on few-shot learning, so the training data size is small

Table 2 Tuned hyperparameters groups

	Training data size	Initialization method	Prefix length
Group1	1024	Transferred	10
Group 2	1024	Random initialization	10
Group 3	2000	Transferred	10
Group 4	2000	Random initialization	10
Group5	1024	Transferred	20
Group 6	1024	Random initialization	20
Group 7	2000	Transferred	20
Group 8	2000	Random initialization	20

4. RESULTS AND EVALUATION

4.1.Main results

The PPL of the validation set recorded during the training process is shown in the following multiple figures:

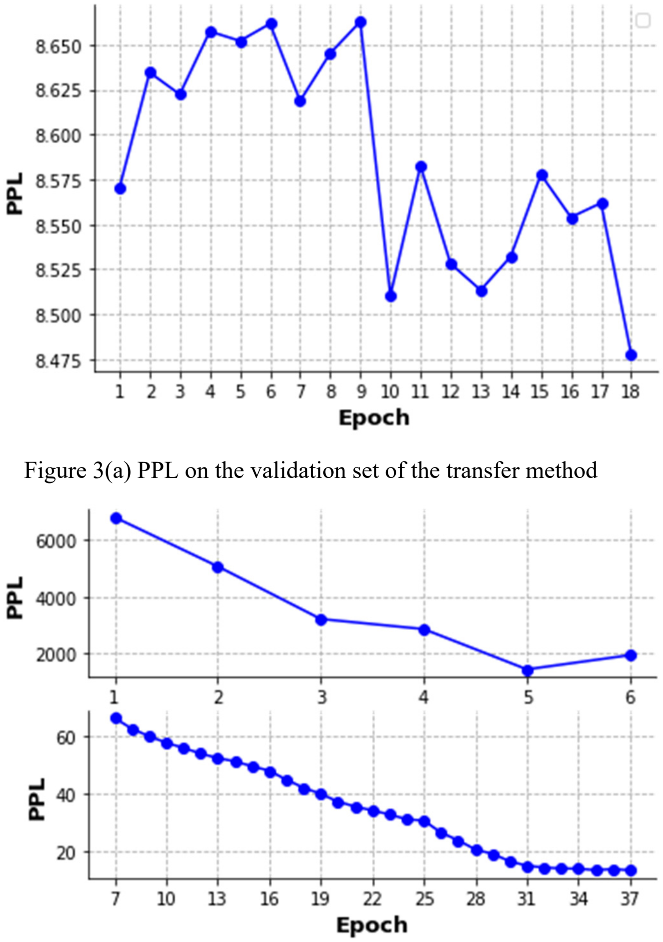


Figure 3(a) PPL on the validation set of the transfer method

Figure 3(b) PPL on the validation set of the random initialization method[1]

Figure 3 PPL on the validation set of different methods, training data size=1024, prefix length=10

[1] Due to the large variance of perplexity caused by random initialization, the curve of PPL is drawn separately, including Figure (4)b, Figure (5)b, and Figure (6)b, to present the result.

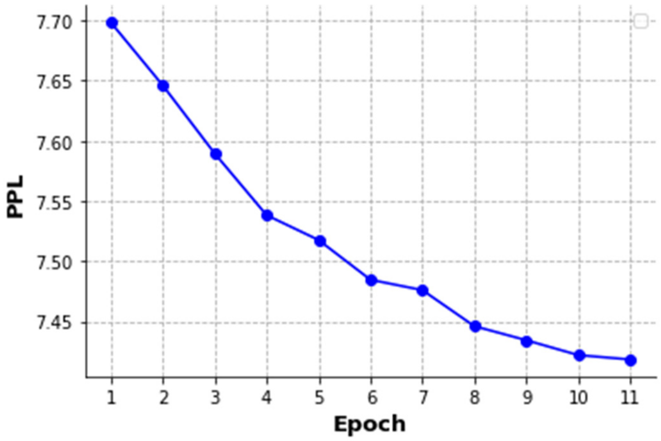


Figure 4(a) PPL on the validation set of the transfer method

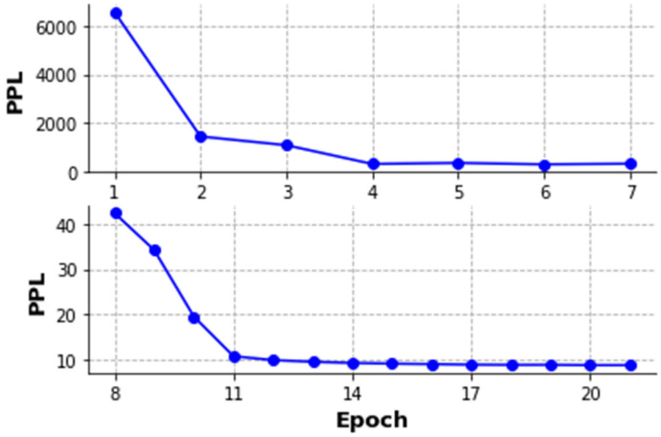


Figure 4(b) PPL on the validation set of the random initialization method

Figure 4 PPL on the validation set of different methods, training data size=2000, prefix length=10

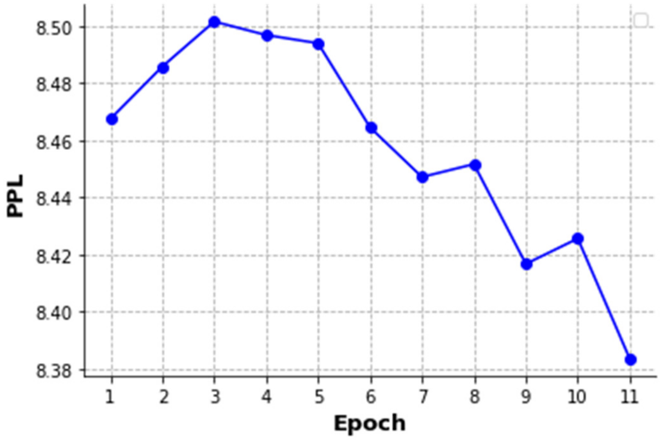


Figure 5(a) PPL on the validation set of the transfer method

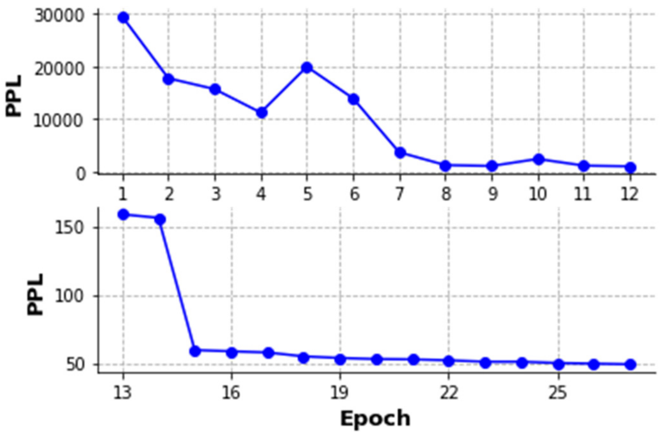


Figure 5(b) PPL on the validation set of the random initialization method

Figure 5 PPL on the validation set of different methods, training data size=1024, prefix length=20

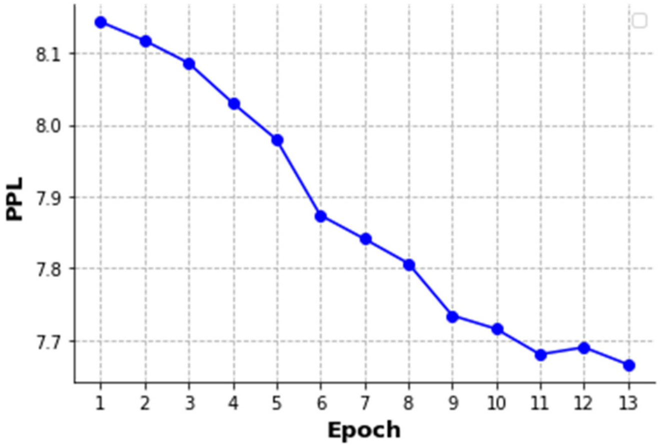


Figure 6(a) PPL on the validation set of the transfer method

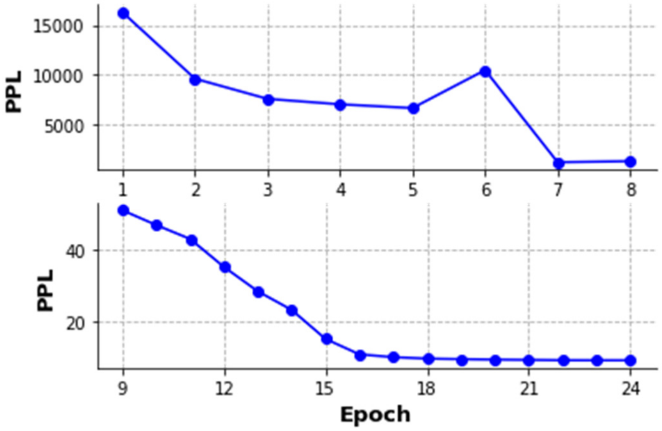


Figure 6(b) PPL on the validation set of the random initialization method

Figure 6 PPL on the validation set of different methods, training data size=2000, prefix length=20

Rouge scores are shown in Table 3 and Table 4.

Table 3 Rouge score on the test set (prefix length=10)

Training data size	**Initialization method**	**Rouge-1**	**Rouge-2**	**Rouge-L**
1024	Transferred	26.41	8.52	22.12
1024	Random initialization	24.47	7.19	20.67
2000	Transferred	27.51	8.78	23.44
2000	Random initialization	27.46	8.85	22.52

Table 4 Rouge score on the test set (prefix length=20)

Training data size	**Initialization method**	**Rouge-1**	**Rouge-2**	**Rouge-L**
1024	Transferred	28.00	8.84	23.49
1024	Random initialization	20.67	5.92	18.41
2000	Transferred	28.30	10.58	24.33
2000	Random initialization	26.27	8.65	22.00

4.2.Analysis

For training data size, when the dataset is small, 1024, the PPL oscillates for the transfer method, but it still converges faster than random initialization. When the training data size is 2000, PPL converges steadily for the transfer method.

For prefix length, the longer the prefix length is, the more training data size is needed for random initialization. In the experiment, when the prefix length is 20 and the training data size is 1024, although the PPL has a downward trend, the minimum value of PPL is 49.09, and the performance is poor. On the contrary, the transfer method can achieve a lower PPL on the same setup.

In the experiments, it’s found that the proposed method is conducive to the rapid convergence of the model. Including the 5 epochs of the prompt tuning phase, the convergence speed is still faster than random initialization. Random initialization will lead to a large PPL at the beginning of the training, and the training process is unstable. PPL will drop sharply in an unpredictable epoch, and the final PPL after convergence is not as small as that using the transfer method.

In terms of the quality of the generated summaries, the rouge scores of the summaries generated by the transfer method are generally higher than those of the random initialization, which indicates that the method helps the model generate higher quality summaries. Among experiments, the highest score is obtained when the prefix length is 20 and the training data size is 2000, which indicates longer prefix length and more training data size can help improve the performance of the model.

5. CONCLUSIONS

In this paper, we present a new transfer method from prompt tuning to prefix tuning. To verify the feasibility of the method, we conduct some experiments based on the method. Through experiments, the proposed transfer method is proven to be effective in text generation tasks, which is much stronger than the random initialization method and can accelerate the training process and improve the quality of text generation.

REFERENCES

[1] Lester, Brian, Rami Al-Rfou, and Noah Constant. "The power of scale for parameter-efficient prompt tuning." arXiv preprint arXiv:2104.08691 (2021).

[2] Li, Xiang Lisa, and Percy Liang. "Prefix-tuning: Optimizing continuous prompts for generation." arXiv preprint arXiv:2101.00190 (2021).

[3] Su, Yusheng, et al. "On transferability of prompt tuning for natural language understanding." arXiv preprint arXiv:2111.06719 (2021).

[4] Shin, Taylor, et al. "Autoprompt: Eliciting knowledge from language models with automatically generated prompts." arXiv preprint arXiv:2010.15980 (2020).

[5] Liu, Xiao, et al. "GPT understands, too." arXiv preprint arXiv:2103.10385 (2021).

[6] Schick, Timo, and Hinrich Schütze. "It's not just size that matters: Small language models are also few-shot learners." arXiv preprint arXiv:2009.07118 (2020).

[7] Vu, Tu, et al. "Spot: Better frozen model adaptation through soft prompt transfer." arXiv preprint arXiv:2110.07904 (2021).

[8] Narayan, Shashi, Shay B. Cohen, and Mirella Lapata. "Don't give me the details, just the summary! topic-aware convolutional neural networks for extreme summarization." arXiv preprint arXiv:1808.08745 (2018).

[9] Gao, Leo, et al. "The pile: An 800GB dataset of diverse text for language modeling." arXiv preprint arXiv:2101.00027 (2020).

Research on video image stabilization decision algorithm based on ROI block matching method

Xiong-juan LING [1], Yun-jiao ZHOU [2,3], Jian-xi Peng [1]
[1] School of Electronic Information, Foshan Polytechnic, Foshan 528000, China;
[2] School of Mechanical and Automotive Engineering, South China
University of Technology, Guangzhou 510640, China
[3] Guangdong Provincial Key Laboratory of Automotive

Abstract

Aiming at the problem that the feasibility and accuracy of image measurement and analysis are reduced due to video stabilization in automobile collision experiment, a video image stabilization decision algorithm based on block matching method based on region of interest (ROI) is proposed in this paper. Firstly, in the first frame, the user selects the ROI region containing the marker identification, uses the block matching method to automatically match the ROI region features frame by frame, and estimates the motion of the video frame. And then determines whether the feature matching fails according to the motion of the video frame. When the feature matching fails, the search range can be reduced and the ROI features can be re matched to improve the matching accuracy and speed, Finally, the video frame sequence after image stabilization is output by frame reverse motion compensation according to the amount of motion. The experimental results show that the proposed algorithm can effectively reduce the video stabilization in collision experiments in the subjective evaluation and the objective evaluation of local peak signal-to-noise ratio.

Keywords: Region of interest; Block matching method; Video stabilization; Policy decision; Peak signal to noise ratio

1. Introduction

In automobile collision experiments, due to the huge impact force generated when a collision occurs, it is very easy to cause high-speed camera shaking, and this undesired shaking will seriously affect the feasibility and accuracy of subsequent image analysis and cannot meet the requirements of complete and clear image information acquisition, which directly affects the effect of automobile safety technology development and test verification and causes significant economic losses [1, 2].

The current video Stabilization technology is divided into mechanical Stabilization, optical Stabilization, and electronic Stabilization[3]. Due to the strong impact of an automobile collision, mechanical and optical stabilization cannot guarantee that the captured video is completely free from stabilization, so video stabilization of crash experiments has been a problem in the field of automobile safety. With the development of image processing technology, electronic image stabilization technology has been widely used in the field of UAVs and robots by using high-speed digital image signal processing technology to remove stabilization from the captured image sequences frame by frame. Zheng Xie [4] et al. used the SURF algorithm to extract features for mobile flight platform video stabilization, used bidirectional matching strategy to obtain matched point pairs, RANSAC to eliminate mismatched points, and finally used affine motion parameters to compensate video frames; Yannan Yin [5] et al. proposed Kalman filter-based visual monitoring algorithm for video surveillance image de-stabilizationing for video surveillance stabilization, and the feature point motion trajectory Kalman filtering is performed to compensate for video stabilization and achieve a steady image.

However, the information of each frame in the car collision video is crucial, the frame rate of high-speed camera shooting is generally above 1000FPS (Frames Per Second), and the use of a common electronic stabilization algorithm although visually feels that the stabilization has been eliminated the frame-by-frame viewing will find a frame de-stabilizationing effect is not good or even the phenomenon of lost frames, can not be applied in the collision video stabilization to remove stabilization. This paper proposes an ROI (Region of Interest) based block matching video stabilization decision algorithm, firstly, in the first frame, the user selects two ROI regions containing the Marker logo, automatically extracts the ROI feature regions, then uses the block matching method to automatically match the feature regions frame by frame, calculates the amount of motion for the feature regions before and after the two frames, and judges the feature based on the size of the motion If the amount of motion is too large, the user can manually select the

International Conference on Computer Graphics, Artificial Intelligence, and Data Processing (ICCAID 2022)
edited by Ruishi Liang, Jing Wang, Proc. of SPIE Vol. 12604, 126040M
© 2023 SPIE · 0277-786X · doi: 10.1117/12.2674608

ROI' region and match the ROI feature marker within the ROI' range to improve the matching accuracy and speed, and finally perform motion compensation frame by frame to output the stabilized video.

2. Algorithm design

2.1 Algorithm flow

The algorithm flow chart of the video stabilization decision algorithm based on ROI block matching is shown in Figure 1.

The process can be described as follows.

1) Before shooting the video, Markers are pasted on the B-pillar, door frame, and other locations with less deformation. Theoretically, the on-board camera and the B-pillar and door frame remain relatively stationary, so Markers that clearly distinguish from environmental features are pasted on the B-pillar and door frame to facilitate the subsequent automatic matching of feature areas, and the common Marker described are shown in Figure 2.

2) Importing the collision experiment video, pre-processing the video with filtering and noise reduction, and converting it into grayscale image video frames, the user manually selects two ROI regions containing Marker in the first frame of the video, and extracts the features within the ROI as the subsequent features to be matched. The user is recommended to select the diagonal area of the video frame for the two selected feature areas, because the further the location of the two feature areas, the better the subsequent motion compensation effect.

3) Matching the ROI features selected in step 2) frame by frame using the block matching method, and calculating the offset and rotation angle of the ROI features between the current video frame and the previous video frame.

4) To solve the problem that the feature area disappears during the collision, is obscured by other splash objects, or the feature area changes or disappears due to depth of field, determine whether the corresponding feature is correctly matched in each video frame, set the thresholds for the offset and rotation angle of the video frame, and when the motion estimation exceeds the threshold, it means that the feature matching is abnormal and manual intervention is needed to correct it, and enter step 5), and if the threshold is not exceeded, enter step 6);.

5) reselect the ROI region with a range greater than that in step 2) defined as ROI', and match the ROI feature region in ROI' using the block matching method to narrow the search range to improve the matching accuracy.

6) Based on the offset and rotation angle of the motion estimation, reverse motion compensation of all video frames of the video frame sequence to eliminate stabilization, exported as low bit rate, high-quality H264 video coded steady image video.

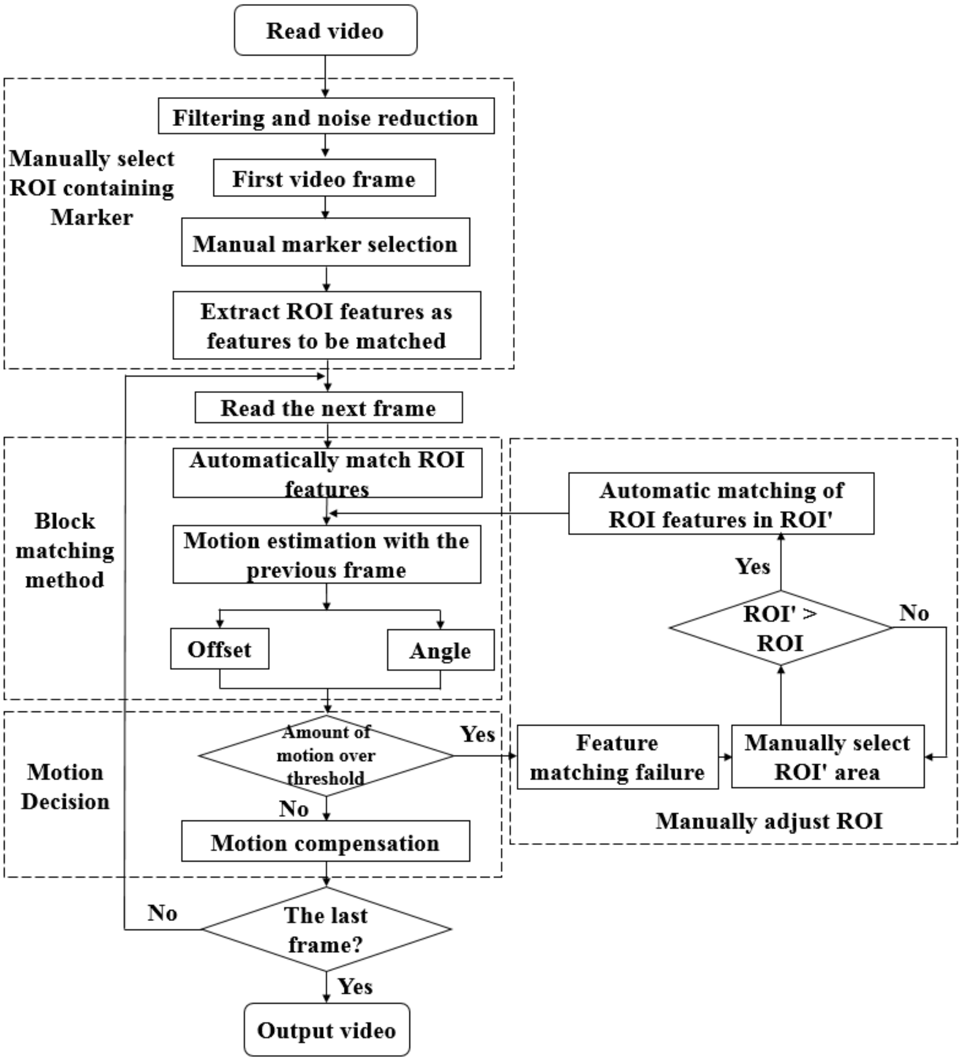

Read video
Manually select ROI containing Marker
Filtering and noise reduction
First video frame
Manual marker selection
Extract ROI features as features to be matched
Read the next frame
Block matching method
Automatically match ROI features
Motion estimation with the previous frame
Offset
Angle
Automatic matching of ROI features in ROI'
Yes
ROI' > ROI
No
Motion Decision
Amount of motion over threshold
Yes
Feature matching failure
Manually select ROI' area
No
Motion compensation
Manually adjust ROI
No
The last frame?
Yes
Output video

Figure 1 Algorithm flow chart

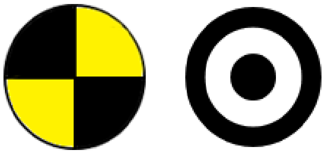


Figure 2 Common Marker

2.2 Block matching method

The block matching method is an electronic image stabilization algorithm with good stability and high accuracy. However, when the block matching method searches from the starting point to the matching point in the global area, it needs to match the image blocks repeatedly, which causes too much computation and decreases the accuracy. Therefore, this paper starts from the idea of narrowing the search range of the block matching method, and when the feature matching fails, the search range is narrowed by selecting ROI' which is slightly larger than the ROI range to reduce the computation and improve the matching accuracy. The size of ROI' is defined as $P \times Q$.

The block-matching method uses the gray value of the image as the feature in the calculation and searches for the best image block with the smallest deviation from the reference image block within the search window of the image. If the upper-left corner pixel coordinate of the ROI area in the k frame is (x_1, y_1), the upper-left corner pixel coordinate of the best matching block of the same size in the search window of the frame $k+1$ is (x_2, y_2), and the difference between the two is the global motion vector of the image (T_x, T_y). The common matching criteria used to calculate the global motion vector are mainly the mean absolute difference criterion (MAD), the mean squared error criterion (MSE), and the sum of absolute differences (SAD) [6, 7].

Among them, the MAD is defined as follows:

$$MAD(x,y)= \frac{1}{PQ}\sum_{x=1}^{P}\sum_{y=1}^{Q}\left|G_k(x,y)-G_{k+1}\left(x+T_x,y+T_y\right)\right| \tag{1}$$

MSE is defined as:

$$MSE(x,y)= \frac{1}{PQ}\sum_{x=1}^{P}\sum_{y=1}^{Q}\left[G_k(x,y)-G_{k+1}\left(x+T_x,y+T_y\right)\right]^2$$

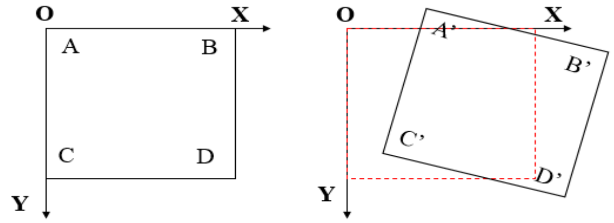
Both of these laws require the calculation of multiplication. To simplify the calculation process of MAD, the size of the ROI region $P \times Q$ is a constant value, so the formula (1) is simplified and the formula of SAD is obtained:

$$SAD(x,y)= \sum_{x=1}^{P}\sum_{y=1}^{Q}\left[G_k(x,y)-G_{k+1}\left(x+T_x,y+T_y\right)\right]^2 \tag{2}$$

Since SAD does not need to do multiplication operations, the hardware consumption is low. Taking into account the speed and accuracy of the algorithm, this criterion is adopted in this paper.

2.3 Steady Image Decision

The position relationship between the dithered video frame and the initial image is shown in Figure 3.



(a) Initial image (b) Dithered video frames

Figure 3 Schematic diagram of the dithered video frame and initial image position

For the selected ROI rectangular area ABCD shown in Figure 3(a), when dithering occurs, it becomes A'B'C'D' in Figure (b), deviating from the red dashed box, which should move $\vartriangle X$ along the X-axis, $\vartriangle Y$ along the Y-axis, when A ' moves to the position of point A. To make A'B'C'D' the same as ABCD, it is also necessary to rotate an angle α :

$$\alpha=\arctan\left(\frac{\vartriangle X}{\vartriangle Y}\right) \tag{3}$$

If the feature region is correctly matched, the offset and rotation angle are directly calculated according to the motion estimation, and the inverse motion-compensated video frame is stabilized. However, the feature matching failure problem often occurs in actual use, so this paper uses the threshold decision method of motion to determine whether the feature matching is successful, and the decision rules are as follows.

$$\begin{cases} \sqrt{\Delta X^2 + \Delta Y^2} > T_1 \\ |\alpha| = \left| \arctan\left(\frac{\Delta X}{\Delta Y}\right) \right| > T_2 \end{cases} \tag{4}$$

Where T_1 is the threshold value of video frame displacement amount and T_2 is the threshold value of video frame rotation amount. Since the frame rate of the collision experiment is above 1000, the motion between two consecutive frames is theoretically not too large. When the offset or rotation of the ROI feature region of the video frame exceeds the threshold, it is judged that the feature matching fails, and manual intervention is needed to select the ROI' region and match the ROI again in the ROI' region In this way, the matching accuracy and speed can be improved by narrowing the search area.

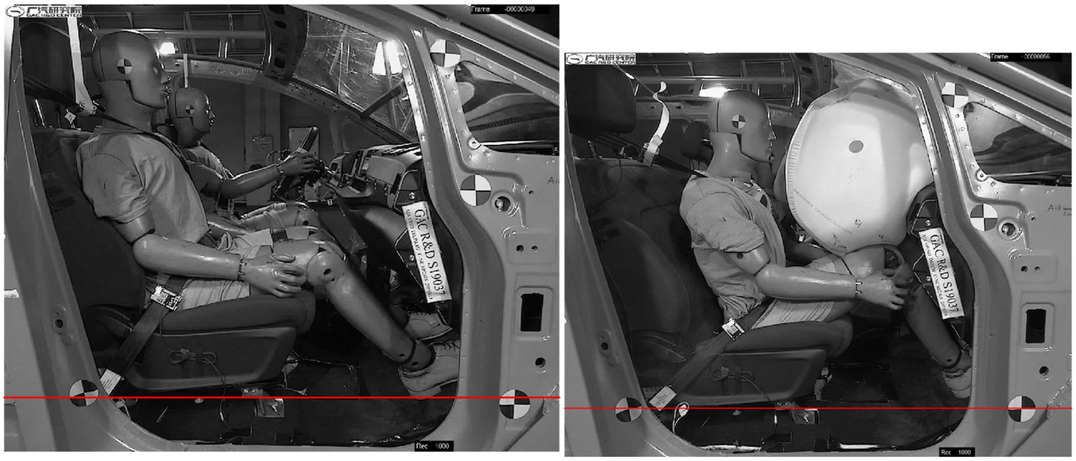
3. Experimental results and analysis

This paper uses the actual collision experiment video taken by the vehicle high-speed camera as the verification material, and the high-speed camera and video information are shown in Table 1.

Table 1 High-speed camera and video information

High-speed camera information		Crash video information	
Brand	NAC Q1m	Duration	350ms (50ms before collision)
Focal length	5mm fixed	Number of frames	351
Frame rate	1000FPS	Video frame width	1280
Memory size	4GB	Video frame height	1024

In general, with the moment of collision T0 as the boundary, the collision experiment video will retain T0 before a period, at this time the car camera and the car maintain relative static, this experiment retained T0 before 50ms, that is, the first 50 frames did not occur stabilization, at this time the video frame as shown in Figure 4 (a), after the collision hair, airbags, seat belts began to act, the collision dummy began to move, as shown in Figure 4 (b).



(a) Original dithered video initial frame (b) Frame 100 of the original dithered video

Figure 4 The original stabilization of the collision experiment video

In order to visualize the amount of stabilization in the collision video, a fixed marker line is drawn at the bottom of the video frame, and it can be found that the Marker near the marker line jumps more in different frames, indicating that the stabilization is more serious. In the following, the method of this paper is used to stabilize the collision experiment video and evaluate the effect of anti-shake from subjective evaluation and objective indexes respectively.

3.1 Subjective evaluation of the algorithm stabilization effect

In this paper, the ROI feature area selected for stabilization is the Marker in the lower left and upper right corners of the video frame. The stabilized video frame is shown in Figure 5.

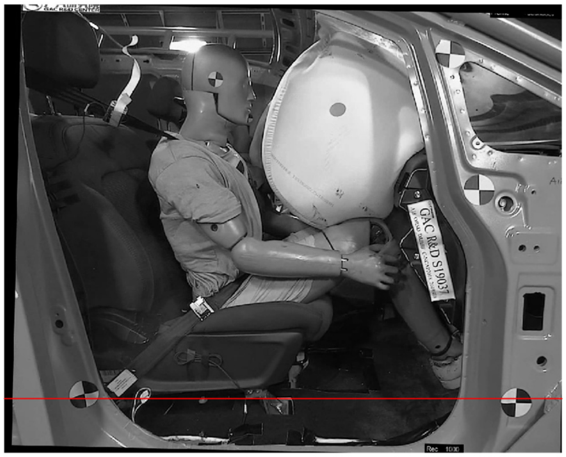


Figure 5 Video of the collision experiment after stabilization image (frame 100)

3.2 Objective evaluation of the stabilization effect of the algorithm

To objectively evaluate the stabilization effect, the Peak Signal to Noise Ratio (*PSNR*) of the current frame and its adjacent previous frame before and after stabilization is usually used as an index to evaluate the accuracy of stabilization [8, 9]. The definition is as follows:

$$PSNR\left(I_K, I_C\right) = 10 \times \lg\left[\frac{255^2}{RMSE\left(I_K, I_C\right)}\right] \tag{5}$$

where I_K and I_C are the previous and next image frames in the image sequence, and the mean square error $RSME$ represents the average difference between each pixel of the two video frames before and after, which can reflect the fluctuation of digital information in the image sequence and is calculated as shown below.

$$RMSE\left(I_K, I_C\right) = \frac{1}{M \times N}\sum_{x=1}^{M}\sum_{y=1}^{N}\left[G_K\left(x, y\right) - G_C\left(x, y\right)\right]^2 \tag{6}$$

where, M and N are the height and width of the video frame, respectively. The smaller the $RSME$ value, the more similar the two video frames are, and the larger the $PSNR$ value, the more stable and less stabilizationed the two video frames are. The values of all the video frames before and after the steady image\ are shown in Figure 6.

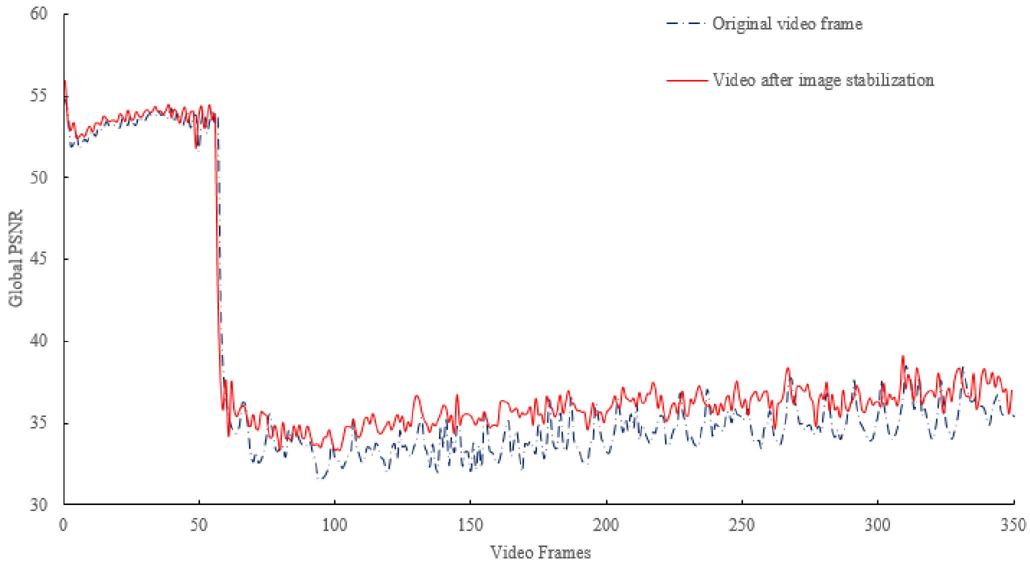


Figure 6 Curve of PSNR values before and after the stabilized image

In Figure 6, the $PSNR$ of the first 50 frames are larger because the first 50 frames of the collision experiment video are the video frames before the collision occurs T0, when the collision has not yet occurred, there is almost no dithering, and the objects in the video are still starting to move, so the $PSNR$ value is larger; after the collision occurs, the overall $PSNR$ value decreases, but the $PSNR$ curve of the original stabilizationed video frames is lower than the curve of the video frames after stabilization, which indicates that the stabilization effect of this paper is good.

Specifically, the average $PSNR$ value of the original stabilizationed video frame sequence is 34.95, and the average $PSNR$ value after stabilization is 36.27, an increase of 1.32 and an increase of 3.78%, which is not obvious after stabilization. This is because the $PSNR$ calculation formula (5) compares the two frames before and after the collision, but the collision dummy, airbag and other objects in each frame will have a large movement, and this part of the movement is not part of the image dithering, so the direct use of the global image $PSNR$ value has its limitations. In this paper, we select a relatively fixed local area in the video and calculate the $PSNR$ value of this local area as the evaluation index. The size of the selected local area is the rectangular area enclosed by the upper left corner (400, 1100) and the lower right corner (950, 1220), as shown in Figure 7.

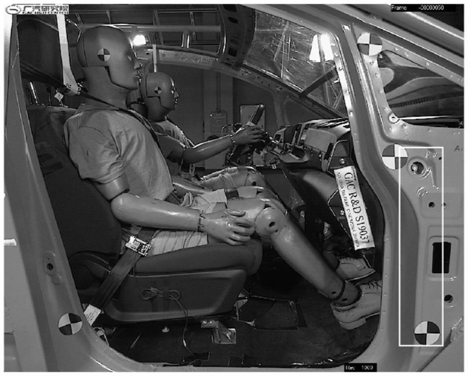


Figure 7 Local area used to calculate PSNR

In the white rectangular area in Figure 7, there is almost no subjective motion of moving objects, only random dithering, so it is easy to evaluate the stabilization effect visually. The calculated $PSNR$ values for all video frames in the local area before and after the steady image are shown in Figure 8.

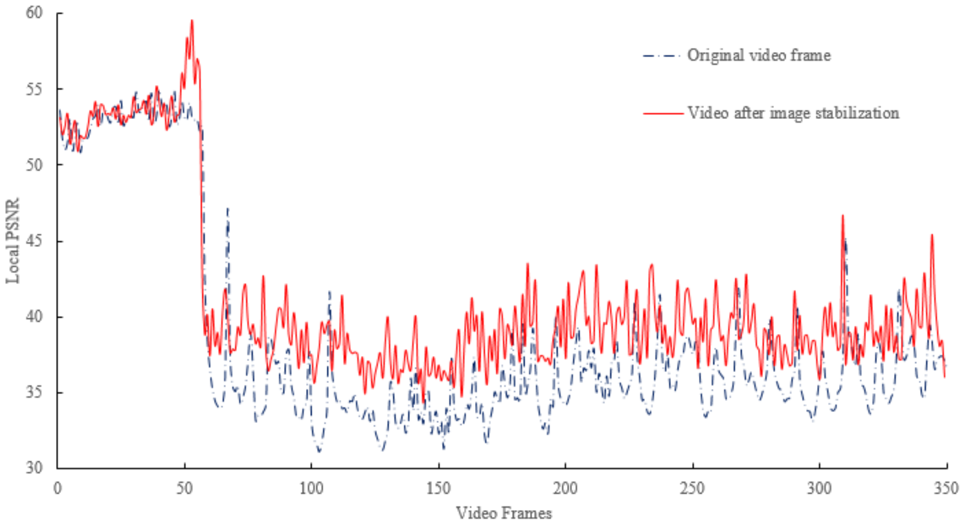


Figure 8 Curve of PSNR values in the local area before and after stabilization

In the above figure, the average $PSNR$ value of the original dithered video frame in the local area is 35.98, and the average $PSNR$ value in the local area after stabilization is 39.15, which increases by 3.17 after stabilization, with an increase rate of 8.81%. The $PSNR$ value increases significantly, and the $PSNR$ value curve in the local area after stabilization is above the value curve in the local area of the original video. It shows that the stabilization algorithm of this paper has good effect on the video stabilization of the collision experiment and can meet the requirements of image measurement in the collision experiment.

4. Conclusion

To address the problem of degradation of feasibility and accuracy of image measurement and analysis due to video shaking in car crash experiments, a block matching video stabilization decision algorithm based on the region of interest (ROI) is proposed in this paper. According to the actual experimental results, the algorithm can effectively suppress the video shaking phenomenon in both subjective and objective evaluations of crash experiments, which can ensure the accuracy of subsequent video detection, automatic tracking, and target judgment as well as visual comfort, thus guaranteeing the feasibility and accuracy of subsequent image measurement and analysis, which is of great significance to the development of automotive safety technology.

Acknowledgement

Grant: Guangdong Provincial Natural Science Foundation (2018A030313727, 2021A1515010912); 2022 Guangdong Province General University Key Research Platform Project "Big Data Technology and Application Innovation Team".

Reference

[1] Wang C, Dai Y, Zhou W, et al. A Vision-Based Video Crash Detection Framework for Mixed Traffic Flow Environment Considering Low-Visibility Condition[J]. Journal of Advanced Transportation, 2020, 2020: 1-11.

[2] Sun Z, Gepner B D, Lee S H, et al. Multidirectional mechanical properties and constitutive modeling of human adipose tissue under dynamic loading[J]. Acta Biomaterialia, 2021, 129: 188-198.

[3] Zou G, He K, He HL, et al. Electronic image stabilization algorithm for jittering video[J]. Journal of Jilin University (Information Science Edition), 2012, 30(5): 487-491.

[4] Xie Zheng, Cui Shaohui. Research on electronic image stabilization technology for imaging matching of flying mobile platforms[J]. Computer Simulation, 2016, 033(2): 138-141.

[5] Yin YAN. Simulation of video surveillance image de-jittering visual monitoring algorithm optimization[J]. Computer Simulation, 2017, 34(9): 450-454.

[6] Zong Yantao, Jiang Xiaoyu, Pei Chuang, et al. Research on electronic image stabilization algorithm based on time series prediction[J]. Journal of Photonics, 2012, 41(2): 244-248.

[7] Wu H, Deng H-B, He S-Y. Video stabilization method for unmanned aerial vehicles based on chunked grayscale projection[J]. Journal of Beijing University of Technology, 2013, 33(4): 385-389.

[8] Wang Chuansheng, Guan Laifu, Tong Lei, et al. A review of video steadicam algorithms[J]. China New Communication, 2019, 21(24): 160-161.

[9] Xue Y, Zhang YF, Yang TY, et al. A motion target detection algorithm for jittering video sequences[J]. Advances in Lasers and Optoelectronics, 2018, 55(9): 326-332.

Proc. of SPIE Vol. 12604 1260401-184

Design of Automobile Virtual Assembly Platform Based on 3D Max

Chuanqi Qin[1], Ruilong Han[1], Lin Fang[1*], Lili Jiao[1]
[1] CATARC Intelligent and Connected Technology Co., LTd., Tianjin, China
*Email: fanglin@catarc.ac.cn

Abstract

To improve the design and development efficiency of automobile project products and ensure product quality, an automobile virtual assembly platform was designed based on 3D Max software. The basic framework and development environment of the platform is built. The automobile virtual model is constructed by using the 3D Max technology. The sequence planning of automobile virtual assembly is realized by using an ant colony distributed algorithm. The functional modules of the automobile assembly platform are analyzed, and the software design of the platform is completed. In order to test the feasibility of the design platform, a simulation experiment is conducted. The results verify that the average running time of the platform's main nodes is 0. 49 s. The function implementation cycle is short, and the platform runs smoothly, with real-time and high efficiency. For the key points of the automobile head assembly, the average deviation of the designed platform assembly's key points is 0. 92 mm, which is within the control range of ±1.0 ~ ±3.0 mm, indicating that the assembly quality and modeling accuracy of the design platform is qualified. For the post-processing of the car head assembly, the rendering completion of the assembly is 100%. The material effect of the car head assembly is realistic. The light and shadow effect is excellent, and the realistic assembly environment and assembly effect are highly restored.

Keywords: 3D Max software, Automobile assembly technology, Automobile virtual assembly technology, Design platform, Virtual technique

1. Introduction

Nowadays, China is in the stage of digital transformation in all aspects of the economy and society. With the wide application of computer networks in people's life, more and more intelligent technologies have been developed and applied in the environment based on computer network [2]. In particular, digital image simulation modeling technology has become a hot research object in the field of computer development [3]. Among them, the development of virtual display technology and 3D Max technology provides intelligent assembly technology for mechanical parts and other related industries, which is to improve the assembly efficiency and quality of mechanical parts [4]. In the automotive industry, the assembly efficiency and precision of automotive components are essential[5]. Previous virtual modeling systems related to automobile assembly have defects such as poor reduction of actual products and low modeling efficiency [6]. Therefore, the main development direction of the automobile industry at the present stage is how to apply relevant technologies based on computers to establish a scheme for precise, realistic, and efficient automobile modeling and assembly operation on the network platform [7]. 3D Max software is a software that applies different plug-ins to realize 3D modeling and 3D animation of products [8]. With the development of science and technology, the software is also gradually developing towards the trend of digitalization to meet the development needs of big data background at this stage [9].

Based on the above background, this paper studies and designs an automobile virtual assembly platform according to the 3D Max software, hoping to provide an efficient and reliable assembly technology of automobile components for the automobile design industry. It also intends to promote the optimized development of virtual assembly technology and 3D Max technology, to improve the design and development efficiency of automobiles and other project products, and to ensure assembly quality, which can increase the economic benefits of the automobile industry.

2. Basic architecture and development environment of the platform

Developing the automobile virtual assembly platform requires the support of virtual reality technology based on the computer network, 3D Max software, 3D graphics texture mapping and other related theoretical basis [10]. Therefore, to meet the requirements of platform's virtual and real modeling function, human-computer interaction function, virtual perception function, and other functions, combined with the technical application required by the platform design, the

International Conference on Computer Graphics, Artificial Intelligence, and Data Processing (ICCAID 2022)
edited by Ruishi Liang, Jing Wang, Proc. of SPIE Vol. 12604, 126040N
© 2023 SPIE · 0277-786X · doi: 10.1117/12.2674518

basic architecture of the automobile virtual assembly platform is built. It is presented in Figure 1.

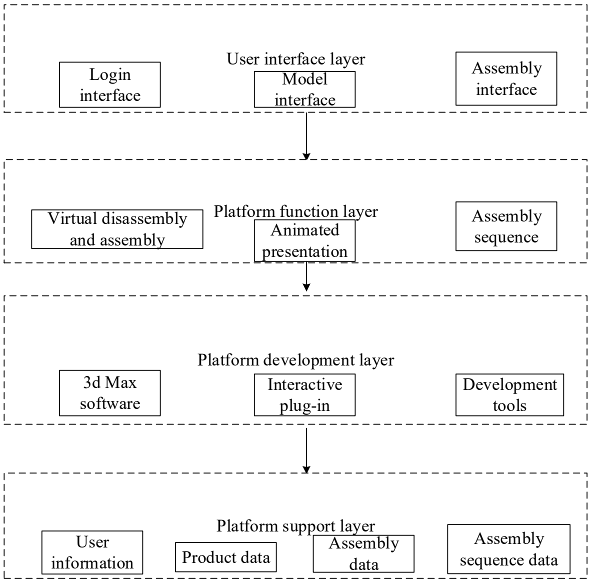


Figure 1 Basic architecture of automobile virtual assembly platform based on 3D Max

According to the basic architecture of the platform shown in Figure 1 above, it can be concluded that the most important part of this design platform is the development layer of the platform. Based on this, the development environment of the automobile virtual assembly platform is designed, as indicated in Table 1 below.

Table 1 Development environment design table of the automobile virtual assembly platform based on 3D Max

No.	Platform development environment	Environment configuration and parameters
1	Operating system (PC side)	Windows7
2	Development platform	Microsoft Visual Studio6.0
3	CPU	64bit
4	Memory	512GB
5	Video card	Gts250
6	Virtual simulation tool	Unity software；VRP software
7	Modeling tools	3d Max software
8	Interactive device	Mouse, touch screen
9	Database	MySQL Server5.7.5
10	Development language	Python+Java+Virtools+JSP

According to the basic structure of the automobile virtual assembly platform mentioned above and the development environment of the platform, combined with the development tools, such as 3dMax, the software, and function of the platform are analyzed and studied. The design of the automobile virtual assembly platform is completed with the support of interactive technology.

3. Software design of the automobile virtual assembly platform

3.1 Vehicle virtual modeling based on 3D max technology

To realize the function of the automobile virtual assembly, 3D Max technology is applied to construct the virtual 3D model of the automobile. The structural data of the automobile and its components are extracted by using the 3D scanning tool. The data is imported into the 3D Max software, and the data processing functions, such as the data editing module and the normal mapping module of the software, are applied to process the extracted structural data of the automobile and its components, which is to obtain accurate automobile modeling data. The modeling data is directly imported into the 3D Max software, and the virtual model of the car is generated by using polygon modeling technology. The model processing function of the 3D Max animation design module, namely V-Ray technology, is adopted to process the material mapping and scene rendering of the car model so as to enhance the authenticity of the virtual car model. The constructed automobile virtual model needs to be saved by the 3ds format. The 3ds format, as the intermediate assembly data of the assembly platform, is to enhance the compatibility of the automobile virtual model in the assembly platform. Depending on the plug-in or code writing in the software, the animation playing and function demonstration of the car model is realized. After completing the preliminary virtual modeling of the automobile, the assembly sequence of the automobile model and the functional design of the assembly platform need to be considered to realize the design and development of an interactive and complete automobile virtual assembly platform.

3.2 Virtual assembly sequence planning based on the ant colony algorithm

Automobile components have the characteristics of large quantity and fine structure, which requires that the assembly sequence of components should be scientific and reasonable in order to ensure the assembly quality of automobile components. Therefore, based on the automobile 3D virtual model, combined with ant colony algorithm, the virtual assembly sequence of automobile components is planned, which provides the core sequence basis for the virtual assembly of automobile components. Ant colony algorithm is a set of robustness, positive feedback, as well as the integration of parallel distributed feature solving algorithm. Its application in the assembly of automotive components can help produce the final time series results. Set the assembly sequence of the automobile as $\eta = \{b_1, b_2, ..., b_r\}$. By default, all automobile assembly components are a set of points, and the connecting sideline of all component points is the relationship of automobile assembly. Based on this, the virtual assembly matrix D of the automobile component is obtained, which is expressed as:

$$D = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1(r-1)} & a_{1r} \\ a_{11} & a_{11} & \cdots & a_{2(r-1)} & a_{2r} \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ a_{(r-1)1} & a_{(r-1)2} & \cdots & \cdots & a_{(r-1)r} \\ a_{r1} & \cdots & \cdots & \cdots & a_{rr} \end{bmatrix} \tag{1}$$

Based on the assembly matrix, each row of nodes in the matrix is selected and combined with each column of nodes. The arrangement and combination are carried out according to the number of automobile components. In order to obtain the assembly sequence of the optimal solution, the concept of ant transition probability is introduced to optimize the path. Calculate the transition probability ω_{pq} of the vehicle component, which is expressed as:

$$\omega_{pq} = \mu e_{pq} + (1-\mu) f_{pq} \tag{2}$$

Where, μ refers to the pheromone coefficient, and $\mu \in (0,1)$; e_{pq} and f_{pq} represent the vehicle component transition probability under global information and the vehicle component transition probability under local information, respectively. When the pheromone coefficient is 0, the assembly path of automobile components is randomized; when the pheromone coefficient is 1, the assembly path planning of automobile components fails. In the process of finding the optimal path, the influence factors of assembly sequence should be taken into account. In the actual vehicle assembly process, the assembly tools, as well as the assembly direction are considered. Based on this, a sequence calculation expression of the automobile virtual assembly is constructed, which is expressed as:

$$\delta(\eta) = \varepsilon s + \varepsilon' s' \tag{3}$$

In the above formula, ε is the number of tool changes of automobile virtual assembly; ε' is the number of direction changes of automobile virtual assembly; s and s' refer to the relative proportions of the two influencing factors respectively. The smaller the $\delta(\eta)$ value, the closer the assembly sequence result is to the optimal solution. The above formula is iteratively calculated, and finally, the optimal sequence planning result of the automobile components' virtual assembly is obtained, which provides a theoretical basis for the design of the automobile virtual assembly platform.

3.3 Function module design of the automobile assembly system

Relating to the automobile virtual model and the automobile virtual assembly sequence, the development and application of the automobile assembly platform are realized by combining the interface presentation module of 3D Max software and the function module of the virtual assembly platform. Interface design is a module based on visualization technology. The car model and its component model generated and stored by 3D Max software are imported into the display structure Virtools of the virtual platform. A group of new projects is established to develop the functional modules of the platform. The functional design of the platform interface is realized by the plug-in of the application software, including the user's login interface, the automobile model interface, the automobile assembly interface, and the disassembly process, disassembly and assembly interface of the automobile components. In the virtual scene of automobile assembly, the collider of the automobile assembly adaptation is added. The interactive actions, such as moving and dragging automobile components, are completed based on Script nodes to ensure the interaction and dynamics between users and assembly actions. A function navigation module of the platform is set to edit the attributes of a two-dimensional frame of a three-dimensional model of the automobile and to describe the three-dimensional model in a picture or text form. It adds mouse plug-ins or touchscreen plug-ins as interactive devices of the platform, so that users can control the platform in real-time. In that case, the virtual assembly and design of automobiles can be realized. The invention provides a virtual technology support for the fields of automobile disassembly and assembly, fault detection, design, and the like.

4. Test and inspection

4.1 Preparation for the test

A simulation experiment is designed in order to test the feasibility of the automobile virtual assembly platform based on 3D Max software. Based on Windows, a test platform is built. The memory of the system is 512GB, and the hard disk memory is 64GB. The CPU selection of the system supports the Python + Java + Virtools multi-language development environment. The MySQL Server database is chosen as the test data source. On the basis of the above test environment, the simulation comparison test of the automobile virtual assembly platform is carried out.

4.2 Test results and analysis

In the running process of the automobile virtual assembly platform, the main running nodes include the automobile virtual structure display module of the platform, the automobile virtual automatic disassembly module, the human-computer interaction module, and the automobile component maintenance module. First, randomly select a group of vehicle model data from the database of the test platform. Then, respectively apply the vehicle virtual assembly platform designed in this paper (test group), the traditional vehicle virtual assembly platform (control group 1), and the CAD-based virtual assembly platform for simulation running. Finally, record the simulation running time of the above main nodes. The results are displayed in Figure 2 below.

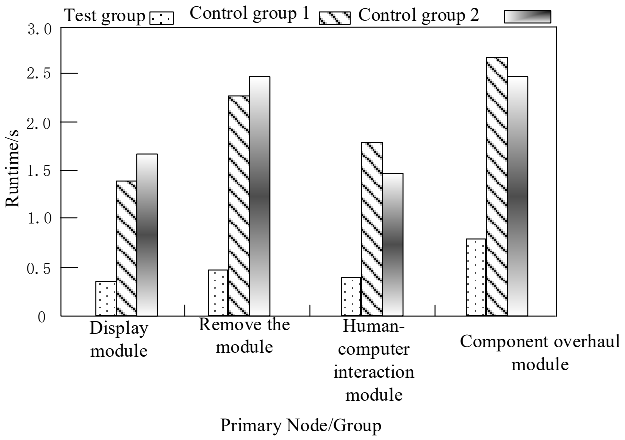


Figure 2 Comparison of main nodes’ running time of different automobile virtual assembly platforms

In Figure 2, the running time of the main nodes of the platform in the test group is lower than that in the control groups 1 and 2. The average running time of all the main nodes in the test group is 0. 49 s, which shows that the functional running cycle of the automobile virtual assembly platform designed in this paper is short, smooth, real-time and efficient. Based on the high efficiency of the design platform, the functional characteristics of the platform are tested. Consistent with the above test, take the assembly part of the vehicle head as an example. The three platforms are adopted to model and assemble the vehicle data, and record the key points of model assembly and the deviation value from the initial set data, as displayed in Figure 3 below.

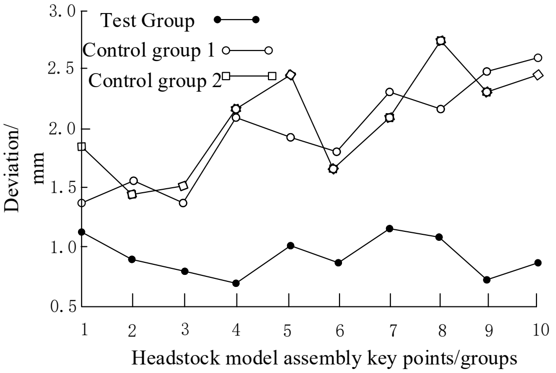


Figure 3 Comparison of key points' deviation values of the scene modeling assembly on different platforms

In Figure 3, for the key points of automobile head assembly, the deviation values of the key points in the test group are lower than those in the control group 1 and control group 2. The average deviation value of all key points is 0. 92 mm, which is within the deviation control range of±1.0-±3.0 mm, indicating that the automobile scene modeling assembly deviation of the design platform in this paper is controlled within a good range. The assembly quality and modeling accuracy are qualified.

After the head assembly is completed, the three platforms are adopted to render and post-process the 3D model of the head. The post-processing effects of the three platforms are recorded for comparative analysis. The results are displayed in Table 2 below.

Table 2 Comparison of post-processing effects of assembly models on different platforms

No.	Test method	Later effect		
		Rendering degree (%)	Assembly effect	Light and shadow effects
1	Test Group	100	Realistic	Great
2	Control group 1	95	Common	Common
3	Control group 2	97	Common	Common

In Table 2, for the post-processing of the automobile head assembly, the rendering completion of the test group is 100%. The material effect of the head assembly is realistic. The light and shadow effect is excellent, and the realistic assembly environment and assembly effect are highly restored. The analysis of the above three groups of experiments verifies that the automobile virtual assembly platform based on the 3D Max software designed in this paper has authenticity and feasibility. It is a kind of automobile virtual assembly product with high precision, which integrates vision and hearing, and lays a firm foundation for the quality and efficiency of the automobile industrial design.

5. Conclusion

Aiming at the present situation of poor reversibility and low efficiency of automobile virtual assembly, the 3D Max modeling software is adopted to strengthen the optimization research of the automobile virtual assembly technology with high precision. It can improve the design and development efficiency of automobile industry and ensure assembly quality. It provides a convenient and diversified assembly platform for automobile developers and designers, promotes the sustainable development of the automobile design industry, and ensures economic benefits.

References

[1] Qu Yunhui, Bai Xinguo. Design and Implementation of Museum Exhibition System Based on 3DMax and Unity3D [J]. Microcomputer Applications, 2022,38(04): 1-3.

[2] Fang Qin Design and Implementation of 3D Modeling of Virtual Laboratory Based on Unity and 3dmax [D]. Beijing University of Posts and Telecommunications, 2015.

[3] Song Hua. Analysis of Automobile Virtual Assembly Based on CATIA [J]. Internal Combustion Engine & Parts, 2020(02): 55-56.

[4] Liu Liyan, Yu Lulu, Chi Yuanyuan. Design of Virtual Assembly System for Vehicle Chassis Steering System Based on Unity3D [J]. System Simulation Technology, 2020,16(02): 131-134.

[5] Song Hua, Yao Hairong. A design-oriented computer-aided automobile assembly process planning system [J]. Automobile Technology, 2020(02): 157-159.

[6] Zhang Guangshu. Virtual Reality Technology and Its Application in Mechanical Design and Manufacturing [J]. Technology Innovation and Application, 2020(18): 151-152.

[7] Ru Xiaoyan. Application of Virtual Reality Technology in Mechanical Design and Manufacturing [J]. South Agricultural Machinery, 2020, 51(09): 136.

[8] Zhang Xin. New generation intelligent vision system guidance technology will help new energy vehicle assembly [J]. New Energy Technology, 2022(07): 35.

[9] Guo Di. Simulation and Analysis of Green Car Assembly Line Layout Based on Intelligent Manufacturing Technology [J]. Logistics Technology, 2022,41(07): 96-99.

[10] Feng Yijiao. Analyze the Application of Three-dimensional Locating System on Automobile Fastener Assembly [J]. Automobile Technology, 2021,46(06): 124-126.

Study on the Application of MF-DMA Model in the Diagnosis of Leaf Nitrogen Nutrient

Jianhui Li[a], Manlan Liu[a], Chengwen Li[a], Han Yang[a], Xuesong Wang*[a]
[a]Foshan Polytechnic, City Foshan, PRC, 528000
* Corresponding author: 694531775@qq.com

Abstract

By analyzing the texture characteristics of rapeseed leaves, this paper proposes a multifractal detrended moving average analysis (MF-DMA) model to identify different nitrogen application levels. The model adopts several global generalized Hurst exponents and some other related multifractal feature parameters of rape leaf images under different λ values, and uses different feature parameter combinations to identify nitrogen nutrition in basal leaves, middle leaves and top leaves, respectively. The results show that the recognition effect is the best when λ=0, and when λ=0, the basal leaves and middle leaves are more sensitive to nitrogen than the top leaves. Through qualitative identification of three parts of rapeseed samples with moderate nitrogen application and deficient nitrogen application, the results showed that the two classification methods of support vector machine and random forest were the best, and the best identification accuracy was 96.01% and 96.83%, respectively. It shows that the model proposed in this paper has good effectiveness.

Keywords: MF-DMA; Leaf texture features; Nitrogen Nutrition Analysis

1. INTRODUCTION

Fractal theory [1-3] has become an effective means to describe the surface texture features of different objects truly and accurately in the field of crop science research. The roughness of the surface texture of most objects in nature is not constant over a wide range of scales, so the single fractal theory cannot explain the complex nature of objects. Therefore, the multifractal method to describe the surface of time objects from different angles and different scales follows. Multifractal is a powerful tool for stationary or non-stationary sequences, and is often used to describe and distinguish many complex graphics in nature. , systems and processes, which provide a means to study the properties of matter in greater depth. In recent years, multifractal processing of images has been widely used. For example, Man H S et al. [4] proposed a new method to extract image features and to evaluate fractality is proposed based on two-dimensional (2D) continuous wavelet transform (CWT). Teknomo et al. [5] developed a micro pedestrian simulation model based on multiracial theory to explain the formula of flow performance or micro pedestrian characteristics. M Anton et al. [6] analyzed digital micro calcification mammography images using a multifractal spectrum-based image segmentation method. However, the above standard multifractal theory is difficult to ensure its accuracy for agricultural images with complex backgrounds and noisy shooting environments. The multifractal detrended fluctuation analysis (MF-DFA) proposed by Kantelhardt [7] can effectively deal with non-stationary objects by removing the local trend and then estimating the singularity index of the object, so they have been widely used in various one-dimensional sequence fields in recent years. In 2006, the two-dimensional (2D) Multifractal Detrended Fluctuation Analysis (MF-DFA) [8] method proposed by Gu and Zhou extended the MF-DFA method from one-dimensional sequences to two-dimensional surfaces, and synthesized the results from fractal Brownian motion's image to verify its effectiveness. Since then, multifractal detrended fluctuation analysis has been used in the study of 2D images [9-12]. In addition, Li, Jian-Hui et al. [13] proposed a new method for image segmentation based on 2D MF-DFA, and the segmentation experiment results were better than the MFS-based method. Detrended Moving Average analysis (DMA) is another method for dealing with non-stationary measures. In 2010, Gu and Zhou promoted a multifractal version of DMA, namely Multifractal Eliminated Average Translation (MF-DMA) [14]. This method can not only accurately estimate the generalized Hurst exponent but also easily describe the multifractal properties of non-stationary series without any assumptions, so it is widely used in time series analysis[15]. Sanz E et al. [16], Wang et al. [17] demonstrated that this method is superior to MF-DFA when used as a feature of two-dimensional grayscale images.

Therefore, this study first used MF-DMA to extract the multifractal features of rapeseed leaves under different nitrogen application levels, and then established an identification and diagnosis model to identify and classify rapeseed leaves at different levels, laying a theory for predicting rapeseed leaves under unknown nitrogen nutrient and practical basis.

International Conference on Computer Graphics, Artificial Intelligence, and Data Processing (ICCAID 2022)
edited by Ruishi Liang, Jing Wang, Proc. of SPIE Vol. 12604, 126040O
© 2023 SPIE · 0277-786X · doi: 10.1117/12.2674723

2. METHOD AND MATERIALS

2.1Multifractal Detrended Moving Average Analysis (MF-DMA)

MF-DMA replaces the polynomial fitting of multi fractal de trend fluctuation analysis method with moving average, which is an effective method to detect whether a non-stationary time measure has multifractality. Because the moving average is simpler than the fitting polynomial and the error is smaller, the MF-DMA operation is faster and more accurate. Assuming a two-dimensional matrix $X(i_1, i_2)$ represents the studied surface, the algorithm can be summarized as follows:

First, calculate the sum $Y(i_1, i_2)$ for a sliding window of size $n_1 \times n_2$, where $n_1 \leq i_1 \leq N_1 - \lfloor (n_1 - 1)\lambda_1 \rfloor, n_2 \leq i_2 \leq N_2 - \lfloor (n_2 - 1)\lambda_2 \rfloor$,$\lambda_1$and λ_2 are position parameters and $\lambda_1, \lambda_2 \in [0,1]$. In particular, it is assumed that the sub matrix of size $n_1 \times n_2$ extracted from X is $\mathrm{Z}(u_1 \times u_2)$, where $i_1 - n_1 + 1 \leq u_1 \leq i_1, i_2 - n_2 + 1 \leq u_2 \leq i_2$, then the sum $Y(i_1, i_2)$ of the sub matrix Z is calculated as follows:

$$Y(i_1, i_2) = \sum_{j_1=1}^{n_1} \sum_{j_2=1}^{n_2} Z(j_1, j_2) \tag{1}$$

Secondly, calculate the moving average function $\tilde{Y}(i_1, i_2)$, where $n_1 \leq i_1 \leq N_1 - \lfloor (n_1 - 1)\lambda_1 \rfloor$, $n_2 \leq i_2 \leq N_2 - \lfloor (n_2 - 1)\lambda_2 \rfloor$. Extract the sub matrix $S(i_1, i_2)$ of size $n_1 \times n_2$ in X, where $K1$ and $K2$ meet $k_1 - \lceil (n_1 - 1)(1 - \lambda_1) \rceil \leq k_1 \leq k_1 + \lfloor (n_1 - 1)\lambda_1 \rfloor$ and $k_2 - \lceil (n_2 - 1)(1 - \lambda_2) \rceil \leq k_2 \leq k_2 + \lfloor (n_2 - 1)\lambda_2 \rfloor$ respectively, and then calculate the sum of $\tilde{S}(i_1, i_2)$.

$$\tilde{\mathrm{S}}(m_1, m_2) = \sum_{d_1=1}^{m_1} \sum_{d_2=1}^{m_2} S(d_1, d_2) \tag{2}$$

where $1 \leq m_1 \leq n_1, 1 \leq m_2 \leq n_2$, the moving average function $\widetilde{\mathrm{Y}}(i_1, i_2)$ is defined as,

$$\widetilde{\mathrm{Y}}(i_1, i_2) = \frac{1}{n_1 n_2} \sum_{m_1}^{n_1} \sum_{m_2}^{n_2} S(m_1, m_2) \tag{3}$$

Third, remove the trend of the matrix $\mathrm{Y}(i_1, i_2)$ through the moving average function $\widetilde{\mathrm{Y}}(i_1, i_2)$ to get the remaining matrix $\mu(i_1, i_2)$,

$$\mu(i_1, i_2) = \mathrm{Y}(i_1, i_2) - \widetilde{\mathrm{Y}}(i_1, i_2) \tag{4}$$

Where $n_1 \leq i_1 \leq N_1 - \lfloor (n_1 - 1)\lambda_1 \rfloor$ and $n_2 \leq i_2 \leq N_2 - \lfloor (n_2 - 1)\lambda_2 \rfloor$.

Fourthly, divide the remaining matrix into $N_{n_1} \times N_{n_2}$ blocks of disjoint sub matrices of size $n_1 \times n_2$, where $N_{n_1} = \lfloor (N_1 - n_1(1 + \lambda_1)/n_1 \rfloor$, $N_{n_2} = \lfloor (N_2 - n_2(1 + \lambda_2)/n_2 \rfloor$, For each sub matrix ϑ_{a_1,a_2}, i.e $\vartheta_{a_1,a_2}(i_1, i_2) = \vartheta(l_1 + i_1, l_2 + i_2)$, $1 \leq i_1 \leq n_{1,} 1 \leq i_2 \leq n_2$, where $l_1 = (a_1 - 1)\, n_1$, $l_2 = (a_2 - 1)\, n_2$, then the de trend fluctuation function of the matrix ϑ_{a_1,a_2} is defined as,

$$\mathrm{F}_{a_1,a_2}^2(n_1, n_2) = \frac{1}{n_1 n_2} \sum_{i_1=1}^{n_1} \sum_{i_2=1}^{n_2} \vartheta_{a_1,a_2}^2(i_1, i_2) \tag{5}$$

Fifth, the q-order global fluctuation function is calculated as follows:

$$F_q(n) = \left\{ \frac{1}{N_{n_1} N_{n_2}} \sum_{a_1=1}^{N_{n_1}} \sum_{a_2=1}^{N_{n_2}} \vartheta_{a_1,a_2}^q(n_1, n_2) \right\}^{1/q} \tag{6}$$

According to Lopida's rule, when q=0, the above equation can be further calculated to obtain the following equation:

$$\ln[F_q(n)] = \frac{1}{N_{n_1} N_{n_2}} \sum_{a_1=1}^{N_{n_1}} \sum_{a_2=1}^{N_{n_2}} \ln[\vartheta_{a_1,a_2}^q(n_1, n_2)] \tag{7}$$

Sixthly, by changing the scale $n_1 \times n_2$, we can get the power rate relationship between the wave function F_q and the scale n, that is, $F_q(m) \sim m^{h(q)}$. $h(q)$ can be obtained by calculating the slope of the fitting straight line of F_q and n in the double logarithmic graph. When $q = 2$, $h(2)$ is the famous Hurst index, so $h(q)$ is called the generalized Hurst index, which describes the roughness of the gray value of a two-dimensional gray surface. According to the standard MFA, the relationship between the quality index $\tau(q)$ and $h(q)$ is as follows:

$$\tau(q) = qh(q) - D \tag{8}$$

D is the geometric topological dimension of the object, $D = 2$ for a 2D grayscale image. So the generalized fractal dimension can be obtained as:

$$D_q = \frac{qH(q)-2}{q-1} \tag{9}$$

In addition, the Hurst index α and multifractal singular spectrum $f(\alpha)$describing the two-dimensional gray surface roughness can also be obtained through the quality index $\tau(q)$, which can be obtained from the Legendre transformation of $\tau(q)$. Combining Formula (9), we can get the relationship between $\alpha, h(q)$ and $f(\alpha)$, which is described as follows:

$$\alpha(q) = h(q) + qh'(q), f(\alpha) = q[\alpha - H(q)] + 2 \tag{10}$$

$\alpha(q)$ is the local roughness index of the two-dimensional grayscale image, and the multifractal spectrum f(α) describes the global roughness of the image. By changing the value of q from q_{min} to q_{max}, the spans $\Delta\alpha$ and Δf of $f(\alpha)$ and $\alpha(q)$ can be obtained:

$$\Delta\alpha = \alpha_{max} - \alpha_{min}, \Delta f = f(\alpha_{max}) - f(\alpha_{min}) \tag{11}$$

Among them, α_{max} and α_{min} represent the maximum probability measure and the minimum probability measure respectively. The larger $\Delta\alpha$ is, the more uneven the distribution of the gray value of the two-dimensional gray image is, that is, $\Delta\alpha = 0$ corresponds to a perfectly uniform distribution. Obviously, the fractal spectrum with a certain width can reflect the characteristics of the non-uniform fractal structure. As the degree of non-uniformity of the fractal structure of the grayscale image increases, it contains more sub-fractals with different singular intensities interacting, resulting in that when the grayscale amplitude of the image is relatively large, the span of the corresponding spectrum is also larger. When the change value is small, the span of the spectrum is relatively small. Δf is often used to describe the degree of confusion in the measurement. The larger the absolute value, the larger the fluctuation of the gray value of the two-dimensional surface, and the more chaotic it looks visually.

2.2 Classifier modeling

Another key issue in constructing a multi-fractal feature-based diagnostic model for nitrogen nutrition in rapeseed leaves is the selection of classification methods. In order to examine the effectiveness of different classifiers on model recognition, this study uses the following six classification methods as classifier modeling, namely Fisher's Linear Discriminant (LDA) [18], Extreme Learning Machine (ELM) [19], Support Vector Machine [20], BP-NN [21] , Random Decision Forests [22], and the K-Nearest Neighbor algorithm [23]. These six classification methods each have their own advantages, and they are several methods commonly used in pattern recognition and processing of big data. Next, the six classification methods are represented by Fisher's LDA, ELM, SVMKM, BP-NN, RF, and KNN, respectively.

3. RESULT AND DISCUSSION

As we know, any color can be represented by three primary colors: red (R), green (G) and blue (B). First, the RGB space of color image is converted into gray image according to formula (12). Each image is regarded as a three-dimensional surface. The first two-dimensional coordinate (x, y) represents the pixel position, and the third three-dimensional coordinate Z represents the corresponding gray value of the pixel point. Then, the MF-DMA method introduced in section 1.1 is used to extract the multifractal feature parameters of all images. In MF-DMA, three special cases are considered when the position parameter λ is 0 (backward sliding average), 0.5 (center sliding average) and 1 (forward sliding average) respectively.

$$Y = 0.2989\mathrm{R} + 0.5866\mathrm{G} + 0.1145\mathrm{B} \tag{12}$$

Y is the two-dimensional sequence $X(i_1, i_2)$ in Section 2.1. In order to explore the relationship between nitrogen application level and the multifractal characteristics of rape leaves, the multifractal parameters obtained from MF-DMA were used as the texture characteristics of rape leaves for nitrogen nutrition diagnosis. 267 rape leaf samples were selected from the middle and late flowering stages, including 159 nitrogen deficient samples. One leaf at the base, middle and top of each rape plant (respectively represented by B, M and T), each image size is 1024×768. The 17 multifractal characteristic parameters of the base, middle and top leaves of 267 rape samples were calculated respectively. Among them, 17 multifractal characteristic parameters represented 11 generalized Hurst indexes, namely, *h(-4), h(-3), h(-2), h(-1), h(0), h(1), h(2), h(3), h(4) and h(5*), and 6 related multifractal parameters, $\Delta\alpha, \Delta f, \alpha_{max}, \alpha_{min}, \Delta h$ and $\Delta\tau$. The random forest and other methods introduced in Section 2.2 were used to identify and diagnose rape nitrogen deficiency and moderate samples under the *K*-fold [24] cross validation method. In the *K*-fold cross validation method, any *1/K* sample is regarded as a test sample, and *(K-1)/K* sample is regarded as a training sample, which is repeated 10 times in the calculation process to eliminate the influence of random factors. Three different position parameters λ The mixed leaf samples from the base, middle, top and three parts of lower rape were classified and identified. Under the 10-fold cross test method, the maximum average recognition accuracy is taken as the target to screen and identify the best three-dimensional characteristic parameter combination of rape leaves at nitrogen application level in each part, and the maximum recognition rate is shown in Table 1-3. It is worth noting that in theory, more feature dimensions will improve the recognition rate. However, through experiments, when the feature dimension is 4 or more, the recognition rate is not significantly improved. Therefore, considering the time complexity, three-dimensional features are selected as the best feature combination.

Table 1 The average diagnostic rate of the two nitrogen application levels of rape leaves in each part when x is 0, 0.5 and 1, respectively under LDA, ELM and SVMKM.

	LDA			ELM			SVMKM		
	λ=0	0.5	1	λ=0	0.5	1	λ=0	0.5	1
Sample B	90.11%	78.65%	75.73%	89.01%	80.43%	80.20%	95.16%	95.73%	95.51%
Sample M	87.87%	82.02%	79.33%	87.51%	86.02%	79.66%	94.61%	97.30%	96.85%
Sample T	78.20%	76.18%	75.73%	82.43%	81.12%	76.02%	93.58%	95.28%	95.96%
Mixed	78.28%	77.23%	72.21%	81.65%	79.80%	73.46%	93.26%	93.48%	96.01%

Table 2 The average diagnostic rate of the two nitrogen application levels of rape leaves in each part when x is 0, 0.5 and 1, respectively under BPNN, RF and KNN.

	BPNN			RF			KNN		
	λ=0	0.5	1	λ=0	0.5	1	λ=0	0.5	1
Sample B	83.42%	74.22%	73.51%	97.75%	95.78%	95.84%	79.33%	80.00%	75.73%
Sample M	77.44%	78.04%	69.75%	97.53%	97.10%	95.75%	84.94%	75.28%	76.85%
Sample T	76.99%	69.37%	66.49%	95.06%	93.87%	94.45%	76.40%	71.69%	72.58%
Mixed	77.53%	74.11%	70.31%	96.83%	96.44%	96.53%	79.25%	79.40%	75.06%

From the above table, it can be concluded that the recognition effect when λ=0 is slightly better than λ=0.5 and λ=1, indicating that the generalized Hurst index obtained by MF-DMA with backward moving average can best reflect different nitrogen application levels. Differences in the leaves of rapeseed. In addition, for the same rape plant, different three parts of rape leaves have different recognition effects on different nitrogen application levels. From the results in the table, it can be seen that the average recognition accuracy of the basal and middle leaves of rapeseed is significantly better than that of the top leaves, indicating that the basal and middle leaves of rapeseed are more sensitive to nitrogen deficiency. Among the 6 classification methods, the average recognition accuracy of the leaves mixed in the three parts is the best with support vector machine and random forest, and the average recognition accuracy is 96.01% and 96.83% respectively, indicating that the model has good performance effectiveness.

Table 3 Nitrogen nutrition analysis results of three parts of rapeseed leaves mixed when λ=0 (K=10)

Classification	Nitrogen level	Medium	Lack	Accuracy
LDA	Medium	87	21	80.56%
	Lack	37	122	70.73%
KNN	Medium	73	35	67.59%
	Lack	20	139	87.42%
ELM	Medium	77	31	71.30%
	Lack	18	141	88.68%
BPNN	Medium	71	37	65.74%
	Lack	23	136	85.53%
RF	Medium	103	5	95.37%
	Lack	4	155	97.58%
SVMKM	Medium	99	9	91.67%
	Lack	9	150	94.34%

The diagnostic and recognition accuracy rates of the leaves of each part of the same rapeseed leaves for the two nitrogen application levels under three different position parameters λ were investigated, and it was found that the best effect was when the position parameter λ=0, and now the rape leaves are considered when λ=0. The diagnostic accuracy when the three parts are mixed together, the diagnostic results are shown in Table 5. Under 10-fold cross-validation, the identification accuracy of random forest method for mixed samples with moderate and deficient nitrogen in rape leaves was 95.37% and 97.48%, respectively.

Finally, the random forest method was used to investigate the effect of K value changes on the mixed leaf samples of rape under two nitrogen application levels. The results show that the average recognition accuracy increases with the increase of the value of K, and the growth rate is first fast and then slow. When the position parameter λ=0, the recognition effect is better than the position parameter λ=0.5 and λ=1.

4. CONCLUSIONS

In this study, MF-DMA was used to obtain multifractal parameters under different λ, and nitrogen nutrition diagnosis modeling was carried out for rape leaves with moderate and deficient nitrogen. The results showed that the diagnosis effect was best at λ=0, and the diagnosis effect of base leaves and middle leaves was better than that of top leaves at λ=0. Then qualitative diagnosis was carried out for the rape samples mixed in three parts by using six classification methods for nitrogen application levels of appropriate and deficient. The results showed that the two classification methods of support vector machine and random forest were the best. The recognition accuracy rates are 96.01% and 96.83%, respectively. It shows that the model has better recognition effect.

This paper mainly studies the multifractal information of rape leaves, and uses it to diagnose nitrogen nutrition. The multifractal information not only considers the mutation color information in the leaf image, but also reflects the singular texture information, so it can better reflect the impact of missing nutrients on the leaves. Although the number of rape leaf samples used is large, the image processing of the collected samples is simple without complex pre-processing, and the extracted multifractal feature parameters not only provide a theoretical basis and implementation method for the diagnosis and classification of rape leaves under different nitrogen levels, but also can be applied to the identification of varieties and diagnosis of nutrient deficiency of different crops. The algorithm is simple, highly intelligent and has good universality, so it has a broad application prospect.

ACKNOWLEDGEMENTS

The authors wish to thank the anonymous reviewers and the handling editor, which led to a great improvement to the presentation of this work. The research work is supported by High-level Talent Research Project of Foshan Polytechnic, China (Grant no.KY2020G01), and Special Projects in Key Areas of Colleges and Universities in Guangdong Province (Grant no.2020ZDZX3104), Natural Science Characteristic Innovation Project of Ordinary Colleges and Universities in Guangdong Province (Grant no. 2020KTSC308). The authors sincerely present thanks to them all.

REFERENCES

[1] Higuchi T. Approach to an irregular time series on the basis of the fractal theory[J]. Physica D Nonlinear Phenomena, 1988, 31(2).

[2] Jacquin A E. Image coding based on a fractal theory of iterated contractive image transformations[J]. IEEE Transactions on Image Processing, 1990, 1(1):18-30.

[3] Jacquin A E. Fractal image coding: A review[J]. Proceedings of the IEEE, 1993, 81(10):1451-1465.

[4] Man H S, Kim C, Jin G H, et al. Image feature extraction and fractality evaluation based on two-dimensional continuous wavelet transform: application to digital elevation model data[J]. Fractals, 2022.

[5] Teknomo K. Microscopic Pedestrian Flow Characteristics: Development of an Image Processing Data Collection and Simulation Model[J]. arXiv e-prints, 2016.

[6] Anton M, Reginatto M, Elster C, et al. The regression detectability index RDI for mammography images of breast phantoms with calcification-like objects and anatomical background[J]. Physics in Medicine and Biology, 2021, 66(22):225015-.

[7] Kantelhardt, Jan W, Zschiegner, et al. Multifractal detrended fluctuation analysis of nonstationary time series. [J]. Physica A, 2002.

[8] Xi C, Zhang S, Xiong G, et al. A comparative study of two-dimensional multifractal detrended fluctuation analysis and two-dimensional multifractal detrended moving average algorithm to estimate the multifractal spectrum[J]. Physica A Statistical Mechanics & Its Applications, 2016:34-50.

[9] Wang J, Yan Y, Kim J. Classification of melanoma images using 2D multifractal detrended cross-correlation analysis[J]. Modern Physics Letters B, 2022.

[10] Wang J, Shao W, Kim J. Combining MF-DFA and LSSVM for retina images classification[J]. Biomedical Signal Processing and Control, 2020, 60:101943.

[11] Nag S. Fractal Arts: a 2D-Mfdfa Approach. 2017.

[12] Fang W, Fan Q, Stanley H E. Multiscale multifractal detrended-fluctuation analysis of two-dimensional surfaces[J]. PHYSICAL REVIEW E, 2016, 93(4):042213.

[13] Li J H, Wang F, Li J W, et al. Multifractal methods for rapeseed nitrogen nutrition qualitative diagnosis modeling[J]. International Journal of Biomathematics, 2016, 9(04):1650064.

[14] Gao-Feng, Gu, et al. Detrending moving average algorithm for multifractals[J]. Physical Review E, 2010.

[15] Wang F, Wang L, Zou R B. Multifractal detrended moving average analysis for texture representation[J]. Chaos, 2014, 24(3):661-674.

[16] Sanz E, Saa-Requejo A , CH Díaz-Ambrona, et al. Generalized Structure Functions and Multifractal Detrended Fluctuation Analysis Applied to Vegetation Index Time Series: An Arid Rangeland Study[J]. Entropy, 2021, 23(5):576.

[17] LIU, Guilin, CHEN, et al. Wave height statistical characteristic analysis[J]. Journal of Oceanology & Limnology, 2019.

[18] Cooke T. Two variations on Fisher's linear discriminant for pattern recognition[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2002, 24(2):268-273.

[19] Huang G B, Zhu Q Y, Siew C K. Extreme learning machine: Theory and applications[J]. Neurocomputing, 2006, 70(1/3):489-501.

[20] Saunders C, Stitson M O, Weston J, et al. Support Vector Machine[J]. Computer ence, 2002, 1(4):1-28.

[21] Shi Z L, Kang J, Sun R. BP NN-based method for lens distortion correction of large-field imaging[J]. Guangxue Jingmi Gongcheng/Optics and Precision Engineering, 2005, 13(3):348-353.

[22] Rahmani H, Mahmood A, Du Q H, et al. Real time action recognition using histograms of depth gradients and random decision forests[C]// Applications of Computer Vision. IEEE, 2014.

[23] Zhu M, Chen W, Hirdes J P , et al. The K-nearest neighbor algorithm predicted rehabilitation potential better than current Clinical Assessment Protocol[J]. Journal of Clinical Epidemiology, 2007, 60(10):1015-1021.

[24] Rubin D B, Van D. A Note on K-fold Least Squares Cross-Validation in Density Estimation[J]. Bepress, 2006.

Artificial Intelligence-Based Traditional Chinese Medicine Assistive Diagnostic System: Analysis of pulse parameters in patients with systemic lupus erythematosus and the factors influencing their Chinese medical evidence patterns

Chengdan Pan[1a], Ai-Min Gong*[1], Fang-Zhi Wei[1], Xuan Zhang[1], Yitian Song[1]
[1] Comprehensive Laboratory of Traditional Chinese Medicine Diagnostics, College of Traditional Chinese Medicine, Hainan Medical College, Haikou, Hainan571799China;
[a]Panchengdan2405@163.com
[*] Corresponding author: 422789075@qq.com

ABSTRACT

Objective: To analyze the characteristics of pulse pattern parameters of systemic lupus erythematosus (SLE) patients and the factors influencing their TCM patterns, and to provide an objective basis for their TCM pulse diagnosis and diagnosis. Methods: The SmartTCM-A1 TCM intelligent detection system was used to collect pulse diagnosis image information from 267 SLE patients (140 in the Yin deficiency internal heat evidence group, 96 in the heat toxin incandescence evidence group, and 31 in the rheumatic heat paralysis evidence group) and 130 healthy individuals, extract the image parameters of pulse diagnosis, and analyze the characteristics of pulse diagnosis parameters of SLE patients and their influencing factors of pulse diagnosis parameters of different TCM evidence types by logistic regression model. Results: ① Compared with the healthy group, the pulse parameters h1 value, h4 value, and t1/t value were significantly higher in the SLE group ($P<0.05$) and the pulse parameter t value was significantly lower in the SLE group ($P<0.05$); logistic regression analysis showed that the independent influencing factors in SLE patients included: h4 value (OR=1.073; 95% CI=1.003-1.148 ;$P<0.05$), t-value (OR=0.003;95% CI=0.000-0.763;$P<0.05$) and t1/t-value (OR=0.000;95% CI=0.000-1.199;$P<0.05$). ② Compared with the group with incandescent heat toxin evidence, the h4 value, h5 value, t4 value, t5 value, t5/t4 value were significantly lower in the group with Yin deficiency internal heat evidence ($P<0.05$), the h4/h1 value and t1/t4 value were significantly higher in the group with Yin deficiency internal heat evidence ($P<0.05$), the h4 value, h5 value, t5 value, h4/h1 value were significantly lower in the group with rheumatic heat paralysis evidence ($P<0.05$), and the t4 value, h5/h1 value, t5/t1 value were significantly lower in the group with rheumatic heat paralysis evidence ($P<0.05$). , h5/h1 values, and t1/t4 values were significantly higher in the rheumatism-heat paralysis group ($P<0.05$); logistic regression analysis. The results showed that the independent influencing factors of TCM evidence in SLE patients included: h3 values in the yin deficiency internal heat evidence group (OR=2.295; 95% CI=1.843-2.858; $P<0.05$) and h3 values in the rheumatic heat paralysis evidence group (OR=2.309; 95% CI=1.87-2.85; $P<0.05$). Conclusion:. Conclusion Pulse diagnosis parameters h4 value, t value and t1/t value are one of the influencing factors for the diagnosis of SLE patients, and pulse diagnosis parameter h3 value is one of the influencing factors for the diagnosis of SLE Chinese medical evidence.

Keywords: Systemic Lupus Erythematosus, Chinese medicine, pulse diagnosis, parameters, artificial intelligence

1. INTRODUCTION

SLE is an autoimmune disease characterized by multiple organ involvement and multiple autoantibody positivity, mostly in women of reproductive age, and the prevalence of SLE varies widely by region, ranging from 0 to 241 per 100,000 worldwide [1], and from 30 to 70 per 100,000 in mainland China [2-3]. Currently, Western medicine treats SLE mainly with hormones and immunosuppressive drugs, and the long-term use of hormones and immunosuppressive drugs brings a lot of side effects and economic pressure to patients, while Chinese medicine treatment of SLE can effectively alleviate the toxic side effects, so the research of Chinese medicine treatment of SLE has received wide attention from the medical community. Pulse diagnosis is an important basis for judging the physical state and diagnosing the condition in TCM clinics, as stated in "Ling Shu - Meridian Pulse": "The meridian pulse is so capable of deciding life and death and dealing with all diseases." This illustrates the importance of pulse diagnosis in diagnosing diseases in Chinese medicine. Pulse diagnosis is characterized by "the pulse is delicate and its body is difficult to distinguish" and is one of the most difficult

International Conference on Computer Graphics, Artificial Intelligence, and Data Processing (ICCAID 2022)
edited by Ruishi Liang, Jing Wang, Proc. of SPIE Vol. 12604, 126040P
© 2023 SPIE · 0277-786X · doi: 10.1117/12.2674642

dialectical tools to master among the four diagnoses in traditional Chinese medicine, which is influenced by the subjectivity and clinical experience of Chinese medicine practitioners and lacks objective and unified diagnostic criteria[4-6]. The authors have not reported in the literature on the use of pulse diagnostic parameters in SLE diseases.Therefore, in this study, the SmartTCM-A1 TCM intelligent detection system was used to collect pulse diagnosis information from SLE patients and healthy individuals, and to statistically analyze the pulse diagnosis information of SLE patients with different TCM evidence types to objectively assess the changes of pulse diagnosis image parameters in SLE patients with different TCM evidence types and to provide an objective basis for TCM evidence-based treatment of SLE.

2. MATERIALS AND METHODS

2.1 General Information

The cases in this study were mainly from SLE patients attending the outpatient and inpatient departments of the Department of Traditional Chinese Medicine and the Department of Rheumatology and Immunology at the First Affiliated Hospital of Hainan Medical College and Hainan Hospital of Hainan Medical College from January 2020 to 07, 2022.The healthy individuals were mainly derived from college students and faculty staff of Hainan Medical College without other diseases. A total of 267 cases of SLE patients were included in this study, including 23 males and 244 females with a mean age of (35.88±14.40) years. Among them, there were 31 cases of rheumatic heat paralysis, 4 males and 27 females with a mean age of (40.61±15.24) years; 96 cases of incandescent heat toxicity, 10 males and 86 females with a mean age of (37.27±14.56) years; and 140 cases of internal heat deficiency, 9 males and 131 females with a mean age of (33.88±13.83) years. One hundred and thirty healthy patients were included, 65 males and 77 females, with a mean age of (24.48±13.97) years.

2.2 Diagnostic criteria

(1) Western medical diagnostic criteria.

Refer to the 2019 revised diagnostic criteria for SLE by the European League Against Rheumatism and the American Rheumatism Association [7].

(2) TCM Diagnostic Criteria.

Refer to the diagnostic criteria in the 2002 Guidelines for Clinical Research on New Chinese Medicines [8]. Yin deficiency internal heat evidence Symptoms: Dark red rash with persistent low or irregular fever, irritable fever in the five hearts, spontaneous sweating and night sweating, puffy red face, joint pain, heel pain, red tongue, thin coating, and fine pulse. Incandescence of heat toxins Symptoms: butterfly-shaped erythema on the face with bright color and purple skin spots, with high fever, irritability and thirst, delirium, convulsions, joint and muscle pain, dry stools, short and red urine, red and vivid tongue, yellow and greasy coating, and flooding or fine pulse. Wind-damp heat paralysis evidence Symptoms: swelling of both hands and fingers, pain in the joints of the extremities, or swelling, or indefinite pain, rash around the body, muscle pain, fever, wind, stiffness of joints, red tongue with yellow coating, smooth or fine pulse. Each case was identified by two to three TCM experts with associate or higher titles with reference to the SLE identification criteria, and the final identification results were determined by two experts who agreed on the diagnosis of the symptom typology.

(3) Inclusion criteria.

Those who met the Western medical diagnostic criteria for SLE; those who met the Chinese medical diagnostic criteria for SLE; those who gave informed consent and signed the informed consent form; and those aged 18-80 years.

(4) Exclusion criteria.

Those with combined psychiatric or neurological disorders; pregnant or lactating women; or those who do not want to cooperate with this test; incomplete data.

2.3 Machine equipment

(1) SmartTCM-A1 type Chinese medicine intelligent detection system.

SmartTCM-A1 Chinese medicine intelligent four-diagnosis detection system provided by Shanghai University of Traditional Chinese Medicine was used .Hardware indexes of the equipment: tongue image acquisition pixels: ≥ 5 million; imaging unit size > 22.5 x 15.0 mm; light source color temperature: 5500k+200k, using DC LED light source, no strobe, life > 50000 hours; with closed dark box, acquisition of tongue images in an airtight environment; the equipment provides physical shooting button, which can be taken by the subject independently; can automatically divide the pulse cycle, number of sensor points. 96 points; pressurization mode: pneumatic automatic pressurization; pulse sensor range: 0~300g; pulse sensor sampling frequency: 500hz; sensitivity: 3.89mV/g; excitation voltage: 12Vdc. (Figure 1)

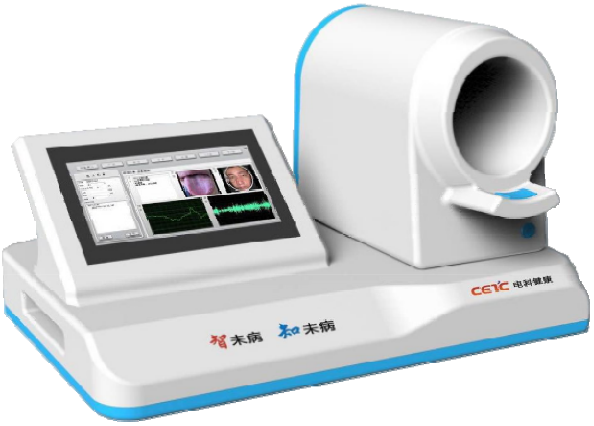


Figure 1: SmartTCM-A1 type Chinese medicine intelligent detection system.

(2)iTCM-I Chinese Medicine Pulse Analyzer.

The iTCM-I TCM pulse analyzer is part of the SmartTCM-A1 TCM intelligent detection system, which is a modern electronic technology, computer technology and pattern recognition technology used in TCM pulse research, which can digitally collect and automatically analyze the human pulse, and come up with objective pulse indicators and diagnostic conclusions in line with TCM symptom theory. The iTCM-I TCM pulse analyzer consists of a pulse sensor and acquisition and analysis software: the pulse sensor is attached to the arm via a wrist strap and connected to the PC via a USB Type-c cable; the dedicated acquisition and analysis software runs on the PC and works with the hardware to achieve pulse acquisition, analysis, and data management functions.(Figure 2)

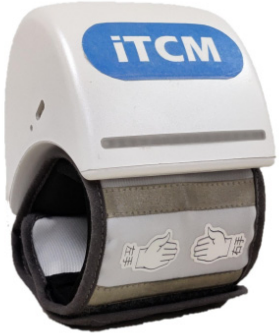


Figure 2: iTCM-I Chinese Medicine Pulse Analyzer.

2.4 Pulse diagnosis parameter acquisition

Observe the subject in a sitting or supine position. The forearm is naturally spread forward and placed at the same level as the heart, the wrist is straight, palm up, fingers slightly bent, and a loose pulse pillow is placed under the wrist joint so that the local qi and blood flow is smooth and easy to diagnose the pulse. Tie the sensor to the pulse area, adjust the pressure through the knob, observe and operate in the software interface to achieve the acquisition, collect the left and right hand off pulse map, take a number of pressure segments (need to include the best pulse pressure pulse map), each pressure segment acquisition time of 10 s, select the best pressure pulse map for analysis. Real-time pulse acquisition display: dynamic pulse waveform display fully automatic acquisition of 96 points of pressure array data .(Figure 3 and Figure 4)

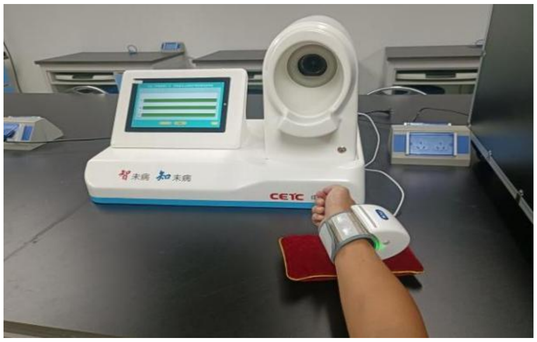


Figure 3: Pulse diagnosis image acquisition using TCM intelligent detection system.

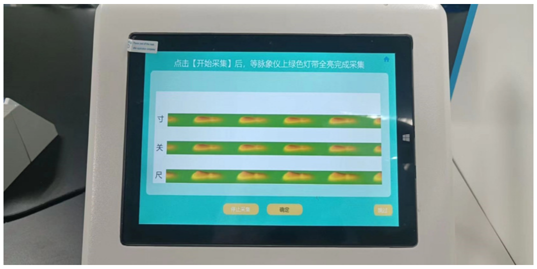

点击【开始采集】后，等脉象仪上绿色灯带全亮完成采集

Figure 4: Optimal pulse taking pressure pulse chart.

2.5 Analysis of pulse diagnosis parameters

The significance of the parameters is as follows [9]:h1: Main wave amplitude, i.e., the vertical distance between the peak of the main wave and the baseline of the PWG. It mainly reflects left ventricular ejection function and large artery compliance.h3, h3/h1: is the pre-repulse wave amplitude, i.e. the vertical distance between the peak of the pre-repulse wave and the baseline of the PWG and its ratio to the main wave amplitude, which mainly reflects the arterial vascular elasticity and peripheral resistance status.h4, h4/h1: The amplitude of the descending isthmus, i.e. the vertical distance between the bottom of the descending canyon and the baseline of the PWG and its ratio to the amplitude of the main wave, mainly reflecting the peripheral resistance of the arterial vasculature and corresponding to the diastolic pressure. h5, h5/h1: The amplitude of the repulse wave, i.e. the vertical distance between the peak of the repulse wave and the bottom of the descending canyon, mainly reflecting the vascular compliance and aortic valve function.h5, h5/h1: is the amplitude of the repulse wave, i.e., the vertical distance between the peak of the repulse wave and the bottom of the descending canyon.t1: The time value between the onset of the pulse and the peak of the main wave, which corresponds to the rapid ejection phase of the left ventricle and mainly reflects the compliance of large vessels and the tension of small vessels. t4: The time value between the onset of the pulse and the descending mid-isthmus, which corresponds to the systolic phase of the heart and reflects the systolic function of the heart. t5: The time value between the descending mid-isthmus and the end of the pulse, which corresponds to the diastolic phase of the heart and mainly reflects the diastolic function of the heart.t: is the pulsation period, i.e. the time value from the beginning to the end of the pulsogram, corresponding to one cardiac cycle in the left ventricle. w: the wave width at 1/3 of the height of the main wave, i.e. the duration of the high level state of intra-arterial pressure. It is related to the appearance of h3 and peripheral resistance. w/t: The ratio of the peak width value in the upper 1/3 of the main wave height to the pulsation cycle, i.e. the proportion of the duration of the high intra-arterial pressure state in the pulsation cycle. It mainly reflects the wall elasticity and the magnitude of peripheral resistance.t1/t: Ratio of the rapid ejection phase of the left ventricle to the whole cardiac systolic cycle. t4/t5: Ratio of the systolic phase of the left ventricle to the whole cardiac systolic cycle. The relative ratios of each amplitude characteristic parameter can better reflect the pulse map characteristics. For example, h4/h1, h3/h1, w/t can be used together with h4 and h3 to reflect the peripheral resistance of the arterial vasculature and arterial vascular elasticity. (Figure 5)

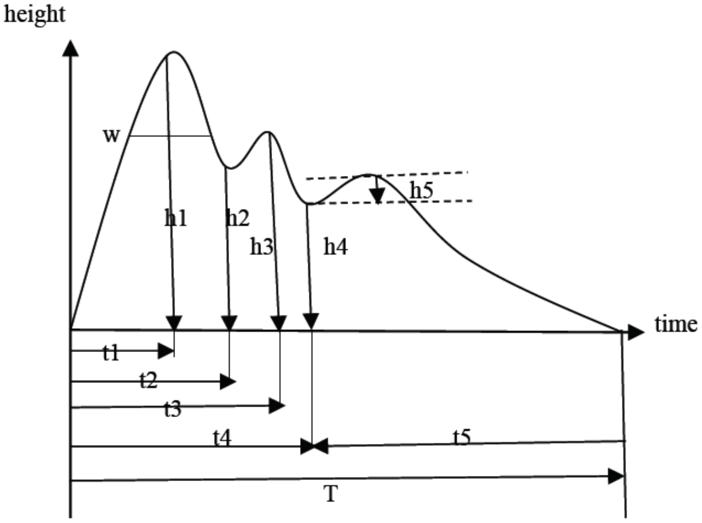


Figure 5: Amplitude and temporal values of pulse maps.

2.6 Statistical methods

IBM SPSS26.0 statistical software was used for statistical analysis, conforming to normal distribution, described by ($\bar{x}$±s), and independent sample t-test or one-way ANOVA was used for comparison between groups; not conforming to normal distribution, described by quartile M (Q25, Q75), and non-parametric MannWhitney U test was used for comparison between groups. p<0.05 indicated that the differences were statistically significant. Binary logistic regression was used to analyze the factors influencing pulse diagnosis in patients with SLE, and multivariate logistic regression was used to analyze the factors influencing pulse diagnosis in patients with SLE TCM evidence. The difference was considered statistically significant at P<0.05.

2.7 Research Process

Study preparation phase: literature review, clinical flow survey, and preliminary questionnaire development. Study refinement phase: small sample validation, establishment of the SLE clinical information questionnaire through expert consultation, discussion, and refinement. Protocol implementation phase: The basic information, four diagnostic information and pulse diagnosis objective parameters of SLE patients were collected simultaneously using the SLE Clinical Information Checklist and the SmartTCM-A1 Chinese Medicine Intelligent Detection System, and subjects with SLE who met the criteria were included in the study. The basic pulse information, four diagnostic information and pulse diagnosis objectification parameter information of healthy individuals were collected and included in the study using the SmartTCM-A1 Chinese medicine intelligent testing system.(Figure 6)

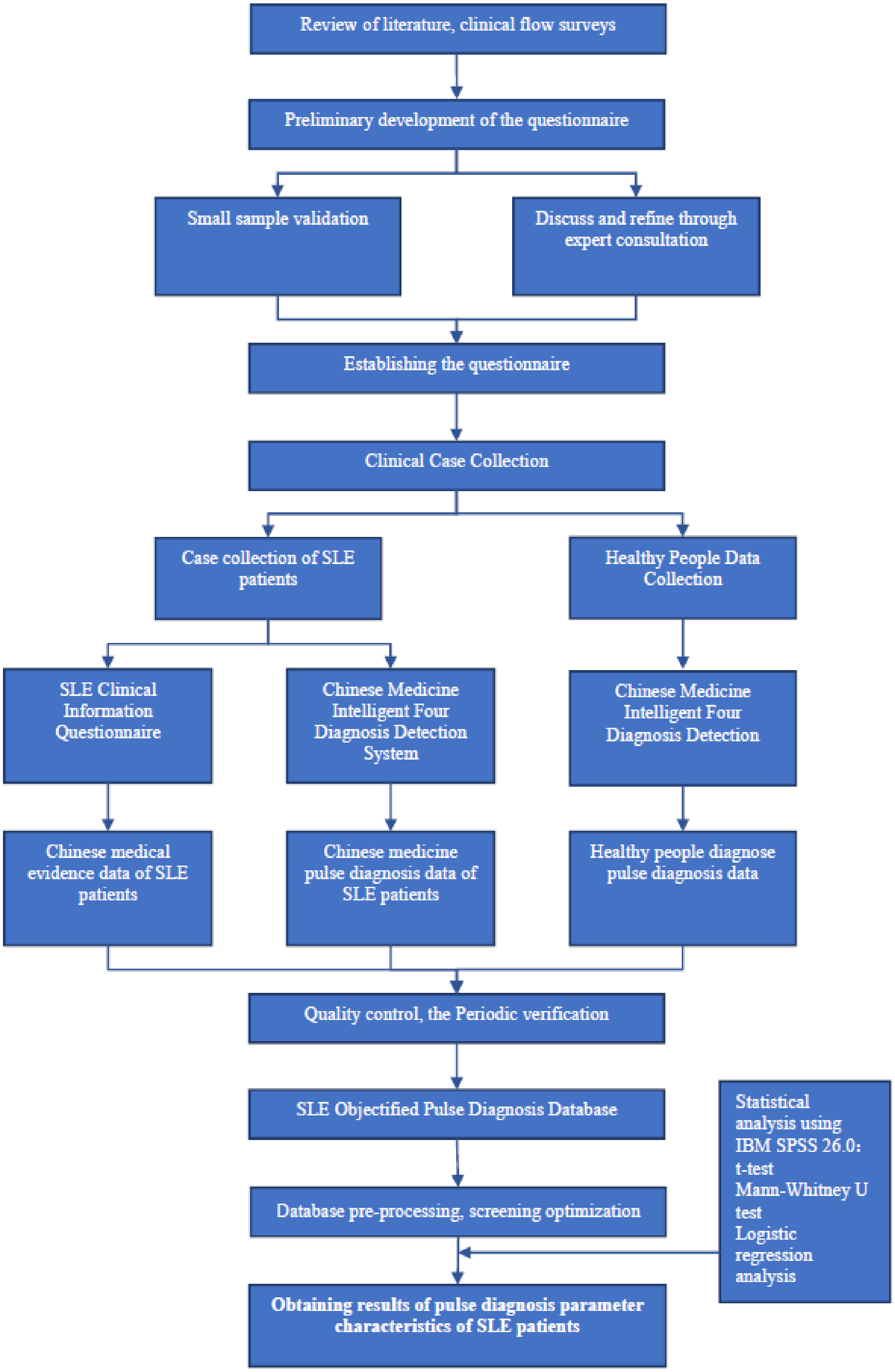

Review of literature, clinical flow surveys
Preliminary development of the questionnaire
Small sample validation
Discuss and refine through expert consultation
Establishing the questionnaire
Clinical Case Collection
Case collection of SLE patients
Healthy People Data Collection
SLE Clinical Information Questionnaire
Chinese Medicine Intelligent Four Diagnosis Detection System
Chinese Medicine Intelligent Four Diagnosis Detection
Chinese medical evidence data of SLE patients
Chinese medicine pulse diagnosis data of SLE patients
Healthy people diagnose pulse diagnosis data
Quality control, the Periodic verification
SLE Objectified Pulse Diagnosis Database
Database pre-processing, screening optimization
Obtaining results of pulse diagnosis parameter characteristics of SLE patients
Statistical analysis using IBM SPSS 26.0: t-test Mann-Whitney U test Logistic regression analysis

Figure 6: Research Methodology Flow Chart

3. RESULTS

3.1 Analysis of pulse parameters in the healthy and SLE groups

Table 1 The results of MannWhitney U test showed that pulse diagnosis parameters h3 value, h5 value, t1 value, t4 value, t5 value, w value, h3/h1 value, h4/h1 value, h5/h1 value, t1/t4 value, t5/t4 value, w/t value were not statistically significant in the diagnosis of pulse diagnosis in SLE patients (P>0.05); pulse diagnosis parameters h1 value, h4 value, t value, t1/t value in SLE patients were influencing factors (P<0.05), where the pulse diagnosis parameters h1 value, h4 value, and t1/t value were significantly higher in the SLE group compared with the healthy group (P<0.05), and the pulse diagnosis parameter t value was significantly lower in the SLE group (P<0.05).

Table 1: Comparison of pulse parameters between the healthy and SLE groups

Parameters	Healthy group (n=130)	SLE group (n=267)	Z	P
h1(mv)	12.65(8.28,19.4)	19.12(11.1,65.31)	-5.250	0.000
h3(mv)	6.53(4.56,8.58)	7.46(4.53,12.29)	-0.475	0.634
h4(mv)	4.22(2.61,7.79)	6.58(3.15,11.34)	-0.877	0.038
h5(mv)	0.21(0.18,0.25)	0.83(0.3,2.2)	-0.982	0.326
t(s)	0.78(0.7,0.85)	0.72(0.65,0.83)	-3.196	0.001
t1(s)	0.16(0.14,0.19)	0.17(0.13,0.21)	-1.386	0.166
t4(s)	0.38(0.33,0.48)	0.37(0.32,0.47)	-0.819	0.413
t5(s)	0.36(0.32,0.41)	0.35(0.3,0.41)	-0.203	0.839
w(s)	0.19(0.16,0.23)	0.15(0.13,0.21)	-1.701	0.089
h3/h1	0.59(0.54,0.68)	0.56(0.44,0.66)	-0.035	0.972
h4/h1	0.39(0.3,0.48)	0.37(0.21,0.5)	-1.500	0.134
h5/h1	0.03(0.01,0.07)	0.05(0.02,0.11)	-1.419	0.156
t1/t	0.21(0.18,0.25)	0.22(0.19,0.27)	-3.029	0.002
t1/t4	0.43(0.35,0.48)	0.44(0.37,0.52)	-0.164	0.870
t5/t4	0.97(0.66,1.21)	0.92(0.67,1.15)	-0.629	0.053
W/t	0.23(0.2,0.27)	0.23(0.2,0.28)	-1.430	0.153

3.2 Logistic regression analysis of pulse parameters in SLE patients

Table 2 The pulse diagnosis parameters h1, h4, t, and t1/t were analyzed by logistic regression using whether the patient had SLE as the dependent variable. The results showed that the independent influences of SLE patients included h4 value (OR=1.073; 95% CI=1.003-1.148; P<0.05) and t value (OR=0.003; 95% CI=0.000-0.763; P<0.05), t1/t value (OR=0.000; 95% CI=0.000-1.199; P<0.05).

Table 2: Logistic regression analysis of pulse parameters in SLE patients

Parameters	B	SE	Wald	P	OR	95%CI Value
h4(mv)	0.07	0.034	4.132	0.042	1.073	1.003-1.148
t(s)	-5.73	2.787	4.23	0.040	0.003	0.000-0.763
t1/t	-12.21	6.324	3.73	0.053	0.000	0.000-1.199

3.3 Comparison of pulse parameters in 3 groups of TCM evidence of SLE

Table 3 The results of Mann Whitney U test showed that pulse diagnosis parameters h1 value, t value, t1 value, w value, h3/h1 value, w/t value were not statistically significant in the diagnosis of different TCM evidence types in SLE patients (P>0.05); pulse diagnosis parameters h3 value, h4 value, h5 value, t4 value, t5 value, h4/h1 value, h5/h1 value, t1/t4 value, t5/t4 value in SLE patients were The h3, h4, h5, t4, t5, and t5/t4 values were significantly lower in the Yin Deficiency Internal Heat Certificate group compared with the Heat Poison Incandescence Certificate group (P<0.05), and the h4/h1 and t1/t4 values were significantly higher in the Yin Deficiency Internal Heat Certificate group (P<0.05); compared with the Heat Poison Incandescence Certificate group, the h3, h4, h5, t5, and t5/t4 values were significantly higher in the Wind Damp Heat Paralysis Certificate group (P<0.05); compared with the Heat Poison Incandescence Certificate group, the h3, h4, h5, and t5/t4 values were significantly higher in the Wind Damp Heat Paralysis Certificate group (P<0.05). h4, h5, t5, and h4/h1 values were significantly lower in the rheumatic heat paralysis group (P<0.05), and t4, h5/h1, and t1/t4 values were significantly higher in the rheumatic heat paralysis group (P<0.05).

Table 3: Comparison of pulse parameters in 3 groups of TCM evidence of SLE

Parameters	Yin deficiency internal heat evidence group (n=140)	Heat poison incandescence evidence group (n=96)	Wind-damp-heat paralysis evidence group (n=31)	P
h1(mv)	24.18(11.23,71.94)	19.67(11.1,30.01)	18.88(9.99,27.01)	0.071
h3(mv)	4.5(3.94,4.86)[a]	9.71(8.58,12.37)	6.34(4.67,11.85)[a]	0.000
h4(mv)	5.39(1.82,19.49)[a]	7.51(3.71,10.44)	4.46(2.44,8.05)[a]	0.005
h5(mv)	0.7(0.24,1.76)[a]	0.73(0.28,1.31)	0.65(0.21,1.57)[a]	0.005
t(s)	0.74(0.58,0.82)	0.74(0.68,0.85)	0.76(0.67,0.86)	0.162
t1(s)	0.17(0.14,0.21)	0.17(0.14,0.22)	0.16(0.13,0.22)	0.653
t4(s)	0.35(0.31,0.47)[a]	0.39(0.33,0.48)	0.41(0.3,0.47)[a]	0.002
t5(s)	0.37(0.32,0.47)[a]	0.38(0.27,0.48)	0.34(0.29,0.41)[a]	0.004
w(s)	0.18(0.14,0.23)	0.18(0.14,0.22)	0.18(0.15,0.22)	0.888
h3/h1	0.52(0.47,0.63)	0.6(0.56,0.68)	0.31(0.18,0.5)	0.099
h4/h1	0.42(0.29,0.66)[a]	0.31(0.17,0.48)	0.32(0.15,0.52)[a]	0.001
h5/h1	0.03(0.02,0.08)	0.03(0.02,0.08)	0.04(0.03,0.15)[a]	0.002
t1/t	0.24(0.2,0.29)	0.22(0.19,0.27)	0.22(0.19,0.27)	0.256
t1/t4	0.43(0.36,0.52)[a]	0.41(0.33,0.46)	0.46(0.39,0.54)[a]	0.029
t5/t4	0.87(0.7,1.14)[a]	0.94(0.55,1.21)	0.91(0.62,1.19)	0.008
W/t	0.26(0.23,0.31)	0.24(0.2,0.3)	0.24(0.19,0.31)	0.052

[a] Compared with the group with incandescent heat toxin evidence, p<0.05.

3.4 Logistic regression analysis of pulse diagnosis parameters in 3 groups of TCM evidence of SLE

Table 4 Using the TCM certificate types of SLE patients (Yin deficiency internal heat certificate group, heat toxin incandescence certificate group, and wind-damp heat paralysis certificate group) as the dependent variables and the reference category of heat toxin incandescence certificate group, the pulse diagnosis parameters h3 value, h4 value, h5 value, t4 value, t5 value, h4/h1 value, h5/h1 value, t1/t4 value, and t5/t4 value as the independent variables, logistic regression analysis was performed. The results showed that the independent influencing factors of TCM evidence in SLE patients included: h3 values in the yin deficiency internal heat evidence group (OR=2.295; 95% CI=1.843-2.858; P<0.05) and h3 values in the wind-damp-heat paralysis evidence group ((OR=2.309; 95% CI=1.87-2.85; P<0.05).

Table 4 : Logistic regression analysis of pulse diagnosis parameters in 3 groups of TCM evidence of SLE

Parameters	B	SE	Wald	P	OR	95%CI Value
Yin deficiency internal heat evidence group h3(mv)	0.831	0.112	55.137	0.000	2.295	1.843-2.858
Wind-damp-heat paralysis evidence group h3(mv)	0.837	0.107	60.644	0.000	2.309	1.870-2.850

4. DISCUSSION

Pulse diagnosis as one of the four diagnoses of Chinese medicine, is an important method and means of judging disease in Chinese medicine, pulse diagnosis can determine the occurrence, development and prognosis of disease, in Chinese medicine is crucial in the diagnosis and treatment, pulse diagnosis information can reflect to a certain extent the changes in the internal organs, to provide the basis for clinical diagnosis of disease, treatment of disease, but pulse diagnosis is the most difficult of the four diagnoses of Chinese medicine is the most difficult to master the diagnostic method, "Pulse Classic" said "The pulse is subtle, its body is difficult to identify. In the heart is easy to understand, the finger is difficult to understand", it can be seen that "cut and know" is not an easy task. And the traditional pulse diagnosis has a large subjectivity, which seriously affects the accuracy of pulse diagnosis judgment. In recent years, there has been a gradual increase in the number of studies on the objectification of pulse diagnosis [10-14], and it has been applied to the study of various diseases, where the pulse characteristics of different TCM evidence types vary for the same disease, reflecting different disease states.

SLE is a kind of intractable systemic autoimmune disease, the cause of which has not yet been elucidated. Western medicine believes that the cause of SLE is related to genetic abnormalities, environmental factors and estrogen[15-16]. According to Chinese medicine, SLE belongs to the category of "yin and yang toxicity" and "warm toxicity hair spot", and "Jin Kui Yao" says: "Yang toxicity is the disease, the face is red and spotted like brocade." Both of them point out that "yang toxin" is the main cause and mechanism of SLE: SLE patients have insufficient congenital endowment and deficiency of true yin, and the toxin accumulates in yangming, and the internal and external heat evil fights with each other, and the heat toxin burns the ying-yin and develops. Chinese medicine believes that the etiology of SLE is mostly related to heat toxin injury, warm toxin invasion, and heat toxin incandescence [17-18]. Therefore, based on the etiology and pathogenesis of SLE, this study analyzed the pulse diagrams of SLE patients with internal heat evidence of yin deficiency, incandescent heat toxin evidence, and heat paralysis of rheumatism. Therefore, in this study, we analyzed the pulse patterns of SLE patients with yin deficiency and internal heat, heat toxicity and heat paralysis, and provided an objective basis for the diagnosis and treatment of SLE from pulse diagnosis [19-20].

Among the pulsogram parameters h1 is the main wave amplitude, reflecting left ventricular ejection function and large artery compliance. h4: is the descending mid-isthmus amplitude, reflecting arterial vascular elasticity. t value is the pulsation period, i.e. one cardiac cycle of the left ventricle. t1: the temporal value between the start of the pulsogram and the peak of the main wave, which is the rapid ejection period of the left ventricle. In this study, the values of pulsogram parameters h1 and h4 in the SLE group were higher than those in the healthy group, t values were lower than those in the healthy group, and t1/t values were higher than those in the healthy group, suggesting that patients with SLE suggest that patients with SLE have higher main wave amplitude of pulsogram, better arterial vascular elasticity, shorter pulsation period, and faster ventricular ejection, indicating that patients with SLE have faster overall pulsation frequency and faster heart rate than normal.SLE patients have a higher pulse count than healthy people, and Chinese medicine believes that the pulse count means that the pulse beats faster than normal, between about 90 and 130 times per minute, mostly due to evil heat agitation and accelerated blood flow, which is consistent with the pathogenesis of SLE "Yang toxicity" [21].

In the comparison of different TCM evidence types in SLE, the h3, h4, h5, t4, t5, and t5/t4 values of patients in the heat toxin incandescence evidence group were higher than those in the yin deficiency internal heat evidence group, suggesting that patients in the heat toxin incandescence evidence group had higher descending isthmus amplitude, higher degree of heavy beat wave amplitude, longer cardiac systolic and diastolic time, and higher state of arterial peripheral resistance than those in the yin deficiency internal heat evidence group, indicating that patients in the heat toxin incandescence evidence group in SLE The astringent pulse characteristic of the patients in the SLE heat and toxin incandescent evidence group, which is considered by Chinese medicine as astringent pulse, thin and late, with a pulse rate of about 60 beats per minute [22], may be related to the pathogenesis of SLE disease in which the toxic evil of heat enters the blood, resulting in poor blood flow and stasis of the veins and collaterals. The h4/h1 and t1/t4 values of the patients in the Yin deficiency internal heat evidence group were higher than those in the heat toxin incandescence evidence group, suggesting a weakened left ventricular ejection function and a prolonged pulse cycle, indicating that the pulse of patients with SLE Yin deficiency internal heat evidence is slow, which is considered by Chinese medicine to be a pulse beating less than 60 times per minute, mostly related to blood deficiency and blood stasis. Compared with the heat-poison incandescent evidence group, the h3, h4, h5, t5, and h4/h1 values were significantly lower in the rheumatic heat paralysis evidence group, and the t4, h5/h1, and t1/t4 values were significantly higher in the rheumatic heat paralysis evidence group , suggesting that patients in the rheumatic heat paralysis evidence group had lower descending isthmus amplitude, lower amplitude of the degree heavy beat wave, longer diastolic time, and shorter systolic time. This indicates that the patients in the rheumatism-heat paralysis group had reduced blood return to the heart, poor blood flow and easy formation of blood plaques, and slow and astringent pulse, further indicating that the patients in the rheumatism-heat paralysis group in SLE had more severe blood stasis than those in the heat toxin incandescence evidence group[22].

The logistic regression analysis found that h4 value, t value, and t1/t value were independent influencing factors of pulse diagnosis parameters in SLE patients, i.e., arterial vascular peripheral resistance, left ventricular rapid ejection time, and pulsation cycle had an effect on pulse diagnosis in SLE patients; logistic regression analysis found that pulse diagnosis parameter h3 value was an independent influencing factor of Chinese medicine evidence diagnosis in SLE, suggesting that the influence of pulse diagnosis parameter h3 value on evidence diagnosis should be paid attention to in future studies of Chinese medicine pulse diagnosis.

5. CONCLUSION

the results of this study suggest that pulse diagnosis parameters have good reference value in the clinical diagnosis of SLE in TCM evidence-based treatment. It is expected that this study will continue to expand the sample size, add more pulse diagnosis parameters, and continue to optimize the extraction method of pulse diagnosis parameters in order to provide more data and objective basis for the early diagnosis and prevention of SLE based on pulse diagnosis parameters.

ACKNOWLEDGMENTS

We thank the National Natural Science Foundation of China (No.82160874), the Natural Science Foundation of Hainan Province (No.2019RC206), and the Postgraduate Science and Innovation Program of Hainan Province (No.Qhys2021-361) for supporting the funding of this research. Thanks are due to all of volunteers for providing invaluable pulse data.

REFERENCES

[1] Chang NC. Rheumatic diseases in China. J Rheumatol Suppl 1983; 10:41-45.

[2] Zeng QY, Chen R, Darmawan J, et al. Rheumatic diseases in China. Arthritis Res Ther. 2008. 10(1): R17.

[3] Rees F, Doherty M, Grainge MJ, et al. The worldwide incidence and prevalence of systemic lupus erythematosus: a systematic review of epidemiological studies. Rheumatology (Oxford). 2017. 56(11): 1945-1961.

[4] Zhang J, Liao J, Wang T, et al. Effects of joy and sorrow on pulse-graph parameters in healthy female college students based on emotion-evoked experiments. Explore (NY). 2021. 17(4): 303-311.

[5] Feng X, Feng L, Gao H, et al. Characteristics of Pulse Parameters in Patients with Polycystic Ovary Syndrome Varied at Different Body Mass Index Levels. Evid Based Complement Alternat Med. 2022. 2022: 7220011.

[6] Shi HZ, Fan QC, Gao JY, et al. Evaluation of the health status of six volunteers from the Mars 500 project using pulse analysis. Chin J Integr Med. 2017. 23(8): 574-580.

[7] Aringer M, Costenbader K, Daikh D, et al. 2019 European League Against Rheumatism/American College of Rheumatology Classification Criteria for Systemic Lupus Erythematosus. Arthritis Rheumatol. 2019. 71(9): 1400-1412.

[8] CFDA. Guiding principles for clinical research of new drugs in Chinese medicine (for trial implementation). Beijing: China Pharmaceutical Science and Technology Press, 2002:111-115.

[9] Fei ZF. Pulse Diagnosis of Modern Traditional Chinese Medicine.1st. Beijing: People’s Medical Publishing House; 2003.

[10] Jin J, Zhang H, Geng X, Zhang Y, et al. The pulse waveform quantification method basing on contour and derivative. Comput Methods Programs Biomed. 2022. 220: 106784.

[11] Pelaez-Coca MD, Hernando A, Lozano MT, et al. Photoplethysmographic Waveform and Pulse Rate Variability Analysis in Hyperbaric Environments. IEEE J Biomed Health Inform. 2021. 25(5): 1550-1560.

[12] de Sá Ferreira A, Lopes AJ. Pulse waveform analysis as a bridge between pulse examination in Chinese medicine and cardiology. Chin J Integr Med. 2013. 19(4): 307-14.

[13] Huang PY, Lin WC, Chiu BY, et al. Regression analysis of radial artery pulse palpation as a potential tool for traditional Chinese medicine training education. Complement Ther Med. 2013. 21(6): 649-59.

[14] Zhang J, Niu X, Yang XZ, et al. Design and application of pulse information acquisition and analysis system with dynamic recognition in traditional Chinese medicine. Afr Health Sci. 2014. 14(3): 743-52.

[15] Mu Q, Zhang H, Luo XM. SLE: Another Autoimmune Disorder Influenced by Microbes and Diet? Front Immunol. 2015 Nov 30;6:608.

[16] Zucchi D, Elefante E, Schilirò D, et al. One year in review 2022: systemic lupus erythematosus. Clin Exp Rheumatol. 2022. 40(1): 4-14.

[17] Hou HY, Lv XL. Discussion on the etiology and pathogenesis of different stages of systemic lupus erythematosus in Chinese medicine. Rheumatism and Guanitis.2021,10(06):61-66.

[18] Agumuda, Chen WW, Su X, et al. Research progress in the treatment of systemic lupus erythematosus from the theory of yin deficiency. Journal of Liaoning University of Traditional Chinese Medicine.2021,23(06):144-149.

[19] Chen LM, Zhu ZY, Bao J,et al. Treatment of systemic lupus erythematosus from the theory of "spots as Yangming heat toxins". Chinese Journal of Traditional Chinese Medicine.2020,35(08):3972-3974.

[20] Zhu YL, Wu F. Overview of research on Chinese medical evidence patterns and changes in systemic lupus erythematosus. Chinese Journal of Traditional Chinese Medicine.2018,33(07):2973-2975.

[21] Yang KP, Wang XC, Gao X F, et al. Discussion on the effect and mechanism of detoxification and elimination of blood stasis and nourishment of Yin formula on lupus activity in systemic lupus erythematosus. Chinese Journal of Traditional Chinese Medicine.2021,39(04):176-179.

[22] Li CD, Fan CY. Diagnostics of Chinese medicine. Beijing: China Traditional Chinese Medicine Publishing House; 2021.

Mining Library Virtual Reference Service Data by using the software NVivo 12

Leilei Peng[a], Ke Chen*[a]

[a]Library, Sichuan University, No.24 South Section 1, Yihuan Road, Chengdu, China, 610065

* Corresponding author: chenkechenke@scu.edu.cn

ABSTRACT

In the information age, virtual reference service (VRS) has gradually replaced the traditional face-to-face (F2F) reference and become an important way of reference for library users. Based on grounded theory, this research uses NVivo12 software to mine the virtual reference service data of users in the Sichuan University Library, the first is high-frequency word analysis, and the second is mining the data by combining open coding, spindle coding, and selective coding. The analysis results of high-frequency words show that in the process of using various services of the library, the contents of the virtual reference service mainly focus on the following aspects: hours, literature resources, space, and facilities. The coding results show that the library virtual reference service data could be encoded into three core categories, they are "library spaces and facilities", "collections and electronic resources" and "Services", where "library spaces and facilities" is the most concern aspect of the library users, which has 359 nodes, accounting for 50.71% of the total, and the proportion of "collections and electronic resources" and "Services" is not much different, 27.26% and 22.03%, respectively.

Keywords: Library virtual reference service, Data mining, NVivo 12

1. INTRODUCTION

With each successive wave of web innovation, there are increased opportunities for using the internet as an online space for communication and interaction [1]. Due to the continuous innovation of information technology, university libraries try to meet the information needs of their patrons by using a variety of online services [2], compared to traditional face-to-face (F2F) reference, library uses information technology to provide readers with many new reference ways, such as web page reference, email reference, WeChat reference, etc., and among them, web-based virtual reference services are one of the most popular consultation methods for readers. According to Pomerantz and Luo's research in 2006, compared with traditional reference services, perceived convenience (the speed, efficiency, and immediacy with which answers were received, as well as the potential for remote access) was the primary reason for users recommend virtual reference services [3]. However, some research shows that a certain number of users prefer traditional consulting services. Sobel found that 69 percent of students would rather engage in face-to-face reference help than use a virtual medium [4]. Virtual reference service builds a network platform for communication between users and libraries, which makes communication between users and the library more convenient. By analyzing the data, libraries can deeply understand the needs of users, to promote the innovation and sustainable development of the library.

2. THEORY AND RESEARCH TOOL

2.1 Grounded theory

Grounded theory was used in this study to investigate the virtual reference service data of Sichuan University Library from January 2019 to October 2022. Grounded theory is a commonly used approach in qualitative research [5], which is a systematic yet flexible methodology, designed to assist with the development of substantive, explanatory models grounded in relevant empirical data. Since it was first described by Barney Glaser and Anselm Strauss [6]. Grounded theory is one of the best methods for exploring and understanding complex and multifaceted issues [7], and has gradually become one of the most influential methodologies in social science research. Grounded theory is a new model for analyzing text data. It is different from other qualitative research methods in that its research aim is to create theories from text data rather than to describe and explain phenomena through existing theories [8]. The steps of grounded theory research include open coding, axial coding, and selective coding, and the encoding methods include word-by-word encoding, line-by-line encoding, and event-to-event encoding [9]. Researchers can adopt appropriate coding methods according to different research materials and purposes.

International Conference on Computer Graphics, Artificial Intelligence, and Data Processing (ICCAID 2022)
edited by Ruishi Liang, Jing Wang, Proc. of SPIE Vol. 12604, 126040Q
© 2023 SPIE · 0277-786X · doi: 10.1117/12.2674684

2.2 Research tool

The systematic and rigorous preparation and analysis of qualitative data is usually time-consuming and labor-intensive, an often-cited criticism of qualitative research is that researchers' personal viewpoints may unduly impact the ways in which they analyze data [10], therefore, Computer Assisted Qualitative Data Analysis Software (CAQDAS) is gradually applied to qualitative data analysis, such as ATLAS. ti, MaxQDA, NVivo, and N6, and NVivo is the one most commonly used by scholars [11]. NVivo is designed for researchers who wish to display and develop rich data in dynamic documents. Documents can be imported and edited in rich text with hyperlinks to sound, image, and other files [12], the analysis includes three steps: The first is to import text, audio, video, E-mail, images, spreadsheets, online surveys, network content, and other original materials, followed by the visualization analysis of high-frequency words, word cloud map, cluster map and so on, and the last is to carry out the automatic or manual encoding of the original materials to find out the potential links between the texts. It is concluded that using such software is a technological tool that makes it easier to organize, visualize, and access research data, something which represents saving time and work.

3. DATA SOURCE

This research takes Sichuan University Library as a case study. Sichuan University Library is the library with the longest history and the largest number of documents in Southwest China, which is a typical comprehensive university. This study used the web crawling tool to crawl 927 users' virtual reference service text from the official website of Sichuan University Library, and the date range was set to January 2019 to October 2022. Then, this research translates the virtual reference service texts from Chinese to English. After completing the collection and translation of the original data, we first sort out the time and content of each reference, then import the data into the EXCEL, and use the "remove duplicate values" function to delete duplicate data. Then, references were checked one by one manually, typos were corrected, and sentences such as "the installation file could not be opened" and "please solve this problem as soon as possible" were deleted, and 920 research data were obtained.

4. RESULTS

4.1 Analysis of monthly number of user virtual references

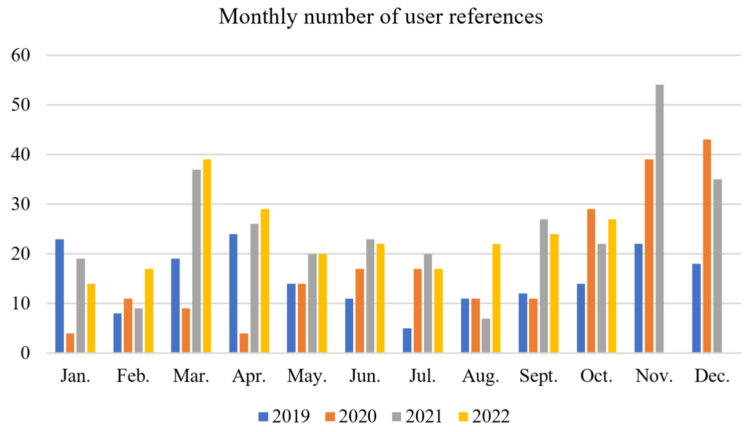


Figure 1. Monthly number of user virtual references

The number of user virtual references shows an upward trend year by year, reaching 181, 209, 299, and 231, respectively, from 2019 to 2022, and in February, August, July and January ranked the bottom, with 45, 51, 59 and 60 respectively. The main reason is that January and February are winter holidays, and July and August are summer holidays. The number of library users in these four months decreases, so the number of references decreased accordingly. The highest number of references occurred in March and November, indicating that the library resources were used more by the users during these two months. The month with the largest number of user inquiries is November 2021, with a total of 54, and the month with the smallest number is April 2020, with only 4. From January to April 2020, there was a significant decrease compared with the same period. The main reason was that COVID-19 broke out in January 2020, and then the library was closed.

4.2 Word Frequency Analysis

High-frequency words represent the most common problems that users encounter when using library services, to analyze the word frequency, we used the "word frequency query" function of NVivo 12, and set word length to three or more letters, based on preliminary analysis, this study puts words such as "now", "one", "ask" and "long", into the "stop words bottle" through manual tagging. The overview of the library virtual reference service can be intuitively displayed through the "words cloud map" function of NVivo 12, as shown in figure 2, the larger the font of the word in the word cloud map, the more frequently the word appears, and the more important it is.

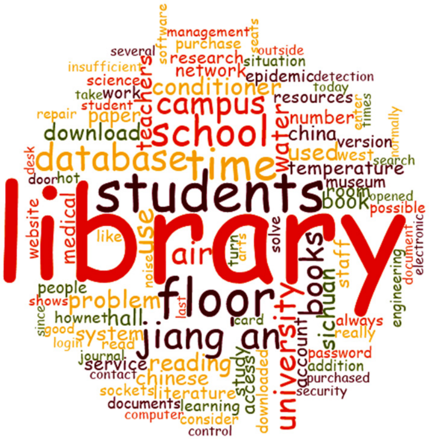


Figure 2: Words cloud map

The most frequently used words and words cloud map shows that, "library", "students", "floor", "time", "school", "Jiang'an", "database", "air", "campus" and "university" is the top 10 most frequently used words, their word frequency has exceeded 150 times, and in words cloud map, the fonts of these 10 words are also significantly larger than other words. It is indicated that Sichuan University Library virtual reference service mainly focuses on two high-frequency words: "library" and "students", the frequency of which is 755 and 293, respectively, much higher than other high-frequency words. The analysis results of high-frequency words show that in the process of using various services of the library, the contents of the virtual reference service mainly focus on the hours, literature resources, space, and facilities.

4.3 Data Encoding Analysis

High-frequency word analysis can directly reflect the main focus of the Sichuan University Library virtual reference service, for further in-depth analysis, this study uses the automatic coding function of NVivo 12 to cluster the text data. The results are shown in figure 3.

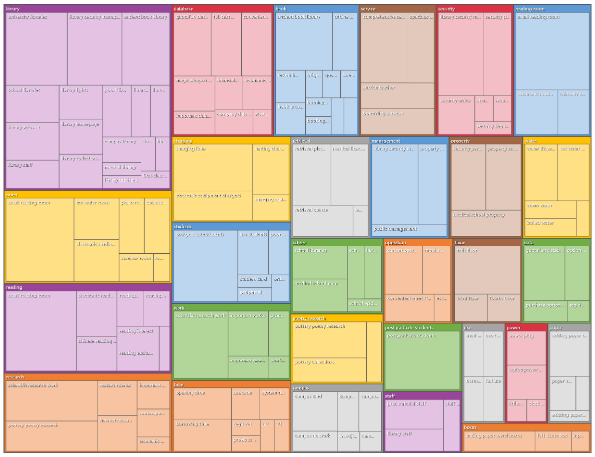


Figure 3. Automatic coding nodes

As shown in figure 3, automatic coding clusters text data into 29 categories, 114 nodes, and 812 coding points, such as "library", "room" and "book". The automatic coding results show that the boundaries of each node are fuzzy and the clustering degree is low. Some of them should be merged into one category, such as "database" and "data", "students" and "postgraduate students". Some categories should not be separated into one category, and the content of the node needs to be sorted out and incorporated into other nodes, such as "use", "water" and "boxes".

Therefore, in this study, all the categories and nodes of the automatic coding were manually analyzed one by one, the inappropriate classifications were adjusted, the nodes of the same category were integrated, and the specific steps included open coding, axial coding, and selective coding. The first step is to carry out open coding, open coding is the interpretive process by which data are broken down analytically, in this step, events, actions, and interactions are compared with others for similarities and differences, and similar events, actions, interactions are grouped to form categories and subcategories. The second step is axial coding, in this step, categories are related to their subcategories, and the relationships are tested against data. The second step is Selective coding, which is the process by which all categories are unified around a "core" category, and categories that need further explication are filled in with descriptive detail [13]. This study encodes the library virtual reference service text data according to the step of open coding, axial coding, and selective coding, the results are shown in table 1:

Table 1. Open coding, axial coding, and selective coding of library VRS texts

Open coding (nodes)	Axial coding (nodes)	Selective coding (nodes)
Liberal arts and science library (14)	Spaces (181)	Library spaces and facilities (359)
Engineering Library (13)		
Medical Library (40)		
Jiang'an Library (60)		
Washroom (21)		
Reading Room (28)		
Water dispenser (21)	Facilities (178)	
Socket (34)		
Air conditioner (72)		
Table (22)		
Lamp (29)		
Databases of liberal arts (38)	Suggest purchase of an E-resources (72)	Collections and electronic resources (193)
Databases of science (12)		
Databases of medicine (16)		
Comprehensive databases (6)		
Access to a database (30)	Services of electronic resources (45)	
How to use a database (15)		
Thesis (13)	Services of Collections (54)	
Borrow, renew, and return books (26)		
Book reserves (15)		
Purchase suggestions (22)	Suggest a purchase of a book (22)	
Hours on weekday (32)	Hours (67)	User services (156)
Hours on weekend (20)		
Hours on Winter and summer vacation (15)		
Librarian's ability (15)	Librarian (36)	
Librarian's attitude (21)		
Literature service (32)	service contents (53)	
knowledge service (21)		

Table 1 shows that, in this study, through layer-by-layer coding, the data is first clustered into 28 nodes by using open coding, and then is clustered into 9 nodes by using axial coding. On this basis, the 9 nodes are encoded to three core categories by using selective coding, they are "library spaces and facilities", "collections and electronic resources" and "user services", where "library spaces and facilities" is the most concern aspect of the library users, which has 359 nodes, accounting for 50.71% of the total, and the proportion of "collections and electronic resources" and "user services" is not much different, 27.26% and 22.03%, respectively. That is, in the process of patrons using the various services of the library, library space and facilities are the most concerning aspect for them.

4.3.1 Library spaces and facilities

Library space and facilities account for 50.71% of the total proportion of library virtual reference service data, including the two main axial codes of space and facilities, the proportion of which has a small difference, respectively 50.42% and 49.58%. Sichuan University Library has been committed to building a smart library in recent years. The first is through various intelligent technologies and facilities with functions of interaction, perception, and capture, to analyze, transmit and master various users' data, analyze users' needs and provide users with personalized, accurate, and diversified services. The second is to transform and upgrade the original single-function space into a smart space that not only contains a variety of advanced technologies to provide users with an intellectual experience but also contains humanistic care to provide users with personalized and diversified services, such as study and discussion space, multimedia space, experience space, information sharing space, maker space, etc., The third is the construction of virtual space. Virtual space is the extension of physical space, but it is not restricted by the physical space of the library and depends on information technology and Internet technology, such as virtual communities, online platforms, virtual libraries, and with the development of the intelligent library, the importance of virtual space will become more and more prominent. Sichuan University Library has four branch libraries, table 2 shows that Jiang'an Library is the focus of users, and in terms of facilities, the air conditioner is the most concerned, with a rate of 40.45%.

4.3.2 Collections and electronic resources

Literature resources are the foundation of a library, including collections and electronic resources, so libraries must put the construction of literature resources in first place. The results of text data mining show that users pay more attention to electronic resources, accounting for 60.62%, indicating that in the information era, users are more likely to use the electronic resources of the library than paper resources. Among them, the number of nodes for recommending libraries to purchase liberal arts databases is 38, accounting for 52.78% of the total, indicating that liberal arts students are more inclined to use library virtual reference services. The number of Access to a database node is 30, accounting for 66.67% of the electronic resource service code. Since COVID-19 in January 2020, the staff of Sichuan University Library has worked at home, and the way users use electronic resources has changed from authorized access to off-campus access. Therefore, the importance of off-campus access is becoming more and more prominent. Libraries should further optimize off-campus access services given the sharp increase in the number of visits and concurrency. In terms of collections, users mainly focus on borrowing, returning, and reserving books, as well as recommending the library buy the books they need.

4.3.3 User services

User services include three aspects: hours, librarian, and service content. As can be seen from the above analysis, users mainly use the library space for self-study, so in terms of library services, the first concern of users is the service hours. In combination with the specific text data, the virtual reference mainly focuses on hours on weekdays (32 nodes), hours on weekends (20 nodes), and hours on Winter and summer vacation (15 nodes). Users will have contact with librarians when using library services, so the ability and attitude of librarians are one of the key concerns of users. However, through text mining, we did not find any readers praising the working ability or service attitude of the librarians. On the contrary, all readers' inquiries are complaints about the poor ability or service attitude of the librarians, which shows that users regard the virtual reference service platform more as a complaint platform. In terms of service contents, knowledge service is the focus of users' attention, and knowledge service mainly focuses on teaching service, research service, and academic evaluation.

5. CONCLUSION

In the information age, the needs of users tend to be personalized and diversified, which brings new challenges to university libraries. Virtual reference service has gradually become an important way for users to express their needs to university libraries. On the one hand, libraries should pay attention to the needs of users, constantly enrich the service content, optimize the service process and improve the service quality. On the other hand, libraries should regularly collect, sort out and analyze the contents of virtual references, and improve the work according to the analysis results, to provide more appropriate library services for users.

REFERENCES

[1] Nicholas, J. Virtual reference, Second Life and traditional library enquiry services. Library Review. 57(6): 417-423 (2008).

[2] Radford M L, Connaway L S. “Screenagers” and live chat reference: Living up to the promise. Scan. 26 (1): 31-39 (2007).

[3] Pomerantz, Luo J, Lili. Motivations and Uses: Evaluating Virtual Reference Service from the Users' Perspective. Library & Information Science Research. 28(3): 350-373 (2006).

[4] Karen Sobel. Promoting Library Reference Services to First-Year Undergraduate Students: What Works?. Reference & User Services Quarterly. 48 (4): 362-371 (2009).

[5] Creswell JW, Poth CN. Qualitative inquiry and research design: Choosing among five approaches (4th ed.). London, England: Sage Publications. (2018).

[6] Hutchison AJ, Johnston LH, Breckon JD. Using QSR-NVivo to facilitate the development of a grounded theory project: an account of a worked example. International Journal of Social Research Methodology. 13(4): 283-302 (2010).

[7] Mishra P, Gupta R, Bhatnagar J. Grounded theory research: exploring work-family enrichment in an emerging economy. Qualitative Research Journal. (14): 289-306 (2014).

[8] Charmaz K. *Constructing grounded theory: A practical guide through qualitative analysis*. London, England: Sage Publications. (2006).

[9] Maher C, Hadfield M, Hutchings M. & de Eyto A. Ensuring Rigor in Qualitative Data Analysis: A Design Research Approach to Coding Combining NVivo With Traditional Material Methods. International Journal of Qualitative Methods. 17(1): 1-13 (2018).

[10] Marie CC. Quantitative and Qualitative Research: A View for Clarity. International Journal of Education. 2(2): 1-14 (2010).

[11] Zamawe FC. The Implication of Using NVivo Software in Qualitative Data Analysis: Evidence-Based Reflections. Malawi Medical Journal. 27(1): 13-15 (2015).

[12] Richards L. Data Alive! The Thinking Behind NVivo. Qualitative Health Research. 9(3): 412-428 (1999).

[13] Corbin JM, Strauss A. Grounded theory research: Procedures, canons, and evaluative criteria. Qualitative Sociology. 13: 3-21 (1990).

Lightweight Refueling Behavior Recognition Algorithm Based on Sequence Diagrams

Dasheng Guan [a], Lei Wang [a], Zhijun Zhang [a], Cong Liu [a]
([a]Shanghai ChengFei Aviation Special Equipment Co., Ltd., Shanghai 201613)

ABSTRACT

Some specific, repetitive actions made by the staffs in the refueling work scenario at the airport can be considered as a way of information transmission, so it is necessary to carry out on-site automatic identification and monitoring of these specific actions to improve the level of supervision. This paper proposes a lightweight refueling behavior recognition algorithm applicable to the field based on video sequences. The algorithm firstly uses the YOLOv3 improved target detection network for human body detection. The resulting human body detection box is tracked using the target tracking algorithm, and the tracked human body sequence maps are input into the behavior classification algorithm based on time-space feature fusion to realize the fast and intelligent analysis of the behavior. The test results of deploying the algorithm to Hi3559A embedded equipment show that the recognition accuracy of the algorithm reached 94.68%, and the inference speed reached 22FPS, which can meet the needs of real-time behavior analysis and processing at the airport refueling site.

Keywords: lightweight, spatio-temporal feature, target detection, behavior classification

1. INTRODUCTION

In some production scenarios such as refueling at the airport, the refuelling staffs will make some behavioral actions with specific significance during the work, indicating the completion of relevant operations, such as raising arms, bending down,etc., and the managers need to supervise the staff's behavior. The method currently used is manual supervision or video surveillance. However, the actual processing is time-consuming and laborious, and it is difficult to meet the requirements of real-time and all-weather. Thanks to the development of deep learning, the behavior recognition algorithm based on deep learning can intelligently analyze and process the surveillance video, and the behavior of the staff can be recognized and monitored 24/7 in real time.

At present, there are two main categories of video-based behavior recognition algorithms: traditional algorithms and deep learning-based methods[1]. Traditional methods characterise behaviour by manually designing features and and use classification methods on statistical learning to classify behaviors. Deep learning-based approaches use neural networks for feature extraction, with two-stream, convolutional 3D (C3D), and long short term memory network (LSTMs) being the three main approaches. At the same time, some scholars use deep learning to recognize human behavior from other angles, such as behavior recognition based on skeletal keypoints[2]. Traditional methods are simple to implement but have average accuracy and robustness, while methods based on deep learning have high robustness, which can automatically extract features and have more advantages in handling complex problems. Therefore, this paper chooses the method based on deep learning to design the algorithm for recognizing behavior.

Currently, the behavior recognition algorithm based on video requires high computing power for GPUs and other hardware devices. In actual operation, it often brings higher costs and power consumption, making it difficult to deploy flexibly into various production environments. To solve these problems, this paper proposes a sequence-based lightweight refueling behavior recognition algorithm. First, this algorithm proposes a lightweight target detection network YOLOv3-SE based on YOLOv3[3]. And then it designs a sequence map-based behaviour recognition algorithm[4], which fuses temporal and spatial features of multi-frame human RGB images.

Main work of this paper: a lightweight sequence graph-based behaviour recognition algorithm is proposed, which consists of a lightweight human detection algorithm and a fast human behaviour recognition algorithm based on spatio-temporal features. The human detection algorithm is optimized based on the YOLOv3, which reduces the network parameters and computation. A human behaviour recognition algorithm based on spatio-temporal features is proposed with reference to the GaitSet algorithm[5] , which enables fast recognition of human behaviour. The algorithm can be deployed on embedded devices to obtain images of specified scenes, and realize real-time recognition of the behaviour of the stuff to reduce the occurrence of accidents.

International Conference on Computer Graphics, Artificial Intelligence, and Data Processing (ICCAID 2022)
edited by Ruishi Liang, Jing Wang, Proc. of SPIE Vol. 12604, 126040R
© 2023 SPIE · 0277-786X · doi: 10.1117/12.2674654

2. RELATED WORK

2.1 Human detection algorithm

Currently, the target detection algorithm based on convolutional neural network is mainly divided into two major categories: one-stage target detection and two-stage target detection[6]. Among them, two-stage target detection appeared earlier. The main idea is to first regress the prediction box and then classify the targets of the prediction box.The relevant algorithms are R-CNN[7]series, SPPNET[8], etc. The one-stage target directly grids the input picture. After feature extraction, the target position and classification prediction are directly carried out. The relevant algorithms are YOLO[9] series, SSD, etc.

The YOLOv3 network has good balance in speed and accuracy, but considering the different application scenario and the different size of the detection target, the network structure need not to be so complex. Besides, on mobile devices, edge devices and other terminals with limited computing power and storage resources, it is difficult for YOLOv3 to achieve real-time. In this case, a portion of the performance loss is acceptable in exchange for faster inference speed[10]. Therefore, this paper chooses to make a lightweight improvement on YOLOv3 from four aspects:

(1) Adjust the Bottleneck module of MobileNetV2[11], and use the adjusted module to reconstruct the feature extraction network;

(2) Add the SPP structure to the end of feature extraction network to improve the detection accuracy;

(3) Adjust 3-layer feature information fusion module to 2-layer;

(4) The output head part is decoupled to improve the convergence speed of the network.

2.2 Destination Tracking Algorithm

Multi-target tracking is a type of task that is widely studied in computer vision. Currently, the research and implementation of target tracking are mostly based on detection tracking. First, the target's positioning box is obtained through the target detection algorithm, and then the next step of tracking is carried out according to the positioning box.

IOUTracker[12] is a simple and efficient target tracking model proposed by Erik et al., which is fast and does not require additional image information. The algorithm is based on two hypotheses, one is that the detector performance is good enough, and the other is that the video frame rate is high. In order to improve the stability of the algorithm, this paper chooses to improve the IOUTracker algorithm by calculating the similarity of the unmatched detection boxes according to its scores and aspect ratio after the IOU matching.

2.3 Behavior classification algorithm

The key to the behavior classification is to establish the relationship between the human image and the behavioural category. Depending on the input data behavior classification algorithms can be divided into those based on video or image sequences and those based on static images. The latter is similar to the image classification in that it uses a static image to determine the classification. Typical examples include the traditional Resnet series, the Inception series, and the lightweight MobileNet series. The behavior classification algorithm based on static images does not utilize the association information between images in the classification process, so it is not effective in the recognition of some continuous actions or behaviors.

Based on the behavior classification algorithm of the image sequence, it is necessary to extract the intermediate features such as the contour map and the bone key points from the RGB map, and then classify the extracted features. However, extracting to the contour map will lose the spatial fine-grained information accompanying with the noise interference of other visual information. What's worse, the multi-stage design will lead to slow inference speed. It is proposed in the literature that end-to-end-based feature extraction from RGB images is better. Therefore, the behavior classification algorithm based on the spatio-temporal feature fusion of the sequence diagram is designed with reference to the gait recognition algorithm GaitSet in this paper.

3. ALGORITHM DESIGN AND IMPLEMENTATION

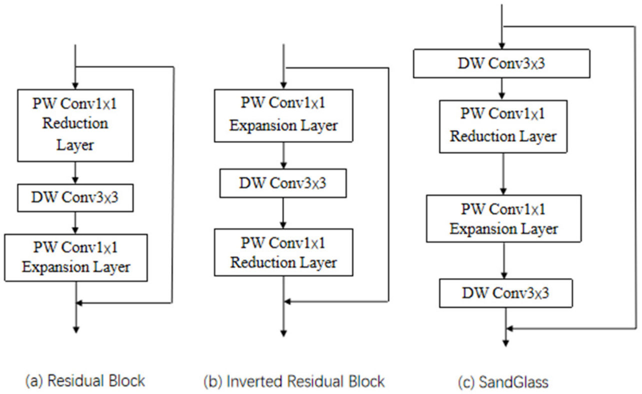
3.1 Improved target detection algorithm based on YOLOv3

3.1.1 Improved feature extraction network based on MobileNetV2

MobileNetV2 network is widely used in mobile or edge devices due to its simple structure, small number of parameters, fast inference speed. The MobileNetV2 network uses depthwise separable convolution instead of ordinary convolutions, and divides ordinary convolutional operations into deep convolutions and pointwise convolutions, which greatly reduces the amount of operations compared to ordinary convolutions.

MobileNetv2 draws on bottleneck in the ResNet[13] using PW convolutions for the purpose of flexibly changing feature dimensions. Compared with the ResidualBlock used in MobileNetv1, shown in Figure 1(a), MobileNetv2 uses the Inverted Residual structure, shown in Figure 1(b), which first expands the input feature to avoid the loss of information cause by reducing the dimension of the input then carries out the deep convolution operation and finally reduces the number of channels.

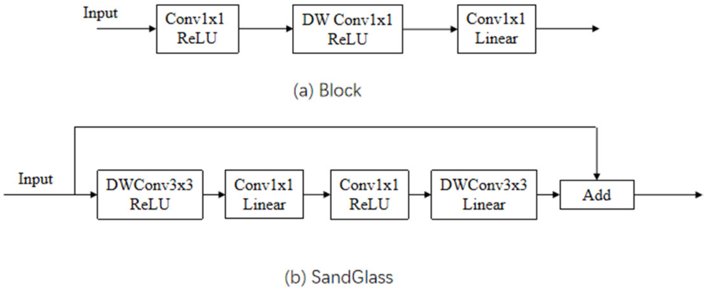
It is known that the short connection in the ResidualBlock may affect the gradient back propagation in the case of less characteristic information. Considering the limitations of ResidualBlock and Inverted Residual Block, the literature[14] presents SandGlass Block. SandGlass Block mainly reconstructs the bottleneck module based on two points: firstly, it improves the effect of gradient back propagation, so as to ensure that the residual structure has more characteristic information when it is transmitting from the bottom to the top; secondly, it adopts two DW convolutions to maintain more space information and improve classification performance. The network structure of the SandGlass is shown in Figure 1 (c).



(a) Residual Block (b) Inverted Residual Block (c) SandGlass

Figure 1 Three types of bottleneck structure

This paper replaces the MobileNetv2 bottleneck with the SandGlass module, then modifies the input dimensions of the main network to 640 × 640, and discards the average pooling layer and pointwise convolution at the end of the MobileNetV2 network. Improved MobileNetV2 network has two bottleneck modules as shown in Figure 2, one is a SandGlass module, and the other is a block combining PW convolution with DW convolution.



(a) Block

(b) SandGlass

Figure 2 Improvement of two modules of the MobileNetV2 network

Under the premise of maintaining accuracy, the network structure of MoblieNetV2 is further adjusted by reducing the number of network layers and channels, limiting the network expansion factor to 4. The adjusted network structure is shown in Table 1, where Input, Operator, t, c, n, s, conv2d respectively indicate input size, the operation type of a layer, network expansion factor, the number of convolutional kernels in a layer, the number of repetitions of the module, the convolutional step size, two-dimensional convolution.

Table 1 Adjusted network structure

Input	Operator	t	c.	n	s
$320^2\times3$	Conv2d	---	32	1	2
$160^2\times32$	Block	4	16	2	1
$160^2\times16$	SandGlass	4	32	1	2
$80^2\times32$	Block	4	32	1	1
$80^2\times32$	SandGlass	4	64	2	1
$80^2\times64$	Block	4	64	2	1
$80^2\times64$	SandGlass	4	128	1	2
$40^2\times128$	Block	4	128	2	1
$40^2\times128$	SandGlass	4	256	2	2
$20^2\times256$	Block	4	256	1	1
$20^2\times256$	SandGlass	4	--	2	--

3.1.2 SPP Spatial Pyramidal Pooling

When the input size of the feature extraction network is fixed, it is necessary to process the input image using cropping, zooming and other operations before training, which may cause image distortion and loss of feature information. Therefore, after the feature extraction, the SPP space pyramid pooling structure[15] is introduced, which can compress and fuse the information of feature maps of different sizes, and finally obtain a fixed size output. Considering the target size of the human body, the SPP network structure introduced in this paper is shown in Figure 3.

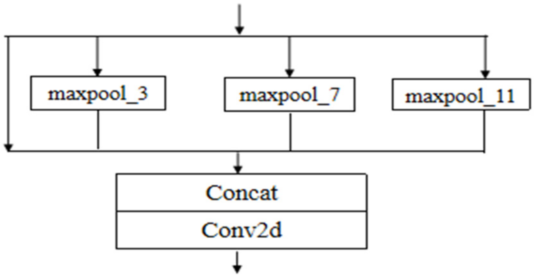


Figure 3 Pooling module of spp space pyramid

Specifically, the SPP module works as follows: first, the input feature map passes through three maximum pooling layers of dimensions 3, 7 and 11; then, the input feature map is concatenated with three pooling layer outputs via a shortcut path; finally, the feature information of four different scales is fused through a convolutional layer.

3.1.3 PANet Feature Fusion

In order to reduce the amount of network computation and improve the speed of inference, this paper uses a 2-layer feature information fusion, and reserves the path aggregation network PANet[16] for feature enhancement. PANet retains two layers with input sizes of 20×20, and 80×80 respectively. The two feature fusion paths, bottom to top and top to bottom, are adopted, so that the multi-scale features are repeatedly fused and mutually reinforced.

3.1.4 Output head decoupling

The YOLO algorithm's original detection head remains coupled to output classification, regression and score simultaneously. By decoupling the detection head, the conflicts between the classification task and the regression task in the target detection can be eliminated. As shown in Figure 4, the original detection head is adjusted by referring to the YOLOX algorithm[17],.The prediction branch is decoupled, which improves the convergence speed. It contains a 1x1 Conv layer to reduce the dimensionality of the channel, after which parallel branches of two 3x3 Conv layers are added separately for classification and regression tasks, respectively.

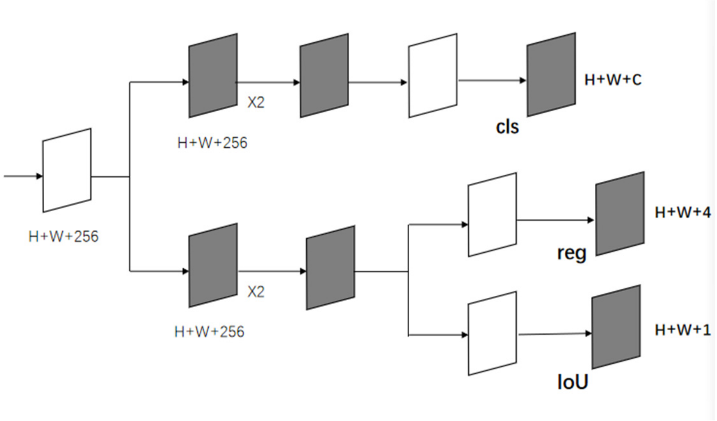


Figure 4 Adjusted decoupling head

3.1.5 Improved human detection network

Combining the above all improvements, the Lightweight YOLOv3 network structure also named YOLOv3-SE is shown in Figure 5.

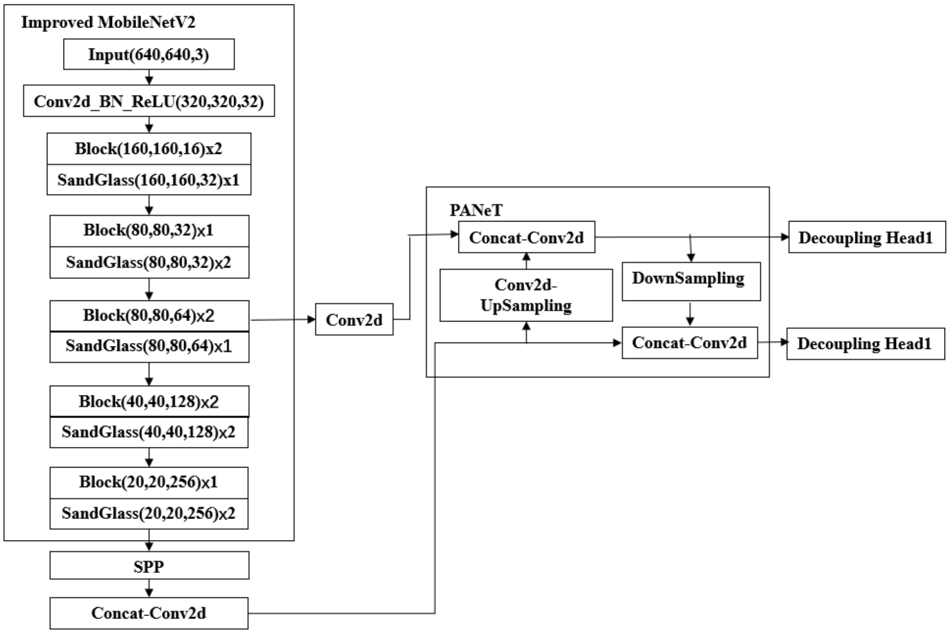


Figure 5 Improved target detection network YOLOv3-SE structure

3.2 Tracking algorithms based on IOU and similarity

The core step of the IOUTracker target tracking algorithm is to find the detection box with the maximum IOU for each active track in the current frame, and determine whether it matches according to a set threshold (the experimental set threshold is 0.25). Unmatched detection boxes are treated as new active tracks.

In order to improve the stability of the tracking algorithm, considering that the detection boxes of a same person obtained by the target detection network has relatively stable aspect ratio and confidence, this paper defines the similarity of different detection boxes based on aspect ratio and the confidence as Sim as shown in Equation 1:

$$Sim = 1 - 0.5abs\frac{R_a + R_b}{R_a R_b} + 0.5abs(C_a - C_b) \tag{1}$$

Among them, R_a、$R_b$represent the aspect ratio of the detection boxes, and $C_a - C_b$ represent the difference of the confidence of the detection boxed, Sim reflects the similarity of the two detection boxes. The closer the Sim is to 1, the more similar the two detection boxes are, and the more likely they belong to the same target.

After the IOU matching, the detection box with the largest Sim to unmatched boxes is found in the current frame and a threshold is set to judge whether it matches or not. The finally unmatched detection box is regarded as a new active track. The structure diagram of the tracking algorithm based on IOU and similarity is shown in Figure 6.

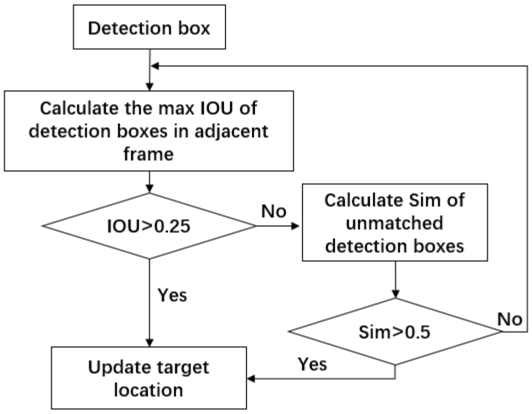

Detection box
Calculate the max IOU of detection boxes in adjacent frame
IOU>0.25
No
Calculate Sim of unmatched detection boxes
Yes
Sim>0.5
No
Yes
Update target location

Figure 6 Structural diagram of the tracking algorithm based on IOU and similarity

3.3 Behavior classification algorithm based on spatio-temporal feature fusion in sequence diagrams

The main difference between current behavior classification algorithms based on video sequences is the different ways of fusing spatio-temporal information but all of them have problem when it comes to practical applications. The dual-flow network structure divides the information into time domain and space domain, but the time domain convolution has the risk of losing important information for long video. The 3D convolutional network has too many parameters and the activation function of CNN-LSTM, which adopts a hybrid network structure, will cause gradient disappearance and gradient explosion. This paper taking the network inference speed and accuracy into account, and referring to the GaitSet algorithm, designs a fast classification algorithm based on the fusion of spatio-temporal features of sequence diagrams.

In order to improve the classification speed, the network structure is adjusted according to the GaitSet algorithm. The input is 30 consecutive human body sequence images with the fixed size of 64 × 64, followed by an 8-layer convolutional neural network to extract the features. Then the features are fused in the time domain firstly,that is to concatenate the features extracted from each sequence diagrams through maximum pooling. The spatial feature fusion superimposes the feature maps of each channel into a value through average pooling and maximum pooling. At last all channel features are spliced into a vector containing the information of time and space domain.Finally, using a fully-connected network complete classification.

The algorithmic structure designed in this paper is shown in Figure 7, noted as Action-Classifier. In the figure, N denotes the batch size; S denotes the number of frames in the sequence; and C denotes the number of channels.

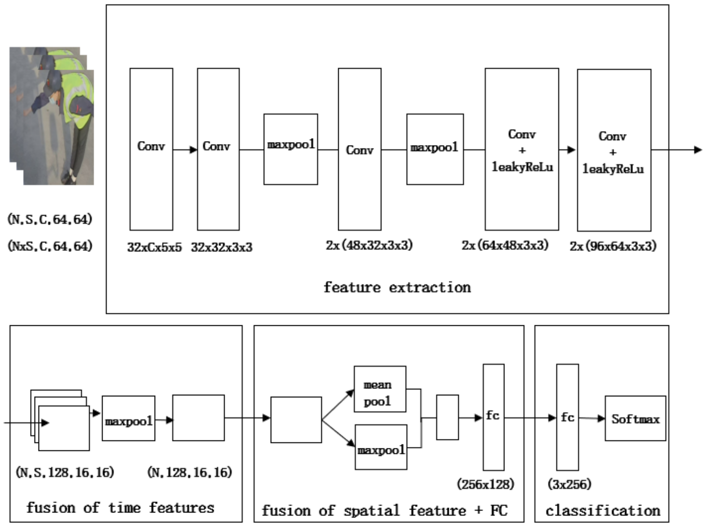


Figure 7 Action-Classifier Algorithm Structure Diagram for Behavior Classification

4. EXPERIMENTAL RESULTS AND ANALYSIS

4.1 Experimental Preparation

4.1.1 Introduction to the Experimental Data Set

The human body dataset used in this paper include live scene images and network images, and the number of images collected is 4,560. To enhance the generalization capability of the network, 4,500 images containing human body are selected from the Coco dataset. So the total dataset has 9060 images, where the training set, validation set, and test set are divided at a ratio of 7: 1: 2.

The behavior recognition algorithm in this paper is to classify the two actions, raising arms and bending down. In this paper, 1220 video samples are produced through collecting video in application scenes or online, and the simulation of raising arms and bending down. In order to enhance the generalization ability of the network, 650 video samples including raising arms and bending down are selected from behavior recognition databases such as BBC Pose and Human3.6M. A total of 1870 videos are included in the dataset, where the training set, validation set, and test set are divided at a ratio of 7: 1: 2.

4.1.2 Experimental platform

In this paper, the algorithm chooses a local PC as the experimental platform, and the PC configuration is shown in Table 2.

Table 2 Local PC Configuration Table

Modules	Parameters
Processor	Intel Core i5-7500 CPU at 3.4 GHz
GPU	NVIDIA GeForce GTX1080Ti @ 12GB
Memory	16GB
Deep Learning Framework	Darknet

The Huawei Hi3559A is chosen as the edge device for the actual deployment in this experiment, which has the characteristics of fast speed,high efficiency and low energy consumption. The parameters of Hi3559A are shown in Table 3.

Table 3 Hi3559A Configuration Table

Modules	Parameters
Operating System	Linux
Processor	Hi3559A 4-Core A53 @ 1.8GHz
Memory	4GB
Inference engine	FPU

4.2 Validation of experimental results

4.2.1 Validation of the results of the human detection network YOLOv3-SE

To verify the detection effect of the YOLOv3-SE algorithm, the algorithm and YOLOv3 are retrained using the same hardware conditions and datasets, respectively. The accuracy, inference time and model size of various algorithms are shown in Table 4.

Table 4 Comparison of results between improved network and YOLOv3 YOLOv4

Algorithm Name	FPS	Model Size	MAP
YOLOv3	81	232.6	95.64%
YOLOv4	74	245.8	96.75%
YOLOv3-SE	152	38.5	94.21%

Experiments show that the YOLOv3-SE algorithm, at a sacrifice of 1.43% accuracy, improves the running speed by 87.65%, and the frame rate can reache 152FPS, with the model size reduced to 1/6.

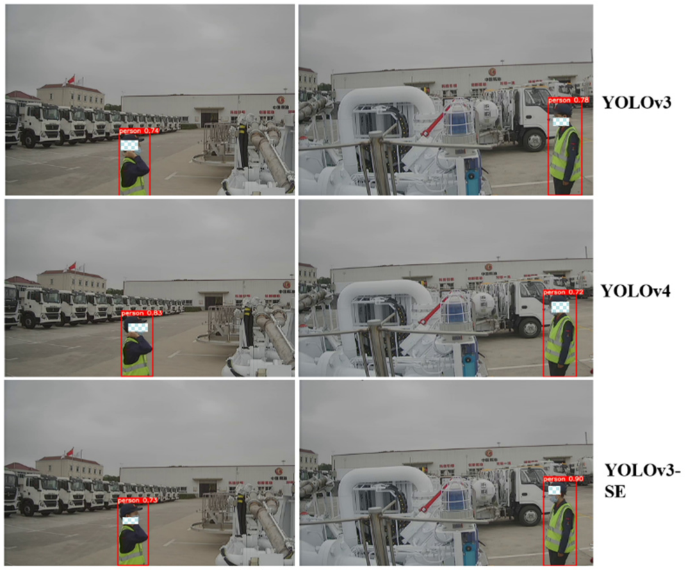


Figure 8 GPU Test Results Diagram

4.2.2 Hi3559A Terminal Test Results

The second experiment is to run the YOLOv3-SE algorithm on PC (GPU) and Hi3559A terminals respectively, to compare the detection speed, power consumption, and accuracy. The result is shown in Table 5.

Table 5 YOLOv3-SE network test results in different environments.

Experiment Type	FPS	MAP	Power consumption/W
GPU/YOLOv3-SE	79	95.69%	65
Hi3559A/YOLOv3-SE	42	94.36%	7.9

When the YOLOv3-SE algorithm is deployed to Hi3559A, the accuracy is reduced by 1.33% due to the loss of model conversion, but the FPS reaches 42, which is sufficient for real-time target detection. In terms of power consumption, Hi3559A is approximately the GPU's1/8. The detection result images on Hi3559A are shown in Figure 9:

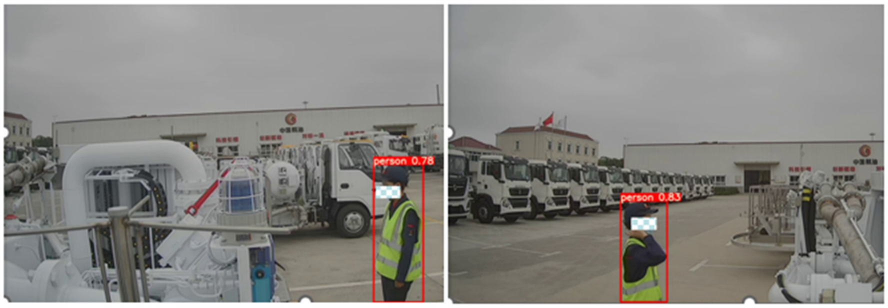


Figure 9 Hi3559A Test Result Diagram

4.2.3 Action-Classifier and comparison with the adjusted GaitSet*

Since the GaitSet algorithm is designed for gait recognition, its output structure needs to be adjusted when used for behavior classification. In order to verify the effectiveness of the behavior recognition algorithm designed in this paper, the action-classifier of this algorithm and the adjusted GaitSet* are retrained using the same hardware conditions and datasets, respectively. 30 frames with the size of 64×64 human sequence images are input every single time to compare the accuracy, model size and inference time of the algorithm. Specific experimental data are shown in Table 6.

Table 6 Comparison Results of Action-Classifier Algorithm and Adjusted GaitSet * Algorithm

Algorithm Name	Time-consuming reasoning	Model Size	MAP
GaitSet*	18ms	6.89	95.84%
Action-Classifier	14ms	4.85	95.47%

Experiments show that the Action-Classifier algorithm and the adjusted GaitSet algorithm are basically the same in accuracy, but the Action-Classifier's running speed is increased by 28% and the frame rate reaches 72FPS with the model size decreases by 29.6% compared to the original model.

4.3 Sequence-based behavior recognition algorithm result validation

The sequence-based lightweight behavior recognition algorithm presented herein is deployed to Hi3559A for testing. The frame rate of the algorithm can reach 22FPS, and the detection results are shown in Table 7.

Table 7 Experimental results of behavior recognition algorithm

Video Type	Identify correctly	Identify errors
Raising arms	111	9
Bending down	105	5
Miscellaneous	87	3

In this paper, the recognition accuracy of all the video clips is chosen as the evaluation index. From the data in the table, it can be seen that the recognition accuracy rate can reach 94.68%. The result of the algorithm is shown in Figure 10.

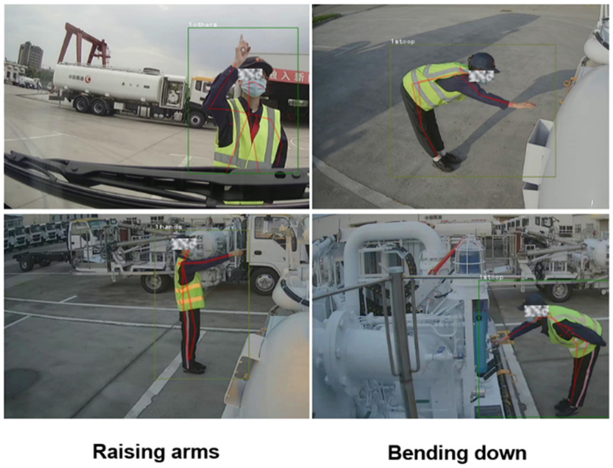


Figure 10 Behavioral Algorithm Test Result Diagram

5. CONCLUSION

This paper proposes a sequence-based lightweight algorithm for refueling behaviour recognition. Experiments on GPUs show that the YOLOv3-SE algorithm has improved the speed by 87.65%, with the model size reduced to 1/6 at the loss of 1.43% detection accuracy.The behavior classification algorithm achieves an inference speed at 14ms which is improved by 28%. Deploying the algorithm to the embedded device Hi3559A, the recognition accuracy rate could reach 94.68% with the inference speed at 22FPS, which can achieve the demand of real-time processing and have great application prospects.

REFERENCES

[1] Pei Lishen., Liu Shaobo., Zhao Xuezhuan. Review of Human Behavior Recognition Research[J]. Journal of Frontiers of Computer Science and Technology, 2022, 16(2):305-322.

[2] Li Menghe, Xu Hongji, Shi Leixin. Multi-person Activity Recognition Based On Bone Keypoints Detection[J]. Computer Science,2021.

[3] Redmon J, Farhadi A. YOLOv3: An Incremental Improvement[J]. arXiv e-prints, 2018.

[4] Chao H, He Y, Zhang J, et al. GaitSet: Regarding Gait as a Set for Cross-View Gait Recognition[J]. 2018.

[5] Ji Xiongwu, Zhang Yonghui, Zhang Jian. Action Recognition Algorithm Based on Optial Flow and Depth Motion Map[J]. Natural Science Journal of Hainan University, 2020, 38(2):8.

[6] Zaidi S, Ansari M S, Aslam A, et al. A Survey of Modern Deep Learning based Object Detection Models[J]. 2021.

[7] Girshick R, Donahue J,D arrell T,et al. Rich Feature Hierarchies for Accurate Object Detection and Semantic Segmentation[J]. IEEE Computer Society, 2013.

[8] Purkait P, Zhao C, Zach C. SPP-Net: Deep Absolute Pose Regression with Synthetic Views[J]. 2017.

[9] Redmon J, Divvala S, Girshick R, et al. You only look once: Unified, real-time object detection[C]//Proceedings of the IEEE conference on computer vision and pattern recognition, 2016: 779-788.

[10] Lei Lei, Tao Qingchuan. Off-position Detection Algorithm Based on Lightweight YOLOv4[J]. Modern Computer, 2022.

[11] SANDLER M, HOWARD A, ZHU M, et al. Mobilenetv2: Inverted residuals and linear bottlenecks [C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE, 2018:4510-4520.

[12] Bochinski E, Eiselein V, Sikora T. High-speed tracking-by-detection without using image information[C]//2017 14th IEEE international conference on advanced video and signal based surveillance (AVSS). IEEE, 2017: 1-6.

[13] He K, Zhang X, Ren S, et al. Deep residual learning for image recognition[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. 2016: 770-778.

[14] Zhou D, Hou Q, Chen Y, et al. Rethinking bottleneck structure for efficient mobile network design[C]//Computer Vision-ECCV 2020: 16th European Conference, Glasgow, UK, August 23-28, 2020, Proceedings, Part III 16. Springer International Publishing, 2020: 680-697.

[15] He K, Zhang X, Ren S, et al. Spatial pyramid pooling in deep convolutional networks for visual recognition[J]. IEEE transactions on pattern analysis and machine intelligence, 2015, 37(9): 1904-1916.

[16] Liu S, Qi L, Qin H, et al. Path aggregation network for instance segmentation[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. 2018: 8759-8768.

[17] Ge Z, Liu S, Wang F , et al. YOLOX: Exceeding YOLO Series in 2021[J]. 2021.

Research On Relation Extraction Model of Overlapping Entity Based On Attention Mechanism

Ling Gan[1], Xiaobin Liu[1,a]
[1]School of Computer Science and Technology, Chongqing University of Posts and Telecommunications, Chongqing 400065, China
[a]Email: lxb_212@163.com

Abstract

Relation extraction refers to get the triple structure composed of semantic relation entity pairs from unstructured text, which is an important part of tasks such as knowledge graphs. At present, the joint extraction model is in common used to avoid the impact of overlapping entities, but there are the following problems. First, the dependencies between text words are not fully considered, and the recognition performance of entities with long spans is low. Insufficient utilization of information makes it difficult to fully extract implicit relationships. In order to address these issues, this text proposes an improved joint learning model, which builds text semantic representation through BERT pre-training, obtains relation type representation as an additional mapping through a multi-label classification method, and sequentially uses multi-layer BiLSTM combined with Highway network to obtain semantic information, and combine The attention mechanism obtains the entity location score, and the pointer network is used to obtain the entity location. The experiments of this method on the common dataset of relation extraction task is effective.

Keywords: Overlapping entity relation extraction; BiLSTM; Attention mechanism; Pointer network

1. Introduction

On text processing, enhance the understanding of sentence semantics and establish an entity semantic knowledge base, it is usually necessary to extract entity relationship triples and extract these structured content. This task belongs to information extraction and has many application in information retrieval, intelligent question answering, knowledge graph construction and other fields.

At present, the task of extracting overlapping entity relations mainly uses the joint model. Zeng et al. [1] designed CopyRE, which first extracts relationships, then extracts entities, and allows entities to participate in different triples by copying entities. Wei et al.[2] proposed the cascade binary marking framework CaseRel to enable the model to learn the mapping function between HE (Head Entity) and TE (Tail Entity) under a given relation, so as to achieve the overall modeling effect of triples. Bai et al.[3] proposed multi-layer neural network coding and combined with self-attention mechanism, designed a dual-pointer network structure to identify the starting and ending positions of entities respectively, so that complete entities could participate in the replication process. Wang et al.[4] tried a graph structure method to handle overlapping entity relationship extraction tasks.

The above methods do not fully consider relational information as additional feature information to guide joint extraction. Ma et al.[5] used a cascaded double decoding joint extraction model. By detecting the relationship type in the text, By detecting relationship types in text, BiLSTM can integrate context, predict head entities, and map them to corresponding tail entities. However, such methods generally have the problems of low performance in recognition of entities with long entity spans and insufficient consideration of the dependence between words in the text, so they are unable to extract the implied relations and corresponding entities more fine-grained.

To this end, we proposes a relation-guided attention mechanism for joint extration. There are some of the main work and contributions:

(1) The multi-layer BiLSTM structure is used to obtain the abstract semantic features and contextual semantic dependencies of sentences.

(2) Multiple attention mechanism combined with relationship type. The attention mechanism can fully consider the dependencies between each word in a more granular manner to guide entity extraction.

International Conference on Computer Graphics, Artificial Intelligence, and Data Processing (ICCAID 2022)
edited by Ruishi Liang, Jing Wang, Proc. of SPIE Vol. 12604, 126040S
© 2023 SPIE · 0277-786X · doi: 10.1117/12.2674559

(3) The Highway network connection can alleviate the gradient vanishing problem of deeper model. Mainly to mitigate the error caused by gradient disappearance, each layer of BiLSTM can be connected by highway network.

Finally, we carried out experimental verification on common data sets of the task, and all of them have been improved accordingly.

2. Model

First of all, BERT-based pre-training is used for text encoding, Then obtain the relation type by setting the threshold value through the corresponding activation function, the BiLSTM-Highway network layer is used for feature extraction, and entity locations are obtained through a fusion of attention score and pointer network. Fig 1 shows the method framework of this article.

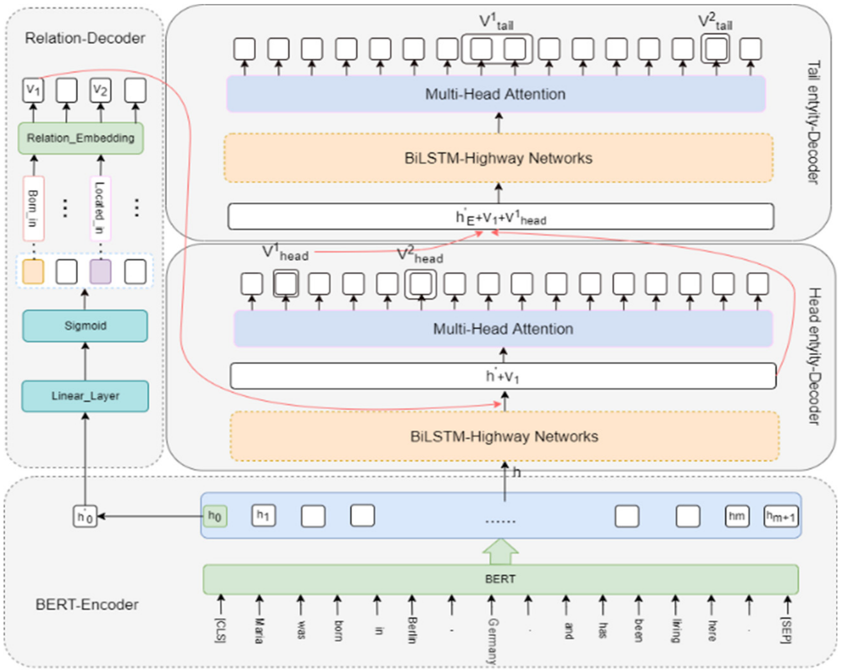


Fig 1 The framework of our model

2.1. BERT Encoder

In this paper, BERT is used to encode the text to get the complete semantic representation of sentence containing n words, BERT concatenates tokens, as shown in Eq.1:

$$x = [c, x_1, x_2, \dots, x_n, s] \tag{1}$$

Where x_i represents the i-th token representation in the sentence. c represents the classifier CLS of the sentence in BERT, and s represents the separator SEP. Therefore, n+2 token representations will be generated after BERT, as shown in Eq.2:

$$h = [h_0, h_1, h_2, \dots, h_n, h_{n+1}] \tag{2}$$

Where h_i represents the last hidden layer representation of BERT, where 0≤ i ≤n+1, h_0, h_{n+1} are denote c and s, respectively.

h_0 obtained through BERT is the input of the relation decoder layer, which is used to generate the representation vector as the relational decoding layer, as shown in Eq.3:

$$h_0' = Pool(h_0) \tag{3}$$

2.2. Relation Decoder

Different sentences have different types of relations, and the detection of candidate relation types in the text can be regarded as a Multilabel classification task.

First define a set R for relation, the BERT pooling process CLS output h_0' as the input of relation decoding, through the linear layer, uses the sigmoid φ is used to calculate the probability of relation categories existing in the sentence, as shown in Eq.4:

$$P_i^r = \varphi(W^r \cdot h_0' + b^r) \tag{4}$$

Where $W^r \epsilon \mathbb{R}^{j \times d}$, $b^r \epsilon \mathbb{R}^j$, d represents the BERT's last hidden layer dimension, and j is the number of relation types.

For the set $(r_{i_1}, \dots, r_{i_g})$ of existing relationship types can be obtained from the text through the relation decoder, g represents the number of relation types of sentences, and we can get the corresponding relation representation by using the lookup table. Similarly, the predefined relation set $R = \{r_1, \dots, r_j\}$ can be encoded as $V_j = [V_1, \dots, V_j]$.

Given a text representation x, relation decoding detects relation r by optimizing the following probabilities, as shown in Eq.5:

$$P_\theta(r|x) = \prod_{i=1}^{K} (P_i^r)^{y_i^r} (1 - P_i^r)^{1-y_i^r} \tag{5}$$

Where y_i^r represents the true label of the ith relation.

2.3. Head-entity Decoder

After BERT obtains the word representation, this paper proposes to use two-layer stacked BiLSTM to extract semantic feature representations of different granularities, and obtain the hidden state representation of each token, as shown in Eq.6:

$$\mathrm{h_i^{H_{sta}}} = \mathrm{BiLSTM([h_i])} \tag{6}$$

For the negative impact of gradient disappearance after neural network stacked, the Highway network of the adaptive gating unit is used to adjust the information flow, and the Transform gate and the carry gate are mainly added to the output of the BiLSTM layer. A typical neural network is an affine transformation plus a nonlinear function, that is, $y = H(h_i, W_H)$, where x represents the network input and W_H represents the network weight. Its definition is shown in Eq.7:

$$\mathrm{y = H}(h_i, \mathrm{W_H}) * \mathrm{Tg}(h_i, \mathrm{W_T}) + h_i * \mathrm{Cg}(h_i, \mathrm{W_C}) \tag{7}$$

Where $Tg(h_i, W_T)$ represents Transform gate Tg, and $Cg(h_i, W_c)$ represents Carry gate Cg.

Each token in a sentence has different scores in terms of importance in identifying entities under different relationship types. Therefore, this paper, attention mechanism is used to fuse relationship type vectors to obtain the importance scores of each word under different relationship types, as shown in Eq.8-10:

$$\mathrm{e_{ij}} = \mathrm{h_i^{H_{sta}} tanh\,(W_r V_j + W_h h_i^{H_{sta}})} \tag{8}$$

$$a_{ij} = softmax(e_{ij}) \tag{9}$$

$$c_j = \sum_{i=1}^{l} a_{ij} h_i^{H_{sta}} \tag{10}$$

Finally, the softmax function σ computes the probability that each word represents a label that starts as a head entity. The probability calculation is shown in Eq 11:

$$P_i^{sta} = \sigma(W^{h_i} h_i^{H_{sta}} + b^{H_{sta}}) \tag{11}$$

A similar operation, marking the end of the header entity, is shown in Eq.12-15:

$$\mathrm{h_i^{H_{end}}} = \mathrm{BiLSTM([h_i^{H_{sta}}; p_i^{se}])} \tag{12}$$

$$h_i^{H_{end}} = Highway\left(h_i^{H_{end}}\right) \tag{13}$$

$$h_i^{H_{end}} = selfAttention(\mathrm{h_i^{H_{end}}}) \tag{14}$$

$$\mathrm{p_i^{H_{end}}} = \sigma\left(\mathrm{W^{H_{end}} \cdot h_i^{H_{end}} + b^{H_{end}}}\right) \tag{15}$$

Finally, the head entity extractor extracts the position span of the head entity by optimizing this probability, as shown in Eq.16:

$$\mathrm{P_\theta(h|r,x)} = \prod_{\mathrm{i=1}}^{\mathrm{m}} \mathrm{(P_i^{H_{sta}})^{y_i^{H_{sta}}} (1 - P_i^{H_{end}})^{1-y_i^{H_{end}}}} \tag{16}$$

The length of the sentence is m, $y_i^{H_{sta}}$ *and* $y_i^{H_{end}}$ are the labels at the beginning and end of the actual head of the i-th word.

2.4. Tail-entity Decoder

After obtaining the head entity start and end positions, the entity representation is obtained by averaging the positions, as shown in Eq.17:

$$V_{head} = avg(\mathrm{h}_\mathrm{i}^{\mathrm{H}_\mathrm{sta}}, \mathrm{h}_\mathrm{i}^{\mathrm{H}_\mathrm{end}}) \tag{17}$$

By fusing By fusing the entity representation and the head entity decoder input, it is used as the input of the tail entity decoder into the BiLSTM-Highway network. A similar operation marks the end position of the tail entity.

Finally, the head entity extractor extracts the location span of the head entity by optimizing this probability, as shown in Eq.18:

$$\mathrm{P}_\theta(\mathrm{t}|\mathrm{r}, \mathrm{h}, \mathrm{x}) = \prod_{\mathrm{i}=1}^{\mathrm{m}} (\mathrm{P}_\mathrm{i}^{\mathrm{T}_\mathrm{sta}})^{\mathrm{y}_\mathrm{i}^{\mathrm{T}_\mathrm{sta}}} (1 - \mathrm{P}_\mathrm{i}^{\mathrm{T}_\mathrm{end}})^{1-\mathrm{y}_\mathrm{i}^{\mathrm{T}_\mathrm{end}}} \tag{18}$$

Where $\mathrm{y}_\mathrm{i}^{\mathrm{T}_\mathrm{sta}}$, $\mathrm{y}_\mathrm{i}^{\mathrm{T}_\mathrm{end}}$ are the labels at the beginning and end of the actual tail of the i-th word.

2.5. Joint Training

For a better joint training model, this paper defines the loss function as:

$$\mathcal{L} = -\underset{(h,r,t),x}{\mathbb{E}}\{\log p_\theta(r|x) + \log p_\theta(h|r,x) + \log p_\theta(t|r,h,x)\} \tag{19}$$

Where $P_\theta(r|x)$, $\mathrm{P}_\theta(\mathrm{h}|\mathrm{r}, \mathrm{x})$, $\mathrm{P}_\theta(\mathrm{t}|\mathrm{r}, \mathrm{h}, \mathrm{x})$ can be calculated from in Eq.(5), Eq.(16), Eq.(18),respectively.

3. Experimental results and analysis

3.1. Experimental environment and model parameters

This paper is based on PyTorch deep learning framework and uses the hardware environment NVIDIA TESLA V100 GPU-32GB to train the model, and the English Bert-Base-cased pre-training model is used as the text encoder. For the relation decoder, this paper uses the 300-dimensional Glove. 840B as the relation vector representation, and the hyperparameters are shown in Table 1.

Table 1 Hyper-parameter setting

Hyper-parameter	Value
lr	2e-5
Heads_nums	2
Drop_out	0.4
Position_emb_dim	20
Word_emb_dim	300
Tokens_emb_dim	768

3.2. Datasets and Evaluation Metrics

The experiments are performed on the NYT dataset and the WebNLG dataset. The text in the NYT dataset comes from the corpus annotated by the New York Times; the WebNLG dataset was originally constructed for natural language generation tasks using triples in DBPedia. The detailed data volume and number of relation types of the two datasets are shown in Table 2.

Table 2 Data volume of commonly used datasets

Category	NYT		WebNLG	
	Train	Test	Train	Test
#type				
Normal	37013	3266	1596	246
EPO	9782	978	227	26
SEO	14735	1297	3406	457
relations	24		246	
All	56195	5000	5019	703

The datasets can be classified into three types by different methods of entity overlap. Normal, there are no shared entities in the sentence; EntityPairOverlap (EPO), with more than two triples sharing an entity pair; SingleEntityOverlap (SEO), refers to more than two triples in a sentence sharing the same entity.

This paper conducts a comprehensive experimental evaluation from two aspects:

(1) Partial Match[6]: In the triple form (Head, Relation, Tail), the last word of the entity is compared when the relation prediction is correct, which is considered correct if it matches.

(2) Exact Match [7]: In the extracted triple form, an entity is considered correct if the relationship is correct and the full name of the entity is correct.

For the above two aspects, the experimental results will be evaluated from three aspects: precision (Pre), recall (Rec) and F1 value (F1).

3.3. Experimental results and analysis

This paper conducts experiments on the NYT dataset and the WebNLG dataset. The results of the proposed model are compared with the model described in Section 4.3. The specific results are shown in Table 3.

Table 3 Results on triple extraction under Partial Match and Exact Match.

Method	**Partial Match**						**Exact Match**					
	NYT(N)			WebNLG(W)			NYT(N)			WebNLG(W)		
	Pre	Rec	F1	Pre	Rec	F1	Pre	Rec	F1	Pre	Rec	F1
CopyRe[1]	61.0	56.6	58.7	37.7	36.4	37.1	-	-	-	-	-	-
CopyRL[8]	77.9	67.2	72.1	63.3	59.9	61.6	-	-	-	-	-	-
WDec[7]	-	-	-	-	-	-	75.7	68.7	72.0	58.0	54.9	56.4
AttentionRE[9]	-	-	-	-	-	-	88.1	78.5	83.0	89.5	86.0	87.7
CasRel[2]	89.7	89.5	89.6	**93.4**	90.1	91.8	89.1	89.4	89.2	87.7	85.0	86.3
DualDec[5]	90.2	90.9	90.5	90.3	91.5	90.9	89.9	**91.4**	90.6	88.0	**88.9**	88.4
Ours	**90.8**	**91.3**	**91.0**	91.5	**92.6**	**92.1**	**91.0**	91.1	**91.1**	**90.2**	88.4	**89.1**

The data in Table 3 can prove that the method in this paper has achieved corresponding improvement in both partial matching and accurate matching. In terms of partial matching, in the NYT and WebNLG datasets, each evaluation index has a certain improvement compared to the comparison model, and the main evaluation index F1 value reached 91.0% and 92.6%, respectively. The value has been improved. Its F1 value has reached 91.1% and 89.1%, respectively, indicating that the proposed method has certain stability in multi-label relation type classification, entity recognition ability and performance.

In order to further explore the robustness of the method in this paper, we used three entity overlapping methods to conduct accurate matching comparison experiments on the dataset, as shown in Fig 2, Fig 3, and Fig 4.

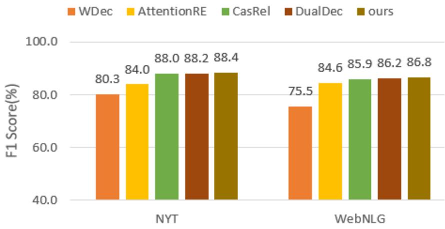


Fig 2. Normal

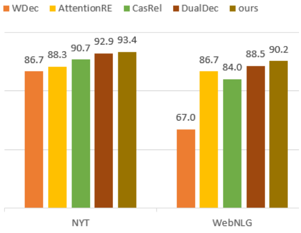


Fig 3. EPO

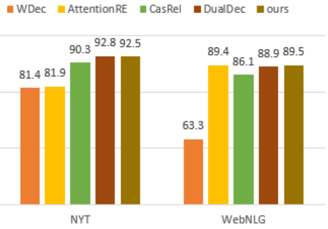


Fig 4. SEO

By comparing Fig 2, Fig 3, and Fig 4, the method in this paper extracts entities under the premise that the relationship type is guided by prior knowledge, and has a good result in processing entity overlapping tasks, whether it is a simple Normal-type or the more difficult EPO and SEO types, all demonstrate the potential of our model in solving overlapping triple problems.

4. Concluding remarks

In this paper, the overlapping entity relationship is extracted based on joint extraction model. Through multi-label classification, the Relation types of sentences are extracted first, and then the entities are extracted to form the final < head, Relation, tail > triplet form.

The experiment in this paper is carried out on the public datasets NYT and WebNLG, and the proposed method is 0.5% and 1.2% higher than that of DualDec in the comprehensive evaluation index F1 value of partial matching in the overlapping entity relation extraction task; 0.5%, 0.7%. Future work can try to classify multi-label tasks, explore the accuracy of relation type extraction, and improve the performance of the overall model.

References

[1] Zeng X, Zeng D, He S, et al. Extracting relational facts by an end-to-end neural model with copy mechanism[C]//Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers). 2018: 506-514.

[2] Wei Z, Su J, Wang Y, et al. A Novel Cascade Binary Tagging Framework for Relational Triple Extraction[C]//Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics. 2020: 1476-1488.

[3] Bai C., Pan L., Luo S., et al. Joint extraction of entities and relations by a novel end-to-end model with a double-pointer module[J]. Neurocomputing, 2020, 377: 325-333.

[4] Wang S., Zhang Y., Che W., et al. Joint extraction of entities and relations based on a novel graph scheme[C]//IJCAI. 2018: 4461-4467.

[5] Ma L., Ren H., Zhang X. Effective Cascade Dual-Decoder Model for Joint Entity and Relation Extraction[J]. arXiv preprint arXiv:2106.14163, 2021.

[6] Zeng X, Zeng D, He S, et al. Extracting relational facts by an end-to-end neural model with copy mechanism[C]//Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers). 2018: 506-514.

[7] Nayak T, Ng H T. Effective modeling of encoder-decoder architecture for joint entity and relation extraction[C]//Proceedings of the AAAI conference on artificial intelligence. 2020, 34(05): 8528-8535.

[8] Zeng X, He S, Zeng D, et al. Learning the extraction order of multiple relational facts in a sentence with reinforcement learning[C]//Proceedings of the 2019 conference on empirical methods in natural language processing and the 9th international joint conference on natural language processing (EMNLP-IJCNLP). 2019: 367-377.

[9] Liu J, Chen S, Wang B, et al. Attention as relation: learning supervised multi-head self-attention for relation extraction[C]//Proceedings of the Twenty-Ninth International Conference on International Joint Conferences on Artificial Intelligence. 2021: 3787-3793.

Finite Element Analysis of the Golden Gate Bridge

Zixuan Wang[1], Runchen Zhu[2, *], Junyi Gao[3], and Junran Jiang[4]
[1]University of California, Davis Davis 95616, CA, USA
[2]University at Buffalo, Buffalo14228, NY, USA
[3]Royal Russell School
[4]Wuhan Britain-China School, WuHan, 430030, China
*Corresponding author email: runchenz@buffalo.edu

ABSTRACT

As a miracle of bridge engineering, it is meaningful to analyze the stability of the Golden Gate Bridge. This paper focuses on the finite element method to analyze the Golden Gate Bridge's performance on loading. In this work, ANSYS is used to build a simplified model of the Golden Gate Bridge without cables. There are six modes of the deformation of the bridge, and there are six different frequencies. The natural frequency of the bridge is 0.055Hz. Because this simplified model does not have cables to reduce the deformation, the average frequency is 0.07737 Hz which is 0.02237 Hz higher than the natural frequency. Moreover, after analyzing the harmonic response, the largest equivalent stress is 1.5487e^7 Pa at joints of the bridge when the frequency is 0.0268 Hz and the sweep phase is -56.649 degrees. As a result, the analysis of the bridge based on the models in Ansys clearly shows the stability of the bridge.

Keywords: the Golden Gate Bridge, finite element analysis, vibration, Harmonic response

1. INTRODUCTION

The Golden Gate Bridge is a cross-sea bridge connecting downtown San Francisco and Marin County north of the United States. It is located over the Golden Gate Strait and is the city's main symbol of San Francisco. The bridge is 2737 meters long and 27 meters wide. The height of each bridge tower is 342 meters, the part of the tower over the water surface is 228 meters, and the weight is 24,500 tons[1](Golden Gate Bridge Research Library). According to the PBS website, the south tower of the Golden Gate Bridge was built over a thousand feet into open ocean[2]. "It took much more than a clever engineer and a relentless promoter to build the Golden Gate Bridge." said in the article *Men Who Built the Bridge*[3]. Also, many parties initially opposed the bridge. Bank of America issued a bond, with the cooperation of Joseph Strauss and the workers, designed the bridge that was considered impossible at the time. After its construction, the Golden Gate Bridge became one of the largest single-hole Suspension Bridges in the world. At the same time, it also has the fourth highest bridge tower in the world, a miracle of modern bridge engineering.

It is important for people to analyze the Golden Gate Bridge's structure and behavior of stability because a large number of loads are caused by heavy traffic every day. According to Nakasone et al.'s article, ANSYS is a software that is a general-purpose, finite-element modeling package, and it is helpful for solving structural analysis[4](Nakasone et al., 2006). Therefore, we use ANSYS as the main software in order to execute the finite element analysis of the Golden Gate Bridge's model. Afterward, the bridge's frequencies, vibration, and loading behaviors are found and demonstrated as graphs. Therefore, in this research, we focused on the structural analysis of the Golden Gate Bridge. The finite element analysis data may be useful for the future bridge engineers, because it is important to know the vibration of the bridge itself.

2. METHODOLOGY

Based on the Golden Gate Bridge Research Library data, its total length is 2737m, and its width is 27m[1]. The model of the bridge is made based on the data from the Library. We chose the two bottom surfaces and two side surfaces as our fixed supports. Our group uses ANSYS as our modeling application. It is a simplified version of the bridge. We ignored the cables and other details of the bridge to make our calculation easier and avoid fatal errors that may influence our final result.

International Conference on Computer Graphics, Artificial Intelligence, and Data Processing (ICCAID 2022)
edited by Ruishi Liang, Jing Wang, Proc. of SPIE Vol. 12604, 126040T
© 2023 SPIE · 0277-786X · doi: 10.1117/12.2674963

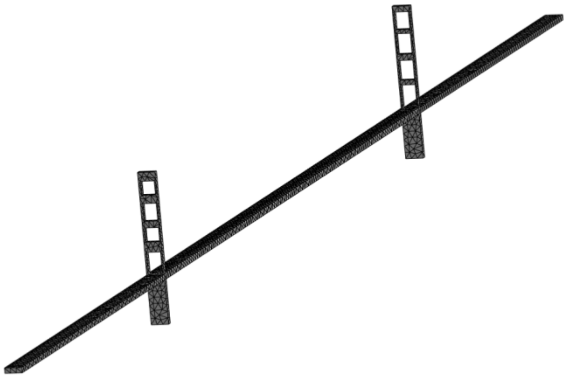
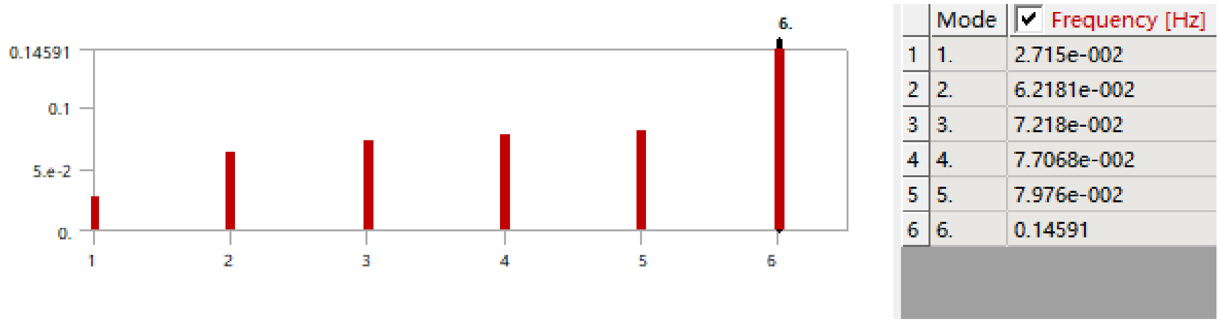


Figure 1. Bridge model with meshes

Then, our group uses mesh independence analysis, which means we put meshes on our model to make our result more accurate. Meshes can separate our model into many different elements. On each element, the computer will calculate based on the functions. The more meshes we have, the more accuracy we will get. A certain mesh density will cause a certain result, so we chose several models with different mesh densities to verify the accuracy of the result. The data on which our model is based is given below. The element size is 6m, 8m, 10m, and 12m. We use the minimum frequency of the largest size minus the minimum frequency of the minimum size and use that value divided by the minimum frequency of the minimum size in order to calculate the error.

$$(0.027179-0.027036)/0.027036 \approx 0.005289 = 0.5289\%$$

This error is acceptable.



	Mode	☑ Frequency [Hz]
1	1.	2.715e-002
2	2.	6.2181e-002
3	3.	7.218e-002
4	4.	7.7068e-002
5	5.	7.976e-002
6	6.	0.14591

Figure 2. The model analysis result with element size of 10m

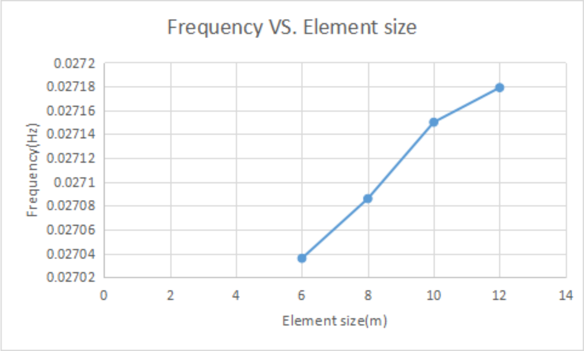


Figure 3. The relationship between mesh density and accuracy

Live load is one of the most important parts when analyzing the load capacity of a bridge. The Live load capacity per linear foot of the Golden Gate Bridge is 4,000 lbs[1](Golden Gate Bridge Research Library, 2012), which means each foot of the bridge can only hold a weight of 4000 lbs or 1814.4 kg simultaneously. Most of the live load is from the cars passing over the bridge every day. In Thomas' article, they used information from Australian standard AS 5100,2, in which the maximum weight of each vehicle is assumed to be 360 kg. The reason why AS 5100,2 made this assumption is that the bridge must be built to avoid accidents. Extreme cases need to be considered in order to make the bridge safe. According to Thomas Game et al.(2016), the density of the load is 6kN/m for all six lanes of the bridge so that the Golden Gate Bridge withstands a 222 Pa stress[5]. The stress in the research is equal to the total load divided by the total surface area.

The harmonic response model is made based on the live load, frequency, and damping ratio. For frequency, values from figure 2 will be used. According to Wikipedia, the damping ratio is a system parameter, denoted by ζ (zeta), that can vary from undamped (ζ = 0), underdamped (ζ < 1) through critically damped (ζ = 1) to overdamped (ζ > 1)(2022)[6]. Thus, for the damping ratio, we randomly chose 0.02, because the bridge is underdamped.

3. RESULT & DISCUSSION

3.1 Mode 1 & 2

Mode 1 shows deformation, and the frequency is 0.02715 Hz. Figure 4 shows three parts of the bridge, and the deformation appears in the middle of part 2. This model also shows the degree of deformation. The center of part 2 is red, which means this part has maximum deformation.

Mode 2 also shows the bridge's deformation, and the frequency is 0.062181 Hz. According to figure 5, the deformation of the bridge is also at part 2 of the bridge. However, the maximum deformation part moves towards part 3 of the bridge.

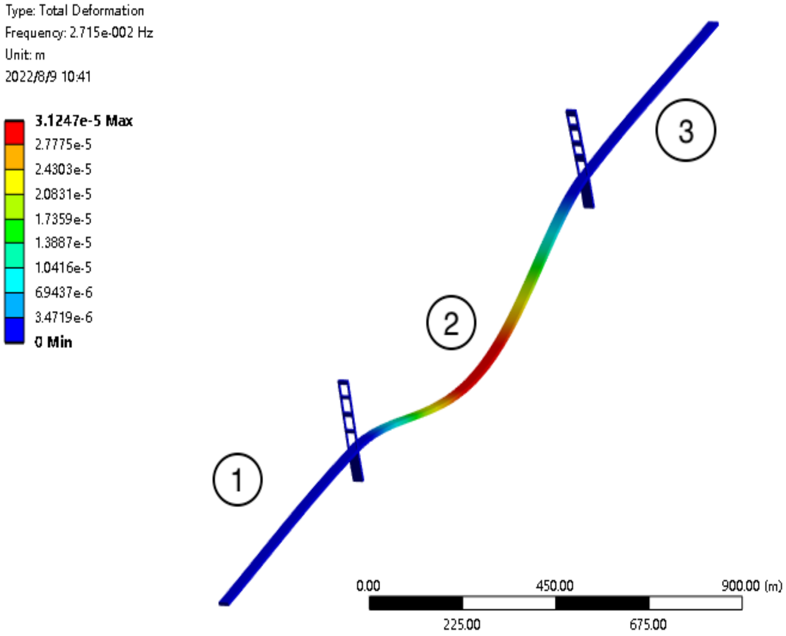


Figure 4. Mode 1

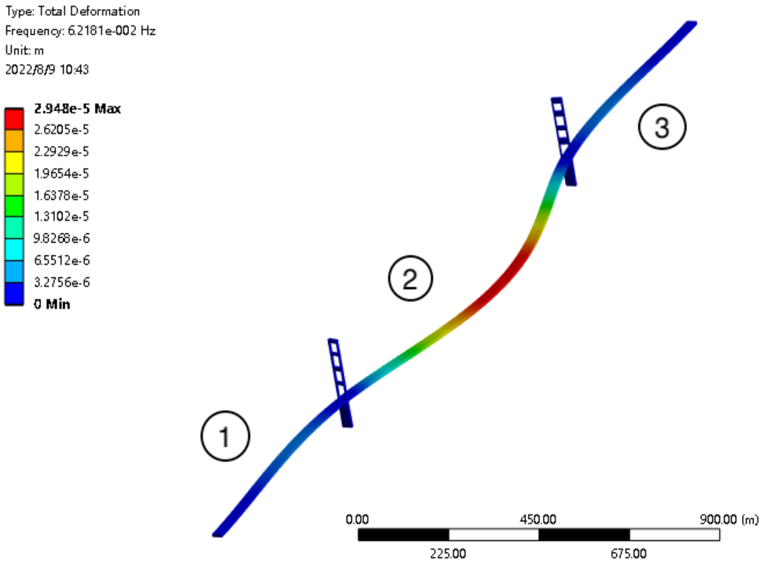


Figure 5. Mode 2

3.2 Mode 3 & 4

Mode 3 shows the total deformation, the frequency is 0.07218 Hz, and the red parts have maximum deformations. The whole bridge has a vibration. Moreover, according to figure 6, each tower between two parts also has light deformation.

Mode 4 shows that the frequency is 0.077068 Hz. Based on figure 7, deformations appear in parts 1 and 3; maximum deformations are in the middle of part 1 and part 3.

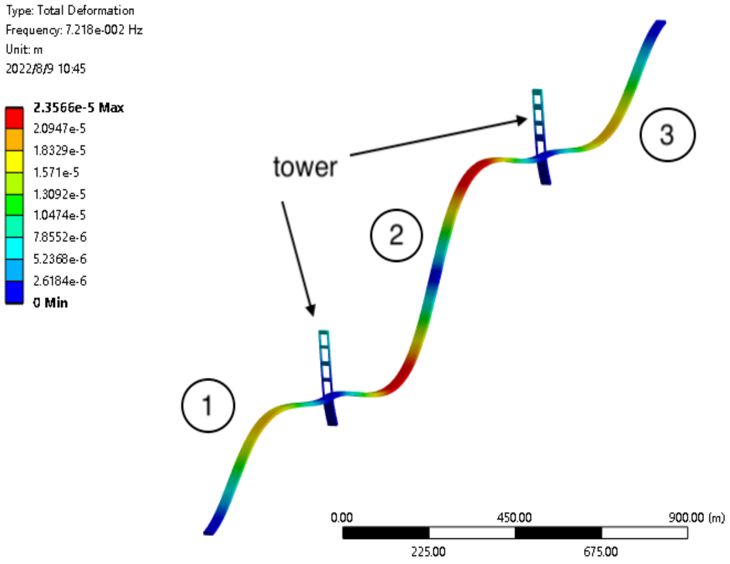


Figure 6. Mode 3

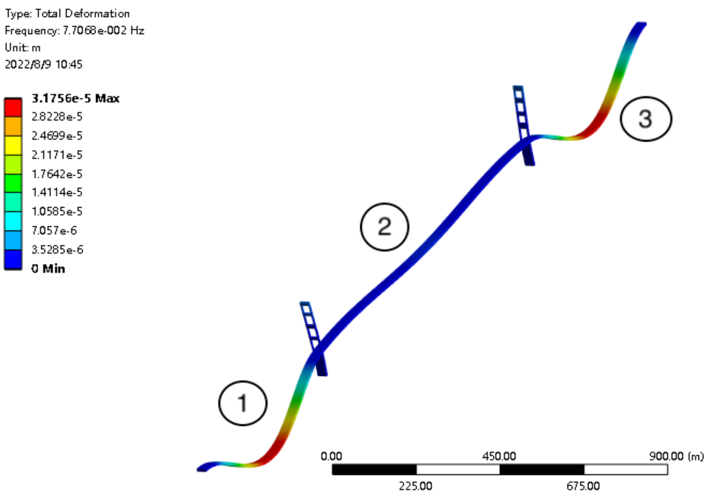


Figure 7. Mode 4

3.3 Mode 5 & 6

For mode 5, the frequency is equal to 0.07976 Hz. Figure 8 shows that the whole bridge has deformations, and the maximum deformations appear in the middle of part 1 and part 3.

Mode 6 shows the deformation of the bridge, and the frequency is 0.14591 Hz. Figure 9 shows that two towers and part 2 have deformations. In this case, the tops of two towers have maximum deformations.

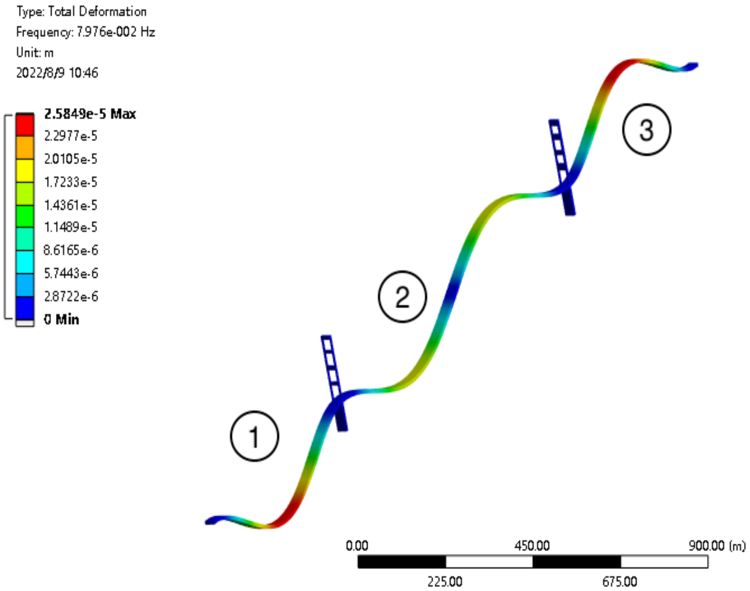


Figure 8. Mode 5

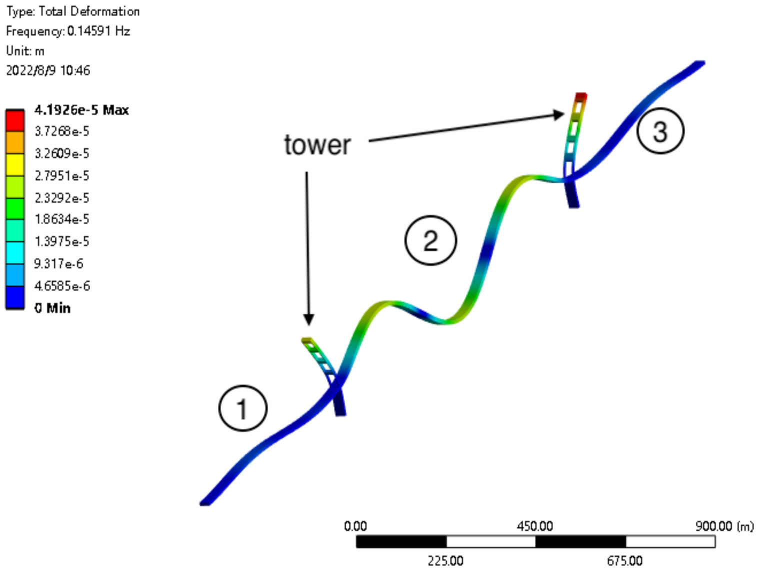


Figure 9. Mode 6

4. MODEL ANALYSIS

After getting all six modes, the following table can be found:

Table 1. Table of mode and maximum frequency

Mode	**Maximum Frequency (Hz)**
1	0.02715
2	0.062181
3	0.07218
4	0.077068
5	0.07976
6	0.14591

Based on the table of mode and maximum frequency, the average value of six modes' maximum frequencies is 0.07737Hz, and the first mode has the lowest maximum frequency which is 0.02715 Hz. On the other hand, when researching the natural frequency of the Gold Gate Bridge, the value of the natural frequency of the Golden Gate Bridge is 0.055 Hz. When comparing the lowest maximum frequency and the natural frequency, the difference is -0.02785 Hz. The difference between the average frequency and the natural frequency is 0.02237 Hz. One possible reason for these two differences is our simplified model ignores the cables of the bridge but does not reduce the mass of cables.

The cables of the Golden Gate Bridge play a role in balancing. According to Vasen, cables of the Golden Gate Bridge are used for balancing, and these cables make the load evenly distributed on the bridge[7](Vasen 1). Thus, the cables can reduce the deformation, and that is the reason that the average frequency of our simplified model is 0.02237 Hz higher than the natural frequency. Moreover, the reason for the difference between the lowest frequency and natural frequency should be ignoring cables as well. Obviously, cables play an important role in bridge engineering.

Thus, the cables can reduce the deformation, and that is the reason why the average frequency of our simplified model is 0.02237 Hz higher than the natural frequency. Moreover, the reason for the difference between the lowest frequency and natural frequency should ignore cables as well. Obviously, cables play an important role in bridge engineering.

5. HARMONIC RESPONSE

According to ANSYS official website, Harmonic response analysis is used to simulate how a structure will respond to sinusoidally repeating dynamic loading, which is what we are trying to analyze[8]. According to figure 10, it can be seen that the maximum amplitude is 0.66021m when the frequency is between 0.025 Hz and 0.03 Hz. Thus, the frequency should be in this domain if people want to get the maximum equivalent bridge stress. In figure 11, it shows the phase response, which is the relationship between the phase of a sinusoidal input and the output signal passing through any device that accepts input and produces an output signal of the bridge[9]. Based on figure 11, it can be found that the red curve (output) is always greater than the blue curve(stress). Then, the difference between the output and stress is about 50 sweeping phases. Moreover, because the output is greater, the sweeping phase difference should be about -55 degrees.

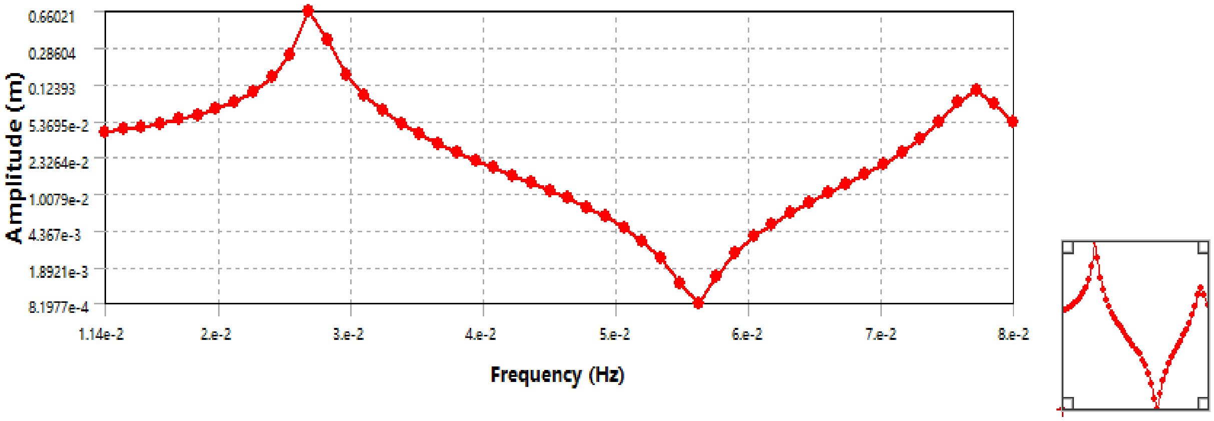


Figure 10. Chart for Amplitude vs. Frequancy

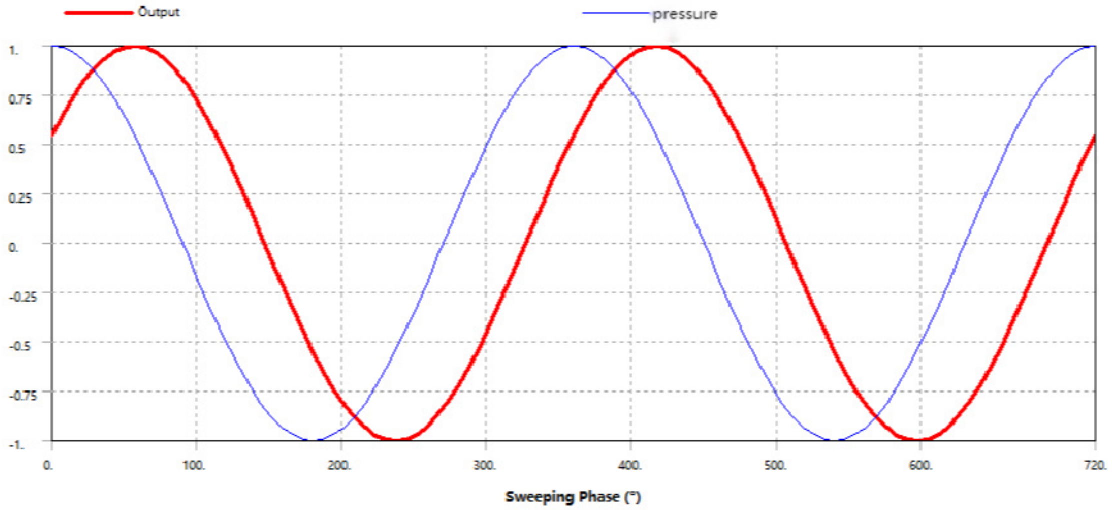


Figure 11. Chart for phase response

The model in figure 12 shows the bridge's distribution of the equivalent stress. Based on figure 12, the frequency of the equivalent stress model is 0.0268 Hz, which is between the domain of maximum amplitude. Moreover, the sweep phase is -56.649 degrees, which is close to -55 degrees; this means that the sweep phase of the model is an expected value.

According to SIMSCALE.com, the equivalent von Mises stress means the value used to determine whether a ductile material will break or not[10](SIMSCALE.com). This value is useful to determine the stability of the Golden Gate Bridge's model because the material of the bridge is steel and is a ductile material. In figure 13, the model is enlarged, and it is obvious that the maximum equivalent stresses are at joints between part 2 of the bridge and two towers. This means that when the frequency is 0.0268 Hz and the sweep phase is -56.649 degrees, the equivalent stress is 1.5487e^7 Pa at two joints; this value will be equal to or greater than the yield limit of the bridge. Therefore, in this condition, Thus the equivalent stress is the largest and the most destructive to the joints of the bridge when the frequency is 0.0268 Hz and the sweep phase is -56.649 degrees. To keep the stability of the Golden Gate Bridge, the joint part should be reinforced.

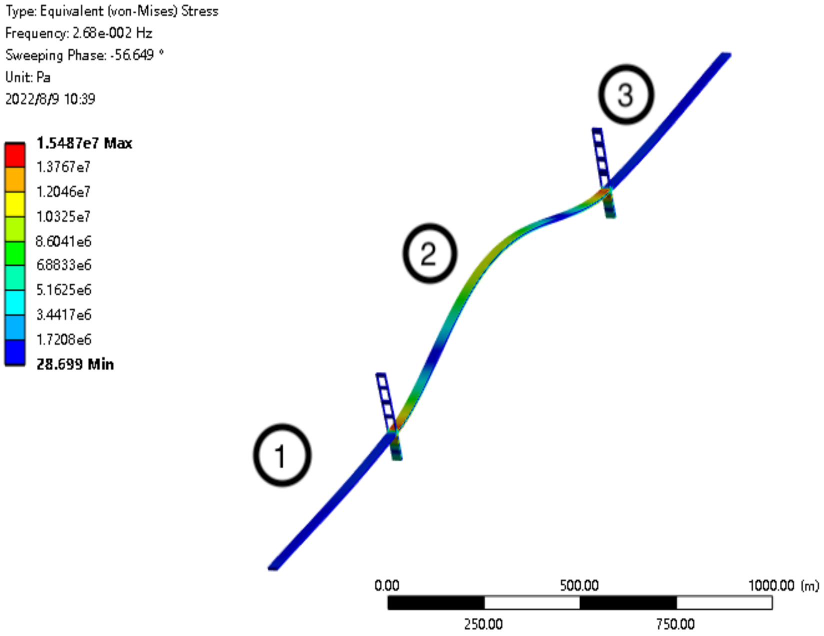


Figure 12 Equivalent Stress

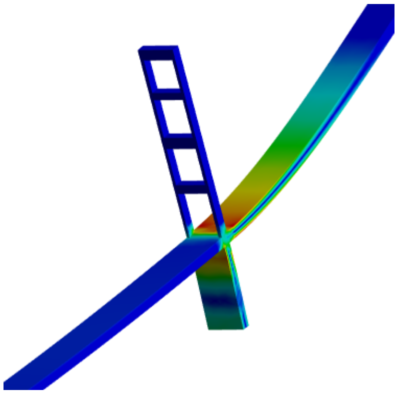


Figure 13. Equivalent Stress at one of the tower

6. CONCLUSIONS

This paper analyzed the stability of the Golden Gate Bridge by doing a finite element analysis. Mesh analysis was adopted, and six modes of bridge deformation were analyzed by building a simplified model. Also, the harmonic response model was made based on the live load, frequency, and damping ratio. Conclusions were obtained in the following.

Mode 1 and Mode 2 showed deformation in different frequencies. Mode 1 showed that the deformation mainly appeared in the middle part of the bridge and the center of it showed the maximum deformation. Mode 2 showed the maximum deformation part moving towards the right part of the bridge. Mode 3 showed the total deformation. The bridge had a vibration, and each tower between the two parts also had light deformation. Mode 4 and Mode 5 showed that the whole bridge has deformations, and the maximum deformations appear in the middle of part 1 and part 3 of the bridge. Mode 6 showed that the two towers and part 2 have deformations, and the tower between part 2 and part 3 has the maximum deformation. In addition, according to the harmonic response, the maximum amplitude is 0.66021m when the frequency is between 0.025 Hz and 0.03 Hz. The frequency of the equivalent stress model is 0.0268 Hz, which is between the domain of maximum amplitude. Moreover, the external load is the most destructive to the joints of the bridge when the frequency is 0.0268 Hz, and the largest equivalent stress is 1.5487e^7 Pa. Accordingly, this work shows the bridge's behaviors of vibrations and the largest equivalent that can influence stability. For further research on the Golden Gate Bridge, people can focus on methods to improve stability through reinforcement, for example, changes in materials or secondary structure.

REFERENCES

[1] Golden Gate Bridge Research Library. (May 2012). Retrieved from: https://www.goldengate.org/bridge/history-research/statistics-data/design-construction-stats/

[2] PBS learning media. (May 2004). Retrieved from: https://www.pbs.org/video/american-experience-building-bridge-ocean/

[3] Men Who Built the Bridge. Retrieved from: https://www.pbs.org/wgbh/americanexperience/features/goldengate-workers/

[4] Nakasone, Y., Yoshimoto, S.(2006). Overview of ANSYS Structure and Visual Capabilities. ScienceDirect. Retrieved from: https://www.sciencedirect.com/science/article/pii/B9780750668750500326

[5] Game, T., Vos, C., Morshedi, R., et al. (2016). Full dynamic model of Golden Gate Bridge. AIP Conference Proceedings. Retrieved from: https://aip.scitation.org/doi/pdf/10.1063/1.4961103

[6] Damping. (July 2022). Retrieved from: https://en.wikipedia.org/wiki/Damping

[7] "Facts on the Golden Gate Bridge." LoveToKnow. Retrieved from: https://sanfrancisco.lovetoknow.com/wiki/Facts_on_the_Golden_Gate_Bridge.

[8] Topics in Harmonic Response Analysis. (2022). Retrieved from: https://courses.ansys.com/index.php/courses/topics-in-harmonic-response-analysis/#:~:text=Harmonic%20response%20analysis%20is%20used,a%20fan%20inside%20a%20laptop.

[9] Phase response. (2019). Retrieved from: https://en.wikipedia.org/wiki/Phase_response

[10] What is von mises stress in fea?: SimWiki. SimScale. (2021, September 2). Retrieved August 11, 2022, from https://www.simscale.com/docs/simwiki/fea-finite-element-analysis/what-is-von-mises-stress/

Calligraphy image processing with stroke extraction and representation

Yi Wang*[a], XiaFen Zhang[b]
[a] Department of Information Engineering, Shanghai Maritime University, Shanghai, China 201306;
[b] Department of Information Engineering, Shanghai Maritime University, Shanghai, China 201306
[*] Corresponding author: 1142942657@qq.com

ABSTRACT

Calligraphy is the art of writing and each writer has his own features, where calligraphy strokes carry the most important features. In order to identify the stroke and the writer, this paper proposes a approach of stroke extraction and presentation. Firstly, skeleton segments of the character are extracted according to the writing rules. Secondly, neighbouring common segment strokes are connected to build a complete stroke, followed by special short stroke backtrack. Thirdly, individual outline strokes are isolated by using curves to build the closed contour for the stroke. Finally, generate and save outline strokes for future stroke feature extraction. Our approach can extract strokes from parts of regular script and official script.

Keywords: calligraphy characters; stroke extraction; contour segmentation; intersection area

1. INTRODUCTION

Each nature has its own language and characters, different people have different writing styles in their creations. Known as “the art of strokes”, Chinese calligraphy consists of a series of strokes. [1] The strokes carry the most important features of calligraphy characters. For the features recognition, we need to extract features from strokes, and then identify the style of different handwriting. In the process of writing characters, the calligraphy strokes can be reproduced in sequence according to the stroke order. Also the calligraphy animation simulation is written according to the stroke order of real calligraphy characters.

This paper proposes a method for stroke extraction and representation. Firstly, extract skeleton strokes, then extract outlines of strokes, get the stroke image information. Sun[2] recorded the shape of strokes by establishing a basic stroke chart. The strokes were merged into a stroke by using the chart. However, it was impossible to classify the stroke categories of Chinese characters with complex structures. In order to solve this problem, Xu Zongyi[3] judged the angle of strokes and connected strokes with similar slopes. The disadvantage was that the thinned skeleton is greatly affected by the strokes, which leaded to errors. For the extract outlines of strokes, Chen[4] divided the strokes according to the font style, selected a representative font, and calculated the characters formed by it. But it took a lot of time to build font styles. Yang Chenxu[5] combined the skeleton information and the binary image to restore the stroke width, and scanned the thick stroke width left and right along the contour points of the stroke binary image. Found four connectable feature points (corner points), and corrected the corner point connection. However, there are no restrictions on the determination conditions of corner points, resulting in low search efficiency.

2. SYSTEM STRUCTURE

In order to obtain the stroke direction, determine the stroke type. Firstly, use thinning algorithm to obtain character skeletons as the start of stroke extraction[6]. Then carry out skeleton merging and special skeleton processing on the skeleton strokes. Finally, extract the stroke outline and store the information. The system structure diagram is as follows.

International Conference on Computer Graphics, Artificial Intelligence, and Data Processing (ICCAID 2022)
edited by Ruishi Liang, Jing Wang, Proc. of SPIE Vol. 12604, 126040U
© 2023 SPIE · 0277-786X · doi: 10.1117/12.2674635

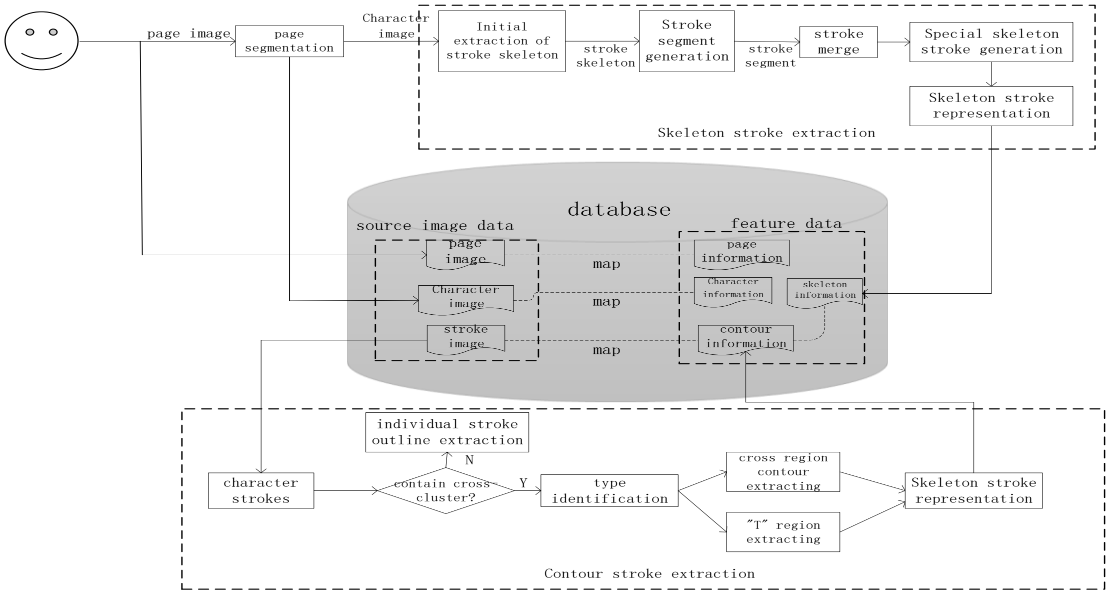


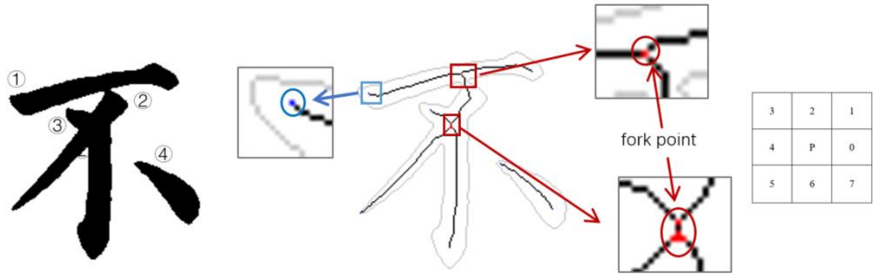
Fig.1 System structure

3. SKELETON STROKE EXTRACTION

Chinese characters are extremely complicated because of the variations in the combination of strokes.[7] In order to remove the interference of other complex factors, calligraphic characters need to be segmented through a thinning algorithm, and then the thinned skeleton strokes are obtained.

3.1 Stroke segment extraction

Chinese characters have an average of 16.1 strokes per character, and many of them cross and connect with each other, as shown in Fig.2(a). And the cross and connection will cause isolated strokes and connected strokes. The strokes are thinned to obtain a single pixel, and then the pixels are divided into 3 categories: endpoints, bunch points and regular points. Among them, endpoints have only one adjacent pixel in 8 neighboring code, as shown in Fig.2(b); Bunch points have more than two adjacent pixel in 8 neighboring code, as shown in Fig.2(d).



(a) original character (b) endpoint (c) skeleton and contour (d) bunch point (e) 8 neighboring code

Fig.2 endpoint, intersection, 8 neighboring code

Take any endpoints or bunch points as start, find the position of the next pixel in 8 neighboring code, and update the position after finding it. When the end points are found, this round of search ends. The obtained direction is stored into the stroke segment information, which is called 8-chain code. The 8-chain code direction encoding is shown in Fig.2 (e) above.

3.2 Regular skeleton generation

Based on the unique shape of Chinese characters, the Chinese characters can be divided into strokes[8], a stroke is composed of one or more stroke segments. If the pixel of the start and the end of stroke segments are both endpoints, then this segment is an individual stroke, shown in Fig.2(④). If the pixel of the start and the end of stroke segments are bunch points, then this segment needs to be merged, shown in Fig.2(①). By traversing all stroke segments, if the stroke directions are similar, then merge the two stroke segments. The specific merging rules are as follows: (1) The start and end of strokes are in the same bunch points; (2) The stroke direction is the same. After the merging is completed, the preliminary stroke skeleton is extracted.

4. SPECIAL SKELETON STROKE GENERATION

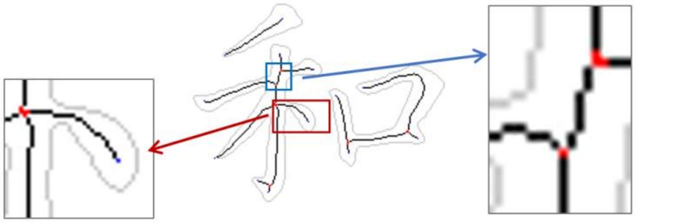
After the common stroke segments are merged, some strokes overlap at intersections, causing abnormal skeletons. Since the abnormality mainly occurs on short strokes, it is necessary to judge the category of short strokes, and use different methods to correct the strokes and correctly extract the skeleton of the strokes.

4.1 short stroke classification

First judge the average length μ and variance ρ of the skeleton strokes, then calculate the length of the short strokes < μ-ρ, when the length of the short stroke is less than the smaller value of the two values, it is judged that the current stroke is a short stroke. Divide short stroke types into 2 categories.

Type A: one end is a fork and the other is an endpoint, shown as Fig.3(a).

Type B: two ends are fork, shown as Fig.3(c).



(a) Type A short strokes (b) Stroke skeleton (c) Type B short strokes

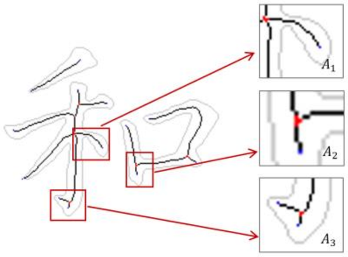
Fig.3 Examples of short strokes of class A and B

(1) Type A can be further divided into 3 categories:

Type A_1: connect a regular stroke, which is a right stroke, as shown with mark A_1 in Fig.4 (b). In this situation, do nothing with the current stroke.

Type A_2: connect with two regular strokes, and the direction of one regular stroke is the same with this short stroke, as shown with mark A_2 in Fig.4 (b). In this situation, strokes in the same direction need to be merged.

Type A_3: connect with one regular stroke and one short stroke, as shown with mark A_3 in Fig.4 (b). In this situation, the stroke need to be rewritten.



(a) Skeleton original image (b) Type A short strokes

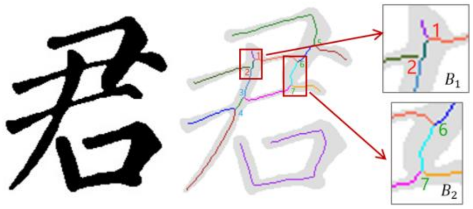
Fig.4 Example of 3 types of A short strokes

(2) Type B can be further divided into 2 categories:

Type B_1: Both end of the stroke are forks, the other stroke connected to the fork can not intersected by three stroke segments, and the other two stroke segments other than themselves can be merged through them, as shown with mark B_1 in Fig.5 (b).

Proc. of SPIE Vol. 12604 126040I-242

Type B_2: One of the stroke endpoints, the two connected strokes can be merged, as shown with mark B_2 in Fig.5 (b).



(a) Original character (b) Skeleton original (c) Type B short strokes

Fig.5 Example of two types of B short strokes

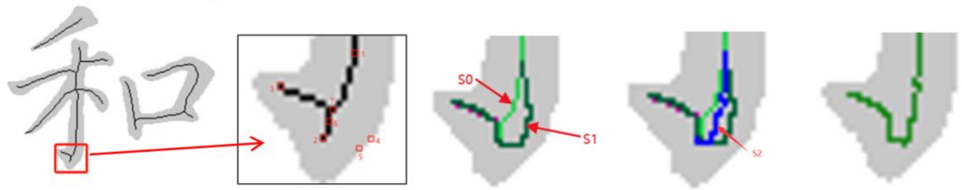
4.2 Deformed skeleton stroke generation

According to the writing characteristics of A_3 type strokes, when writing vertical hooks and horizontal hooks, there will be a slight pause at the twisting point, resulting in the pause of the stroke in the current area, which will increase the width of the stroke. According to the writing characteristics of B_2 strokes, when two strokes intersect, the thinned skeleton will be deformed due to the large difference in thickness between the two strokes.

(1) A_3 stroke generation

A_3 type strokes region can be called "hook" region. In this area, it is necessary to find 7 control points first, and divide these 7 points into two groups for bezier curve fitting generation, as shown in Fig.6(b). The calculation formula of curve fitting is as follows.

$$y = y_{(S1)} * (1.0 - t * t\_count) + y_{(S0)} * t * t\_count; \tag{1}$$

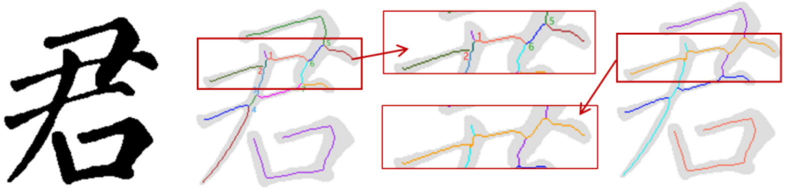


(a) Skeleton original (b) Skeleton original image (c) Area to be modified (d) fitted curve (e) corrected skeleton path

Fig.6 A_3 stroke generation

(2) B_2 stroke generation

B_2 type strokes region can be called "bridge" region. By traversing the trends of the four stroke segments at both ends of the "bridge", and merging the stroke segments with the same trend, the complete strokes combined in pairs are obtained. The final result is shown in Fig.7(d).



(a) original image (b) Skeleton extraction (c) Partial indication (d) Correct Skeleton

Fig.7 B_2 stroke generation

4.3 Skeleton stroke representation

The extracted skeleton strokes contain trajectory information, which needs to be expressed and stored. define the structure array to store the skeleton stroke information as follows.

$$\mathrm{Ske}_i = \{\mathrm{begin}_i, \mathrm{end}_i, \mathrm{snake}_i, \mathrm{bunchID}\} \tag{2}$$

Begin_i represents the start of the stroke, end_i represents the end of the stroke, bunchID represents the cross area information, and direction is represented by snake_i. The stroke information of Fig.7(c) is shown in the Table 1.

Table 1. skeleton stroke information in Fig.7

begin_x	begin_y	end_x	end_y	snake	bunchID
103	9	75	125	00700000…00770777	1,4
26	48	69	100	01000100…766666	2,3
40	65	161	8	6676676…55555556	1,2
67	14	47	148	701001…077770	3,4
121	55	120	82	676676…5444445	/

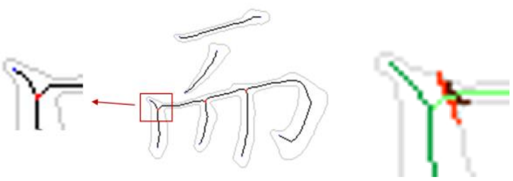
5. OUTLINE STROKE EXTRACTION

Based on the thickness transformation of strokes, it is necessary to extract width contour. Each stroke requires a closed loop to represent the contour. Calligraphers always like to use some specific actions when creating, such as frivolous, pressing, etc.[9] These subtle changes will make different outline of the strokes. When extracting the contour, it is necessary to focus on dividing the contour of the intersecting area. Divide crossovers into 2 categories: "T" region and "cross" region.

5.1 "T" region contour extracting

The two strokes form an intersection at the intersection, and the center point of the intersection was found in the previous step. Then you need to select the contour inflection point, that is, the blue point in the figure below represents the selected contour inflection point closest to the center point. The following takes the word "er" as an example to extract the outline of the intersection area.

It is judged that the current "T" intersection type, and the center point of the intersection is found. In the first and fourth quadrants with the center point as the coordinate origin, select the contour point closest to the center point as the inflection point, and select the contour point three pixels away from the inflection point as the control point of curve fitting. The specific positions are as follows as shown in the Fig.8 (a), take the green point as the start and end point of the Bezier curve B, and make two Bezier curves as the dividing line of the stroke outline, as shown in the following Fig.8 (c).



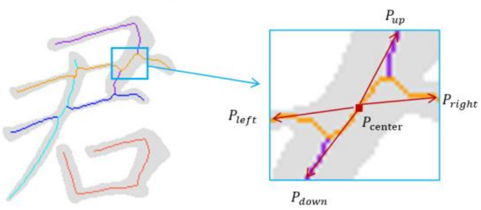
(a) zoom in (b) "Er" character (c) contour dividing line

Fig.8 "Er" character intersection

5.2 "cross" region contour extracting

After identifying the "cross" intersection, the calculation of the center point of the intersection area is completed, a square scan area with the center point as the center and two widths as the width is established. Scan the coordinates of the pixel points in the area, and find the contour point closest to the center point as the inflection point.

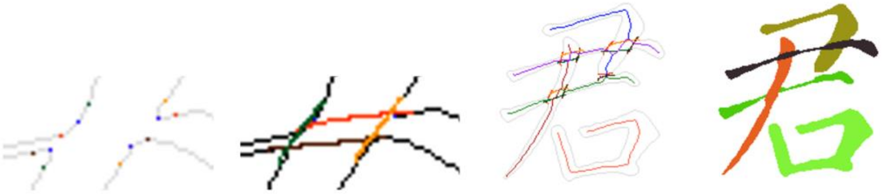
For the deformed sweep area stroke segment, the angle value of the partition needs to be modified. The red line in the figure below represents the connection between the stroke and the intersection area. Taking the four red lines as the limiting slope, the inflection point to be found not only needs to be within the corresponding quadrant, but also the angle between the connection line from the inflection point to center and the positive direction of the x-axis needs to meet the corresponding slope requirements.



(a) "Jun" character (b) Quadrant Partition

Fig.9 "Jun" character quadrant

After the inflection point is determined, it is necessary to select the start point or end point of the third-order Bezier curve at the same distance from the inflection point on the contour line, and select eight points in two groups, blue The inflection point of the color is used as the control point, and the other two points are used as the starting point for Bezier curve fitting. As shown in Fig.10 (b).



(a) inflection point (b) Contouring of the intersection area (c) contour segmentation (d) contour fill

Fig.10 The outline of the character "Jun" and outline extraction

5.3 Closed stroke outline extraction

Traversing the merged thinning strokes, first find a stroke contour point closest to the starting point as a starting point for the stroke contour extraction algorithm. When the contour is extracted to the end point of the Bezier curve, the Bezier curve is directly added to the contour line, and then the search for the 8-neighborhood contour points is continued from the other end of the bezier curve until all strokes are searched . With a certain contour point as the starting point, a single stroke contour is extracted. And through the filling algorithm, the following strokes of the character "Jun" are obtained. As shown in Fig.10 (d) above.

6. EXPERIMENT AND ANALYSIS

This experiment uses Visual Studio 2019 as the development platform, uses C++ as the programming language, uses QT for program interface implementation, and the background database version is MySQL8.0. The testing samples are chosen from the character set GB2312. There are 150 single-character pictures in the whole experiment. The results are as follows.

Table 2. accuracy of stroke extraction

category	total	correct number of skeleton	correct number of contour	Accuracy of skeleton（%）	Accuracy of contour（%）
regular script	50	47	45	94.0	90.0
semi-cursive script	50	45	42	90.0	84.0
official script	50	35	27	70.0	54.0

As can be seen from the Table 2 above, among all calligraphy types, regular script has the highest extraction accuracy rate, and official script has the lowest extraction accuracy. It can be seen through experiments that the extraction can handle some strokes, use bezier curves to supplement stroke outlines, and maintain calligraphy shapes, but it is not good for curves in some intersection areas. Compared with other stroke extraction methods, it can recognize more stroke outline structures.

7. CONCLUSION

This paper proposes a way of calligraphy stroke extraction and representation. In the generation of skeleton strokes, the skeletons of calligraphy characters have different shapes and complex structures. For the case where there is a shape deviation in the generated strokes, we use the method of further detection and classification. Determine the category of the current stroke, process the stroke, implement different stroke correction measures, and correctly extract the skeleton of the stroke. Due to the variable width of calligraphy strokes, resulting in different contour shapes, it is necessary to judge the stroke intersection type, then extract the contour of the intersection area. Finally, the correct closed stroke outline is extracted, and the stroke information with width is obtained.

Our future work include (1) Extraction of other word structures, such as "mouth". (2) Dealt with unclear glitches. (3) Increase the number of calligraphy stroke images and further optimizing the algorithm of contour segmentation. It is also necessary to complete a series of work such as stroke style, and complete the judgment of stroke flatness.

REFERENCES

[1] Li W. "Chinese writing and calligraphy. Hawaii University Press," Honolulu, 2009.

[2] Sun, Y.,Qian, H.,Xu, Y. "A geometric approach to stroke extraction for the Chinese calligraphy robot[C]," 2014.

[3] Xu Zongyi. "Research and implementation of Chinese character stroke automatic extraction system[D]," University of Electronic Science and Technology of China, 2015.

[4] Chen X, Lian Z , Tang Y , et al. "An automatic stroke extraction method using manifold learning," 2017.

[5] Yang, Chenxu, Zhang, Hongyun, Miao, "Duoqian, Calligraphy Characters Dynamic Reproduction Algorithm Based on Principal Curve," Pattern Recognition and Artificial Intelligence, 2019, 32(09): 835-843.

[6] X. Wang, X. Liang, L. Sun and M. Liu, "Triangular Mesh Based Stroke Segmentation for Chinese Calligraphy," 2013 12th International Conference on Document Analysis and Recognition, 2013.

[7] Y. Sun, H. Qian and Y. Xu, "A geometric approach to stroke extraction for the Chinese calligraphy robot," 2014 IEEE International Conference on Robotics and Automation (ICRA), 2014.

[8] X. Zhou, Z. Zhang, X. Chen and M. Qin, "Chinese Calligraphy Character Generating via CGAN with a Multi-subnet Parallel and Cascade Generator," 2020 39th Chinese Control Conference (CCC), 2020.

[9] L. Chen, "Research and Application of Chinese Calligraphy Character Recognition Algorithm Based on Image Analysis," 2021 IEEE International Conference on Advances in Electrical Engineering and Computer Applications (AEECA), 2021.

Integrating Users' Global and Local Interest for Session-based Recommendation

Yiqiu Fang, Ning Wu *, Junwei Ge
College of Computer Science & Technology, Chongqing University of Posts and Telecommunications, Chongqing 400000, China
* Corresponding author: 996372817@qq.com

ABSTRACT

Conventional recommendation systems mostly recommend based on users' global stable interest, while SR focuses on learning local active interest from the current session. Relevant research points out that integrating users' global and local interest can improve model accuracy. However, most SR models are deficient in interest mining and fusion, and do not fully utilize the time interval content between sessions, result in inaccurate recommendations. In the paper, we propose a new session-based recommendation method called Global and Local Interest Integration Model (GLI-GNN). GLI-GNN introduces the concept of session time and attention mechanism to obtain global preferences from users' previous sessions; it uses GNN to obtain local interest from current sessions; it uses attention mechanism to integrate both. The results indicate that GLI-GNN realizes new optimal performance in next-shot interactive recommendation. Experimental results show that GLI-GNN is superior to the existing SR models.

Keywords: session based recommendations, graph neural networks, integrate user interest

1. INTRODUCTION

Recommendation system can solve the problem of user information overload, and it is the basis for users to select the information they are interested in. Session-based recommendation (SR) is one of the important research branches. According to a recent study [1], most SR methods do not consider users' global interests ignore the user's global interests and make recommendations only based on the current session content, resulting in unreliable or inaccurate recommendations. Users with different global interest may take different actions even if they show consistent interest in the current session. Even if users with dissimilar global interest show consistent interest in the current session, they probably take different interact at the end. It is crucial to capture users' local and global interests and integrate them according to effective rules. Related studies such as [2] - [3] integrate the global interest extracted from the user's historical sessions with the current interest according to the same impact factor, without considering the time decay of the global interest, resulting in inaccurate recommendation.

To overcome the shortcomings in existing models, we design a new SR model called GLI-GNN. It effectively fuses visitors' global and local interest. Its core ideas include: 1) GLI-GNN introduces the concept of time interval to learn the global interest, and also uses aggregators such as self-attention mechanism, average pool and maximum pool in the training of the model; 2) GLI-GNN introduces GGNN to capture local interest from user's current session; 3) GLI-GNN effectively integrates global interest and local interest through an attention layer. We complete extensive experiments on representative dataset, the related experiments show that the performance of this paper proposed model is optimal. This paper has the following two contributions:

- We design a new model GLI-GNN, that can calculate the impact of the user's global interest on local interest based on time information, so as to obtain accurate user interest.
- We conducted empirical research on Yoochoose and Tmall, GLI-GNN achieved the best performance in both MRR and Recall.

2. RELATED WORK

This section briefly reviews related work, including traditional SR models, SR methods based on recurrent neural network, and SR methods based on graph neural network. There are two types of graph neural network models: models that only focus on local interests and models that fuse global and local interests

International Conference on Computer Graphics, Artificial Intelligence, and Data Processing (ICCAID 2022)
edited by Ruishi Liang, Jing Wang, Proc. of SPIE Vol. 12604, 126040V
© 2023 SPIE · 0277-786X · doi: 10.1117/12.2674571

2.1 Conventional SR Models

In traditional session recommendation methods, item similarity is defined as the frequency of appearance in the same session. Most of these methods are based on Markov chains [4], using serialized data of a given user's last click behavior to judge which items that customer probably clicks in next session. Literature [5] proposed a Markov chain-based serialized recommendation method, and explored how to use the probabilistic decision tree model to extract serialized patterns to learn the user's next behavioral state.

2.2 DNN SR Models

Lately, the methods of deep learning have been extensively used for session recommendation. Among them, the GRU4Rec method proposed in literature [6] used RNN network in conversational recommendation for the first time. The NARM method proposed in literature [7], the mechanism improves upon GRU4Rec by adding an attention mechanism to the recurrent neural network. So as to further diminish the bias made by time series, relevant literature [8] proposed an effective method STAMP, which is optimized by adding an attention layer.

2.3 GNN SR models

Recently, graph neural network has achieved many successes in computer vision, object recognition, natural language processing and other fields, more and more researchers have begun to consider how to integrate GNN into SR in recent years. Among them, the SR-GNN method proposed in [9] is used to capture the item features on the session graph, and then integrate accurate session features through focusing on every learned feature. The FGNN [10] model considers the intrinsic order of items in a session, and utilizes the weighting session graph and attention layer to learn hidden content among items. The GCE-GNN [11] method is proposed to capture two layers of item embeddings from the global graph and local session graph.

3. METHODOLOGY

3.1 Problem Definition

The symbols in this paper are defined as follows: Let $V = \{v_1, v_2, ..., v_n\}$ denote the collection of items and n is the number of items. Each project v_i in V is encoded into the unitary embedding space $\mathrm{h} \in \mathbb{R}^d$, embedding dimension of item embedding is d. Let the user's session sequence $S^u = \{v_{s,1}, v_{s,2}, ..., v_{s,n}\}$, $v_{i,t}$ represents the user's clicked item in the session S^u. Items are sorted in order of user interaction. The task of the GLI-GNN model is to learn the user's historical session S_h^u and current session S_c^u, and predict the most likely interactive item $v_{s,t+1}$. The GLI-GNN model will calculate the probability of all projects, where $\hat{y}$ is the recommended score of the corresponding project, and generally select the top n projects from $\hat{y}$ for recommendation.

3.2 Overview

Figure 1 is the model architecture diagram of GLI-GNN, which consists of three layers. User interest learning layer. This layer learns the user's global preferences based on historical sessions and time interval information between sessions. Interest integration layer. This layer fuses visitor's global interest and local interest through soft attention to get user's final interest. Prediction layer. This layer generates recommended items by computing score for each item.

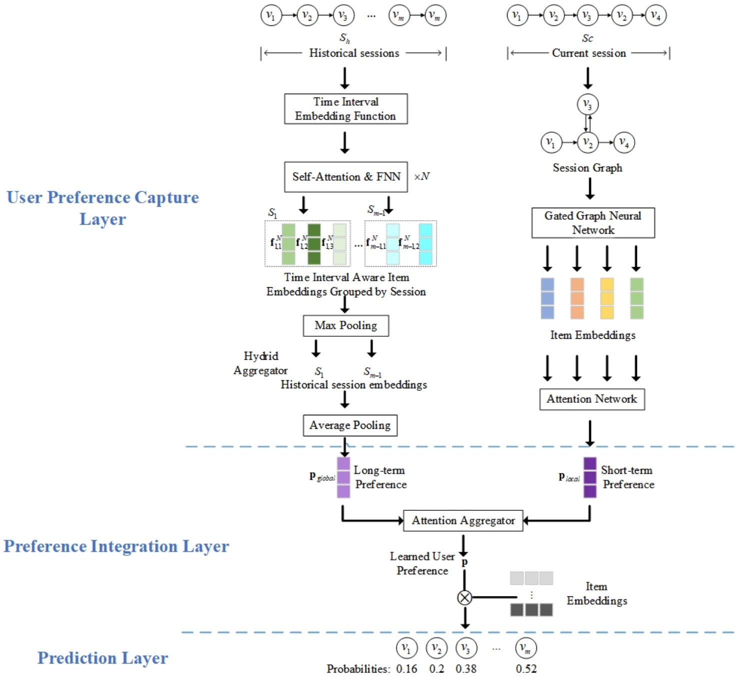

v_1 v_2 v_3 ... v_m v_m
S_h
Historical sessions
v_1 v_2 v_3 v_2 v_4
Sc
Current session
Time Interval Embedding Function
Self-Attention & FNN
×N
v_3
v_1 v_2 v_4
Session Graph
User Preference Capture Layer
S_1
S_{m-1}
$\mathbf{f}_{11}^N$ $\mathbf{f}_{12}^N$ $\mathbf{f}_{13}^N$ ... $\mathbf{f}_{m-11}^N$ $\mathbf{f}_{m-12}^N$
Time Interval Aware Item Embeddings Grouped by Session
Gated Graph Neural Network
Item Embeddings
Max Pooling
Hydrid Aggregator
S_1
S_{m-1}
Historical session embeddings
Attention Network
Average Pooling
$\mathbf{p}_{global}$
Long-term Preference
$\mathbf{p}_{local}$
Short-term Preference
Attention Aggregator
Preference Integration Layer
Learned User Preference
p
Item Embeddings
Prediction Layer
v_1 v_2 v_3 ... v_m
Probabilities: 0.16 0.2 0.38 0.52

Figure 1. The model architecture diagram of GLI-GNN

3.3 User interest learning layer: Historical session

In GLI-GNN, we learn the method in literature [12] to introduce time interval information, set the training sequence of interacted items in historical sessions as TRSQ, and the time sequence as TISQ, and the two are truncated or padded according to the maximum length x. Therefore, TRSQ can be expressed as $R=\{\mathbf{v}_1,\mathbf{v}_2,\mathbf{v}_3,...,\mathbf{v}_x\}$, and TIQS can be expressed as $T=\{t_1,t_2,t_3,...,t_x\}$. The time interval embedding vector $u(\Delta t_{ij})$ is calculated according to (1).

$$u(\Delta t_{ij})=\text{sigmoid}(\frac{1}{u(\Delta t_{ij})+0.005}) \tag{1}$$

For TRQS, this paper employs self-attention algorithm to compute H, and then get the embedding $\mathbf{h}_i\in\mathbb{R}^d$ of each item, as defined in Eq. (2).

$$\mathbf{z}_i = \sum_{j=1}^{l} a_{ij}\mathbf{v}_j \tag{2}$$

where we apply a softmax function on the attention score r_{ij} to compute the weight system a_{ij}, as defined in Eq. (3).

$$r_{ij} = \frac{\mathbf{W}_Q\mathbf{v}_i(\mathbf{W}_K\mathbf{v}_j)^{\mathrm{T}}}{\sqrt{c}}u_{ij}(\Delta t_{ij}) \tag{3}$$

where $\mathbf{W}_Q,\mathbf{W}_K$ are learnable model parameters. We use scale factor $\sqrt{c}$ to prevent the model from overfitting.

Only linear combinations are used in above calculations. Therefore, GLI-GNN imparts nonlinearity to the model through feedforward neural network.

$$\mathbf{b}_i = \text{ReLU}(\mathbf{z}_i\mathbf{W}_1+\mathbf{d}_1)\mathbf{W}_2+\mathbf{d}_2,\ i=1,2,3,...,l \tag{4}$$

where $\mathbf{W}_2, \mathbf{W}_1$ and $\mathbf{d}_1, \mathbf{d}_2$ are learnable model parameters. As shown in the investigation [13], the higher-level representation of further learning projects is useful. To make recommendations more accurate, we use multi-layer attention to aggregate item information.

$$\mathbf{B}^{N} = \mathrm{Ti}(\mathbf{B}^{N-1}) \tag{5}$$

We compute the per-item embedding $\mathbf{b}_1$ in the historical sessions, and then get the session embeddings by element-wise maxima. In order to more efficiently obtain the user's global interest $\mathbf{I}_{global}$ from historical sessions, we also use average pooling. as defined in Eq. (6).

$$\mathbf{I}_{global} = \frac{1}{m-1}\sum_{n=1}^{m-1}\mathbf{S}_n \tag{6}$$

3.4 User interest learning layer: Current-session

In this session, we convert the current conversation into a conversation graph using the method proposed in SR-GNN [9], and then utilize GGNN to learn local interest.

$$\begin{aligned}
\mathbf{a}_{s,i}^{t} &= \mathbf{A}_{s,i}[\mathrm{v}_1^{t-1},\ldots\mathrm{v}_n^{t-1}]^{\mathrm{T}}\mathbf{H} + b \\
\mathbf{z}_{s,i}^{t} &= \sigma(\mathbf{W}_z \mathrm{a}_{s,i}^{t} + \mathbf{U}_z \mathbf{v}_i^{t-1}) \\
\mathbf{r}_{s,i}^{t} &= \sigma(\mathbf{W}_r \mathrm{a}_{s,i}^{t} + \mathbf{U}_r \mathbf{v}_i^{t-1}) \\
\tilde{\mathbf{v}}_i^{t} &= \tanh(\mathbf{W}_o \mathrm{a}_{s,i}^{t} + \mathbf{U}_o(\mathbf{r}_{s,i}^{t} \odot \mathbf{v}_i^{t-1})) \\
\mathbf{v}_i^{t} &= (1\text{-}\mathbf{z}_{s,i}^{t}) \odot \mathbf{v}_i^{t-1} + \mathbf{z}_{s,i}^{t} \odot \tilde{\mathbf{v}}_i^{t}
\end{aligned} \tag{7}$$

where t is the number of training steps. $\mathbf{H} \in \mathbb{R}^{d\times 2d}$ is determined by parameter learning. $\mathbf{A}_s$ is the concatenation of the outgoing and incoming adjacency matrices. $\sigma(\cdot)$ is a nonlinear function, $\odot$ is element multiplication. $\mathbf{r}_{s,j}$ is reset gate, and $\mathbf{z}_{s,j}$ is update gate.

The local interest embedding $\mathbf{I}_{local}$ and global interest embedding $\mathbf{I}_{global}$ are calculated based on item node.

$$\begin{aligned}
\mathbf{I}_{global} &= \sum_{i=1}^{n} a_i \mathbf{v}_i \\
\mathbf{I}_{local} &= \mathbf{W}_2[\mathbf{I}_{local} \,||\, \mathbf{I}_{global}]
\end{aligned} \tag{8}$$

where $\mathbf{W}_2 \in \mathbb{R}^{d\times 2d}$ is a weight coefficient in algorithm training.

3.5 Interest integration layer

In this section, we use soft attention algorithm to integrate local interest $\mathbf{I}_{local}$ and global interest $\mathbf{I}_{global}$ to obtain user interest $\mathbf{I}$, as defined in Eq. (9).

$$\mathbf{I} = \beta_1 \mathbf{I}_{global} + \beta_2 \mathbf{I}_{local} \tag{9}$$

where β_1, β_2, is calculated by employing α_i, which is learned through the attention mechanism in the model. (10).

$$\alpha_2 = \mathbf{w}_3^{\mathrm{T}} \tanh(\mathbf{W}_3 \mathbf{I}_{global})$$
$$\alpha_3 = \mathbf{w}_4^{\mathrm{T}} \tanh(\mathbf{W}_4 \mathbf{I}_{local}) \tag{10}$$

where $\mathbf{w}_3, \mathbf{w}_4 \in \mathbb{R}^{d \times d}$ and $\mathbf{W}_3, \mathbf{W}_4 \in \mathbb{R}^{d \times d}$ are weight parameters.

3.6 Prediction Layer

In this subsection, this paper employs the learned visitor interest $\mathbf{I} \in \mathbb{R}^d$ and interacted item embeddings $\mathbf{v}_i$ to compute the recommended score $\hat{z}_i$ of all items, and use the softmax function to get the output vector of the model $\hat{y}_i$.

$$\begin{aligned} \hat{z}_i &= \mathbf{I}^{\mathrm{T}} \mathbf{v}_i \\ \hat{y}_i &= \mathrm{softmax}(\hat{z}_i) \end{aligned} \tag{11}$$

The cross-entropy loss function commonly used in recommendation systems is selected as the learning aim to optimize this algorithm.

$$L(\hat{y}) = -\sum_{i=1}^{N} y_i \log(\hat{y}_i) + (1\text{-}y_i)\log(1\text{-}\hat{y}_i) \tag{12}$$

4. EXPERIMENTS

4.1 Datasets

The public Tmall and Yoochoose are used to perform tests to confirm the validity of the raised model. Yoochoose dataset includes user click records on ecommerce sites over a period of 6 months. The Tmall dataset originates from IJCAI 2015 contest, which includes the purchase records of users on the Tmall. For fairness of the comparison, the two datasets are preprocessed with reference to the work of literature [2]-[3]: firstly, the conversations with length 1 on the two datasets are filtered, and items with occurrences less than 5 are filtered. The input conversation sequence is then split to produce sequences and commensurable tags. Set current session data of the last week on the Tmall dataset as the test suit, and the remaining data as the training suit. Since the extensive data in the Yoochoose dataset, this paper only selects the 1/64 session sequence closest to the time of the test set as training suit.

4.2 Baselines

In this paper, we contrast the GLI-GNN model with 8 existing mainstream methods.

Item–KNN [14]: It makes recommendations according to the similitude between present session items and other items.

FPMC [15]: It is a sequence prediction method based on matrix decomposition and first-order Markov chain, which can capture time information and user interest at the same time.

GRU4Rec [6]: It is a Recurrent Neural Network based method which employs GRU to simulate session sequences.

NARM [7]: It promotes upon GRU4Rec by adding an attention mechanism to the RNN.

STAMP [8]: It employs attention layer to obtain the customer's long-range interest from historical clicks as well as the present interest from the last interact in the session.

SR-GNN [9]: It constructs sessions as session graphs and uses GNN to extract session topology features.

FGNN [16]: It learns item embeddings by designing a weighted attention layer.

GCE-GNN [11]: It expands the session graph into a global graph, and aggregates high-order neighbor node features on the basis of the global graph.

4.3 Evaluation Metrics

$P@K$ (Precision) is a key indicator value, which is widely used as an indicator to consider the accuracy of trend analysis. For example, $P@10$ indicates the ratio of precisely recommended candidates among the roof 10 items.

$MRR@K$ (Mean Reciprocal Ranking) is employed to evaluate rank of correct item in list of candidate items. The MRR measure takes recommendation order into account, and a larger MRR figure suggests that the accurate item recommendation is at the roof of the candidate list.

4.4 Hyperparameter Settings

The hyperparameters of the GLI-GNN method in our paper mainly includes three points: (1) The embedding dimension of item and user interest is 128; (2) The learning attenuation factor η is chosen among {0.05, 0.1, 0.15, 0.2}; (3) the regularization coefficient λ is chosen among {0.00001, 0.0001, 0.001, 0.01, 0.1}; the fixed length of TRQS is chosen among {60, 70, 80, …, 150}.

4.5 Overall Comparison

Table 1 implies the data of performance comparison in the below datasets. The best conclusion of DNN model and GNN model is with the horizontal line below or the number with *. Improvement represents that GLI-GNN is superior to the best GNN model. We can get the following results from the conclusion:

- GLI-GNN achieved excellent performance on all evaluation metrics on both datasets. Improvement shows that GLI-GNN has greatly improved than the better GNN-SR model, indicating that unifying users' global and local preferences is beneficial to recommendation performance.
- The DNN-based model has been greatly improved over the traditional session-based recommendation model, illustrating the advantages of DNN in the field of session recommendation.

Table 1. Performance comparison on Yoochoose1/64 and Tmall.

Models	**Yoochoose1/64**				**Tmall**			
	Pre@10	***Pre@20***	***MRR@10***	***MRR@20***	***Pre@10***	***Pre@20***	***MRR@10***	***MRR@20***
ItemKNN	11.85	14.67	5.42	5.62	30.32	38.85	12.88	13.49
FPMC	2.42	3.27	0.50	0.37	34.31	44.32	6.56	4.54
GRU4Rec	14.36	17.64	7.78	8.83	42.38	50.33	26.88	27.44
NARM	14.86	17.21	8.68	8.74	46.23	54.69	27.44	28.73
STAMP	15.51	17.98	9.47	9.80	47.84	56.53	27.75	28.63
SR-GNN	16.62	18.99	9.88	10.06	52.95	61.12	32.65	33.01
FGNN	16.80	19.71	10.09	10.24	53.44	61.80	33.29	33.88
GCE-GNN	17.06*	19.98*	10.71*	10.91*	59.43*	68.00*	34.92*	35.52*
GLI-GNN	**17.82**	**20.47**	**11.03**	**11.35**	**60.77**	**69.98**	**36.01**	**36.61**
Improvement	4.45%	2.45%	2.99%	4.03%	2.25%	2.91%	3.12%	3.07%

5. CONCLUSION

In order to cope with challenge of integrating customers' global and local interest embeddings, we proposed a new GLI-GNN model. It adopts time interval function formula, hybrid aggregator and gate graph neural network (GGNN) to efficiently integrate global and local interest of users to promote the correctness of recommendations. Experimental results show that GLI-GNN significantly outperforms 8 baseline models on Yoochoose and Tmall datasets. The basis for future work is to combine the coordination and scalability in dynamic GNN.

REFERENCES

[1] Wang, S., Wang, Q. Z., Cao, L. Y. and Sheng M. A., “A Survey about Session-based Recommend Systems,” ACM Comput, 1-38 (2021).

[2] Ruocco, M., Skrede, S. L. and Langseth, H., “Inter Session Modeling in Session-Based Recommendation,” DLRS@RecSys, 24–31 (2017).

[3] Hidasi, B., Karatzoglou, A. and Cremonesi, P., “Personalizing Session-based Recommendation with Hierarchical Recurrent Neural Network,” RecSys, 130–137 (2017).

[4] Freudenthaler, C., Rendle, S. and Schmidt-Thieme, L., “Factorizing personalized markov chain for next basket recommendation,” Proceedings of the 19th International Conference on World Wide Web, Now York, 811-820 (2010).

[5] Chickering, D. M., Zimdars, A. and Meek, C., “Using temporal data for making recommendation,” arXiv, 1301-1320 (2013).

[6] Karatzoglou, A., Hidasi, B. and Baltrunas, L., “Session-based recommendation with recurrent neural network,” arXiv, 1511.06939 (2015).

[7] Ren, P., Li, J., Lian T. and Ren, Z., “Neural attentive session-based recommendations,” Proceedings of the 2017 ACM on Conference on Information and Knowledge Management. Singapore, 1419-1428 (2017).

[8] Zeng, Y., Liu, Q. and Mokhosi, R., “Stamp: short-term attention memory priority models for session-based recommendations,” Proceeding of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining, London, 1831-1839 (2018).

[9] Tang, Y., Wu, S. and Zhu, Y., “Session-based recommendations with graph neural network,” Proceedings of the AAAI Conference on Artificial Intelligence, 346-353 (2019).

[10] Qiu, R., Huang, Z. and Li, J., “Exploiting cross-session information for session-based recommendations with graph neural networks,” ACM Transactions on Information System, 1-23, (2020).

[11] Wei, W., Wang, Z. and Cong, G., “Global context enhanced graph neural network for session-based recommendations,” Proceedings of the 43rd International ACM SIGIR Conference on Research and Development in Information Retrieval, 169-178 (2020).

[12] Wang, Y., Li, J. and McAuley, J., “Time Interval Aware Self-Attention for Sequential Recommendation,” WSDM, 322–330 (2020).

[13] Chiang, D. and Anastasopoulos, A., “Tied Multitask Learning about Neural Speech Translations,” NAACL-HLT, 82–91 (2018).

[14] Karypis, G., Sarwar, B. and Konstan, A., “Item-based collaborative filtering recommendation algorithm,” WWW, 285–295 (2001).

[15] Freudenthaler, C., Rendle, S. and Schmidt-Thieme L., “Factorizing personalized markov chains for next-basket recommendations,” WWW, 811–820 (2010).

[16] Jingjing, L., Qiu, R. and Hongzhi Y., “Rethinking the Items Order in Session-based Recommendations with Graph Neural Network,” CIKM, 579–588 (2019).

Attribution analysis on the changing trend of sesame yield data in southern Henan under climate change

Meng Zhou[a,b,c], Haijiao Liu[*a,b,c], Jie Zhang[a,b,c], Guoqiang Li[a,b,c], Hecang Zang[a,b,c], Yajie Qiu[a,b,c], Qing Zhao[a,b,c], Xiuzhong Yang[a,b,c], Jianta Zhang[a,b,c], Guoqing Zheng[a,b,c]

[a] Institute of Agricultural Economics and Information, Key Laboratory of Huang-Huai-Hai Smart Agricultural Technology, Henan Academy of Agricultural Sciences, Zhengzhou, Henan, China, 450002;

[b] Henan Technology Innovation Strategic Alliance for Intelligence Agriculture Industry, Zhengzhou, Henan, China, 450002;

[c] Henan Engineering and Technology Research Center for Intelligent Agriculture, Zhengzhou, Henan, China, 450002

ABSTRACT

Analyzing the characteristics of climate change and clarifying the critical climate factors affecting sesame yield are significant for sesame production to cope with future climate change and ensure stable yield and income increase. In this study, trend analysis and mutation analysis were used to study the changing trends of climatic factors, yield per unit area, and climatic yield. The data included meteorological and sesame yield data in Zhumadian City of Henan Province from 1989 to 2018. The stepwise regression model was used to extract critical climate factors, and multiple regression models of sesame climate yield and critical climate factors were established. The contribution rates to sesame yield of the critical climate factors affecting sesame yield were determined. The results showed as follows: from 1989 to 2018, the temperature during the sesame growing period in Henan Province showed an increasing trend, and the minimum temperature showed the most apparent increasing trend, with an average annual increase of 0.03℃. The decreasing trend of precipitation and sunshine duration in the sesame growing period was insignificant, but the inter-annual fluctuation was significant. The coefficient of variation was 40.8% and 15.41%, respectively. The yield per unit area of sesame showed a significant increasing trend, and the annual growth rate was 34.08kg·hm^{-2}. The rising trend of climate yield is not adequate. The critical climate factors affecting sesame climatic yield were sunshine duration and average temperature. The contribution rate of each climate factor to climate yield in each growth period was ranked from the largest to the smallest as the average temperature at the seedling stage, the sunshine duration at the maturity stage, and the sunshine duration at the emergence stage. We have developed a future climate change adaptation strategy for the region based on the findings. The following measures should be taken to produce high-yield and high-quality sesame in south Henan Province. Local production departments should conduct training on high-yield cultivation techniques such as early sowing of sesame and risk management of waterlogging. Research institutions accelerated the selection, demonstration, and promotion of improved sesame seeds. The government has issued preferential agricultural policies.

Keywords: Sesame, Climate change, Climate yield, Mann-Kendall trend test, Sen's slope estimate, Stepwise regression model

1. INTRODUCTION

Crop model methods and statistical models using historical observations can be used to estimate the effects of climate change on crop yields. The crop model is a process simulation based on the complex system of the crop, climate, soil, and management, which requires many parameters. Compared with the statistical model method, it has low cost and flexible operation in determining the climate-yield relationship. It has been widely used to analyze the relationship between climate change and crop yield[1]. Shi Xiaoli[2] established multiple linear regression models of precipitation, average degree-days, extreme degree-days, and winter wheat yield. The results showed that high temperatures adversely affected the formation of winter wheat yield in the Huang-Huai-hai Plain. Zhao Dandan[3] established multiple regression models of the climatic yield of winter wheat and corn in Henan Province and climatic factors in different growth periods and analyzed that the

[*] Corresponding author: liuhaijiao@hnagri.org.cn

International Conference on Computer Graphics, Artificial Intelligence, and Data Processing (ICCAID 2022)
edited by Ruishi Liang, Jing Wang, Proc. of SPIE Vol. 12604, 126040W
© 2023 SPIE · 0277-786X · doi: 10.1117/12.2674617

increase of temperature and precipitation was not conducive to the formation of yield and the influence of temperature change on yield was much higher than that of precipitation. Wang Yan[4] established multiple regression equations between winter wheat yield and meteorological factors. The critical period and limiting meteorological factors that affected wheat yield. In the regression analysis, it is necessary to consider whether there is multicollinearity between two or more independent variables. If there is multicollinearity, it will affect parameter estimation and expand model error[5]. The stepwise regression model can select the most critical variable from many available variables and establish a regression analysis equation. Xu Xiangying [6] used the stepwise regression model to screen the meteorological factors affecting wheat climatic yield. Gao Juan[7] used the stepwise regression model to determine the critical growth periods and key meteorological factors affecting maize yield. The above studies mainly reflect the impact of climate change on the yield of wheat, corn, and other major food crops.

Sesame is a traditional characteristic oil crop in China[8], and Henan Province is the first of the three main sesame-producing areas in China, as well as a large consumption province of sesame [9,10]. Zhumadian City is a concentrated white sesame production area in the south of Henan Province [11], located in the Huang-Huai-hai Plain. In the context of climate change, climate warming in the Huang-Huai-hai region leads to the delay of crop seeding time and the shortening of the growth period. Extreme temperature tends to extend the duration of high temperature and shorten the duration of low temperature, which has an important impact on crop production [12,13]. Southern Henan has a north subtropical climate. Since 1960, the temperature has increased significantly, the precipitation and the number of precipitation days are more, the inter-annual fluctuation is significant, and the sunshine duration is relatively small [14,15]. Previous studies have shown that climate change affects sesame yield through climate factors such as temperature, precipitation, relative humidity, and sunshine duration[16-19]. However, spatial heterogeneity exists in the variation characteristics of limiting climate factors of sesame yield in different regions.

Select sesame-concentrated producing areas for climate change analysis and screening critical limiting climate factors for yield becomes the critical core issue. This paper selects the sesame-concentrated producing regions in southern Henan Province, where sesame is the anchor crop in summer and has had a planting habit for many years. Therefore, it is representative and reliable to analyze the impact of climate change on sesame yield in this region.

This article provides the following methods to solve the above problems. We selected the meteorological and statistical data of sesame yield in the south of Henan Province. The first step was to study the climate change rule during 1989-2018 and analyze the variation characteristics of sesame yield per unit area and climate yield. In the second step, the Lasso regression screened 20 climate factors closely related to light, temperature, and water required for the growth and development of sesame in different growth stages and excluded unimportant climate factors that caused multicollinearity. The third step is to establish multiple linear regression models of critical climate factors and sesame yield and then effectively quantify the contribution rate of climate factors in different growth periods to the yield. Finally, regional climate adaptation strategies are formulated for sesame production. In this study, we investigated data on many sesame production management. The method is suitable for analyzing climate yield-limiting factors in sesame-concentrated producing areas.

2. MATERIALS AND METHODS

2.1 Experimental data

According to the survey data, the specific growth dates of summer sesame were determined in this study, which were the emergence stage (from June 5th to June 10th), seedling stage (from June 11th to July 14th), flowering stage (from July 15th to August 24th), and maturity stage (from August 25th to September 10th). The meteorological data were obtained from the China Meteorological Data Network (http://data.cma.cn/). It includes daily precipitation (*P*, mm), sunshine duration (*SD*, h), relative humidity (*RH*,%), daily mean temperature (T_{mean},℃), maximum temperature (T_{max},℃), and minimum temperature (T_{min},℃) at Zhumadian meteorological Station during 1989-2018. We use the mean value of adjacent points to calculate the missing test data. Sesame yield data of Zhumadian City statistics are from the Statistical Yearbook of Henan Province. (http://www.henan.gov.cn/zwgk/zfxxgk/fdzdgknr/tjxx/tjnj/).

2.2 Research Methods

2.2.1 Climatic yield

The crop yield can be decomposed into trend yield and climatic yield if the influence of random fluctuation is ignored. Trend yield reflects the long-term yield changes caused by factors such as farming methods and agricultural science and technology level. In contrast, climate yield represents crop yield's short-term fluctuation caused by climatic conditions changes [20]. The formula of Climatic yield is:

$$Y_w = Y - Y_t \tag{1}$$

Y represents the actual crop yield, kg·hm^{-2}; Y$_t$ represents the trend yield, kg·hm^{-2}; Y$_w$ represents climatic yield, kg·hm^{-2}. The moving average method is a common and universal method for trend fitting[1]. After the sliding average, the period shorter than the sliding length in the production sequence is weakened. In this study, a 3-year moving average was used for trend fitting. Since the average sliding method will cause losses at both ends of the production series, the data analysis in this paper selects the production data from 1988 to 2019 for the sliding average to obtain the trend production from 1989 to 2018.

2.2.2 Mann-Kendall test

Mann-Kendall test is often used to test the significance of time series trends. It is the most widely used analysis method for time series change characteristics in meteorological, hydrological, and other fields [21]. It does not require the measured values to follow the normal distribution and can avoid the interference of a few outliers and missing values [22]. The calculation formulas are as follows:

$$Z = \begin{cases} \frac{S-1}{\sqrt{Var(S)}}, & S > 0 \\ 0 \quad , & S = 0 \\ \frac{S+1}{\sqrt{Var(S)}}, & S < 0 \end{cases} \tag{2}$$

$$S = \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} sign(x_j - x_i) \tag{3}$$

$$sign(x_j - x_i) = \begin{cases} 1 \ , & x_j - x_i > 0 \\ 0 \ , & x_j - x_i = 0 \\ -1 \ , & x_j - x_i < 0 \end{cases} \tag{4}$$

$$Var(S) = \frac{n(n-1)(2n+5)}{18} \tag{5}$$

sign is a sign function, and *n* is the number of time series data. In the bilateral trend test, for a given confidence level of α, if $|Z| \geqslant Z_{1-\alpha/2}$, the time series data has an obvious upward or downward trend at the confidence level. A positive value of *Z* indicates an increasing trend, while a negative value of *Z* indicates a decreasing trend. When the absolute value of *Z* is greater than or equal to 1.65, 1.96, and 2.58, it indicates that the confidence level passes 90%, 95%, and 99%, respectively.

2.2.3 Theil-Sen Median

Theil-SenMedian slope estimation, also known as Sen's slope estimation, can offset the influence of outliers and is a non-parametric statistical trend calculation method [23]. The calculation formula is as follows:

$$\beta = Median(\frac{x_j - x_i}{j-i}) \tag{6}$$

Median means the median value, x_{j} and x_{i} are the time series values of the j and i years, respectively (j > i). If $\beta > 0$, it means an upward trend; if $\beta < 0$, it means a downward trend.

2.2.4 Coefficient of Variation

Coefficient of Variance reflects the degree of fluctuation between different time series and eliminates the effects of measurement scale and dimension.

$$C_v = \frac{\sigma}{\mu} \tag{7}$$

σ is the standard deviation of time series, and μ is the multi-year average of time series.

2.2.5 Attribution analysis of sesame climatic yield

In this study, five typical climate factors closely related to light, temperature, and water required by sesame growth and development were screened at each growth stage of sesame (1:seedling stage, 2: seedling stage, 3: flowering stage, 4: maturity stage), including precipitation (*P*, mm), sunshine duration (*SD*, h), average temperature (T_{mean}, ℃), active accumulated temperature (A_a, d·℃) and relative humidity (*RH*, %).They are respectively abbreviated as *P*1, *SD*1, T_{mean}1, A_a1, *RH*1; *P*2, *SD*2, T_{mean}2, A_a2, *RH*2; *P*3, *SD*3, T_{mean}3, A_a3, *RH*3, *P*4, *SD*4, T_{mean}4, A_a4, *RH*4. The above 20 climate factors were used as dependent variables of the stepwise regression model. According to the degree of influence of climate factors on sesame climate yield, the correlation significance test coefficient was successfully used to eliminate climate factors without significant significance. The optimal model of key climate factors and sesame yield was established. *Variance Inflation Factor* (*VIF*) can characterize the degree of collinearity between independent variables. Generally, *VIF*≥10 indicates that there is serious collinearity between independent variables [24]. The calculation formula of multiple linear regression analysis is:

$$Y_S = a + \sum_{i=1}^{m} b_i\, x_i \qquad (i=1,2,\ldots,\mathrm{m}) \tag{8}$$

$$C_i = \frac{|b_i|}{\sum_{i=1}^{m} b_i} \qquad (i=1,2,\ldots,\mathrm{m}) \tag{9}$$

Y_s is the standardized value of sesame climate yield, *a* is the constant term, b_i is the standardized partial regression coefficient, and x_i is the standardized value of each key climate variable. c_i is the contribution rate of each climatic variable to climatic yield.

3. RESULTS

3.1 Analysis on the change characteristics of climate factors in sesame concentrated producing areas in south Henan Province

The minimum temperature and relative humidity significantly vary among all climatic factors, and the annual increase rate of minimum temperature is 0.03℃. The annual decline rate of relative humidity was 0.17% (Tab.1). Compared with the minimum temperature, the increasing trend of the mean and maximum temperatures is insignificant, and the rising annual rate is 0.01℃. The increase in the minimum temperature is the primary determinant of the rise of the average temperature in the region. The precipitation and sunshine duration showed an overall decreasing trend, but the decreasing trend was insignificant, and the annual rate of decline were 1.95mm and 1.89h, respectively. The average annual precipitation of sesame during its growth period was 505.72mm, and the inter-annual variation range was much higher than other climatic factors. The coefficient of variation was 40.80%. The average annual sunshine duration was 510.39h, and the coefficient of variation was 15.41%. The variation coefficients of other climate factors were relatively low, ranging from 3.01% to 5.29%.

Table 1. Variation trend of climate factors during sesame growth period

Statistical Index	Precipitation (mm)	Sunshine duration (h)	Relative air humidity (%)	Mean temperature (℃)	Maximum temperature (℃)	Minimum temperature (℃)
σ	505.72	510.39	77.48	26.24	31.03	22.29
C_v	40.80	15.41	5.29	3.01	3.19	3.65
Z	-0.36	-1.03	-1.93*	0.70	0.25	1.71*
β	-1.95	-1.89	-0.17	0.01	0.01	0.03

*,** and*** represent the significance test at P < 0.1, P＜0.05 and P＜0.01

3.2 Trend analysis of sesame yield and climate yield

The yield per unit area of sesame showed an overall significantly increasing trend from 1989 to 2018 (Tab. 2). The annual growth rate was 34.08 kg·hm^{-2}. In contrast, although climatic yield has an increasing trend, but it is not significant, and the annual growth rate was only 1.84 kg·hm^{-2}. The fluctuation range of sesame per unit yield and climate yield was extensive, and the fluctuation of sesame per unit yield was consistent with climate yield. Since 2011, sesame per unit yield has entered a period of rapid growth, and the fluctuation range of climate yield has weakened in this stage. The climatic yield of sesame per unit yield was 290 kg·hm^{-2} in 1990, which was 293 kg·hm^{-2} higher than the annual average climatic yield. The climatic yield of sesame in 2003 was -425 kg·hm^{-2}, which was 421 kg·hm^{-2} lower than the yearly average (Fig.1).

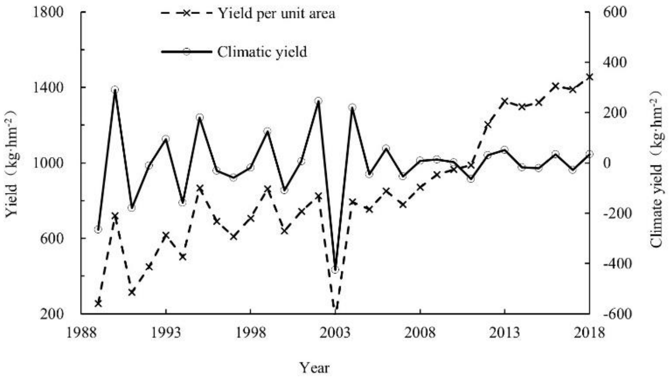


Figure 1. Variation curve of sesame yield and climate yield in Zhumadian from 1989 to 2018

Table 2. Variation trend of sesame yield and climate yield

Statistical Index	**Yield（kg·hm^{-2}）**	**Climate yield（kg·hm^{-2}）**
σ	843.55	-3.27
Z	5.82***	0.64
β	34.08	1.84

*,** and*** represent the significance test at $P < 0.1$, $P<0.05$ and $P<0.01$.

3.3 Attribution analysis of sesame climatic yield

The results of the step-up regression analysis and F test are shown in Table 3. F-test data showed that the P value was 0.001, reaching the highly significant level, so the null hypothesis that the regression coefficient is 0 is rejected. The results showed that VIF values were all less than 10, so there was no multicollinearity problem in the model. The above results prove that the model is well constructed. The critical climate factors selected by stepwise regression were SD4, Tmean2, and SD1(Table 4). The contribution rate of climate factors to climate yield in the sesame growth period was ranked as $T_{mean}2 > SD4 > SD1$.

Table 3. Stepwise regression model analysis

Variable	$\boldsymbol{b_i}$	**t**	**P**	**VIF**	**R^2**	**F**
a	0	-3.924	0.001***	-	0.481	F=8.043， P=0.001***
$SD4$	0.389	2.735	0.011**	1.012		
$T_{mean}2$	0.532	3.434	0.002***	1.204		
$SD1$	0.432	2.794	0.010***	1.198		

*,** and*** represent the significance test at $P < 0.1$, $P<0.05$ and $P<0.01$

Table 4. Partial regression coefficient and contribution rate of key climatic factors

Contribution rate	**Variable**		
	***SD*1**	***Tmean*2**	***SD*4**
c_i	28.75	39.32	31.93

4. DISCUSSION

The climate change in the main sesame-producing area of Henan Province showed that the minimum temperature increased significantly and the humidity decreased significantly from 1989 to 2018. The decreasing trend of rainfall and sunshine duration was insignificant but varied considerably from year to year. The above analysis results were consistent with the research conclusions of Li Wenxu [15]. The main sesame-producing area in Henan Province is distributed in the rainfed agricultural area in southern Henan Province, which belongs to the northern subtropical climate. The sesame growing period is in flood season, with frequent and high precipitation and relatively less solar radiation and sunshine duration. In this study, the annual average precipitation during the growth period of sesame is 505.72mm, with a coefficient of variation of 40.8%, which fluctuates significantly between years, indicating the uneven distribution of precipitation. It is one of the main factors causing the inter-annual fluctuation of sesame climatic yield. Different growth stages of sesame have different temperature requirements. Gurjinder[16] studied the three-point temperature of the vegetative growth and reproductive growth of sesame, and they found that the vegetative growth stage is more sensitive to low temperature (< 15℃), while the reproductive growth stage is sensitive to both low and high temperature, and 25℃ treatment has the highest yield. Bakhshandeh[25] determined that sesame seed germination's three basis point temperatures were 14, 36, and 47℃, respectively. The temperature range of three basis points for the growth and development of Chinese sesame is roughly 10, 25-27, and 40℃[26]. Previous studies have predicted that the temperature will increase by 0.3 to 0.6℃ every decade [27]. Under the background of global warming, the temperature in the main sesame-producing areas of Henan Province is showing a significant warming trend, with an annual temperature increase rate ranging from 0.01 to 0.03℃. In this study, the annual average of minimum temperature, average temperature, and maximum temperature during the growth period of sesame in the past 30 years were 22.29, 26.24 and 31.03℃, respectively, with an inter-annual fluctuation range of about 3%. The temperature conditions in the main producing areas of sesame in Henan Province were suitable for a certain period, basically within the optimal temperature range for the growth and development of sesame. Early sowing of sesame is beneficial in the main sesame-producing areas of Henan Province. On the one hand, sesame seedlings will not meet the local flood season. On the other hand, the maturity period is advanced to avoid the production reduction caused by low temperatures.

According to the yield attribution analysis of sesame climate factors, the formation of sesame climate yield is closely related to sunshine duration and average temperature. The critical climate factors and their contribution rates differ in different growth periods. Temperature and sunshine duration play a vital role in sesame's vegetative and reproductive growth. Sunshine duration at the emergence stage has the most significant influence on climate yield. Increasing sunshine duration can improve the stability of soil temperature, and the seedling formation rate of sesame. The sesame seedling stage is in the vegetative growth stage. Dry matter accumulation is distributed to the leaves. At this stage, photosynthesis produces photosynthates to provide their material basis. The maturity stage is in the reproductive growth stage, a critical period for sesame fruit and seed development—the distribution and accumulation of photosynthates to fruits and seeds[28]. Insufficient sunshine in the maturity stage, weakened photosynthesis, blocked dry matter accumulation and increased blurry seed rate also causes nutrient deficit, increased disease, insect pests, 1000-grain weight decline, and reduced yield.

The final formation of crop yield is restricted by climate factors and related to breeding, cultivation management techniques, and favorable policies. In this paper, the sesame yield per unit area in Zhumadian showed a highly significant rising trend, and the increasing rate was much higher than that of the climate yield. The sesame yield per unit area was affected by the fluctuation of climate yield, and the fluctuation trend of the two was consistent. Since 2011, the sesame yield per unit area entered a period of rapid growth, while the climate yield at this stage had a weakened fluctuation range and an insignificant rising trend. Therefore, the main reason for the increase in yield per unit at this stage may be related to local policy support. In recent years, the ability to identify and evaluate sesame germplasm resources, genetic breeding, and research of high-yield and high-efficiency planting technology in Henan Province has been at the top level in China. The above technology provides scientific and technological support for local sesame's high-yield and high-quality production[29].

5. CONCLUSIONS

From 1989 to 2018, the temperature in the main sesame-producing areas of Henan Province was significantly warmer, mainly driven by the increase in minimum temperature, with the annual increase rate of 0.03℃. The relative humidity has a significant decreasing trend, and the yearly decrease rate is 0.17%. The decreasing trend of precipitation and sunshine duration was insignificant, and the annual decreasing rates were 1.95 mm and 1.89 h, respectively. However, the inter-annual variation was considerable, and the inter-annual variation of precipitation was the largest, with the coefficient of variation being 40.8% and 15.41%, respectively.

The yield per unit area of sesame showed an overall trend of fluctuation and the annual growth rate was 34.08 kg·hm^{-2}, which fluctuated under the influence of climate change in the short term. Under the long-term trend, a high and stable yield of sesame could be achieved through variety breeding and supporting high-yield cultivation management technology, formulating reasonable promotion measures, and supporting policies regarding planting norms and promoting improved varieties. Reduce yield loss due to climate change.

The formation of sesame climatic yield was closely related to sunshine duration and average temperature. The critical climatic factors and their contribution rates in sesame growth stages were $T_{mean}2 > SD4 > SD1$.

ACKNOWLEDGMENT

This work was supported by the Open Fund of the Key Laboratory of Agricultural Information Service Technology of the Ministry of Agriculture and Rural Affairs and the Independent Innovation Project of Henan Academy of Agricultural Sciences(2022ZC49)

REFERENCES

[1] Gammans, M., Mérel, P., & Ortiz-Bobea, A. Negative impacts of climate change on cereal yields: statistical evidence from France[J]. Environmental Research Letters, 12(5), 054007(2017).

[2] Shi X L, Shi W J. Impacts of Extreme High Temperature on Winter Wheat Yield in the Huang-Huai-Hai Plain[J].Journal of Ecology and Rural Environment, 32(2), 259-269 (2016).

[3] Zhao D D, Zhai S Y. Influence of Climate Change on the Production of Winter Wheat and Corn in Henan Province from 1951 to 2012[J]. Chinese Agricultural Science Bulletin, 31(29):152-157(2015).

[4] Wang Y, Zhang X L, Shi J L, Shen Y J. Climate change and its effect on winter wheat yield in the main winter wheat production areas of China[J]. Chinese Journal of Eco-Agriculture, 30(5), 723-734(2022).

[5] Graham, M. H. Confronting multicollinearity in ecological multiple regression[J]. Ecology, 84(11), 2809-2815(2003).

[6] Xu X Y. Prediction and analysis of wheat yield changes based on an integrated climatic assessment indicator for wheat production in Jiangsu Province [D]. Yangzhou University (2019).

[7] Gao J, Feng LZ, Zhang JK, Wang Y. Effect of climate change on corn yield in Yulin, North Shaanxi Province of China[J]. Chinese Journal of Agricultural Resources and Regional Planning, 37(12):118-124+135(2016).

[8] Li F, Gao T M, Wei S L. Optimal ratio of nitrogen basal application and top-dressing for N uptake, distribution and yield of sesame [J]. Journal of Plant Nutrition and Fertilizers, 25(05):756-764(2019).

[9] Cui Y Q, Xu J, Guo Y Z, Guan Z B, Jian J L. Comprehensive analysis and trend of sesame breeding based on the regional test in northern China[J]. Chinese Journal of Oil Crop Sciences, 42(03):401-410(2020).

[10] Zhang W L. High quality development in direction and countermeasures of specialized oil industry in China[J]. Chinese Journal of Oil Crop Sciences,42(02):167-174(2020).

[11] Sun H X. Constraints and countermeasures of the development of white sesame industry in Zhumadian City[J]. Henan Agriculture, (22):8-9(2019).

[12] Sun X S, Long Z W, Song G P, Chen C Q. Effects of Climate Change on Cropping Pattern and Yield of Summer Maize-Winter Wheat in Huang-Huai-Hai Plain [J]. Scientia Agricultura Sinica, 50(13):2476-2487(2017).

[13] Guan Y, He Q J, Liu J H, Li R C, Hu Q, Huang B X, Pan X B. Variation Characteristics of Extreme Temperature and Its Earliest and Latest Day Sequence in Huang-Huai-Hai Region During the Period 1961 to 2015 [J]. Research of Soil and Water Conservation, 28(01):147-152+2(2021).

[14] Jiao J L, Kang W Y, Wang J, Dou H Y. Temporal and spatial Change Analysis of Sunshine Hour in Henan[J]. Meteorological and Environmental Sciences, (S1):4-6(2008).

[15] Li W X, Wu Z Q, Lei Z S, Jiang G Y. The Characteristics of Climate Factors Change and Its Effectson Main Grain Crops Yield per Unit Area in Henan Province[J].Crops, (01):124-134(2021).

[16] Baath, G. S., Kakani, V. G., Northup, B. K., Gowda, P. H., Rocateli, A. C., & Singh, H. Quantifying and Modeling the Influence of Temperature on Growth and Reproductive Development of Sesame[J]. Journal of Plant Growth Regulation, 41(1), 143-152(2022).

[17] Ibrahim, A. M. The impact of rainfall on the yields of staple crops-sorghum and sesame in Sudan[J]. J. Plant Sci. Res, 2, 1-4(2015).

[18] Deepthi, P., Shukla, C. S., Verma, K. P., & Reddy, S. S. Yield loss assessment and influence of temperature and relative humidity on charcoal rot development in sesame (Sesamum indicum L.) [J]. Bioscan, 9, 193-195(2014).

[19] Meena, H. M., & Rao, A. S. Growing degree days requirement of sesame (Sesamum indicum) in relation to growth and phonological development in Western Rajasthan[J]. Curr. Adv. Agric. Sci, 5(1), 107-110(2013).

[20] Li X Y, Zhang Y, Zhao Y X, et al. Comparitive study on main crop yield separation methods[J]. Journal of Applied Meteorological Science,31(01):74-82(2020).

[21] Ahmed, K., Shahid, S., & Nawaz, N. Impacts of climate variability and change on seasonal drought characteristics of Pakistan[J]. Atmospheric research, 214, 364-374(2018).

[22] Nashwan, M. S., & Shahid, S. Spatial distribution of unidirectional trends in climate and weather extremes in Nile river basin[J]. Theoretical and Applied Climatology, 137(1), 1181-1199(2019).

[23] Chervenkov, H., & Slavov, K. Theil-Sen estimator vs. ordinary least squares–trend analysis for selected ETCCDI climate indices[J]. Comptes Rendus Acad. Bulg. Sci, 72, 47-54(2019).

[24] Salmerón, R., García, C. B., & García, J. Variance inflation factor and condition number in multiple linear regression[J]. Journal of Statistical Computation and Simulation, 88(12), 2365-2384(2018).

[25] Bakhshandeh, E., Jamali, M., Afshoon, E., & Gholamhossieni, M. Using hydrothermal time concept to describe sesame (Sesamum indicum L.) seed germination response to temperature and water potential[J]. Acta Physiologiae Plantarum, 39(11), 1-9(2017).

[26] Xue L. Study on simulation model of sesame growth and development [D]. Nanjing Agricultural University(2012).

[27] Singh, B., Chastain, D. R., Jumaa, S., Wijewardana, C., Redoña, E. D., Gao, W., & Reddy, K. R. Projected day/night temperatures specifically limits rubisco activity and electron transport in diverse rice cultivars[J]. Environmental and Experimental Botany, 159, 191-199(2019).

[28] Zhou M, Li G Q, Zhang J T, et al. Sesame dry matter and yield after waterlogging during full flowering[J]. Chinese Journal of Oil Crop Sciences, 38(05):598-604(2016).

[29] Zhang Gende, Wang Baoqin, Du Zhenwei, et al. Review and prospect of sesame breeding achievements in Henan Province[J]. Journal of Henan Agricultural Sciences, 46(10):32-37(2017).

Edge Temporal Anti-Aliasing

Zhi Xu[1,2,3], Yuhang Luo[1], Daojing He[3], Yiping Qin[1], Xin Wang[1]*, SiTao Peng[4], JianFan Yao[4]
[1]School of Computer Science and Information Security, Guilin University of Electronics Technology, No.1, Jinji road, Guilin, China, 541004;
[2]Guangxi Key Laboratory of Precision Navigation Technology and Application, Guilin, China, 541004;
[3]School of Computer Science and Technology, Harbin Institute of Technology, Shenzhen, China, 518055.
[4]China Nuclear Power Technology Researc Institute Co. Ltd., Shenzhen, China, 518028;
* Xin Wang: wxin@guet.edu.cn

Abstract

Temporal Anti-Aliasing (TAA) is a popular method for eliminating temporal aliasing problems. However, the images simply processed by TAA become blurred and lose some details. In this paper, an improved TAA algorithm named Edge Temporal Anti-Aliasing (ETAA), is proposed. A time iterative edge detection method is designed to enhance the detection accuracy of pixels with temporal aliasing and spatial aliasing. These pixels are blurred and blended, and other pixels are replaced with the pixels of the current frame. Furthermore, an approximate minimum filter is used to eliminate the flickering phenomenon of high-energy noise. Compared to TAA, ETAA outperforms TAA in terms of image detail preservation when the rendering camera moves. Meanwhile, the time cost of proposed method also can satisfy the requirement of real-time rendering. Experiment results show that ETAA can effectively and quickly eliminate the time aliasing when the camera moves, and achieves a better trade-off in flicker, ghosting and blurring.

Keywords-Temporal aliasing, Edge detection, Approximate minimum filter

1. Introduction

In computer graphics, the process of image rendering is indispensable for a sampling restoration task, which samples points in 3D space and restores them into 2D images. In 1983, Thornhill et al.[5] provided theoretical support for temporal aliasing. According to the Nyquist sampling theorem, the signal can be completely reconstructed only if the sampling frequency is higher than twice the highest frequency of the signal. In digital image processing, image aliasing can be generated as the sampling mode does not consistent to the variation of the image signal. When the appearance of a signal is not clear, it is impossible to accurately control the sampling frequency and sampling mode. Therefore, aliasing problems are inevitable for sampling tasks. Increasing the image resolution is an effective way to relieve the aliasing problem, but the burden for real-time rendering also significantly increases, thus it is necessary to perform the anti-aliasing method.

The aliasing generated in the rendering process is usually divided into two categories according to the performance, including Geometric Aliasing and Shading Aliasing. Geometric Aliasing is usually caused by insufficient sampling in the rasterization stage, which is reflected in the aliasing of the geometric edges of objects in image. Shading Aliasing is generated in the fragment shader stage and mainly reflects the flickering of edge pixels in successive frames. In the process of calculating the color of each pixel, due to the under-sampling of rendering equation (insufficient data precision, formula error, etc.), the difference value of pixel color between frames is too large. At the same time, if the rendering frame rate is insufficient, the flickering will be exacerbated. The temporal anti-aliasing algorithm is mainly used to solve the Shading Aliasing. Shading Aliasing tends to occur in Physically Based Rendering (PBR) lighting that uses low roughness values. Lower roughness values tend to produce narrower frequency cycles, which makes PBR's specular reflections a high-frequency signal, so it is readily to result in under-sampling. Therefore, we propose ETAA algorithm, our contributions are concluded as follows:

- A temporal edge detection algorithm is proposed. Compared with the traditional single-frame image edge detection algorithm, the time-aliased pixels can be more accurately identified, and edge up-sampling is used to improve the edge detection accuracy.

International Conference on Computer Graphics, Artificial Intelligence, and Data Processing (ICCAID 2022)
edited by Ruishi Liang, Jing Wang, Proc. of SPIE Vol. 12604, 126040X
© 2023 SPIE · 0277-786X · doi: 10.1117/12.2674546

- An approximate minimum filter is designed, instead of tone mapping, in order to better mitigate flickering of high-energy noise.

- Discarding the use of motion vectors as edge detection information, which eliminates the problem of flickering when the rendering camera is still.

2. Related Work

TAA algorithm is derived from motion blur, which is actually a photography technique in photography. Back in 2002, Sung et al.[6] used a motion blur algorithm to solve geometry and shadow aliasing. If the scene captured by the camera changes as the camera is exposing, it will produce a blurry picture. In general, the rendering camera in the computer does not have the physical phenomenon of exposure, and the rendered image has sharp edges and corners without motion blur. In recent years, however, with the development of physically-based cameras, this limitation has begun to disappear. Before that, there was a way to accurately simulate this effect by rendering more frames than output and applying a temporal filter to the multi-frame results, called cumulative buffering. But, like the spatial super-sampling, this is very time-consuming. A common improvement is to generate a per-pixel motion vector[14-17] for the historical frame and the current frame, use it to sample the historical frame to obtain samples of the historical pixels, and then perform cumulative blending. The results may be quite reasonable under certain conditions, but in many cases some artifacts are produced due to the lack of sampling points for occluded objects. So far, many anti-aliasing methods have been proposed and verified to be effective, which can be roughly divided into two categories: single-frame anti-aliasing and continuous-frame (temporal) anti-aliasing.

2.1. Single frame anti-aliasing

This kind of methods can be further divided into two classifications: spatial anti-aliasing and post-processing anti-aliasing. The spatial anti-aliasing method is inspired by the sampling theorem, increasing the number of samples per pixel and using super-sampling to alleviate aliasing. For example, Super Sample Anti-Aliasing (SSAA) proposed by Beets et al.[9], which renders the result image by pre-computing several times the resolution of it, and down-sampling it before display. Each pixel color is obtained by mixing with multiple samples. Although this method performs well, it is time-consuming and not suitable for deferred rendering. The commonly used hardware anti-aliasing method Multi-Sample Anti-Aliasing (MSAA) is an improved version of SSAA. Unlike SSAA, the color of the pre-computed multi-resolution image pixel is copied from the color of the parent pixel according to the coverage of the sub-pixel, which can reduce a lot of computation. Post-processing anti-aliasing[18-19] is inspired by techniques commonly used in image processing. Generally, when the rendered image is obtained, in the post-processing stage, the aliasing pixels of the geometric edge are found by borrowing the cache data and the high-pass filter, and then these aliased edge pixels are mixed to generate an anti-aliased image. Such as Morphological Anti-Aliasing[10-12] (MLAA) and Fast Approximate Anti-Aliasing [13] (FXAA). This kind of methods performs well in terms of performance, but is generally less effective than spatial antialiasing. However, all of the above methods only alleviate geometric aliasing in rendering, and have little improvement in shading aliasing.

2.2. Temporal anti-aliasing

Temporal anti-aliasing is also known as Temporal Super-Sampling, and is widely used in today's 3D game engines. TAA is an anti-aliasing technique that performs super-sampling by collecting sub-pixel samples across multiple consecutive frames. By reprojecting the shading results of historical frames with motion vectors, blending multiple samples per pixels in successive frames can effectively alleviate geometric and shading aliasing. It produces images, which is comparable to single-frame anti-aliasing without sacrificing too much performance, but TAA may loss some details, resulting in more blurry images than single-frame anti-aliasing. There have been many works on the comprehensive elaboration of TAA. References[1-3] elaborate on the technical methods involved in temporal anti-aliasing and the related issues involved. Podee N et al.[4] also used TAA to solve the scintillation problem of water wave reflection. The Spatiotemporal Variance-guided Filtering[16] (SVGF) proposed by Schied et al. utilizes spatiotemporal filtering to remove noise, which has many similarities with temporal anti-aliasing, and also shows that temporal anti-aliasing can also be used for image denoising. Marrs A et al.[20] proposed the Adaptive Temporal Antialiasing (ATAA) algorithm, which detects various aliasing areas for images, and uses different anti-aliasing algorithms to comprehensively perform antialiasing. Since TAA accumulates samples in time, it is usually implemented as a single post-processing iteration. This not only works for deferred rendering, but also works with single frame anti-aliasing.

3. ETAA

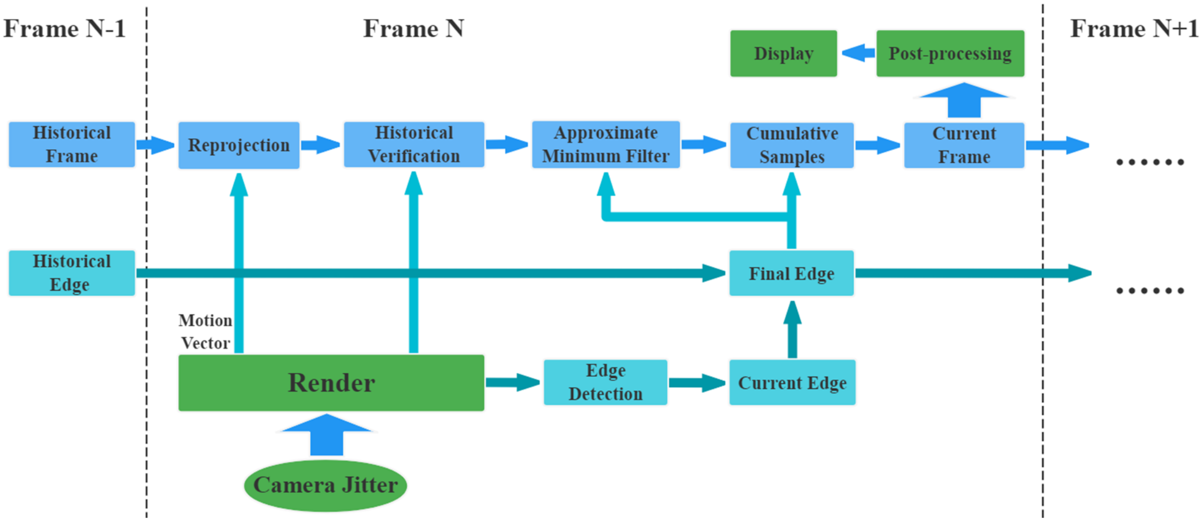


Figure 1.Flow diagram of ETAA

The ETAA algorithm process and module are shown in Figure 1, which can be roughly divided into two sub-processes. First, in order to sample each pixel uniformly, the jitter offset of the sub-pixels extracted from the sampling sequence is used to shift the viewport for each frame. The screen coordinates of the historical frame are obtained by reprojecting the screen coordinates and motion vectors of the current frame, and then sampling to obtain historical sample. History validation is a key module of the TAA algorithm as it identifies outdated historical data and rejects or corrects the data to avoid introducing errors into the current frame. Approximate minimum filtering on samples of the current frame suppresses high-energy noise pixels before accumulating samples. For each pixel, historical samples and current frame samples are blended by using edge blending weights. Finally, the resulting image is cached and post-processed for display. Second, using the motion vector to reproject the edge of the historical frame to obtain the historical edge information (edge weight, edge direction, etc.), and then performing edge detection on the current frame image to obtain the current frame edge information. The combines to form the final edge, and the edge result is cached for the next frame.

3.1. Edges based on time iterations

In order to accurately detect spatial and temporal aliasing pixels, we assume that temporal aliasing pixels appear at the intersection of successive frame edge pixels. In this paper, we propose an edge detection strategy based on time iteration. For the current frame, thin and thick edges are detected separately. Thin edges are mainly used to detect spatially aliased pixels of the current frame. Thick edges are used to verify historical edges, in order to improve edge accuracy and prevent edges from spreading over time. Additionally, temporal aliasing pixels are detected in conjunction with historical edges. The purpose of edge verification is to improve edge accuracy while prevent the spread of edges. Finally, the edge detection results are cached and used for the next frame. The edge detection idea is shown in the Figure 2.

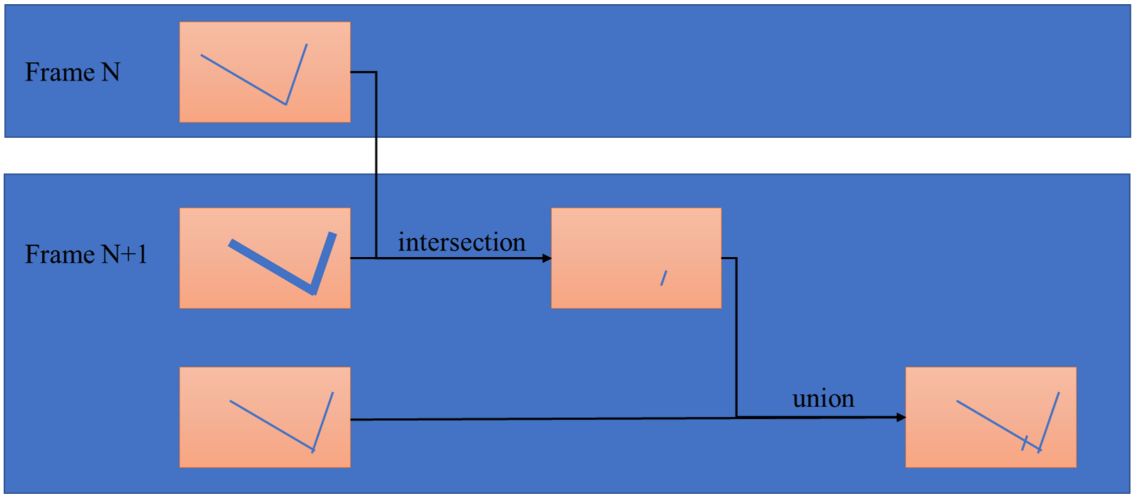
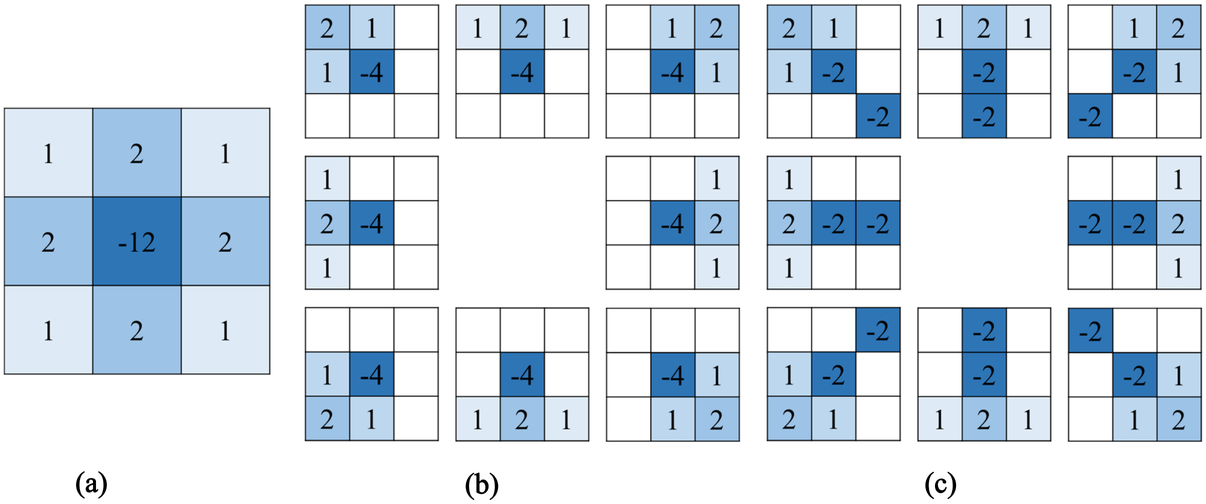


Figure 2. Edge detection strategy based on time iteration

In the edge detection module, the detection operators are shown in Figure 3, including point detection operators and line detection operators in 8 directions, and it is divided into fine edge detection and thick edge detection operators according to the pixel span.



(a)

$$\begin{bmatrix} 1 & 2 & 1 \\ 2 & -12 & 2 \\ 1 & 2 & 1 \end{bmatrix}$$

(b)

$$\begin{bmatrix} 2 & 1 & \\ 1 & -4 & \\ & & \end{bmatrix} \quad \begin{bmatrix} 1 & 2 & 1 \\ & -4 & \\ & & \end{bmatrix} \quad \begin{bmatrix} & 1 & 2 \\ & -4 & 1 \\ & & \end{bmatrix}$$

$$\begin{bmatrix} 1 & & \\ 2 & -4 & \\ 1 & & \end{bmatrix} \quad \begin{bmatrix} & & 1 \\ & -4 & 2 \\ & & 1 \end{bmatrix}$$

$$\begin{bmatrix} & & \\ 1 & -4 & \\ 2 & 1 & \end{bmatrix} \quad \begin{bmatrix} & & \\ & -4 & \\ 1 & 2 & 1 \end{bmatrix} \quad \begin{bmatrix} & & \\ & -4 & 1 \\ & 1 & 2 \end{bmatrix}$$

(c)

$$\begin{bmatrix} 2 & 1 & \\ 1 & -2 & \\ & & -2 \end{bmatrix} \quad \begin{bmatrix} 1 & 2 & 1 \\ & -2 & \\ & -2 & \end{bmatrix} \quad \begin{bmatrix} & 1 & 2 \\ & -2 & 1 \\ -2 & & \end{bmatrix}$$

$$\begin{bmatrix} 1 & & \\ 2 & -2 & -2 \\ 1 & & \end{bmatrix} \quad \begin{bmatrix} & & 1 \\ -2 & -2 & 2 \\ & & 1 \end{bmatrix}$$

$$\begin{bmatrix} & & -2 \\ 1 & -2 & \\ 2 & 1 & \end{bmatrix} \quad \begin{bmatrix} & -2 & \\ & -2 & \\ 1 & 2 & 1 \end{bmatrix} \quad \begin{bmatrix} -2 & & \\ & -2 & 1 \\ & 1 & 2 \end{bmatrix}$$

Figure 3. Edge detection operators. (a) Point detection operator. (b) Thin edge detection operator. (c) Thick edge detection operator.

In order to improve the accuracy of edge detection, we also perform edge up-sampling, which use a rendering pass in advance to cache the detected edge information into a rendering texture larger than the current screen resolution. Figure 4 compares the blur mask of TAA and ETAA. Compared with the TAA algorithm that directly uses the motion vector as the blur mask of the image, the blur mask calculated by ETAA can accurately blur the pixels with flickering edges and reduce unnecessary blur, improve the clarity of the resulting image.

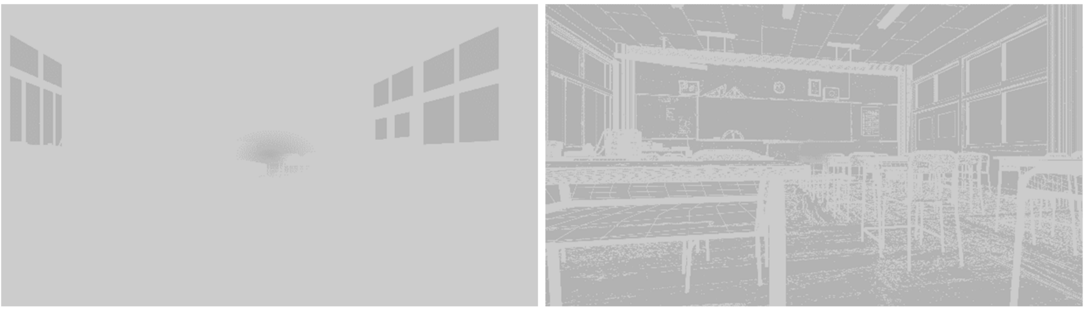

(a) (b)

Figure 4. (a) TAA’s blur mask. (b) ETAA’s blur mask.

3.2. Approximate minimum filter

The flickering of high-energy noise pixels is one of the most difficult problems in TAA. Traditional TAA algorithms solve it by using tone mapping, compressing colors in HDR space, then transform the result to the original color space. This method is cumbersome and difficult to control, requires a reasonable mapping function, and is not necessarily suitable for all scenarios. Sometimes the effect of suppressing high-energy noise is not ideal.

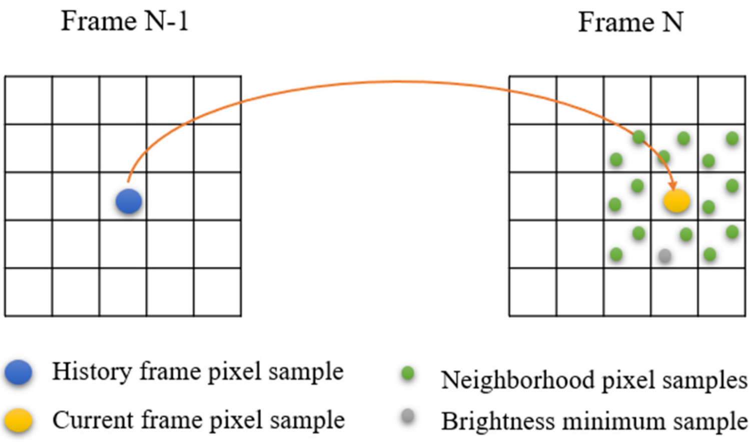

Frame N-1
Frame N
● History frame pixel sample
● Current frame pixel sample
● Neighborhood pixel samples
● Brightness minimum sample

Figure 5. Approximate minimum filter schematic.

ETAA proposes an approximate minimum filter. The pixels involved in the calculation are shown in Figure 5. It first gets the pixel samples with the smallest brightness (gray dots) according to the neighboring pixels of the current pixel, and blends the current pixel samples (big yellow dots) with them according to the edge weights of the current frame. Then the historical pixel samples (big blue dots) are mixed with them according to their edge weights and the result is used in the subsequent sample accumulation module. By introducing edge weights at the same time of filtering, the filtering calculation can be limited to the edge region and unnecessary filtering calculations can be reduced. We compare the proposed method with the currently commonly used tone mapping methods, and the effect is shown in Figure 6. We can see high-energy noise in the corners of the window, and the image on the right is cleaner than the image on the left. The results show that our method can significantly remove high-energy noise and can significantly reduce flickering edge pixels when the camera moves frequently.

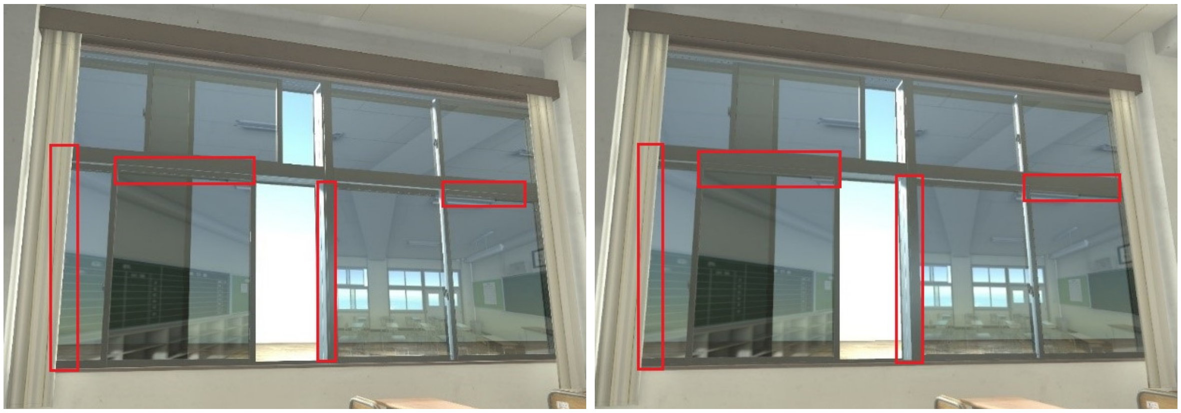

(a) (b)

Figure 6. (a) Tone mapping. (b) Approximate minimum filter.

4. Experimental Results

In this section, the proposed method is compared with TAA in terms of both image quality and real-time.

4.1. Image quality

We used two image quality evaluation parameters, namely peak signal-to-noise ratio (PSNR) and structural similarity (SSIM). The larger the PSNR, the more similar it is to the original image; The closer the SSIM is to 1, the more similar it is to the original image. Figure 7 shows three corners from the classroom scene, the projector, the schedule and the window corner. It can be seen that TAA has stronger blurring of geometric edges, more obvious edge overflow, higher image blurring, and weaker suppression of high-energy noise than ETAA.

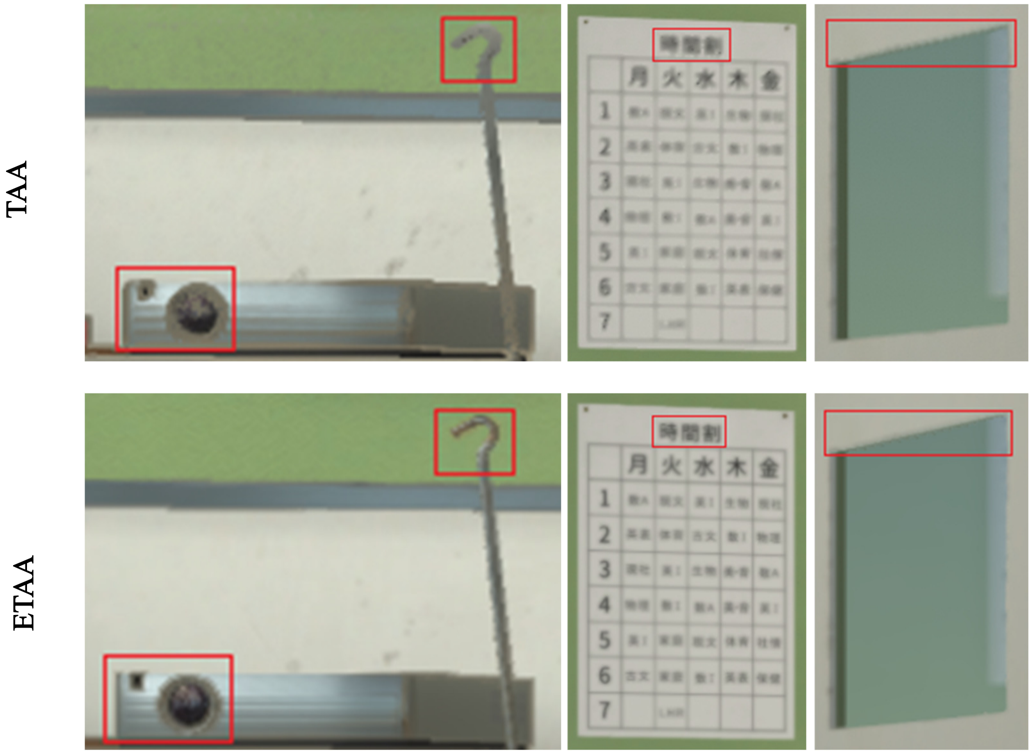


Figure 7. Detail of a classroom scene, TAA and ETAA. (3840x2160)

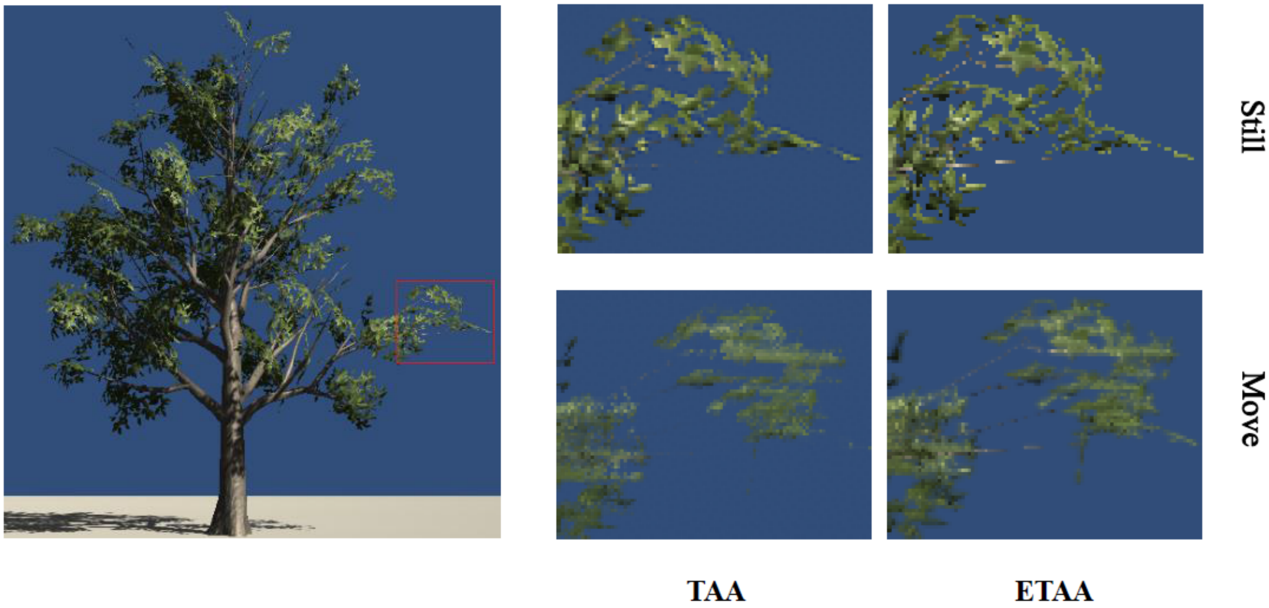


Figure 8． Visual representation of noisy leaves at still and move, TAA and ETAA.

The scene in Figure 8 contains a tree that records foliage when the render camera is still and move. When the rendering camera is still, because TAA uses the motion vector of each pixel as the blur parameter, but due to camera jitter, the motion vector is not 0 when it is still, so a certain blur will also occur. Blur is stronger when rendering camera motion, TAA causes some detail to disappear, ETAA performs better.

Table 1. Average image quality assessment at 1366x768 resolution.

Method	ETAA	TAA	CTAA
PSNR	**21.0439**	19.6957	20.1024
SSIM	**0.6640**	0.6133	0.6220

Table 2. Average image quality assessment at 3840x2160 resolution.

Method	ETAA	TAA	CTAA
PSNR	**22.2340**	20.1779	21.8926
SSIM	**0.7593**	0.7289	0.7468

Tables 1 and Tables 2 present the average performance of image quality at two different resolutions. It can be seen that ETAA is better than TAA in the preservation of image details while eliminating flickering. For 4K resolution, SSIM improves by about 0.03 and PSNR improves by about 2.0. Furthermore, as the resolution decreases, ETAA gets better.

4.2. Real-time analysis

Real-time comparison results are based on DirectX11 API built into Unity engine, hardware devices: CPU (12th Gen Intel(R) Core (TM) i7-12700H 2.30 GHz, RAM 32G) and GPU (Nvidia GeForce RTX 3060 Laptop (6 GB)). We summarize the total time to render the image and its individual GPU time under different anti-aliasing algorithms. The specific data are shown in Table 3.

From Table 3, it can be seen that the individual GPU time of the algorithm is not affected by the number of scene triangles, because the algorithm is executed in the post-processing stage and is independent of the scene. Due to the additional edge detection computation, the individual GPU time of ETAA increases the cost by about 0.4ms compared to TAA. When there are too many triangles in the scene, their individual GPU time accounts for a small proportion of the total time, which has little effect on the total time of rendering the scene, and can also meet the needs of 33ms real-time rendering.

Table 3. Time-consuming per frame in scenes with different triangle counts. (3840x2160)

Triangles	Method	Total Times(ms)	GPU Times(ms)
46K	NoAA	1.8-2.0	ref
	TAA	2.7-3.0	0.7-0.9
	CTAA	2.6-2.9	0.9-1.1
	ETAA	3.1-3.3	1.2-1.4
2.9M	NoAA	23.5-24.5	ref
	TAA	25.0-26.5	0.7-0.9
	CTAA	25.0-26.5	0.9-1.1
	ETAA	25.0-26.5	1.2-1.4

5. Conclusion

In this paper, a time iterative edge detection algorithm is designed based on the TAA algorithm. Compared with TAA, it can accurately eliminate edge aliasing and improve image quality. In addition, the approximate minimum filter is used to replace the tone mapping, which can better suppress the flicker of high energy noise pixels. For future work, we will introduce a certain blending method at the edge to replace the corresponding pixel blending, and optimize the existing edge detection algorithm to reduce the time complexity of ETAA algorithm.

Acknowledgments

This work was supported by Guangxi Science and Technology Major Project No. AA19254016, Guangxi Natural Science Foundation No.2020GXNSFAA297186, Jiangsu Province Agricultural Science and Technology Innovation and Promotion Special Project No. NJ2021-21, Guilin Key Research and Development Program No.20210206-1 and Project supported by Guangxi Key Laboratory of Precision Navigation Technology and Application, Guilin University of Electronic Technology (No. DH202227).

References

[1] Yang, L., Liu, S., & Salvi, M. 2020. A survey of temporal antialiasing techniques. *In Computer graphics forum* (Vol. 39, No. 2, pp. 607-621).

[2] Karis, B. 2014. High-quality temporal super sampling. *Advances in Real-Time Rendering in Games, SIGGRAPH Courses*, 1(10.1145), 2614028-2615455.

[3] Pedersen, L. J. F. 2016. Temporal reprojection anti-aliasing in INSIDE. *In Game Developers Conference* (Vol. 3, No. 4, p. 10).

[4] Podee, N., Max, N., Iwasaki, K., & Dobashi, Y. 2021. Temporal and spatial anti-aliasing for rendering reflections on water waves. *Computational Visual Media*, 7(2), 201-215.

[5] Thornhill, R. J., and Smith, C. C. 1983. "Time Aliasing: A Digital Data Processing Phenomenon." *ASME. J. Dyn. Sys., Meas., Control*. 105(4): 232–237.

[6] Sung, K., Pearce, A., & Wang, C. 2002. Spatial-temporal antialiasing. *IEEE Transactions on Visualization and Computer Graphics*, 8(2), 144-153.

[7] Oliveros Labrador, C. A. 2018. Improved Sampling for Temporal Anti-Aliasing (A Sobel Improved Temporal Anti-Aliasing). *LU-CS-EX* 2018-02.

[8] Dachille, F., & Kaufman, A. 2000. High-degree temporal antialiasing. *In Proceedings Computer Animation 2000* (pp. 49-54). IEEE.

[9] Beets, K., & Barron, D. 2000. Super-sampling anti-aliasing analyzed.

[10] Reshetov, A. 2009. Morphological antialiasing. *In Proceedings of the Conference on High Performance Graphics 2009* (pp. 109-116).

[11] Navarro, F., & Gutierrez, D. 2011. Practical Morphological Antialiasing. *GPU Pro 2*, 2, 95.

[12] Biri, V., Herubel, A., & Deverly, S. 2010. Practical morphological antialiasing on the GPU. *In ACM SIGGRAPH 2010 Talks* (pp. 1-1).

[13] Lottes, T. 2009. Fxaa. *White paper, Nvidia*, Febuary, 2.

[14] Lauritzen, A., & Maps, S. A. V. S. 2007. GPU Gems 3.

[15] Hanika, J., Tessari, L., & Dachsbacher, C. 2021. Fast Temporal Reprojection without Motion Vectors. *Journal of Computer Graphics Techniques* Vol, 10(3).

[16] Schied, C., Kaplanyan, A., Wyman, C., Patney, A., Chaitanya, C. R. A., Burgess, J., ... & Salvi, M. 2017. Spatiotemporal variance-guided filtering: real-time reconstruction for path-traced global illumination. *In Proceedings of High Performance Graphics* (pp. 1-12).

[17] Zeng, Z., Liu, S., Yang, J., Wang, L., & Yan, L. Q. 2021. Temporally Reliable Motion Vectors for Real - time Ray Tracing. *In Computer Graphics Forum* (Vol. 40, No. 2, pp. 79-90).

[18] Huimin, D. U. , Qinqin, D. U. , Kaibo, J. I. , Jiang, B. , & Guo, C. . 2016. Survey on the post-processing anti-aliasing techniques. *Journal of Xi'an University of Posts and Telecommunications.*

[19] Malan, H. 2018. Edge Antialiasing by Post-Processing. *In GPU Pro 360 Guide to Image Space* (pp. 33-58).

[20] Marrs, A., Spjut, J., Gruen, H., Sathe, R., & McGuire, M. 2018. Adaptive temporal antialiasing. *In Proceedings of the Conference on High-Performance Graphics* (pp. 1-4).

[21] Banterle, F., Ledda, P., Debattista, K., & Chalmers, A. 2006. Inverse tone mapping. *In Proceedings of the 4th international conference on Computer graphics and interactive techniques in Australasia and Southeast Asia* (pp. 349-356).

Structural design and finite element analysis of a milling machine beam

Weiwei Zhang*[a], Xiaowei An[a]
[a] School of Mechanical Engineering, Shandong Huayu University of Technology, Dezhou, China
* Corresponding author: zhang8412860@163.com

ABSTRACT

With the rapid development of the modern manufacturing industry, the requirements for milling machine products are getting higher and higher, and then the working accuracy of milling machines has higher requirements. The cross beam of the milling machine is a key component of the milling machine. The cross beam is connected to the top of the machine tool bed by bolts, which mainly play the role of connecting and supporting the saddle and the spindle sleeve. The beam of the milling machine bears most of the working load of the milling machine, so the rigidity and strength of the beam of the milling machine directly affect the working accuracy and service life of the milling machine. In this paper, three-dimensional software is used to establish the geometric model of the milling machine beam and establish a finite element model. The stiffness and strength of the saddle are analyzed by the finite element method according to the load condition of the saddle in the middle of the beam. Through analysis, it can be seen that the sliding saddle of the milling machine is in the worst working condition in the middle of the beam. According to the strength theory, it can be checked that its strength meets the requirements. This static analysis can provide reliable data for the design optimization of the milling machine beam.

Keywords: milling machine beam,finite element,static analysis

1. INTRODUCTION

The milling machine has many components and complex structures. The beam, saddle, and spindle sleeve are the key components of the milling machine. The beam is bolted to the top of the machine bed, which is mainly used to connect and support the saddle and spindle sleeve [1]. The spindle sleeve moves up and down on the cross beam of the milling machine through the guide rail, and the sliding saddle can drive the spindle sleeve to move horizontally left and right on the cross beam to achieve the cutting movement of the milling machine. The cross beam of the machine tool is mainly subjected to the cutting force of the tool, and the gravity of the saddle, spindle sleeve, and spindle. It can be seen from this that as the key components of the milling machine, the milling machine beam, sliding saddle, and spindle sleeve bear most of the working load of the milling machine, so the rigidity and strength of the milling machine beam, sliding saddle and spindle sleeve directly affect many key indicators such as the working accuracy and service life of the milling machine [2].

It can be seen that the static analysis of the milling machine beam, saddle, and spindle sleeve is required to verify whether their strength and stiffness meet the requirements.

2. THREE-DIMENSIONAL MODELING OF BEAM

2.1Modeling concept of the cross beam

The front view and rear view of the beam are shown in Figure 1 and Figure 2 respectively. To model the beam, you can first create two cuboids as the main body of the beam, then create rib plates on both sides, pull out the internal space with a sketch, and then use the sketch command to process other details.

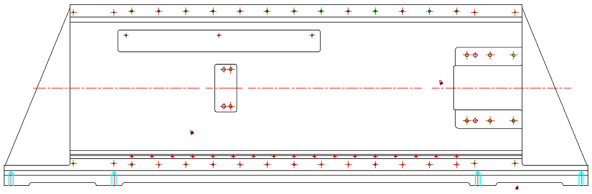


Fig.1 Front view of the beam

International Conference on Computer Graphics, Artificial Intelligence, and Data Processing (ICCAID 2022)
edited by Ruishi Liang, Jing Wang, Proc. of SPIE Vol. 12604, 126040Y

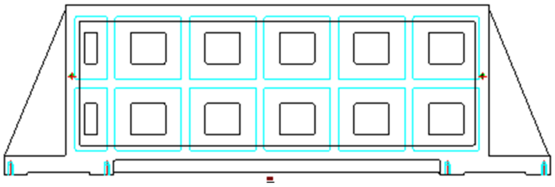
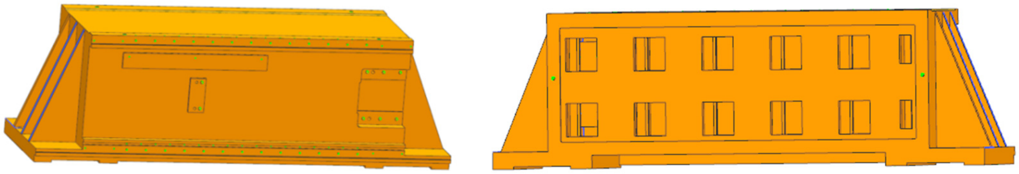


Fig.2 Rear view of the beam

2.2 Beam modeling process

First, wee use the cuboid command to create a cuboid with a length of 1337 mm, a width of 479 mm, and a height of 472 mm. Then, we establish the boss, rib plate, groove, hole, and other details at the corresponding position of the cuboid. The final beam geometric model is shown in Figure 3.



(a) Elevation geometric model of the beam (b) Rear view geometric model of beam

Fig.3 Three-dimensional geometric model of beam

3. FINITE ELEMENT MODELING AND STATIC ANALYSIS OF CROSS BEAM

3.1 Element type and material characteristics of the beam

The selection of the element type of the cross beam is an important preliminary preparation for its gridding [3]. Because the cross beam of the milling machine is complex in structure and irregular in spatial structure, the tetrahedral element is selected to better divide the grid and adapt to the complex structure of the cross beam of the milling machine. The 10-node tetrahedral element is more accurate than the 4-node tetrahedral element, so the 10-node tetrahedral element is finally selected for grid division.

There are 10 node tetrahedral elements. Each node has 3 degrees of freedom that translate along the x, y, and z coordinates respectively. The element degree of freedom is 30. The displacement array of element nodes is shown in Formula (1) [4].

$$\{\delta\}^e = [u_1 v_1 w_1 \ldots u_{10} v_{10} w_{10}]^T \tag{1}$$

The displacement mode of the 10-node tetrahedral element is to express the displacement u, v, w of any point in the element as a function of the coordinates x, y, z, and the displacement mode is shown in Formula (2).

$$\begin{cases} u(x,y,z) = a_1 + a_2 x + a_3 y + a_4 z + a_5 x^2 + a_6 y^2 + a_7 z^2 + a_8 xy + a_9 yz + a_{10} zx \\ v(x,y,z) = a_{11} + a_{12} x + a_{13} y + a_{14} z + a_{15} x^2 + a_{16} y^2 + a_{17} z^2 + a_{18} xy + a_{19} yz + a_{20} zx \\ w(x,y,z) = a_{21} + a_{22} x + a_{23} y + a_{24} z + a_{25} x^2 + a_{26} y^2 + a_{27} z^2 + a_{28} xy + a_{29} yz + a_{30} zx \end{cases} \tag{2}$$

During the static analysis of the beam, the material properties of the beam should be defined and assigned to the corresponding structure. The material of the beam is HT300, the material of the cutter bed is 45, and the material properties and mechanical properties of HT300, 20, and 45 are shown in Table 1 and Table 2[5].

Table 1 Material Properties

name	density/kg/m^3	Poisson's ratio	elastic modulus/Pa
HT300	7.3×10^3	0.270	1.43×10^{11}
20	7.8×10^3	0.282	2.13×10^{11}
45	7.89×10^3	0.269	2.09×10^{11}

Table 2 Mechanical Properties of Materials

name	elongation/%	yield strength/MPa	ensile strength/MPa
HT300	<5	—	300
20	25	245	—
45	16	355	—

3.2 Beam grid division

According to experience, the stress of the beam is very small, and the range of change of the stress is small. It is mainly to analyze the displacement of the beam, so uniform discretization can be adopted to facilitate grid division.

Meshing's MeshTool tool is used to set the element attributes of the divided structure. The side length of the control unit is 0.04 mm. The beam is discretized separately. The structure after grid division is shown in Figure 4. The final finite element model has 59398 elements and 108888 nodes.

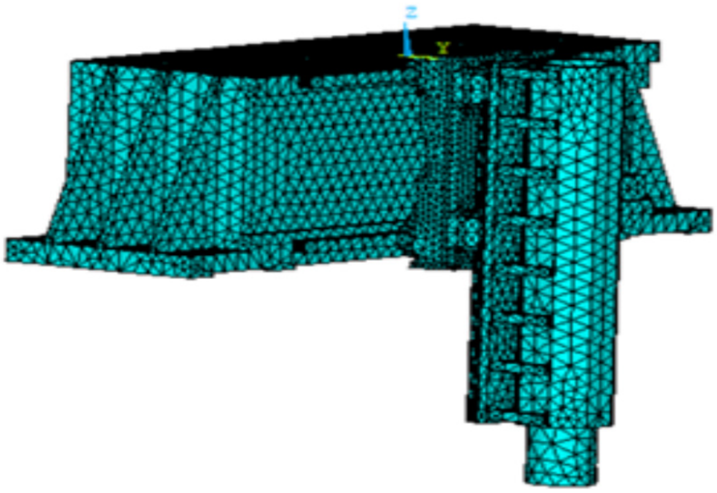


Fig.4 Finite element model

3.3 Analysis of calculation results

3.3.1 Displacement result analysis

The displacement vector sum of the beam is shown in Figure 5. It can be seen from the calculation results that the maximum displacement of the beam is 0.00189 mm, and the maximum displacement is at the middle position of the upper guide rail.

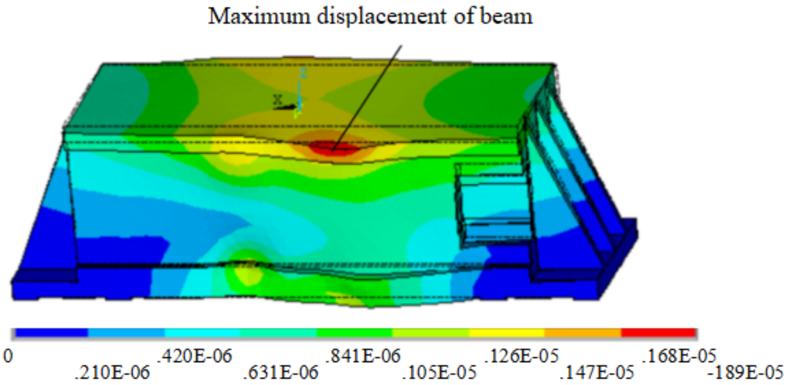


Fig.5 Beam displacement vector and cloud chart

The sum of displacement vectors of the overall structure is shown in Figure 6. It can be seen from the calculation results that the maximum displacement is 0.031 mm, and the maximum displacement is at the bottom of the tool apron.

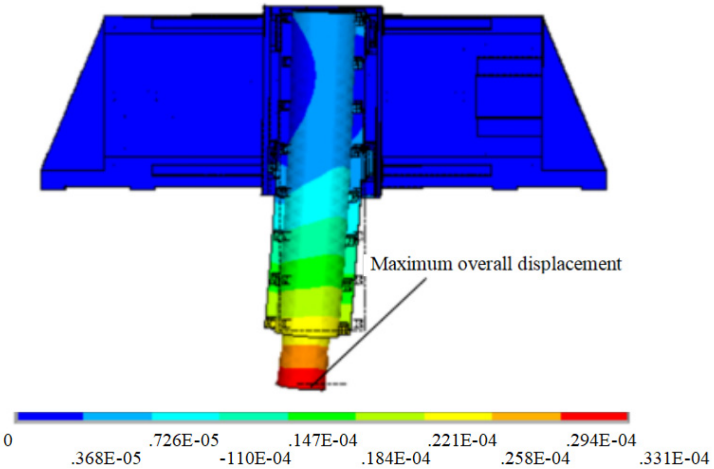


Fig.6 Displacement vector and cloud chart of the overall structure

3.3.2 Stress result analysis

The first principal stress of the beam is calculated as shown in Figure 7. It can be seen from the calculation results that the maximum first principal stress is 1.02 MPa, which is located in the middle of the guide rail placed on the beam.

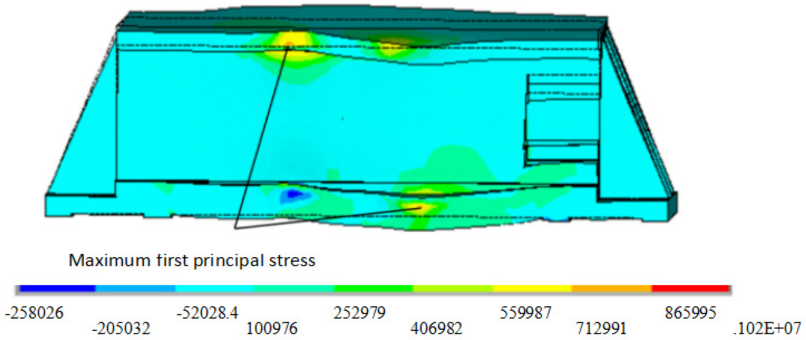


Fig.7 Cloud chart of the first principal stress of the beam

3.4 Cross-beam check

The first strength theory holds that when the maximum tensile stress of the material reaches the limit value of the tensile stress of the material, the material will fracture, which is mainly applicable to brittle materials. The material of the beam is HT300, the elongation of HT300 is less than 5%, which is a brittle material, and the strength failure form is a fracture, which is consistent with the first strength theory. Therefore, the first strength theory is applied to check the strength of the beam and saddle. The first principal stress of the beam and saddle should be checked in Ansys.

The beam is made of HT300 with tensile strength $\sigma_b = 300\mathrm{MPa}$. Take the safety factor n = 1.5. The allowable stress of the beam can be obtained $[\sigma] = \frac{\sigma_b}{n} = 200\mathrm{MPa}$. The maximum first principal stress of the beam $\sigma_1 = 1.02\mathrm{MPa} < [\sigma] < \sigma_b$, which meets the strength conditions.

3.5 Calculation, analysis, and comparison of multiple schemes

The multi-scheme calculation is carried out for the working condition of the sliding saddle in the middle of the beam and the main shaft sleeve at the low end of the sliding saddle. Different element side lengths are used for the discretization to study the influence of different node sizes on the calculation accuracy and cost.

By adopting different side lengths to discretize the beam, the solution time and overall maximum displacement under different node sizes are obtained, as shown in Table 3. With the increase in the number of nodes, the maximum displacement vector sum increases continuously and finally tends to be flat. As the number of nodes increases, the time of finite element calculation increases, and the cost of calculation increases. Therefore, we should take into account the accuracy and cost of calculation.

Table 3 Calculation Results of Different Node Numbers

number of nodes	maximum displacement vector sum/mm	absolute displacement error/mm	Relative displacement error/%	Solution time/s
42893	0.0307	0.0031	9.17	5
61264	0.0324	0.0014	4.14	8
108888	0.0331	0.0007	2.07	18
158217	0.0338	0	0	85
205087	0.0338	—	—	222

4. CONCLUSION

In this paper, according to the drawing of the cross beam of the milling machine, 3D modeling of the cross beam of the milling machine is carried out by using UG (Unigraphics NX), and the cross beam is simplified. The finite element model of the cross beam of the milling machine is established, and the finite element static analysis of the cross beam is carried out. The following conclusions are obtained:

The stresses of the cross beam of the milling machine are far less than the allowable stresses, so the stresses of the cross beam of the milling machine meet the strength requirements.

This paper analyzes the strength and stiffness of the cross beam of the milling machine, which provides a favorable basis for the modal analysis and research of the cross beam of the milling machine and the structural optimization of the cross beam.

ACKNOWLEDGMENTS

Shandong Huayu University of Technology School-level scientific research platform "Intelligent Manufacturing Equipment Design Engineering Technology Research and Development Center”.

The 2022 school-level scientific research plan project "Design of stainless-steel lower cover stamping manipulator of soybean milk machine".

REFERENCES

[1] Zhang, J.P. (2016) Finite Element Analysis and Structural Optimization of CNC Milling Machine [D]. Changchun University of Technology.

[2] Zhang, Z.J. (2020) Structural Analysis and Optimization Design of Moving Beam of Milling turning Composite Machining Center [D]. Lanzhou University of Technology.

[3] Chen, Q.T. (2021) Finite element analysis and optimization design of XK713 CNC milling machine based on ANSYS [D]. Southeast University.

[4] Liu, J.H. (2014) Structural Analysis and Optimization of XK2423 CNC Milling Machine Based on Finite Element Method [D]. Changchun University of Technology.

[5] Wang, W.B. (2001) Mechanical Design Manual. Volume 1 [M]. Beijing: China Machine Press.8:262-284.

Intelligent Prediction and Key Factor Analysis to Lost Circulation from Drilling data based on Machine Learning

Guangyao Wen[a], Huailong Chen[a], Tuo Zhou[*b], Chengwu Gao[a], Bahedaer Baletabieke[b], Haiqiu Zhou[b], Shan Wang[c]
[a] CNPC Amu Darya River Gas Exploration & Development Company, Beijing, 102299, China
[b] CNPC Engineering Technology R&D Company Limited, Beijing, 102206, China
[c] Beijing Chen Yu Jin Yuan Science and Technology Company Limited, Beijing, 100010, China
*Corresponding author email: zhoutuodr@cnpc.com.cn

Abstract

Lost circulation during drilling wells is very detrimental since it greatly increases the non-productive time and operational cost, also seriously lead to wellbore instability, pipe sticking, blow out, etc.. However, in the process of drilling wells, geological characteristics and operational drilling parameters all may have impacts to the lost circulation. This makes the establishment of the relations between the lost circulation and drilling factors very challenging. In this paper, we tested five different kernel function (linear, quadratic, cubic, medium Gaussian and fine Gaussian) derived support vector regression (SVR) models and four-layer artificial neural network (ANN). By combining their accuracy and time efficiency, the ANN is regarded as the optimal predictor of lost circulation. By training ANN using different combination of drilling features, we concluded that depth, torque, hanging weight, displacement, entrance density and export density are the key factors to accurate predict the lost circulation. The corresponding trained ANN network can achieve 99.2% accuracy and evaluate whether a drilling feature vector corresponds to lost circulation or not in milliseconds.

Keywords: Lost circulation, Machine learning, Artificial neural network, Support vector regression.

1. Introduction

Lost circulation refers to a kind of complex downhole situation in which various working fluids (including drilling fluid, cement slurry, completion fluid and other fluids, etc.) directly enter the formation under the effect of differential pressure in the process of downhole operations including drilling, cementing, testing or workover [1]. It is not only a common problem in drilling operations, but also a problem that has puzzled petroleum engineering for many years. Most drilling processes have different degrees of leakage, which causes the loss of drilling fluid and a large number of plugging materials, and affects the drilling speed and subsequent construction. Serious lost circulation will also lead to complex downhole accidents, causing serious losses, and even the abandonment of the entire well and the pollution of the formation. In the early stage, the problem of lost circulation is usually handled after it occurs, but it will greatly increase the non-production time during drilling [2]. Therefore, it is a challenge for industry to accurately predict the occurrence of lost circulation in real time, minimize the loss and harm caused by lost circulation, or prepare the required remedies to stop the risk.

In recent years, some domestic and foreign related lost circulation prediction and diagnosis technologies, such as fine pressure control drilling technology, comprehensive logging technology, separator liquid level detection method and downhole micro flow detection method [3], have made good development, which can be applied to intelligent early warning in oil drilling engineering. However, these methods for lost circulation mainly involve experts proposing solutions in combination with the site conditions of the lost circulation, relying too much on the experience of field experts to achieve pre-drill diagnosis and ignoring other lost circulation decision-making schemes. Machine learning and other artificial intelligence methods can better realize lost circulation prediction and make better decisions, which are now widely used in the field of drilling engineering. Al Hameedi et al. [4][5] established a multiple linear regression model to predict the lost circulation of Rumaila Oilfield in Iraq; In addition, literature [6] also uses partial least squares (PLS) regression to predict the lost circulation before drilling; Reza et al. [7] used ANN to predict the lost circulation in naturally fractured reserves, and analyzed the influence of geo-mechanical parameters and drilling operation variables according to two different ANN models developed, but some geological parameters are difficult to obtain before drilling; Alkinni et al. [8] also used artificial neural networks to predict the lost circulation in induced fractures formations before drilling, using the lost circulation data extracted from more than 1500 wells around the world; In addition, Li et al. [9] used BP neural network,

International Conference on Computer Graphics, Artificial Intelligence, and Data Processing (ICCAID 2022)
edited by Ruishi Liang, Jing Wang, Proc. of SPIE Vol. 12604, 126040Z
© 2023 SPIE · 0277-786X · doi: 10.1117/12.2674534

support vector machine and random forest algorithm for supervised learning and established a risk prediction model for lost circulation during drilling.

The purpose of this paper is to investigate the relation between the lost circulation and drilling parameters and provide an accurate predictor for lost circulation with high time-efficiency. To realize this, we firstly collected the drilling data from 59 wells and built a drilling dataset with labels which refers to lost circulation or not. On the basis of the dataset, we employed six different famous machine-learning algorithms to establish the relationship between drilling data and lost circulation. By comparing their final prediction accuracy and time efficiency, we regarded the best algorithm as our optimal predictor. Based on the optimal predictor, we computed and compared the accuracy when using different drilling features. By balancing the accuracy and time efficiency of predictor, the corresponding factors are the key factors of lost circulation. The diagram of the proposed model is shown in figure 1. The main contribution of this work is to validate whether it is possible to predict the lost circulation accurately in real-time when there is not enough recorded drilling data.

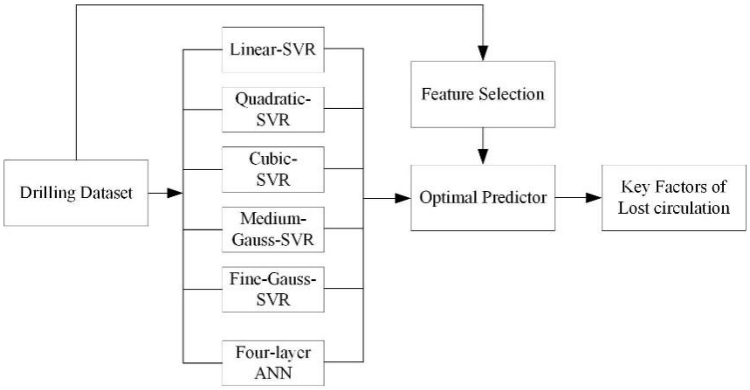


Figure 1. Overview of the proposed method.

2. Methodology

2.1 Data preparation

Dataset utilized to predict the lost circulation from drilling datas is from 59 different wells. In drilling the wells, 22 different feature values are recorded. The prediction of lost circulation in this paper is just to classify the recording datas into two classes, i.e. with lost circulation and without lost circulation. The distribution of the 22 features values are summarized in Table 1, as follows:

Table 1. Distribution of the recorded feature values in drilling wells

Features	Minimum	Maximum	Features	Minimum	Maximum
Depth (m)	14	4398	Pool volume (m^3)	23	306.54
Drilling time (min/m)	0.0083	665.2	Diff. volume (m^3)	-1579.14	209.6
WOB (KN)	0.1614	2080.3	Hydrocarbons (%)	0	102.27
Torque (KN.m)	0.001	55	C1 (%)	0	100
Hanging weight (KN)	122.4	3999	C2 (%)	0	23
Pump pressure (MPa)	0.0034	29.71	C3 (%)	0	23
Displacement (L/S)	0.06	637.7	IC4 (%)	0	23
Entrance density (g/cm^3)	0.03	23	nC4 (%)	0	23
Export density (g/cm^3)	0.0041	23	IC5 (%)	0	23
Inlet temp. (^{o}C)	0.1	111.3	nC5 (%)	0	58
Outlet temp. (^{o}C)	0.1	119.4	Lithology	0	241

2.2 Multi-Kernel derived SVR Algorithms

The prediction of lost circulation is just to classify the recorded drilling data into two parts. Considering lost circulation is a high-dimensional and multi-factor influenced problem, it is difficult to exploit a simple linear hyperplane to separate the data. In this section, we utilize kernel function derived SVR algorithms.

In the traditional SVR model, we can express $f(x) = \boldsymbol{w^T x} + b$ as a minimization problem and donate it as follows:

$$\min_{\boldsymbol{w},b} \left(\frac{1}{2} \|\boldsymbol{w}\|^2 + C \sum_{i=1}^{N} \ell_\epsilon(f(\boldsymbol{x_i}) - y_i) \right) \tag{1}$$

$$\ell_\epsilon(\boldsymbol{x_i}) = \begin{cases} 0 & |\boldsymbol{x_i}| \le \epsilon \\ |\boldsymbol{x_i}| - \epsilon & otherwise \end{cases} \tag{2}$$

where C is a regularization parameter and a pre-defined constant value. ℓ_ϵ is an insensitive loss function. After introducing the slack variables ξ_i and ξ_i', Eq.(1) can be re-formulated as follows:

$$\min_{\boldsymbol{w},b,\xi_i,\xi_i'} \left(\frac{1}{2} \|\boldsymbol{w}\|^2 + C \sum_{i=1}^{N} (\xi_i + \xi_i') \right) \tag{3}$$

and Eq. (3) is conditioned the following equations:

$$\begin{cases} f(\boldsymbol{x_i}) - y_i \le \epsilon + \xi_i \\ y_i - f(\boldsymbol{x_i}) \le \epsilon + \xi_i' \\ \xi_i \ge 0, \xi_i' \ge 0, i = 1,2,\cdots,N \end{cases} \tag{4}$$

By using Lagrangian multipliers, Eqs.(3) and (4) can be expressed as follows:

$$\mathrm{L}(\boldsymbol{w}, b, \mu_i, \mu_i', \alpha_i, \alpha_i', \xi_i, \xi_i') = \begin{array}{c} \frac{1}{2} \|\boldsymbol{w}\|^2 + C \sum_{i=1}^{N} (\xi_i + \xi_i') - \sum_{i=1}^{N} \mu_i \xi_i - \sum_{i=1}^{N} \mu_i' \xi_i' \\ + \sum_{i=1}^{N} \alpha_i (f(\boldsymbol{x_i}) - y_i - \epsilon - \xi_i) + \sum_{i=1}^{N} \alpha_i' (y_i - f(\boldsymbol{x_i}) - \epsilon - \xi_i') \end{array} \tag{5}$$

By computing the partial derivative of Eq. (5) to $\boldsymbol{w}, b, \xi_i, \xi_i'$, respectively, we can obtain the dual representation, as follows:

$$\begin{gathered} \max_{\alpha_i,\alpha_i'} \sum_{i=1}^{N} y_i(\alpha_i' - \alpha_i) - \epsilon(\alpha_i + \alpha_i') - \frac{1}{2} \sum_{i=1}^{N} \sum_{j=1}^{N} (\alpha_i' - \alpha_i)(\alpha_j' - \alpha_j) \boldsymbol{x_i^T x_j} \\ s.t. \sum_{i=1}^{N} (\alpha_i' - \alpha_i) = 0, 0 \le \alpha_i, \alpha_i' \le C \end{gathered} \tag{6}$$

where α_i,$\alpha_i'\alpha_j$,α_j' can be solved by sequential minimal optimization [10].

By substituting partial deviation results of Eq. (5) to $\boldsymbol{w}$, we can get $f(\boldsymbol{x}) = \sum_{i=1}^{N} (\alpha_i' - \alpha_i) \boldsymbol{x_i^T x} + b$, and according to Karush-Kuhn-Tucker conditions [11], we can obtain $b = y_i + \epsilon - \sum_{j=1}^{N} (\alpha_j' - \alpha_j) \boldsymbol{x_j^T x_i}$.

To make the classification much more accurate, we utilize six different kernel functions to map the input drilling data into higher feature space. The kernel function is just utilized to compute the in the higher feature space. The final classification function can be denoted as follows:

$$f(\boldsymbol{x}) = \sum_{i=1}^{N} (\alpha_i' - \alpha_i) \kappa(\boldsymbol{x}, x_i) + b \tag{7}$$

where $\kappa(\cdot)$ is the kernel function and $\kappa(\boldsymbol{x_i}, \boldsymbol{x_j}) = \phi(\boldsymbol{x_i})^T \phi(\boldsymbol{x_j})$, $\phi(\cdot)$ is the mapping function from current space to higher feature space.

2.3 4-Layer Artificial Neural Network

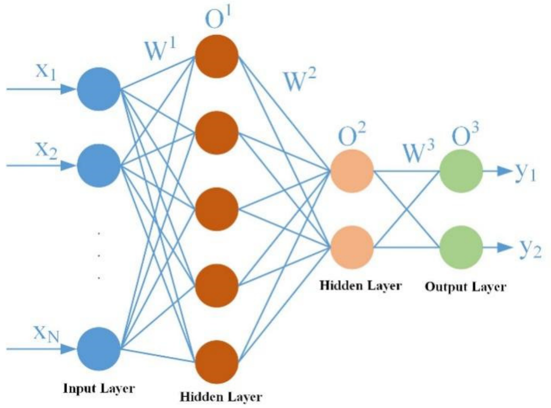


Figure 2. Sketch of four-layer back propagation neural network.

To better search for the complicated relations between the lost circulation and drilling data, back propagation (BP) neural network is furtherly employed through self-study, self-applicability and nonlinear function approximation. In this section, we utilize the neural network with four layers which include the input layer, two hidden layers and output layer. Figure 2 shows the details of the four-layer BP network. In BP process, the network iteratively updates the connection weights and thresholds according to the learning rule and finally reaches out the minimum errors between the predicted output and the real output. In the back propagation, for the output layer, the weight and bias term can be updated, respectively, as follows:

$$W_{ij}^{l} = W_{ij}^{l} - \eta \delta_j^l O_i^{l-1} \tag{8}$$

$$b_j^l = b_j^l - \eta \sum_{k=1}^{n_l} \delta_k^l \tag{9}$$

where O_i^{l-1} is the output feature map of $(l-1)^{th}$ of layer at i^{th} neuron. W_{ij}^l is the weight connected i^{th} neuron at $(l-1)^{th}$ layer and j^{th} neuron at l^{th} layer. b_j^l is the bias at l^{th} layer. η is learning rate. n_l is the number of neurons at l^{th} layer.

In our model, we utilize sigmoid function as the activation function $f(\cdot)$ and summarization of squares of errors between predicted outputs and real outputs as the loss function E. Thus, for the final output layer, δ_j^l can be calculated as follows:

$$\delta_j^l = \frac{\partial E}{\partial f\left(x_j^l\right)} \frac{f\left(x_j^l\right)}{\partial x_j^l} = \left(O_j^l - y_j\right) O_j^l \left(1 - O_j^l\right) \tag{10}$$

And for the intermediate layers, δ_j^l can be computed as follows:

$$\delta_j^l = \frac{f\left(x_j^l\right)}{\partial x_j^l} \left(\sum_{k=1}^{n_{l+1}} \delta_k^{l+1} W_{jk}^{l+1}\right) = O_j^l \left(1 - O_j^l\right) \left(\sum_{k=1}^{n_{l+1}} \delta_k^{l+1} W_{jk}^{l+1}\right) \tag{11}$$

By iteratively updating the weights and bias thresholds, the model can be trained until the loss reaches at pre-defined threshold value.

3. Experimental Results

3.1 Dataset and evaluation criterion

In this study, we collected 212.681k samples from drilling data from 59 wells for training and testing our classifier. In the dataset, 3709 of them are labelled as lost circulation while 208.972k samples are labelled without lost circulation. For each sample, 22 dimensional feature vector is utilized to express it. In multi-kernel derived SVR algorithms, we utilized linear, quadratic, cubic functions and medium, fine Gaussian functions. In medium and fine Gaussian functions, the width is $\sqrt{2}/4$ and $\sqrt{2}$, respectively. Regularization parameter $C = 1$. In four-layer BP network, learning rate $\eta = 0.001$, Adam is utilized as the optimization function. The ration between training set and test set is 4:1.

Before utilizing the dataset to train the classifier, we standardize the features using the following equation:

$$\hat{x} = \left(x - x_{mean}\right) / x_{std} \tag{12}$$

where x_{mean} and x_{std} are the mean value and standard deviation of feature maps of the drilling data, respectively. To qualitatively evaluate the accuracy of the prediction results, we introduce the basic metrics for evaluating the confusion matrix, including the true positive (tp), false negative (fn), false positive (fp) and true negative (tn). More specific, tp refers to the correctly predicted positive samples, fn is the falsely predicted negative samples, fp represents the falsely predicted positive samples while tn refers to the correctly predicted negative samples. Here, in order to better evaluate the improvement of different models, we also compute the sensitivity and specificity, as follows:

$$\text{sensitivity} = \frac{tp}{tp+fn} \tag{13}$$

$$\text{specificity} = \frac{tn}{fp+tn} \tag{14}$$

3.2 Results

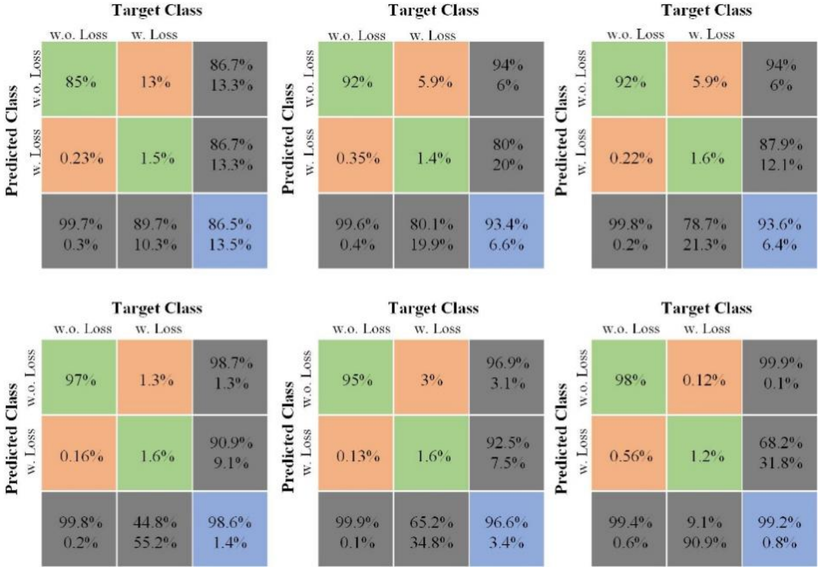


Figure 3. Confusion matrixes of test data based on linear-, quadratic-, cubic-, mediumGaussian-, FineGaussion-SVR and four-layer ANN algorithms. W.loss refers to with lost circulation, and w.o. loss means without lost circulation.

Figure 3 shows the confusion matrixes by the five kernel-based SVR, including linear, quadratic, cubic, mediumGaussian and fineGaussian, and four-layer ANN algorithms. In the figure, green boxes refer to the percentage of correctly classified classes by the trained models. While orange boxes represent the mis-classified classes by the trained models. The blue boxes are the final accuracy of the prediction based on the trained models, respectively. As can be seen from the blue boxes, ANN achieves the highest accuracy which means it can predict the lost circulation most accurate. Besides this, as can be seen from the *fn* percentage, ANN can more accurately estimate the real situation of lost circulation than all the other five kernel-based SVR models. Though *fp* of ANN model is slightly higher than other models, it is much more important not to mis-predicting the real lost circulation. With regard to the calculated specificity and sensitivity of all the models, ANN also achieves the best results.

To furtherly evaluate the effectiveness of different models, we also list their final accuracy and time needed to test a new drilling data in Table 2. In this table, we can find the 99.2% predictions based on ANN model are correct and only 0.8% of the final predictions are mis-classified. Though MediumGaussian SVR also obtains comparative accuracy with ANN model, it is very time-consuming. Gaussian-derived kernel based classifiers achieve much higher accuracy by comparing with other three polynomial-derived kernel based classifiers. Though FineGaussian-SVR has comparative time efficiency with ANN model, ANN achieve 2.6% improvement than it at accuracy. Considering the accuracy and time efficiency, ANN can not only predict the lost circulation but also almost obtain the predictions in real-time pattern which is of great significance to the field drilling wells.

Table 2. Qualitatively comparison between multi-kernel SVR and four-layer ANN algorithms.

Models	SVR					ANN
	Linear	Quadratic	Cubic	MediumGaussian	FineGaussian	Four-layer
Accuracy	86.5%	93.4%	93.6%	98.6%	96.6%	99.2%
Time (s)	**108.72**	**60.73**	**19.71**	**379.82**	**4.42**	**5.97**

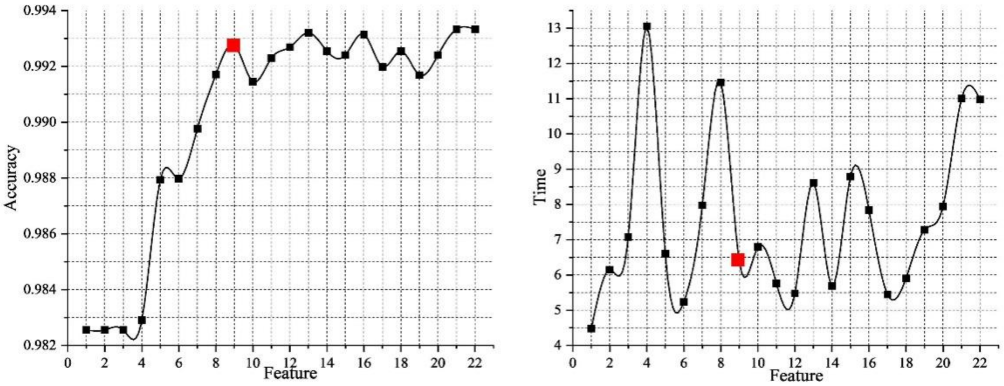


Figure 4. Accuracy and time efficiency by choosing different combination of drilling features under four-layer ANN model. Left: accuracy variation by changing the combination of drilling features. Right: time needed to predict lost circulation from a drilling feature vector based on the trained models.

Figure 4 shows the influence of different combined features to the accuracy of ANN-based predictor. As can be seen from left figure, horizontal axis refer to the features listed in Table 1 from first to fourth columns. According to the accuracy changing along with the increase of drilling features, we can conclude that the depth, torque, hanging weight, displacement, entrance density and export density contribute more to the prediction of lost circulation. From the last several points of the accuracy curve, the accuracy almost retains the same. From the right figure, we can also find that the time efficiency of the predictor based on the concluded key factors is almost real-time. Hence, based on the concluded key factors, we can not only obtain the high prediction accuracy, but also retain the real-time prediction of lost circulation.

4. Conclusion

In this paper, we tested five different kernel function (linear, quadratic, cubic, medium Gaussian and fine Gaussian) derived support vector regression (SVR) models and four-layer artificial neural network (ANN). Based on the huge dataset, we investigate the possible prediction of lost circulation using the drilling recorded datas. By comparing the accuracy of different algorithms, ANN can better establish the complex and non-linear mapping functions. According to the trained ANN based on different drilling features, we demonstrated that depth, torque, hanging weight, displacement, entrance density and export density have essential influence to the lost circulation. In our future work, more geological characteristics will be integrated into our network to furtherly evaluate the relations between the lost circulation and various factors. This will effectively decrease the possible non-productive time and operational cost.

Acknowledgement

This work was supported in part by grants from Research on Key Technologies for Sustainable and Stable Production of Complex Carbonate Gas Reservoirs (2021DJ3301).

References

[1] Magzoub, M I, Salehi, S, Hussein, I A, & Nasser, M S. (2020) Loss circulation in drilling and well construction: The significance of applications of crosslinked polymers in wellbore strengthening: A review. *Journal of Petroleum Science and Engineering*, 185, 106653.

[2] Krishna, S, Ridha, S, Vasant, P, Ilyas, S U, & Sophian, A. (2020) Conventional and intelligent models for detection and prediction of fluid loss events during drilling operations: A comprehensive review. *Journal of Petroleum Science and Engineering*, 195, 107818.

[3] Kulikov S, Veliev G, Bakhtin A, et al. (2013) Secure Drilling Services for Safe and Effective Drilling. *SPE Arctic and Extreme Environments Technical Conference and Exhibition*, Moscow, SPE-166846-MS.

[4] Al-Hameedi, A T, Alkinani, H H, Dunn-Norman, S, Flori, R E, Hilgedick, S A, & Amer, A S. (2017). Limiting Key Drilling Parameters to Avoid or Mitigate Mud Losses in the Hartha Formation, Rumaila Field, Iraq. *J Pet Environ Biotechnol*, 8, 345.

[5] Al-Hameedi, A T, Dunn-Norman, S, Alkinani, H H, Flori, R E, & Hilgedick, S A (2017). Limiting Drilling Parameters to Control Mud Losses in the Dammam Formation, South Rumaila Field, Iraq. *In 51st US Rock Mechanics/Geomechanics Symposium*. OnePetro.

[6] Al-Hameedi, A T, Alkinani, H H, Dunn-Norman, S, Flori, R E, Hilgedick, S A, Amer, A S, & Alsaba, M. (2019) Mud loss estimation using machine learning approach. *Journal of Petroleum Exploration and Production Technology*, *9*(2), 1339-1354.

[7] Jahanbakhshi, R, Keshavarzi, R, & Jalili, S. (2014) Artificial neural network-based prediction and geomechanical analysis of lost circulation in naturally fractured reservoirs: a case study. *European journal of environmental and civil engineering*, *18*(3), 320-335.

[8] Alkinani, H H, Al-Hameedi, A T, Dunn-Norman, S, Alkhamis, M M, & Mutar, R A. (2019) Prediction of lost circulation prior to drilling for induced fractures formations using artificial neural networks. In *SPE Oklahoma City Oil and Gas Symposium*. OnePetro.

[9] Li, Z, Chen, M, Jin, Y, Lu, Y, Wang, H, Geng, Z, & Wei, S. (2018) Study on intelligent prediction for risk level of lost circulation while drilling based on machine learning. In *52nd US Rock Mechanics/Geomechanics Symposium*. OnePetro.

[10] Gordon, G, & Tibshirani, R. (2012) Karush-kuhn-tucker conditions. *Optimization*, *10*(725/36), 725.

[11] Platt, J. (1998) Sequential minimal optimization: A fast algorithm for training support vector machines.

Regional power supply capacity and industrial layout based on big data

Ying Shang *, Muxin Zhang, Xinran Liu, Liying Liao, He Cui
State Grid Liaoning Electric Power Co., Ltd. Marketing Service Center. 19 / f, hunnan east road, hunnan new district, Shenyang,110000, China
*email: 476473516@qq.com

Abstract

Through the analysis of power big data and according to the regional power grid facilities development level, load margin, power supply quality and other aspects of research, evaluate the level of power supply capacity level of each region, for the quantitative power supply capacity level of power supply companies. And through the power data comprehensive analysis of each industry of each region, to determine the most suitable industry recommendation.

Keywords: regional power supply capacity; entropy weight method; ID3 algorithm.

1. Introduction

Electricity is the "thermometer" and "barometer" of the national economy. The change of power demand reflects the activity of economic operation, and the level of regional power supply capacity also greatly affects the local investment attraction and regional industrial layout.

Due to the regional imbalance of power grid development in China, the development of power grid construction lags behind the rapid economic growth, the infrastructure is weak, the power grid transformation is not strong, the power grid operation economy is poor, the national unified coordination and dispatching automation level is not high, and to some extent, there are power failure or power rationing of users. Before November, the regional power grid paid more attention to the power supply reliability, power supply continuity and smart grid construction[1], The assessment of the power supply capacity of the regional power grid is limited to some power supply enterprises and some regions, In the evaluation of the power supply capacity, they are more concerned about the study of the power supply capacity of the step-down substation, the calculation method of the maximum power supply capacity of the distribution network, the model research of some factors of the power supply capacity, the realization of the computer program of the distribution network power supply capacity evaluation system, etc., Few people treat the urban power supply capacity evaluation system as a comprehensive and complex system with many influencing factors and mutual constraints, The existing distribution network evaluation system considers few evaluation factors,The evaluation results cannot completely and accurately reflect the actual power supply capacity of the urban distribution network[2].

For the calculation method of the maximum power supply capacity of the power grid, many researchers have proposed using the linear planning method, the internal point method, the trial method and the maximum load multiple method. However, in these methods, some are used in distribution network error, such as linear planning method; some calculation results have high accuracy, but the process is complex and the calculation time is too long, such as internal point method: some methods can not guarantee the accuracy, such as try method; some methods are fast calculation speed, but the accuracy is lacking, such as the maximum load multiple method. Therefore, according to the characteristics of the power grid, the maximum power supply capacity calculation method that can ensure both the calculation speed and the accuracy of the results should be proposed.

This paper through the big data for regional power supply situation, such as power grid development level, electricity consumption information factors analysis of the industrial layout of the region, identify the different areas of electricity market pillar industries and emerging industries, and evaluate the development trend, put forward their region suitable development industry category, can support the relevant industrial policy and provide guidance for the government to carry out investment promotion and capital introduction.

International Conference on Computer Graphics, Artificial Intelligence, and Data Processing (ICCAID 2022)
edited by Ruishi Liang, Jing Wang, Proc. of SPIE Vol. 12604, 1260410
© 2023 SPIE · 0277-786X · doi: 10.1117/12.2674552

2. Regional power supply capacity

2.1. Overall thinking

Based on business understanding and expert experience, the substation capacity ratio, transformer load rate and transformer open capacity factors have a great impact on the power supply capacity of the region, which directly determines the load margin and future scalable capacity of a certain area. Based on the expert judgment method, the influence factors of substation capacity ratio ①, transformer load rate ② and transformer open capacity ③ on regional power supply capacity are set as 0.2,0.15 and 0.2, respectively. In order to better quantify and calculate the contribution degree of each factor to the comprehensive score of regional power supply capacity level, the corresponding scoring rules are set.

load ratio is ④, the distribution transformer heavy load ratio of 3 years is ⑤, the line meeting N-1 ratio is ⑥, the line connection rate is ⑦, and the average power connection time of industrial expansion and installation is ⑧ on the regional power supply capacity is not clear, which can be calculated through the entropy weight method.

In information theory, entropy is a measure of uncertainty. The greater the uncertainty, the greater the entropy and contains more information, with less uncertainty and less information. According to the characteristics of entropy, the randomness and disorder degree of an event can be judged by calculating the entropy value, and the entropy value can also judge the dispersion degree of an index. The greater the dispersion degree of the index, the greater the influence (weight) of the index on the comprehensive evaluation[7].

2.2. Model Logic

For the specific implementation method, see the flow chart 1:

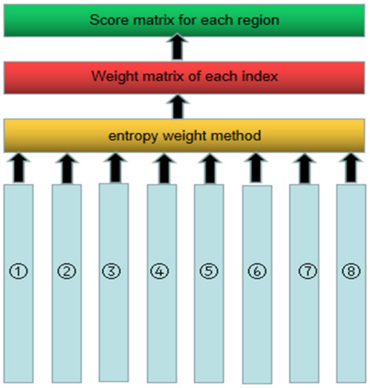


Figure 1 Flow chart of the regional power supply capability algorithm

2.3. Model specific implementation

The algorithm steps are as follows:

For n samples and m indicators, xij is the value of the j th indicator of the i th sample (i=1,2... n, j=1,2... m). Since the number of measurement units of each indicator is not unified, the comprehensive indicators should be standardized, that is, the absolute value of the indicators is converted into the relative value, so as to solve the problem of homogenization of different qualitative indicators. The data format is as follows:

Table 1 Sampling of this index

Sample	**Metric 1**	**Metric 2**	**....**	**Metric m**
Sample 1	X11	X12	....	X1m
Sample 2	X21	X22	....	X2m
....	....	....	....	....
Sample n	Xn1	Xn2	....	Xnm

1. Calculate the proportion of the i th sample value under the j th index in the index:

$$p_{ij} = \frac{x_{ij}}{\sum_{i=1}^{n} x_{ij}}, \quad i = 1, \cdots, n, j = 1, \cdots, m \tag{1}$$

Table 2 The proportion of each index

Sample	Index 1 proprotion	Index 2 proprotion	....	Index m proprotion
Sample 1	P11	P12	....	P1m
Sample 2	P21	P22	....	P2m
....	....	....	....	....
Sample n	Pn1	Pn2	....	Pnm

2. Calculate the proportion of the i th sample value under the j th index in the index:

$$e_j = -k \sum_{i=1}^{n} p_{ij} \ln(p_{ij}), j = 1, \cdots, m \tag{2}$$

among which $k = 1/\ln(n) > 0$

$$d_j = 1 - e_j, j = 1, \cdots, m \tag{3}$$

$$w_j = \frac{d_j}{\sum_{j=1}^{m} d_j}, j = 1, \cdots, m \tag{4}$$

Table 3 Weight of each index

parameter	Index 1 weight	Index 2 weight	....	Index m weight
w	w1	w2	....	wm

Replacing the sample with the actual district and county level power supply area, the final calculation score is shown in the figure below.

Table 4 Score for each power supply area

District and county power supply area	Supply capacity level score value
Prefecture 1	S1
Prefecture 2	S2
....	....
....	....
Prefecture n	SN

3. Industrial layout

3.1. Overall thinking

This technical scheme is based on machine learning, modeling and analysis of different statistical dimensions of different industries in different regions and different regions to obtain the results of the industry division (divided into emerging industries and pillar industries). Then it is divided into different business quadrants according to the attribute scores of different industries, and finally the final results of the industry planning are obtained according to the ranking within the business quadrant, so as to provide guidance for decision makers.

See the flowchart 2 for the specific steps:

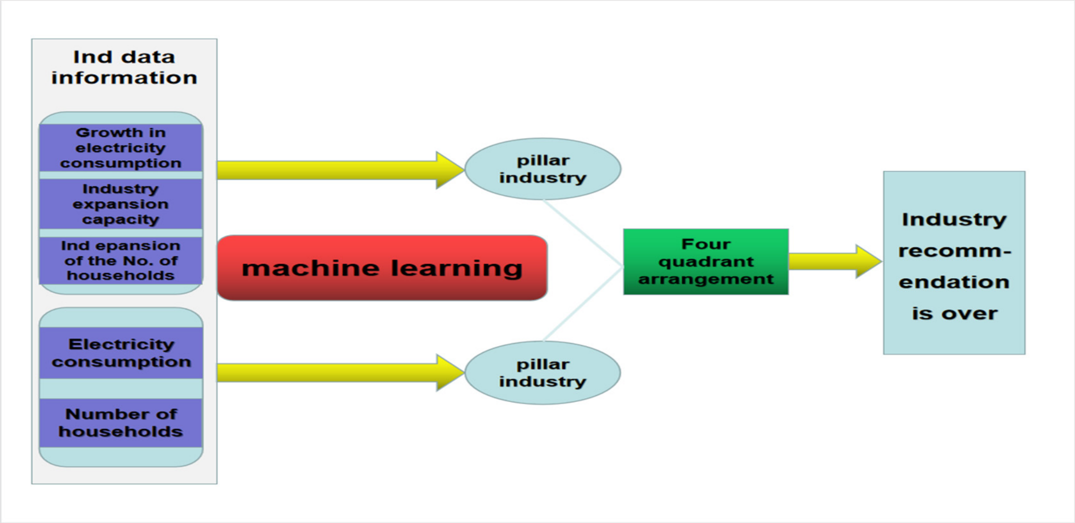
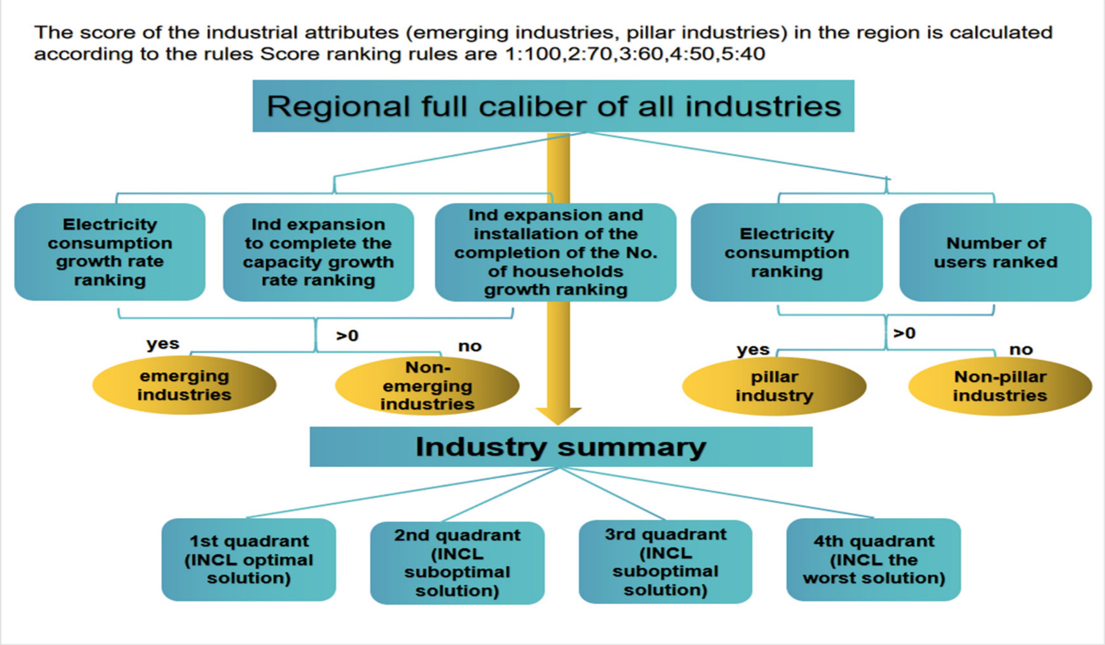

Ind data information
Growth in electricity consumption
Industry expansion capacity
Ind epansion of the No. of households
Electricity consumption
Number of households
machine learning
pillar industry
pillar industry
Four quadrant arrangement
Industry recomm-endation is over

Figure 2 Flowchart of vehicle screening of regional advantageous industries

3.2. Logical model

The algorithm and process of regional industry scoring and competitive industry screening are shown in the figure below.


The score of the industrial attributes (emerging industries, pillar industries) in the region is calculated according to the rules Score ranking rules are 1:100,2:70,3:60,4:50,5:40

Regional full caliber of all industries
Electricity consumption growth rate ranking
Ind expansion to complete the capacity growth rate ranking
Ind expansion and installation of the completion of the No. of households growth ranking
Electricity consumption ranking
Number of users ranked
yes
>0
no
emerging industries
Non-emerging industries
yes
>0
no
pillar industry
Non-pillar industries
Industry summary
1st quadrant (INCL optimal solution)
2nd quadrant (INCL suboptimal solution)
3rd quadrant (INCL suboptimal solution)
4th quadrant (INCL the worst solution)

Figure 3 Figure of industry scoring and competitive industry screening algorithm and process

3.3. Specific model implementation

The main splitting criterion of the decision tree is the ID3 algorithm. Use the characteristics with the largest information gain to establish the current split node of the decision tree. Specifically speaking:

The above division of emerging industries, non-emerging industries, and the division of pillar industries and non-pillar industries can be described as the following formulas (take the division of emerging industries as an example):

$$Y = \alpha * \mathrm{Param}(yi) + \beta * \mathrm{Param}(yi) + \gamma * \mathrm{Param}(yi) \quad (5)$$

It can be abbreviated as: $$Y = \mathrm{func}(yi) \tag{6}$$

among which yi corresponding to different industries.

Table 5 Industry judgment and score value

	industry	**Pillar industry score**	**Emerging industries score**
0	Other manufacturing	35	71
1	stock raising	46	12
2	traffic	21	20
3	Air transport industry	0	42
4	city article	0	41
5	Water production	0	18
6	Resident life	0	12

The above data are then quadrant divided by the softmax multi-classification model:

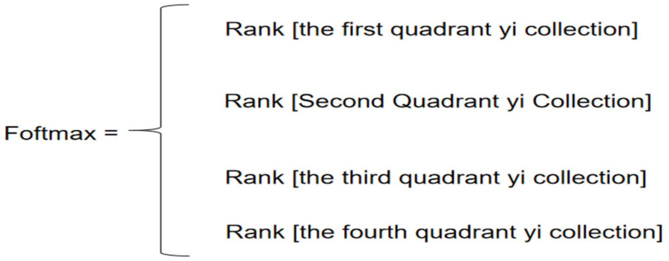

Foftmax =
Rank [the first quadrant yi collection]
Rank [Second Quadrant yi Collection]
Rank [the third quadrant yi collection]
Rank [the fourth quadrant yi collection]

Figure 4 Four-quadrant set division

Finally, the optimal solution is selected in the first to fourth quadrant order.

The mathematical symbols and parameters are described as follows:

Table 6 Mathematical symbols as well as a parameter description

Symbolic annotation		1 Y = func(yi) represents the scoring function of emerging industries, and Y contains the scoring ranking of each industry. 2 The Rank function assigns different weights to the pillar industries and to the emerging industries to calculate the score rankings.
$Param(x1)$	The growth rate of electricity consumption in the past three years is ranked	
$Param(x2)$	In the past three years, the industry expansion completed capacity growth ranking	
$Param(x3)$	In the past three years, the growth rate of the number of households completed in the industry expansion was ranked	
α	Electricity growth rate impact factor	
β	Industry expansion and installation completed capacity growth impact factor	
γ	The impact factor of the growth rate of the number of households completed by industry expansion	
yi	Represents the different industries	
softmax	Multiclassification function	
Rank	Sort function	

4. Conclusion

With the rapid development of social production and the rise of new technology revolution, social life dependence on power energy is more and more big, also for power quality is more and more high, how to accurately, comprehensive assessment of distribution network power supply capacity, strive for the future regional distribution network planning, transformation to provide effective reference and guidance, is the existing county distribution network power supply ability to check, reasonable comprehensive evaluation, has the very important theoretical and practical significance.

This paper is a recommendation of the entropy right algorithm using power big data to comprehensively evaluate the regional power supply capacity, and combining machine learning and ID3 algorithm with the practical characteristics and industrial development situation of each region. For optimizing the industrial allocation, this method has certain guidance for the development of regional economy. To guide the government to attract investment decisions, and to guide the company's power grid transformation and power grid configuration.

References

[1] Kang Chao. (2021). Research on power supply capacity based on Grid Planning and distribution Network (Master's thesis, Xinjiang University). https://kns.cnki.net/KCMS/detail/detail.aspx?dbname=CMFD202201&filename=1021800037.nh

[2] Ge Shaoyun, Han Jun, Liu Hong & Liu Yang. (2012). Optimization of the connection structure between the main transformer stations based on the power supply capacity. Power Grid Technology (08), 129-135. doi:10.13335/j.1000-3673.pst.2012.08. 023.

[3] Wang Liming, Li Xu, Cao Bin, et al. (2020) "Prediction of Leakage Current on Insulator Surface of Transmission Line Based on BP Neural Network." High Voltage Apparatus 56.02 (2020): 69-76.

[4] Liu Zhihai, Xue Yuan, Zhou Chen, et al. (2019) "Population Initialization Improvement of Robot Path Planning Based on Genetic Algorithm. " Machine Tool and Hydraulics 47.21 (2019): 5-8.

[5] Deng Yijie, Liu Kesheng, Zhu Kailong, et al. (2019) "Research on fuzzing technique based on dynamic fitness function." Application Research of Computers 36.05 (2019): 1415-1418,1427.

[6] Li Ya, Chen Minyou, Li Bo, et al. (2020) "Regulating ability analysis of distributed energy storage system based on fuzzy comprehensive evaluation." Journal of Chongqing University 43.05 (2020):1-10.

[7] Lei Yue, Cao Xiangkun, Yang Xiulong & Wang Cheng. (2022). Construction Safety Evaluation based on SNA-Structure entropy Method. Journal of Hubei Institute of Technology (05), 9-13.

Proc. of SPIE Vol. 12604 1260401-289

A Novel Semi-supervised Learning Method for Power Transformer Fault Diagnosis with Limited Labeled Data

Guolin Zhou[a], Dazhi Wang[a]*, Yuqian Tian[a], Jiaxing Wang[a] , Shuo Cao[b]
[a]College of Information Science and Engineering, Northeastern University，Shenyang 110819, China
[b]College of Information Engineering, Shenyang Polytechnic College, Shenyang 110045, China.
* Corresponding author.
E-mail address: ProDZW@126.com (D. Wang).

Abstract

Identifying power transformer faults accurately is critical to maintaining the stable operation of power system. Intelligent fault diagnosis algorithms based on dissolved gases have been extensively researched and implemented. However, in practice, collecting labeled data is time-consuming and costly. Therefore, it is necessary to establish a valid diagnostic model with limited labeled data. To solve this problem, a novel semi-supervised learning method for power transformer fault diagnosis is proposed in this paper. First, all the dissolved gas samples are constructed as a weighted K-nearest neighbor (KNN) graph to initially describe association among all samples. Then, a semi-supervised random multireceptive field propagation graph convolutional network (SSRMFPGCN) is designed for fault feature extraction and classification. Finally, the collected power transformer fault data are used to validate the proposed method. The experimental results show that the method proposed in this paper can still achieve 94.06% accuracy with only 20% of labeled training samples, which is significantly superior to the traditional intelligent diagnosis methods.

Keywords: Power transformer; Fault diagnosis; Semi-supervised learning; Graph convolutional neural network；

1. Introduction

As a key equipment of power system, the safe and stable operation of power transformer is related to the reliability of power supply. How to use the data generated in the operation of power transformer to evaluate the current health status of the equipment and achieve intelligent fault diagnosis is an unavoidable trend in the background of smart grid. When a potential fault occurs in power transformer, the dissolved gases in insulating oil will change accordingly. [1] By analyzing the concentration and corresponding proportion of each dissolved gas such as hydrogen (H_2), methane (CH_4), ethane (C_2H_6), ethylene (C_2H_4), acetylene (C_2H_2) and other characteristic gases, the potential fault of power transformer can be effectively judged, thus timely avoiding irreversible insulation damage and improving the reliability of the power system [2].

In order to reduce losses and improve the stability of power transformer in the production process, many scholars have studied various fault diagnosis methods around dissolved gases in insulating oil. Among them, traditional methods such as IEC Ratio, Duval Triangle, and Rogers Ratio have been widely used [3-5], but their threshold settings are not objective enough and often require experienced professional to make judgments, so it is difficult for these methods to ensure the accuracy and timeliness of diagnosis. To solve such problems, machine learning methods based on feature selection and classification algorithms such as extreme learning machine (ELM) [6], decision tree (DT) [7] etc., have been applied to power transformer fault diagnosis. These methods have complex feature selection engineering and do well with small datasets, but have poor performance with large data.

Deep learning (DL) has also been widely used in power transformer fault diagnosis due to its powerful feature representation capability. DL architectures such as deep residual shrinkage network (DRSN) [8] and deep belief network (DBN) [9] have been successfully applied to power transformer fault diagnosis. Most of these traditional DL fault diagnosis methods are fine-tuned and classified under supervision. With sufficient labeled data, these methods can fully exploit the complex nonlinear relationships between dissolved gases and fault types. However, since power transformer operate relatively stable and are not prone to fault, collecting labeled data is difficult and expensive, which make it difficult to further improve the performance of these methods. Therefore, the methods based on semi-supervised learning (SSL), which allows training classification models with few labeled samples and plenty of unlabeled samples, has become a popular research topic in power transformer fault diagnosis in recent years. For example, Yang et al. [10] proposed a novel

International Conference on Computer Graphics, Artificial Intelligence, and Data Processing (ICCAID 2022)
edited by Ruishi Liang, Jing Wang, Proc. of SPIE Vol. 12604, 1260411
© 2023 SPIE · 0277-786X · doi: 10.1117/12.2674508

SSL method based on a double-stacked autoencoder for power transformer fault diagnosis. Tan et al. [11] proposes a two-stage semi-supervised transformer fault diagnosis system based on improved support vector machine. However, these SSL methods only learn fault features from the input in Euclidean space and do not fully utilize the local geometry property between all samples, thus making it difficult to mine potential associations among samples. Although the authors in literature [12] applied graph convolutional neural network (GCN) to power transformer fault diagnosis with a view to better mining deep features in a non-Euclidean space, this traditional GCN, similar to the fixed kernel in convolutional neural network (CNN) that take a deterministic propagation approach, is prone to overfitting to the scarce label data [13].

In order to overcome above limitations and fully utilize the labeled and unlabeled dissolved gas samples, in this paper, a novel semi-supervised random multireceptive field propagation graph convolutional network (SSRMFPGCN) is designed for power transformer fault diagnosis. Inspired by the fact that homogeneous fault samples are those with the same characteristics, in this method, we represent each sample in the dataset with a node and all the samples are constructed into an undirected weighted K-nearest neighbor (KNN) graph to initially describe the association among all samples. Then, a random multireceptive field propagation mechanism based on GCN is designed to extract fault features so as to avoid overfitting to the scarce label information. The main contributions of this paper are as follows.

(1) A novel SSL method, namely, the semi-supervised random multireceptive field propagation graph convolutional network (SSRMFPGCN) is proposed and applied to power transformer fault diagnosis for the first time. Different from the traditional GCN, it can effectively extract fault features even under the condition of limited labeled data.

(2) A weighted KNN graph is constructed to quantify the correlation among all dissolved gas samples.

(3) Extensive experiments were conducted on real datasets collected from the State Grid Corporation of China and previous publications to validate the effectiveness of proposed method in transformer fault diagnosis.

The rest of this paper is organized as follows. In Section 2, we briefly introduce GCN. Section 3 details the proposed SSRMFGCN. In Section 4, a case study is conducted to verify the superiority of the proposed method. Finally, Section 5 concludes the paper.

2.Related Work

This section briefly introduces GCN. As the name suggests, GCN is inspired by the CNN, which is widely used in industrial system equipment fault diagnosis in recent years because of its ability to better explore the deep correlation between samples. Unlike traditional neural networks, its input sample object is graph-structed data composed of nodes and edges. Assuming that there are *N* nodes in the graph *G*, the graph convolution is defined as the product of the node features $X \in R^N$ and the convolution kernel g_θ in the Fourier domain, as shown in Eq. (1).

$$h = g_\theta * X = U g_\theta U^{\mathrm{T}} X \tag{1}$$

where *h* denotes the result of the graph convolution operation, * represents the convolution operation, *U* is the eigenvector matrix of the normalized Laplacian matrix *L*, and *L* satisfies Eq. (2).

$$L = I - D^{-0.5} A D^{-0.5} \tag{2}$$

where *D* and *A* are the degree matrix and adjacency matrix of graph *G*, respectively. In Eq. (1), g_θ is usually fitted using the *K*-order truncated expansion of the Chebyshev polynomial, so the graph convolution is usually written as the Eq. (3).

$$h = g_\theta * X = \sum_{k=0}^{K} \vartheta_k' T_k(\tilde{\Lambda}) X \tag{3}$$

where $\widehat{\Lambda} = 2\Lambda / \lambda_{\max} - I$ and Λ is the diagonal matrix consisting of the eigenvalues of the Laplacian matrix *L*, $\lambda_{\max}$ is the maximum eigenvalue of *L*, $\vartheta_k' \in R^{K+1}$ denotes the Chebyshev coefficients, and $T_k(\tilde{\Lambda})$ represents the Chebyshev polynomial of *K*-order.

In practical application, the adjacency matrix *A* and node features *X* are usually adopted as inputs to the GCN, while the weight parameters are represented by *W*, so the Eq. (3) can be simplified to Eq. (4).

$$h = A^{K-1}XW \tag{4}$$

Therefore, for the GCN node classification model containing *t* hidden layers, the propagation process can be expressed as follows

$$\begin{cases} h^0 = h(X, A) \\ h^t = \sigma(h^{t-1}, A),\ t > 1 \end{cases} \tag{5}$$

where h^t denotes the feature output of the *t*-th hidden layer, σ is the activation function, which usually uses the rectified linear unit (ReLU) function.

3.Proposed Method

3.1 Constructing KNN graph from selected input feature

According to the actual power transformer fault statistics, CH_4/H_2 , C_2H_4/C_2H_6 , C_2H_4/C_2H_2 , $H_2/(C_1+C_2)$, $H_2/(H_2+C_1+C_2)$, $C_2H_4/(C_1+C_2)$, $CH_4/(C_1+C_2)$, $C_2H_6/(C_1+C_2)$ and $(CH_4+C_2H_4)/(C_1+C_2)$ are considered relevant to the failure modes and these ratio combinations are named as Non-code ratio [14]. Among them, C_1 represents CH_4 and C_2 represents the sum of C_2H_4, C_2H_6 and C_2H_2. In this paper, the above combinations were used as input features for the diagnostic model to avoid the deficiencies of the IEC Ratio, Duval Triangle, and Rogers Ratio.

Due to the large differences in the values of the above nine features, the performance of the model will be adversely affected and even the loss function will be difficult to converge if they are directly used as input variables. Therefore, it is necessary to normalize the above nine features by Eq. (6).

$$f_i = \frac{f_i - f_{i,\min}}{f_{i,\max} - f_{i,\min}}, i = 1, 2, ...9 \tag{6}$$

where f_i is the *i*-th feature, $f_{i.\max}$ is the maximum value, and $f_{i.\min}$ is the minimum value.

The construction of graph can reflect the geometric properties between the dissolved gas samples. In literature [12], each dissolved gas sample in the dataset is considered as a node and an undirected graph is constructed using KNN, where each edge between two nodes in the graph has a binary weight (i.e., 0 or 1). However, no given rule exists for the choice of the adjacency parameter *m*. In addition, when plenty of unlabeled samples exist in the dataset, it is difficult to quantify the association between samples using binary weights [14]. Inspired by the literature [15], we use $\log_2 N$ (*N* is the total number of samples) to calculate the adjacency parameter *m*, and use the heat kernel [16] to calculate the similarity between the adjacency samples. Suppose there is a data set $X = (x_1, x_2, ..., x_N)^{\mathrm{T}}, (x_i \in R^{1\times 9})$ consisting of *N* gas samples, and the element a_{ij} of weighted adjacency matrix *A* constructed in this paper is defined as Eq. (7).

$$a_{ij} = \begin{cases} e^{-\frac{\|x_i - x_j\|_2^2}{2t^2}}, & x_j \in \varphi_m(x_i) \\ 0, & x_j \notin \varphi_m(x_i) \end{cases} \tag{7}$$

where, $\varphi_m(x_i)$ is the *m* neighbours of sample x_i selected by KNN, *t* is the width of the heat kernel, which can be calculated by Eq. (8). If x_i is connected to x_j , the more similar they are, the more the values of a_{ij} and a_{ji} converge to 1.

$$t = \frac{2}{N-(N-1)} \sum_{i=1}^{N} \sum_{j=i+1}^{N} \| x_i - x_j \|_2 \tag{8}$$

3.2 Construction SSRMFGCN for node classification

When the number of layers of GCN increases, the features between nodes tend to homogenize, which is also called the over-smoothing phenomenon [17]. The presence of over-smoothing can hinder model training by separating the output representation from the input features. The literature [18] effectively alleviates this phenomenon by the idea of random feature perturbation, but it is difficult to learn multi-domain information when aggregating features, which can easily cause feature loss. To avoid this problem, inspired by literature [13], SSRMFPGCN is designed in this paper, as shown in Fig. 1.

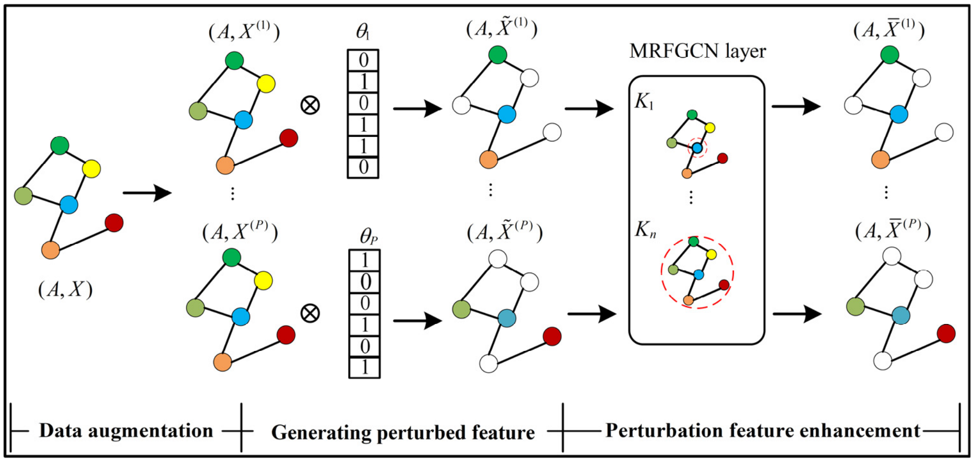


Fig. 1. The structure of SSRMFPGCN

The central idea of SSRMFPGCN is to generate a perturbed feature matrix for data augmentation by randomly eliminating some node features before training. Then, the perturbed features are put into multireceptive field graph convolutional network (MRFGCN) [13] for feature augmentation. By doing so, multiple augmented representations can be randomly generated for each node, since the lost information of a node can be compensated by its neighbours. Specifically, the implementation details of SSRMFPGCN for node classification can be described as follows.

Firstly, the node feature X is copied to obtain $X^{(1)}, \ldots, X^{(P)}$ and the number of copies is controlled by the hyperparameter P. Then, some nodes are randomly deactivated to generate the perturbed feature matrix $\tilde{X}^{(1)}, \ldots, \tilde{X}^{(P)}$. Unlike Dropout [19] which deactivates some dimensions of the features in the hidden layer, we deactivate all the features of the selected nodes, which can be expressed by Eq. (9)

$$\tilde{X}^{(i)} = \theta_i X^{(i)} \tag{9}$$

where, θ_i is a binary mask vector generated by $Bernoulli(1-\delta)$, δ is the hyperparameter that controls the probability of deactivated nodes and the perturbation feature matrix $\tilde{X}^{(i)}$ is obtained by multiplying the feature vector of each node with its corresponding mask.

Then, $\tilde{X}^{(i)}$ is propagated in the MRFGCN to achieve feature augmentation to obtain $\bar{X}^{(i)}$, and $\bar{X}^{(i)}$ can be calculated by Eq. (10)

$$\bar{X}^{(i)} = \left[A^{K_1-1}\tilde{X}^{(i)}, A^{K_2-1}\tilde{X}^{(i)}, \cdots, A^{K_n-1}\tilde{X}^{(i)} \right] \tag{10}$$

where, $[\cdot]$ represents the concatenation operator, n is the amount of receptive filed, K_n represent receptive filed size.

Finally, the obtained augmented feature matrix $\bar{X}^{(i)}$ is put into the fully connected layer to obtain a series of classifier parameters, and the average of which is taken as the final output, as shown in Eq. (11).

$$\hat{Y} = \frac{\sum_{i=1}^{P} \operatorname{softmax}(W_i \overline{X}^{(i)} + b_i)}{P} \tag{11}$$

where, W_i , b_i are the learnable parameters and $\hat{Y} = \{\hat{y}_1, \hat{y}_2, \cdots, \hat{y}_n\}$ is the output node classification information.

3.3 Construction the objective function for semi-supervised learning

The SSL-based fault diagnosis method can improve the diagnostic performance by taking full advantage of the scarce labeled samples and plenty of unlabeled samples. Assuming that the number of labeled nodes in the node feature matrix X is l and the number of unlabeled nodes is u, the SSL-based objective function is usually expressed as Eq. (12) [20].

$$\min \sum_{i}^{l} \left(f\left(x_i\right) - y \right)^2 + \lambda \sum_{i,j}^{l+u} \left(f\left(x_i\right) - f\left(x_j\right) \right)^2 \tag{12}$$

where, $f\left(x_i\right)$ is the output mapping and y is the actual class of the node. The first term is the supervised loss function, which reflects the error of the model on the labelled nodes. The second term is a consistency regularization term to ensure that similar nodes have same outputs as much as possible.

For graph neural networks, the cross-entropy loss over all labeled nodes is often utilized to back-propagate and update parameters. Thus, the supervised loss function used in the training process of the SSRMFPGCN, constructed in this paper, is shown as Eq. (13).

$$\mathcal{L}_1 = -\frac{1}{P} \sum_{p=1}^{P} \sum_{i=0}^{l-1} y_i \log \widehat{y}_i^{(p)} \tag{13}$$

where, y_i denotes the true label of the i-th node, and $\widehat{y}_i^{(p)}$ represents the output label of the model.

In order to make all enhanced features have similar outputs as much as possible, the consistency regularization term is constructed as Eq. (14), and the final objective function of the SSRMFPGCN is expressed as Eq. (15).

$$\mathcal{L}_2 = \frac{1}{P} \sum_{p=1}^{P} \sum_{i=0}^{l+u-1} \| \hat{y}_i - \hat{y}_i^{(p)} \|_2^2 \tag{14}$$

$$\mathcal{L} = \min \left(\mathcal{L}_1 + \lambda \mathcal{L}_2 \right) \tag{15}$$

where λ is the hyperparameter. Subsequently, the weight parameters of SSRMFPGCN are updated by a back-propagation algorithm.

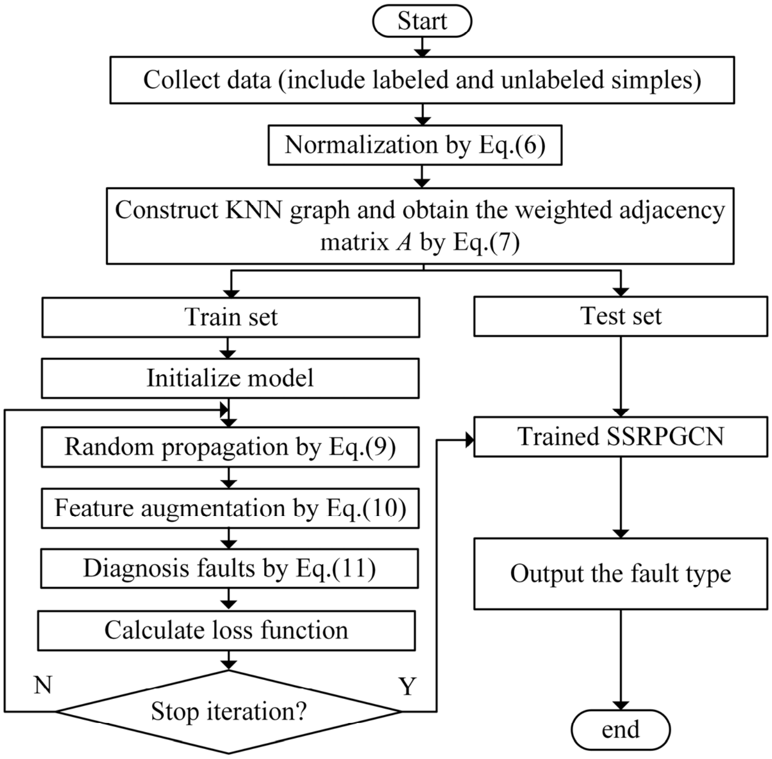

Start
Collect data (include labeled and unlabeled simples)
Normalization by Eq.(6)
Construct KNN graph and obtain the weighted adjacency matrix A by Eq.(7)
Train set
Test set
Initialize model
Random propagation by Eq.(9)
Trained SSRPGCN
Feature augmentation by Eq.(10)
Diagnosis faults by Eq.(11)
Output the fault type
Calculate loss function
N
Y
Stop iteration?
end

Fig. 2. The overall flowchart of the fault diagnostic framework

3.4 Intelligent fault diagnosis framework based on SSRMFPGCN

Finally, in order to establish an effective diagnosis model with limited labeled data, an SSRMFPGCN-based fault diagnosis framework for power transformer is proposed in this paper, as illustrated in Fig. 2. The detailed diagnostic process can be briefly summarized as follows.

1) Data pre-processing: Obtaining the labeled and unlabeled dissolved gas signals of the power transformer and normalizing them by Eq. (6).

2) Constructing KNN graph: Considering each sample as a node, the adjacency relationship among all simples is determined by KNN. In which, the adjacency parameter m is calculated by $\log_2 N$ and the degree of similarity between nodes is quantified by Eq. (7).

3) Model Training: The labeled nodes and some unlabeled nodes are randomly selected as the training set, and the remaining unlabeled nodes are used as the test set. Then, the training set are used for the training of the SSRMFPGCN model as described in Section 3.

4) Fault diagnosis: The trained SSRMFPGCN model is implemented to recognize the samples in the test set and return fault diagnosis results.

4.Case Study

In order to verify the effectiveness of the proposed method, a case study is conducted in this section. The hardware is a computer with a i5-10400F CPU and a NVIDIA 1660Ti GPU. The related model is programmed in Python 3.7 and Pytorch 1.7.

4.1 Data description and processing

The dissolved gas data used in this case were obtained from the State Grid Corporation of China and published literature [21-23]. These samples relate to power transformer of each common voltage class and contain the concentrations of H_2, CH_4, C_2H_4, C_2H_6 and C_2H_2 in seven condition types: normal, thermal fault of low temperature (LT), thermal fault of medium temperature (MT), thermal fault of high temperature (HT), partial discharge (PD), low-energy discharge (LD), and high-energy discharge (HD). We believe that combining these samples collected from different sources into one overall dataset can effectively improve the generalization ability of the algorithm. When converting the collected data to Non-coded ratio, the ratio is set to "0" for 0/0, and s/0 is set to 20, where s is a non-"0" value [24]. The size of each state type in the final obtained dataset is shown in Table 1.

Table 1. Description of dataset.

Status	Number of samples	Label
Normal	360	0
LT	360	1
MT	360	2
HT	360	3
PD	360	4
LD	360	5
HD	360	6

Table 2. Detailed structure of SSRMFPGCN

Layer	Structure and parameters	Output Shape
1-st	H_0=Input (A, X)	1×9
2-rd	H_1=SSRMFPGCN Layer (H_0, K_1=1, K_2=2, K_3=3, δ=0.25, P=4)	1×27
3-th	H_2=SRMFPGCN Layer (H_1, K_1=1, K_2=2, K_3=3, δ=0.25, P=4)	1×81
4-th	H_3=Fully connected layer (H_2, Softmax)	1×36
5-th	H_4=Fully connected layer (H_4, Softmax)	1×7

4.2 Parameter selection of SSRMFPGCN

Compared with GCN, the SSRMFPGCN model constructed in this paper adds three hyperparameters, which are the drop-node probability δ, the number of data enhancements P, and the coefficient λ in calculating the consistency regularization loss. According to the above fault diagnosis flow chart, all hyperparameters can be obtained by grid search method. Specifically, we first search for P from {2,4,6,8,10}. With the best choice of P, then, we search for δ from {0.1,0.25,0.5}. Finally, we keep P, δ constant and search for hyperparameters such as λ and learning rate l in turn. Finally, the optimal set of main hyperparameters for the algorithm is $\{\delta, P, \lambda, l\}$ = {0.25 ,4,0.5, 0.01} and the structure of SSRMFPGCN is detailed in Table 2.

4.3 Performance analysis of SSRMFGCN

To validate the performance of SSRMFPGCN on the few labeled datasets, in this experiment, we first randomly select 50% of the samples from the dataset as the training set and the remaining 50% as the test set. In the training set, we randomly select 20% of the samples as labeled samples, and the remaining samples are considered as unlabeled samples to train the model. The network is trained with 400 iterations on the basis of the above network parameter settings. Fig. 3 shows the loss curves and the diagnostic accuracy of the labeled and unlabeled samples during the model training. As described in Fig. 3, in the early stage of training, the values of the loss function of the labeled sample set and the unlabeled sample set drop sharply and their diagnostic accuracy rises rapidly. The loss values of both began to stabilize after 100 iterations, and the fault diagnosis accuracy of both stabilized after 120 iterations. These fully demonstrate the rapid convergence and excellent robustness of the model proposed in this paper.

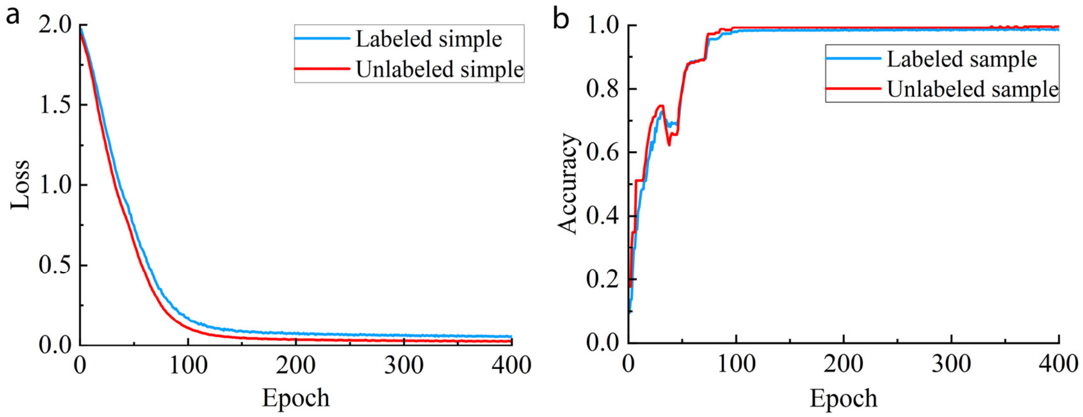
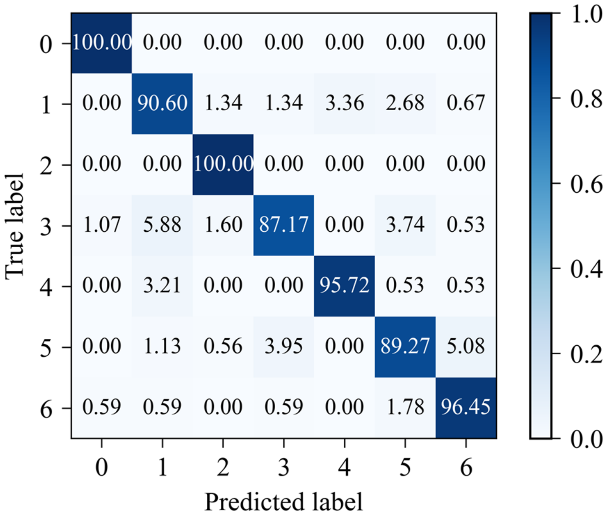


Fig. 3. (a) the loss curves of SSRMFGCN; (b) The diagnostic accuracy curves of SSRMFGCN

Then, the test set was fed into the converged SSRMFPGCN model, and the model performed on the test set as indicated by the confusion matrix in Fig. 4.



True label \ Predicted label	0	1	2	3	4	5	6
0	100.00	0.00	0.00	0.00	0.00	0.00	0.00
1	0.00	90.60	1.34	1.34	3.36	2.68	0.67
2	0.00	0.00	100.00	0.00	0.00	0.00	0.00
3	1.07	5.88	1.60	87.17	0.00	3.74	0.53
4	0.00	3.21	0.00	0.00	95.72	0.53	0.53
5	0.00	1.13	0.56	3.95	0.00	89.27	5.08
6	0.59	0.59	0.00	0.59	0.00	1.78	96.45

Fig. 4. Confusion matrix of test set of SSRMFGCN

From Fig. 4, we can observe that the diagnostic accuracy of the proposed model in this paper for each state of the power transformer can still be maintained at a high level under the premise of training with limited labeled samples. Specifically, the model was able to completely and accurately identify normal state and MT faults in the test set, but was the lowest accurate in identifying HT faults in power transformer with 87.17%, and most of these misidentified samples were identified as LT fault. This is because there is a great similarity between the dissolved gas data for HT fault and LT fault in power transformer, so it is difficult to achieve accurate identification by a model trained with only 20% labeled samples. However, when the labeled samples in the training set rise to 60%, the overall recognition rate of the SSRMFPGCN model proposed in this paper can stabilize around 100%. These sufficiently demonstrate that the power transformer fault diagnosis model proposed in this paper can identify various common states well with limited labeled samples.

In order to better analysis the powerful fault extraction capability of the method proposed in this paper, t-distributed stochastic neighbour embedding (t-SNE) [25,26] is applied to convert the input data and output features of the method proposed in this paper into two-dimensional vector distributions, respectively, and the results of their visualization are shown in Fig. 5. As illustrated that the features extracted by the SSRMFPGCN can perfectly isolate all faults.

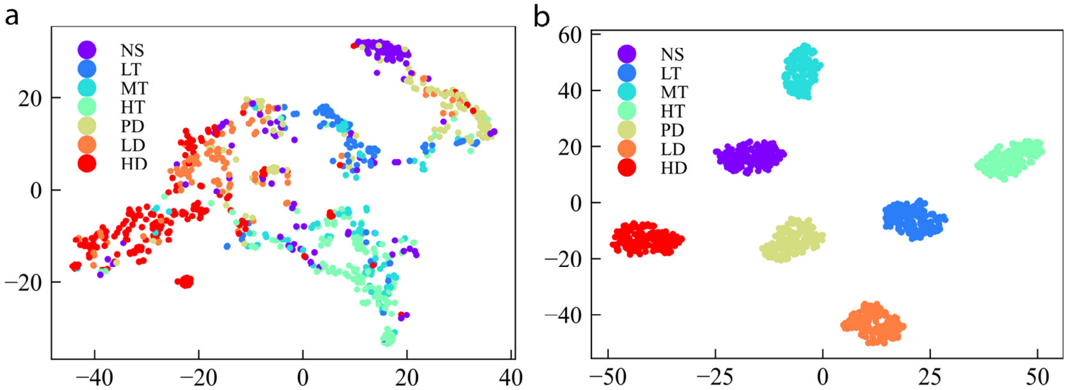


Fig. 5. (a) Visualization of input data; (b) Visualization of the out layer out feature in SSRMFGCN.

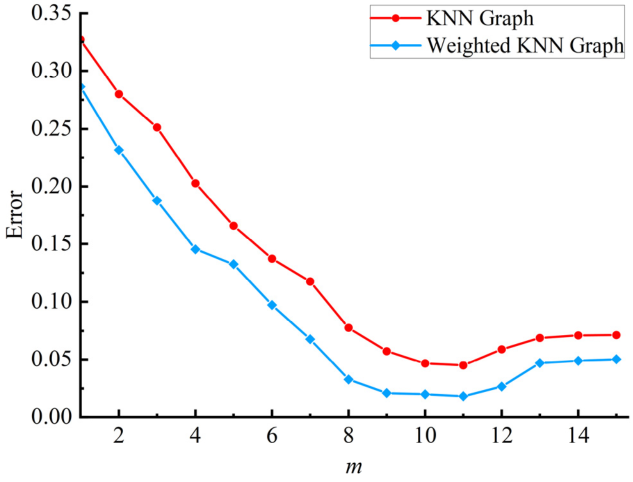


Fig. 6. SSRMFPGCN fault recognition error rate under different KNN graphs

In addition, the selection of adjacency argument *m* and the calculation of edge weights are crucial for the construction of KNN graph in this paper, and the quality of constructed graph has a great impact on the fault diagnosis performance, therefore, it is necessary to analyse them. To verify this, we first set different values of *m* to construct a series of unweighted KNN graphs as well as weighted KNN graphs. Then, these graphs were input into the model proposed in this paper for comparison respectively, and the comparison results are shown in Fig. 6. From Fig. 6 we can learn that the accuracy of the weighted KNN graph is higher than that of the unweighted KNN graph regardless of the choice of *m* values, and fault recognition error rate is minimized when *m*=11, which is consistent with the results calculated by $\log_2 N$.

4.4 Comparison with different methods

To further validate the superiority of the proposed method in this paper, four benchmark algorithms, support vector machine (SVM) [27], extreme gradient boosting tree (XGBoost) [28], CNN [29] and GCN [12] were used to process the same power transformer fault dataset. In this case, for GCN, it includes two graph convolution layers, and two fully connected layers. For CNN, it consists of two convolutional layers with convolution kernel size is 3, two average pooling layers, and two fully connected layers. For the SVM, the kernel function is chosen as a sigmoid function, and penalty factor is set as 1.0. For XGBoost, the gamma value is 0.5, the maximum depth is set as 5, the subsample rate is 0.6, and the min child weight is 3. Considering the influence of the number of labeled samples on the fault diagnosis performance, according to the literature [30], we define the labeled rate of the training set as $lr = l/(l+u)$. Then, the data sets with label rates of 0.2, 0.4, 0.6, and 0.8 were input into the proposed method and above four comparison methods, respectively. Besides， in order to make the comparison fairer and more convincing, five trials were conducted for each of the above methods and their average diagnostic accuracy on the test set is shown in Table 3.

Table 3. Average accuracy with different ratio of labeled training samples

Algorithms	0.2	0.4	0.6	0.8
SVM	0.4423	0.4683	0.6543	0.7322
XGboost	0.4872	0.5341	0.6732	0.8243
CNN	0.5863	0.6482	0.8214	0.8564
GCN	0.6782	0.7543	0.8453	0.8863
SSRMFPGCN	0.9406	0.9732	100	100

As can be seen from Table 3, the diagnostic accuracy of all five methods improved with the increase of labeled samples, but the accuracy of the method proposed in this paper was always above 90%. This fully illustrates that the proposed method in this paper has a strong feature extraction capability even with limited labelled data. In addition, the diagnostic accuracy of the method proposed in this paper is much better than that of the other four methods. The underlying cause of these merits is that the method proposed in this paper can fully exploit the relationship between labeled and unlabeled samples.

5. Conclusion

In order to improve the accuracy of power transformer fault diagnosis with limited labeled data, this paper proposes a semi-supervised intelligent fault diagnosis algorithm based on SSRMFPGCN, which can fully utilize labeled and unlabeled data and graph convolution operation to achieve feature extraction and fault identification. The analysis of real cases shows that the proposed method can extract fault features of power transformer adaptively with rapid convergence speed and excellent stability, and the fault diagnosis accuracy is better than SVM, XGBoost, CNN, GCN under the condition of few labeled samples. Furthermore, the limitations and further research work of this paper can be concluded as follows.

1) The distribution of normal data and various types of fault data in the power transformer data collected in this paper is approximately same, but the fact is that the normal data is much more than the fault data. In future work we will further explore the impact of imbalanced data.

2) In this paper, we use a simple grid search method for rough estimation in the selection of parameters for SSRMFPGCN. In the future work, we will explore the intelligent optimization of the model parameters by algorithms such as particle swarm optimization algorithm.

Acknowledgements

This work is supported by the Natural Science Foundation of China under Grant 52077027, and the Liaoning Province Science and Technology Major, China Project No. 2020020304-JH1/101.

References

[1] Jiang, Jun, et al. "Dynamic fault prediction of power transformers based on hidden markov model of dissolved gases analysis." *IEEE Transactions on Power Delivery* 34.4 (2019): 1393-1400.

[2] Wang, Yuan, et al. "Detection of dissolved acetylene in power transformer oil based on photonic crystal fiber." *IEEE Sensors Journal* 20.18 (2020): 10981-10988.

[3] Duval, Michel. "A review of faults detectable by gas-in-oil analysis in transformers." *IEEE Electrical Insulation Magazine* 18.3 (2002):8-17.

[4] Duval, Michel, and A. DePabla. "Interpretation of gas-in-oil analysis using new IEC publication 60599 and IEC TC 10 databases." *IEEE Electrical Insulation Magazine* 17.2 (2001) :31-41.

[5] Rogers, R. R. "IEEE and IEC codes to interpret incipient faults in transformers." *IEEE Transactions on Electrical Insulation* 13.5 (1978): 349-354.

[6] Malik, Hasmat, and Sukumar Mishra. "Selection of most relevant input parameters using principle component analysis for extreme learning machine based power transformer fault diagnosis model." *Electric Power Components and Systems* 45.12 (2017):1339-1352.

[7] Menezes, Abraão GC, et al. "Induction of decision trees to diagnose incipient faults in power transformers." *IEEE Transactions on Dielectrics and Electrical Insulation* 29.1 (2022): 206-223.

[8] Hu, Hao, Xin Ma, and Yizi Shang. "A novel method for transformer fault diagnosis based on refined deep residual shrinkage network." *IET Electric Power Applications* 16.2. (2022): 111162-111170.

[9] Dai, Jiejie, et al. "Dissolved gas analysis of insulating oil for power transformer fault diagnosis with deep belief network." *IEEE Transactions on Dielectrics and Electrical Insulation* 24.5 (2017): 2828-2835.

[10] Yang, Dongsheng, et al. "A novel double-stacked autoencoder for power transformers DGA signals with an imbalanced data structure." *IEEE Transactions on Industrial Electronics* 69.2 (2021): 1977-1987.

[11] Tan, Xuemin, et al. "A novel two-stage dissolved gas analysis fault diagnosis system based semi-supervised learning." *High Voltage* (2022):1-16.

[12] Liao, Wenlong, et al. "Fault diagnosis of power transformers using graph convolutional network." *CSEE Journal of Power and Energy Systems* 7.2 (2021): 241-249.

[13] Li, Tianfu, et al. "Multireceptive field graph convolutional networks for machine fault diagnosis." *IEEE Transactions on Industrial Electronics*. 68.12 (2020): 12739-12749.

[14] Liu, Hongying, et al. "Deep fuzzy graph convolutional networks for POLSAR imagery pixelwise classification." *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* 14 (2020): 504-514.

[15] Von Luxburg, Ulrike. "A tutorial on spectral clustering." *Statistics and Computing* 17.4 (2007): 395-416.

[16] Gao, Yiyuan, Mang Chen, and Dejie Yu. "Semi-supervised graph convolutional network and its application in intelligent fault diagnosis of rotating machinery." *Measurement* 186 (2021): 110084.

[17] Li, Qimai, Zhichao Han, and Xiao-Ming Wu. "Deeper insights into graph convolutional networks for semi-supervised learning." *Thirty-Second AAAI Conference on Artificial Intelligence* (2018): 3538-354.

[18] Feng, Wenzheng, et al. "Graph random neural networks for semi-supervised learning on graphs." *Advances in Neural Information Processing Systems* 33 (2020): 22092-22103.

[19] Srivastava, Nitish, et al. "Dropout: a simple way to prevent neural networks from overfitting." *The Journal of Machine Learning Research* 15.1 (2014): 1929-1958.

[20] Zhao, Xiaoli, Minping Jia, and Zheng Liu. "Semi-supervised graph convolution deep belief network for fault diagnosis of elector mechanical system with limited labeled data." *IEEE Transactions on Industrial Informatics* 17.8 (2020): 5450-5460.

[21] Li, Enwen, Linong Wang, and Bin Song. "Fault diagnosis of power transformers with membership degree." *IEEE Access* 7 (2019): 28791-28798.

[22] Ibrahim, Saleh I., Sherif SM Ghoneim, and Ibrahim BM Taha. "DGALab: an extensible software implementation for DGA." *IET Generation, Transmission & Distribution* 12.18 (2018): 4117-4124.

[23] Rao, U. Mohan, et al. "Identification and application of machine learning algorithms for transformer dissolved gas analysis." *IEEE Transactions on Dielectrics and Electrical Insulation* 28.5 (2021): 1828-1835.

[24] Li, Xiaohui, Huaren Wu, and Danning Wu. "DGA interpretation scheme derived from case study." *IEEE Transactions on Power Delivery* 26.2 (2010): 1292-1293.

[25] Van der Maaten, Laurens, and Geoffrey Hinton. "Visualizing data using t-SNE." *Journal of Machine Learning Research* 9.11 (2008).

[26] Zhiyi, He, et al. "Transfer fault diagnosis of bearing installed in different machines using enhanced deep auto-encoder." *Measurement* 152 (2020): 107393.

[27] Bacha, Khmais, Seifeddine Souahlia, and Moncef Gossa. "Power transformer fault diagnosis based on dissolved gas analysis by support vector machine." *Electric power systems research* 83.1 (2012):73-79.

[28] Wang, Bo, et al. "The applications of XGBoost in fault diagnosis of power networks." *IEEE Innovative Smart Grid Technologies* (2019): 3496-3500.

[29] Taha, Ibrahim BM, Saleh Ibrahim, and Diaa-Eldin A. Mansour. "Power transformer fault diagnosis based on DGA using a convolutional neural network with noise in measurements." *IEEE Access* 9. (2021): 111162-111170.

[30] Zhiyi, He, et al. "Transfer fault diagnosis of bearing installed in different machines using enhanced deep auto-encoder." *Measurement* 152 (2020): 107393.

Kazakh-Chinese Neural Machine Translation Based on Data Augmentation

Hao Wu[a], Beiqiang Ma[b]
Key Laboratory of China's Ethnic Languages and Information Technology of Ministry of Education,
Northwest Minzu University, Lanzhou, Gansu 730000, China
[a]moiraen@163.com
[b]bqmasms@163.com

Abstract

Machine translation is an important research field in natural language processing and artificial intelligence, which studies how to use computers to automatically convert languages. We experimented with attentional neural network machine translation for Chinese Kazakh language pairs and used data enhancement methods to alleviate the serious shortage of parallel corpus. The experimental results show that the proposed method is simple and effective, realize the mutual translation between Chinese and Kazakh with extremely low resources. At the same time, our proposed the method of two-way iterative back translation improves +4.47 BLEUs on Chinese to Kazakh and +5.97 BLEU on Kazakh to Chinese respectively.

Kewords: nerual machine translation; data augment;neural network;low-source machine translation

1. INTRODUCTION

Machine translation is an important research field in natural language processing and artificial intelligence, which studies how to use computers to automatically convert different languages. In recent years, with the continuous development of neural network deep learning technology and the continuous improvement of machine translation methods, the performance of neural machine translation models has been continuously improved, gradually surpassing the traditional statistical machine translation method [1]. In the use of large-scale and high-quality parallel corpus tasks, neural machine translation has achieved excellent results, but there are often no rich parallel corpora among most languages in the world. This also leads to the unsatisfactory use of neural machine translation method in some language pair tasks [2]. Chinese-Kazakh is one of the representatives.

Kazakh is a lexical change relatively rich gelling language with fewer speakers, so the Chinese Kazakh corpus bilingual parallel literature is very scarce, which is a big obstacle for data-driven neural machine translation and makes the translation of the translation quality serious decline [3][4]. In order to facilitate the processing of data and avoid confusion, in the experiment, the Kazakh text we used composed of Cyrillic letters.

The Data Augmentation method mainly uses the existing monolingual data to increase the training data [5], so as to better train the translation model to improve the performance. Because the number of bilingual parallel corpora involving Kazakh is extremely scarce, its monolingual corpus is relatively sufficient. In order to make full use of these monolingual corpora and effectively improve the quality of scarce language NMT, we intend to use several improved methods of Back-Translation to expand the number of parallel corpora, that is, use the target language monolingual corpus to construct a synthetic bilingual corpus [6], combined with the original corpus for model training. This method can significantly improve the translation effect [7][8].

2. RELATED WORK

The Data Augmentation method was initially widely used in the field of computer vision, such as performing operations like flipping or transforming the image to be processed. This method has been proven effective in this field and has become one of the standard paradigms in image processing technology [9]. In recent years, data augment technology has also begun to be gradually applied to natural language processing quest such as machine translation [5]. In low-resource language machine translation, due to the lack of sufficient bilingual data, data augment methods mainly use existing monolingual data or bilingual data with pivoting languages to achieve the purpose of increasing training data, so as to better train translation models and improve translation effect [6].

International Conference on Computer Graphics, Artificial Intelligence, and Data Processing (ICCAID 2022)
edited by Ruishi Liang, Jing Wang, Proc. of SPIE Vol. 12604, 1260412
© 2023 SPIE · 0277-786X · doi: 10.1117/12.2674637

2.1.Back-Translation

The method of back translation is one of the most effective methods to alleviate the lack of parallel corpus. In essence, it is to construct artificially synthesized parallel corpus and transform the method of unsupervised learning into a method of supervised learning [10][11]. Use the target language monolingual corpus to construct synthetic bilingual data, and add it to the training corpus, so as to expand the number of corpora.

2.2.Back-Translation with a pivot language.

Among most language pairs, there is often little adequate number of available parallel corpora, but it is often possible to find parallel corpora containing a third language with the source or target language. For example, the parallel corpus between German and Russian is very scarce, but the parallel corpora of German-English and English-Russian is quite sufficient.

In the method of back translation with a pivot language, the source language text needs to be translated into pivot language text, and then the pivot language is translated into the target language. This method can realize the translation between zero-resource language pairs and can be significantly Improve the performance of the translation model [12]. Pivot languages are very effective in zero-resource or low-resource NMT [12][13][14], because they can potentially utilize all feasible training data, including parallel corpus and monolingual corpus. The base process is shown in the figure1.

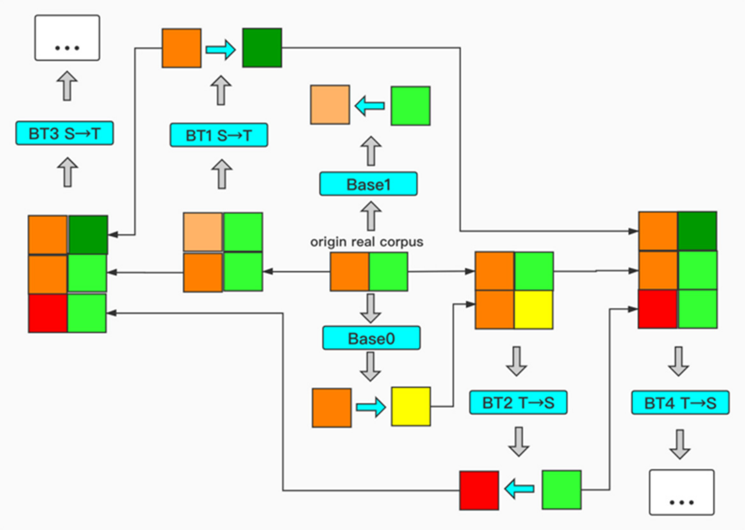


Figure 1. The two-way iterative Back-Translation process

The corpus data resources including English are very rich, so it has become the most commonly used pivot language. But the parallel corpus between English and Kazakh is still very scarce. The sentence number of English-Chinese corpus is very sufficient. Therefore, we will use English-Kazakh corpus to improve the Chinese-Kazakh NMT model.

First, we use the trained English→Chinese NMT model to translate the sentences of English part in English-Kazakh parallel corpus into Chinese sentences, then we can construct a synthetic Chinese-Kazakh corpus and mix it with the real corpus for data augment to train China-Kazakh NMT model.

2.3.Iterative Back-Translation

Iterative Back-Translation strategy is different from conventional back-translation methods [15][16]. We first train a conventional back translation system to obtain more synthetic corpus and mix it with the real corpus. Then we train a more advanced translation system to obtain more synthetic corpus for a larger mixed training data [17][18]. We need to repeat this process over and over again until the performance of model stops improving. The base process is shown in the figure2.

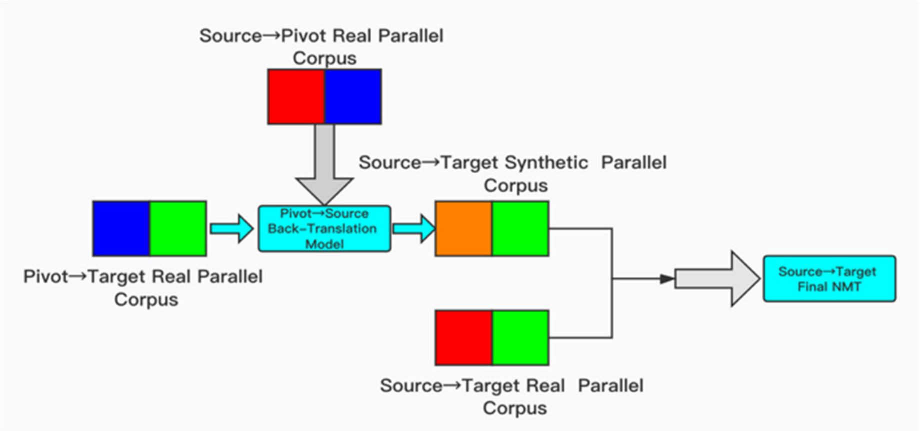

Source→Pivot Real Parallel Corpus
Source→Target Synthetic Parallel Corpus
Pivot→Source Back-Translation Model
Pivot→Target Real Parallel Corpus
Source→Target Final NMT
Source→Target Real Parallel Corpus

Figure 2. The two-way iterative Back-Translation process

We combine both back translation with a pivot language and Iterative Back-Translation approaches and improve the the latter method to operate it simultaneously in both source-target and target-source directions, so that twice data can be obtained with each update, while avoiding the aggregation of synthetic data at only one side.

3. EXPERIMENT AND ANALYSIS

The corpus used in this experiment is as TABLE 1:

TABLE 1. CORPUS WE USED IN EXPERIMENTS

Corpus	***Size***	***Source***
ZH-KK parallel	100k	WMT2019
ZH monolingual	700k	LDC
KK monolingual	700k	WMT2019
EN-ZH parallel	1250k	LDC

3.1 Back-translation via pivot language.

First, we need to train an English→Chinese translation model with the enough EN-ZH parallel from LDC then translate the English part of EN-KK parallel into Chinese sentences. Then the original Chinese-Kazakh corpus was combined with the corpus obtained in the previous step to obtain a total of 150,000 corpus in order to train the original translation model, which we recorded as BASE0 and BASE1.

The synthetic corpus, some of the sentences in the corpus are shown in TABLE 2:

TABLE 2. EXAMPLES OF CORPUS

Source（real）	***Pivot（real）***	***Target（synthetic）***
Ауырсынуды кімнің оңайырақ шыдай алатынын бәріміз білеміз ғой.	It’s anybody’s guess who can stand the pain better.	有人猜测谁能更好地承受痛苦。
Олар жұмыс орындарын емес, жұмысшыларды қорғап жатыр.	They protect the workers, not the jobs.	他们保护工人，而不是工作。
Бұлайша кең ауқымда бағалау өте сирек болады.	Such broad evaluations tend to be rare.	这种广泛的评估往往很少。
Бірақ ақыр соңында сиқыр жойылады.	But the spell will eventually be broken.	但是该咒语最终将被打破。

3.2 Iterative Back-Translation

A total of 150,000 corpus were obtained by combining the original Chinese-Kazakh corpus with the one obtained in the previous step to train the original translation model, which we refer to as BASE0 and BASE1.

Training with the model adopts the open source based on the structure of Transformer OpenNMT system training, the encoder and decoder are set to 6 layers, sharing, the parameters of three of the Dropout probability is set to 0.1, vector set to 0.1%, checkpoint step length is set to 5000, the iteration set to 50 epoches, warmup_steps set to 8000, batch_size set to 2048, using Adam optimization algorithm for optimization.

In the next iteration, 100,000 new synthetic parallel corpora are generated in each iteration for the mixed corpora obtained by the combination of the real parallel corpora and the synthetic parallel corpora. Since the process is bidirectional, a total of 200,000 synthetic parallel corpora are generated in each round. A total of 6 iterations were carried out in 3 cycles, and 8 models were trained. The last training used 738,000 sentences of corpus.

TABLE 3. EXPERIMENTAL RESULTS OF DIFFERENT STEPS

Model	***Direction***	***size***	***BLEU***
Base-0	KK→ZH	150k	10.36
Base-1	ZH→KK	150k	4.18
BT-1	KK→ZH	350k	12.13
BT-2	ZH→KK	350k	7.46
BT-3	KK→ZH	550k	13.47
BT-4	ZH→KK	550k	8.94
BT-5	KK→ZH	750k	14.73
BT-6	ZH→KK	750k	10.15

TABLE 4. EXAMPLES OF KAZAKH → CHINESE TRANSLATION RESULTS

Side	***Sentence1***
Source	Гонконг пен Сингапурдағы жағдай да осымен шамалас.
English	The situation is similar in Hong Kong and Singapore.
Original	香港和新加坡的情况也差不多。
Base-0	在香港和新加坡通常就是这种情况。
BT-1	这种情况通常发生在香港和新加坡。
BT3	这种情况香港和新加坡相似。
BT5	香港和新加坡的情况与此相似。
Side	***Sentence2***
Source	Бүгінгі қарқынды өрлеу мен қор нарықтарының қайта жандануы 2016 жылдың жазынан басталды.
English	Today's rapid growth and revival of stock markets began in the summer of 2016.
Original	最近的增长和股市上升趋势自 2016 年夏天以来一直保持强劲。
Base-0	今天的快速增长和股市复兴始于 2016 年夏天。
BT-1	今天股市的快速上涨和反弹始于 2016 年夏季。
BT3	今天股市的快速增长和复兴始于 2016 年夏季。
BT5	今天的快速增长和股市反弹开始于 2016 年夏天
Side	***Sentence3***
Source	Гонконг пен Сингапурдағы жағдай да осымен шамалас.
English	The situation is similar in Hong Kong and Singapore.
Original	众所周知的大数据无疑将带来重要的科学、技术和医疗成果。
Base-0	很清楚，大数据将推动科学、技术和医学。
BT-1	已知的大数据无疑将提供重要的科学，技术和医学成果。
BT3	很明显，这个大数据正在推动科学，技术和医学的进步。
BT5	香港和新加坡的情况与此相似。显然，大数据将推动科学，技术和医学的发展。
Side	***Sentence3***
Source	Бағымызға орай, қиын жағдай болған жоқ.
English	Fortunately, there were no difficulties.
Original	幸运的是，当时我们避免了最糟糕的情况。
Base-0	令我们高兴的是，没有什么困难的事情发生。
BT-1	令我们高兴的是，没有发生任何困难
BT3	幸运的是，没有困难。
BT5	幸运的是，没有遇到任何困难。

In the process of iterative back translation, the pseudo corpus obtained from the previous round of back translation is added to the training data of machine translation before each iteration. It can be seen from the TABLE III that adding corpus in each round will make the performances of KK→ZH and ZH→KK NMT models has been significantly improved.

After these 6 rounds of iterative back translation, the BLEU value of the Kazakh-Chinese translation system has increased from 10.36 to 14.73, an increase of 4.47 BLEU values, and the BLEU value of the Chinese-Kazakh translation system has increased from 4.18 to 10.15. 5.97 BLEU values.

From the translation results in TABLE IV, we can find that as the number of iterations increases, the translation also undergoes significant changes. However, there are also cases in which translation 2 does not change much in several iterations. This may be due to the fact that this type of text in translation 2 is closer to the literal translation of the original without too much rhetoric of manual translation.

4. CONCLUSIONS

We introduced how to use the back-translation method in data augment to improve the NMT system of the Chinese-Kazakh language pairs and explored the use of English as the pivot language for data augment of the Chinese-Kazakh language corpus, while using the method of two-way iterative back translation improves the performance of the Chinese→Kazakh and Kazakh→Chinese translation model. From the results, the translation performance of the two translation models in the process of mutual training is significantly improved. The performance of the Kazakh→Chinese translation model increased by 4.47 and 5.97 BLEU values respectively. The data augmentation methods we used don’t require any special modification to the model structure, which is conducive to further extension to other tasks.

ACKNOWLEDGMENT

We would like to thank the anonymous reviewers for their valuable comments. We would also like to thank Linjun Zhang, Jiahao Meng, and Beiqiang Ma for their useful suggestion and help with experiments and this paper.

REFERENCES

[1] Sennrich R, Haddow B, Birch A. Edinburgh neural machine translation systems for wmt 16[J]. arXiv preprint arXiv:1606.02891, 2016.rizhevsky A , Sutskever I , Hinton G . ImageNet Classification with Deep Convolutional Neural Networks[C]// NIPS. Curran Associates Inc. 2012.J. Clerk Maxwell, A Treatise on Electricity and Magnetism, 3rd ed., vol. 2. Oxford: Clarendon, 1892, pp.68–73.

[2] Luong M T, Pham H, Manning C D. Effective approaches to attention-based neural machine translation[J]. arXiv preprint arXiv:1508.04025, 2015..

[3] Krizhevsky A , Sutskever I , Hinton G . ImageNet Classification with Deep Convolutional Neural Networks[C]// NIPS. Curran Associates Inc. 2012.J. Clerk Maxwell, A Treatise on Electricity and Magnetism, 3rd ed., vol. 2. Oxford: Clarendon, 1892, pp.68–73.

[4] Hartmann W , Ng T , Hsiao R , et al. Two-Stage Data Augmentation for Low-Resourced Speech Recognition[C]// Interspeech 2016. 2016.

[5] Gao F, Zhu J, Wu L, et al. Soft contextual data augmentation for neural machine translation[C]//Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics. 2019: 5539-5544.

[6] Zhang J J, Zong C Q. Exploiting source-side monolingual data in neural machine translation. In: Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing. Austin, Texas. 2016. 1535–1545

[7] Sennrich R , Haddow B , Birch A . Improving Neural Machine Translation Models with Monolingual Data[C]// Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers). 2016.

[8] Currey A , Heafield K . Zero-Resource Neural Machine Translation with Monolingual Pivot Data[C]// Proceedings of the 3rd Workshop on Neural Generation and Translation. 2019.

[9] Xia M , Kong X , Anastasopoulos A , et al. Generalized Data Augmentation for Low-Resource Translation[C]// Meeting of the Association for Computational Linguistics. 2019.

[10] Yang Z , Chen W , Wang F , et al. Unsupervised Neural Machine Translation with Weight Sharing[J]. 2018.

[11] Edunov S, Ott M, Auli M, Grangier D. Understanding back-translation at scale. In: Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing. Brussels, Belgium.2018.489–500

[12] Firat O, Sankaran B, Al-Onaizan Y, et al. Zero-resource translation with multi-lingual neural machine translation[J]. arXiv preprint arXiv:1606.04164, 2016.

[13] Lakew S M, Lotito Q F, Turchi M, et al. FBK’s Multilingual Neural Machine Translation System for IWSLT 2017[C]//14th International Workshop on Spoken Language Translation (IWSLT 2017). 2017: 35-41.

[14] Jaehong Park, Jongyoon Song, and Sungroh Yoon. 2017. Building a neural machine translation system using only synthetic parallel data. arXiv preprint arXiv:1704.00253.

[15] Hoang V C D, Koehn P, Haffari G, et al. Iterative back-translation for neural machine translation[C]//Proceedings of the 2nd Workshop on Neural Machine Translation and Generation. 2018: 18-24.

[16] Y. Yorozu, M. Hirano, K. Oka, and Y. Tagawa, “Electron spectroscopy studies on magneto-optical media and plastic substrate interface,” IEEE Transl. J. Magn. Japan, vol. 2, pp. 740–741, August 1987 [Digests 9th Annual Conf. Magnetics Japan, p. 301, 1982].

[17] M. Young, The Technical Writer's Handbook. Mill Valley, CA: University Science, 1989.

[18] Jaehong Park, Jongyoon Song, and Sungroh Yoon. 2017. Building a neural machine translation system using only synthetic parallel data. arXiv preprint arXiv:1704.00253.

Study on load calculation of net clothes and applicability of hydrodynamic coefficient under Screen model

Zhaolei Li[a], Fei Wang*[ab], Peng Li[a], Qiang Fu [b], Chunxiao Wang[a]
[a]College of Civil Engineering and Architecture, Shandong University of Science and Technology, Qingdao, China,266590;
[b]CIMC Raffles Offshore Engineering Limited, Yantai, China, 264670
* Corresponding author: wangfei_ouc@163.com

Abstract

In marine ranching, as an important part of marine ranching, the force of netting is extremely complex. The calculation of hydrodynamic load of netting is particularly important for the design and construction of marine ranching. In this paper, the theory of calculating the force of netting at home and abroad is sorted out, the MATLAB software is used to program, and the Screen calculation model is used to study the applicability of the Aarsnes formula used in the current Chinese ' Guidelines for the Inspection of Marine Fishery Facilities '. The applicability of the Aarsnes formula formula between PE material netting, metal netting and nylon netting and the influence of the presence or absence of nodules on the numerical calculation results of the netting are judged respectively, and the applicable conditions of the Aarsnes formula are summarized.

Keywords; Marine ranching, Net hydrodynamic load, Hydrodynamic coefficient, Screen model

1.Introduction

With the increasing demand for seafood, the aquaculture industry is developing rapidly, and the development of artificial marine ranching has become an inevitable trend. Deep water cage is an important facility of artificial marine ranching, and the stress calculation of cage net is an important part of marine ranching construction.

At present, many scholars have studied the hydrodynamic calculation formula of net. Through their experiments with different materials, different shapes and different nodal forms of net, a large number of empirical formulas for calculating the hydrodynamic coefficient of net have been obtained on the basis of different theoretical assumptions. In 1934, Japanese scholar Tauti[1][2] assumed that the hydrodynamic forces between the twine and the knot did not affect each other. The hydrodynamics on the net was experimentally studied, and the formula for calculating the drag of the net with a hanging radio of 0.707 was obtained. The Japanese scholar Miyamoto[3]carried out the test with the net with the horizontal hanging ratio of 0.707 according to the Tauti theory, and obtained the drag coefficients of the twine and the knot of the net with different knot types. Baranow [4]carried out a test in natural waters, with the test speed of 0.042 ~ 0.51m/s. Using the net made of cotton thread with a horizontal hanging ratio of 0.5 and a vertical hanging ratio of 0.87, the drag calculation formula of the net perpendicular to the incoming direction under the action of water flow was obtained. Levien [3]of the Soviet Union considered the influence of the camber formed by the net on the net under the action of water flow, and proposed the resistance calculation formula when the net was perpendicular to the incoming flow. Norway Aarsnes [5]proposed the calculation formulas for the lift coefficient and drag coefficient of nylon or polyethylene mesh with a solidity ratio of 0.13~0.3, a Reynolds number of 1400~1800, and d/a<0.15. In 1991, Løland [6] introduced the application of the Screen calculation model in detail in his doctoral dissertation, and used a mesh with a solidity ratio of 0.13 to 0.317 in a water tank to conduct tests in the range of speed of 0.156 to 0.996 m/s. The formulas for calculating the drag coefficient and lift coefficient of the net under any attack angle are presented.In 2012, kristianse and Faltinsen[7]fitted the drag coefficient and lift coefficient in the experiment under different angles and compactness to obtain the calculation formula for calculating the drag coefficient and lift coefficient of the net. Dong et al. [8]conducted hydrodynamic experiments in a large water tank under pure flow load for metal mesh and PE mesh under the condition of speed of 0.4~1.0m/s, and obtained the result under any attack angle. Calculation formulas for net drag coefficient and lift coefficient related to Reynolds number and compactness. Chinese scholars have also studied the hydrodynamic calculation formula of mesh. In 2005, Li Yucheng and Gui Fukun [9][10] used the knotless hexagonal mesh of nylon material to experimentally derive the calculation formula of the drag coefficient of the knotless hexagonal mesh in the standard state and the tension state, and used the PE mesh with a knot coefficient of 0.707, a compactness of 0.256, and a

International Conference on Computer Graphics, Artificial Intelligence, and Data Processing (ICCAID 2022)
edited by Ruishi Liang, Jing Wang, Proc. of SPIE Vol. 12604, 1260413
© 2023 SPIE · 0277-786X · doi: 10.1117/12.2674661

filling rate of 0.65 to improve the calculation formula of Miyazaki Fangfu. In 2006, Zhan et al. [11]carried out the drag experiment of netting in the flume when the netting is perpendicular to the incoming flow. The knotless nylon netting with a solidity ratio of 0.128 ~ 0.223 was used. The experiment was carried out at a speed of 0.25 ~ 1.0m/s, and the empirical formula of the netting drag coefficient was derived.

The formula used in China 's " Guidelines for the Inspection of Marine Fishery Facilities "[12]is the calculation formula of drag coefficient and resistance coefficient proposed by Aarsnes[5] in 1990. The calculation accuracy of Aarsnes formulaunder different materials and different net forms is compared by using Screen calculation model. The applicable conditions of Aarsnes formula are summarized to provide reference for the design of marine ranching.

2. Net load hydrodynamic calculation

2.1 Screen model introduction

In his doctoral dissertation, Løland[6] introduced in detail the calculation method of the screen model to calculate the force of the net under the action of water flow. The mesh is a plane model consisting of many truss units, which divide the mesh into several panels. The solidity ratio Sn of Screen model is an important parameter, which is calculated by formula (1).

$$S_n = \frac{2d_w}{l_w} - \left(\frac{d_w}{l_w}\right)^2 \tag{1}$$

where, l_w is twine diameter; d_w is twine length.

As shown in Fig. 1. the load on the Screen model under the action of flow is divided into drag along the direction of incoming flow velocity and lift perpendicular to the direction of fluid motion

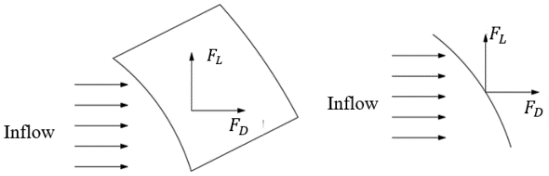


Fig. 1 Stress diagram of Screen model

The drag F_D and lift F_L calculation formula of the net in the Screen model:

$$F_D = \frac{1}{2} C_D \rho A U^2 \tag{2}$$

$$F_L = \frac{1}{2} C_L \rho A U^2 \tag{3}$$

where, C_D is the drag coefficient, C_L is the lift coefficient, A is the equivalent area of the mesh, ρ is the fluid density, U is the relative velocity of the mesh and the fluid

2.2 Hydrodynamic coefficient formula

The selection of hydrodynamic coefficients plays a key role in the calculation of the force of the netting. This paper selects the calculation formulas of lift coefficient and drag coefficient proposed by Aarsnes [5] in 1990, which are used in the 2019 edition of the "Guidelines for the Inspection of Marine Fishery Facilities" [12], to verify the applicability of these three formulas in different materials, different netting structures and different load conditions. The calculation formula of hydrodynamic system is as follows:

$$C_D = 0.04 + \left(-0.04 + S_n - 1.24{S_n}^2 + 13.7{S_n}^3\right) cos\theta \tag{4}$$

$$C_L = \left(0.57S_n - 3.54{S_n}^2 + 10.1{S_n}^3\right) sin2\theta \tag{5}$$

In the formula, θ is the angle of attack of the net, that is, the angle between the net and the flow direction.

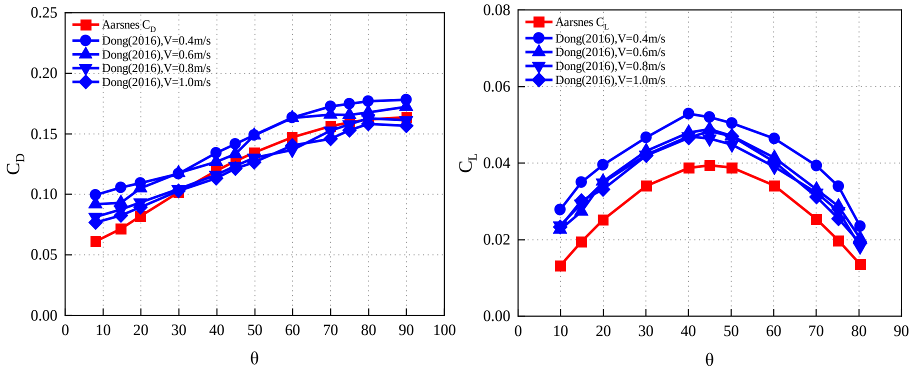
3. Result analysis

3.1 Analysis of hydrodynamic calculation results of different materials netting

In this paper, Dong [8] ' s metal mesh experimental data, knotless PE mesh experimental data and Cheng zhou [14] ' s knotless nylon mesh experimental data are used to verify the applicability of Aarsnes ' hydrodynamic coefficient formula to calculate different material mesh.

As shown in Fig. 2(a), the experimental results of the drag coefficient of Dong [8] metal mesh are compared with the numerical results. The experimental data of metal mesh by Dong [8] at any incidence angle are used. The comparison results show that the error between the calculation results of Aarasnes hydrodynamic coefficient formula and the experimental results is small, and the average relative errors are 15.3 %, 11.6 %, 6.7 % and 6.9 % respectively when the flow velocity is 0.4m / s, 0.6m / s, 0.8m/sand 1.0m/s. When the flow velocity is less than 0.6m / s, the calculated value of drag coefficient is less than the experimental value. With the increase of flow velocity, the calculated value of hydrodynamic coefficient will gradually be greater than the experimental value.

In Fig.2(b), the experimental results of the lift coefficient of Dong[8]metal net are compared with the numerical results. By comparison, it is found that the calculated values are less than the measured values. When the flow velocity is 0.4~0.6m/s, there is a large error between the calculated results and the measured values, and the average relative error is greater than 25%. When the flow velocity is 0.8m/s and 1.0m/sand the angle of attack is in the range of 30°~70°, the values calculated by Aarsnes formula are close to the measured values, and the average relative error is less than 20 %.

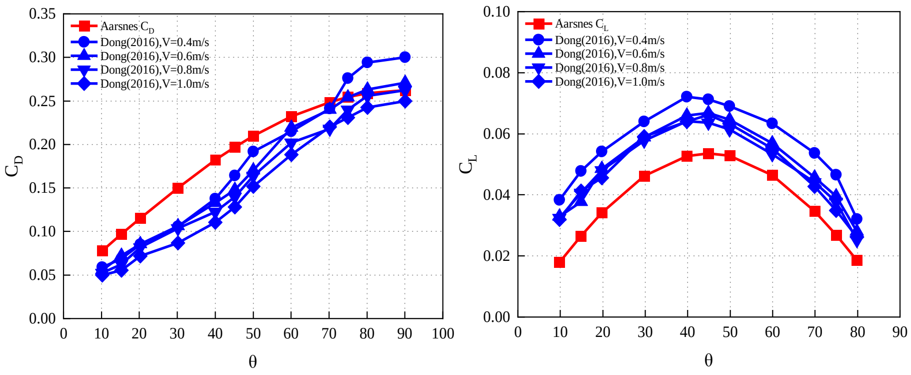


(a)Drag coefficient (b)Lift coefficient

Fig. 2 Comparison of Experimental and Calculated Hydrodynamic Coefficients of Dong ' s Metal Mesh

As shown in Fig. 3(a), the test results of drag coefficient of knotless PE netting are compared with the calculation results of hydrodynamic coefficient. The experimental data were selected from Dong[8]' s experimental data of the net resistance coefficient of the knotless PE net at any angle of attack under pure current load. By comparison, it was found that when the angle of attack was greater than 45 ° and the flow rate was less than 1.0m / s, the results calculated by the Aarsnes hydrodynamic coefficient formula were close to the experimental results. When the flow rates were 0.4m / s, 0.6m/sand 0.8m / s, the average alignment errors were 8.8 %, 6.2 % and 10.6 %, respectively.

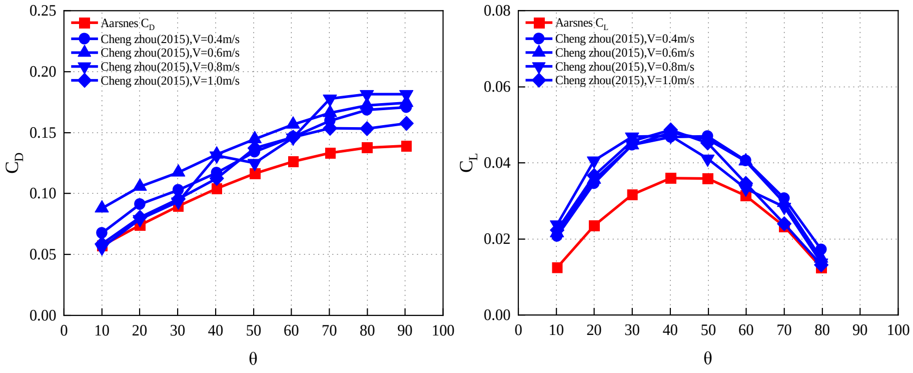
Fig. 3(b) shows the comparison between the experimental results and the numerical results of the lift coefficient of Dong[8] PE non-knotted net. It is found that the results calculated by Aarsnes hydrodynamic coefficient formula are close to the experimental results. When the flow velocity is 0.4m / s, 0.6m / s, 0.8m/sand 1.0m / s, the corresponding average relative errors are 16 %, 8.7 %, 17.7 % and 15.9 % respectively. It is found that the calculation results are closer to the experimental values when the incidence angle is greater than 30 °.



(a)Drag coefficient (b)Lift coefficient

Fig. 3 Comparison of experimental and calculated hydrodynamic coefficients of Dong ' s knotless PE net

Fig.4(a) shows the comparison between the numerical calculation results and the experimental results of the drag coefficient of the knotless nylon net. Fig.4(b) shows the comparison between the numerical calculation results and the experimental results of the lift of the knotless nylon net. The data are based on the experimental data of drag coefficient and lift coefficient measured by Cheng Zhou et al.[14] in the hydrodynamic experiment of knotless nylon net in 2015. By comparison, it is found that the calculated drag force is closer to the experimental value. When the water velocity is 0.4m / s, 0.6m / s, 0.8m/sand 1.0m / s, the average relative errors are 15.3 %, 23.1 %, 14 % and 9.7 % respectively. However, when the flow velocity is 0.4 m/sand 0.6 m / s, the overall error between the lift calculation results of the knotless nylon net and the experimental results is large, and the average relative error is greater than 26 %. When the flow velocity is 0.8m/sand 1.0m / s, and the angle of attack is in the range of 50 ° ~ 80 °, the lift coefficient calculated by Aarsnes formula is closer to the experimental value.

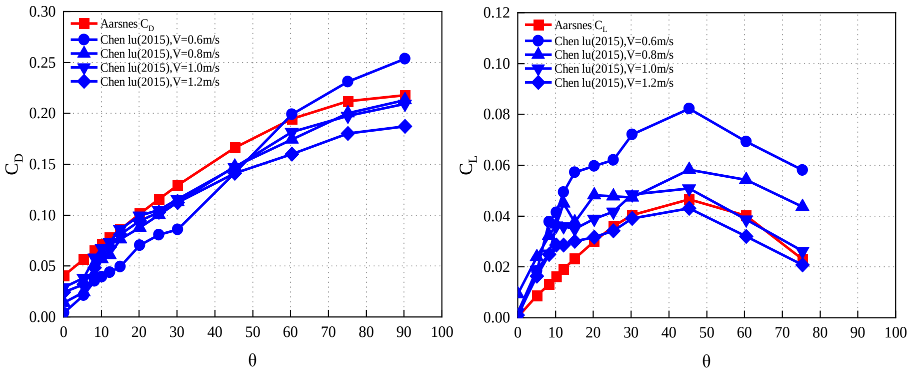


(a)Drag coefficient (b)Lift coefficient

Fig.4 Comparison of experimental and calculated hydrodynamic coefficients of knotless nylon net by Cheng Zhou

3.2 Analysis of hydrodynamic calculation results of knot nets

As shown in Fig. 5 (a), the experimental data and calculation results of knotted nylon net in Chen lu [14]' s 2015 master 's thesis are compared. Comparing the calculation results of the drag coefficient with the experimental results, it is found that when the incoming flow velocity is 0.6 m/sand the angle of attack is greater than 45 °, the average relative error of the formula calculation result is closer to the experimental value, which is 10.6 %. When the flow velocity is 0.8 m / s, the angle of attack is 15 ° ~ 90 ° and the flow velocity is 1.0 m/sand 1.2 m / s, the angle of attack is 10 ° ~ 90 °, the calculation results of the Aarsnes hydrodynamic coefficient formula are closer to the experimental values, and the maximum relative error is less than 18.8 %.



(a)Drag coefficient (b)Lift coefficient

Fig. 5 Comparison of Experimental and Calculated Hydrodynamic Coefficients of Chenlu Knotted Nylon Net

Fig. 5 (b) compares the experimental results and numerical results of the lift coefficient of the knotted nylon net under pure current load in Chen[14]. When the flow velocity is 1.0 m/sand 1.2 m / s, and the angle of attack is 20 ° ~ 75 °, the Aarsnes hydrodynamic coefficient formula is closer to the experimental value, and the maximum relative error is less than 20 %.

Fig. 6 shows the comparison between the numerical results and the experimental results of the resistance coefficient of the knotted PE net. The experimental data of knotted PE net by Li and Gui [9][10]are used. The calculation results of the Aarsnes hydrodynamic coefficient formula are very close to the experimental values. When the flow rates are 0.3 m / s, 0.45 m/sand 0.6 m / s, the average relative errors between the calculation results of the Aarsnes hydrodynamic coefficient formula and the experimental results are 5.7 %, 7.9 % and 12.2 %, respectively, and there is a tendency to increase with the increase of the flow velocity.

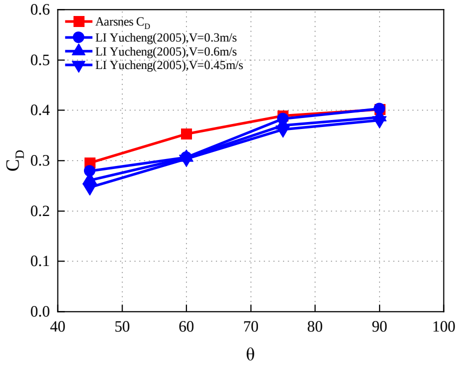


Fig. 6 Comparison between experimental and calculated drag coefficient of Li Yucheng PE net

4. Conclusion

In this paper, the Screen model is used to calculate the force of the net. By comparing with the experimental data, the scope of application of the Aarsnes hydrodynamic coefficient formula is obtained. The following conclusions are obtained:

The Aarsnes drag coefficient calculation formula is suitable for calculating metal netting, knotless nylon netting and knotted PE netting. It also has good accuracy for knotless PE netting with flow velocity less than 0.8 m/sand knotted nylon netting with flow velocity greater than 1.0 m / s.

The Aarsnes lift coefficient formula is suitable for calculating the knotless PE network.

Because the Aarsnes formula is a function related to solidity ratio, it will not change with the change of flow velocity, but from the experimental data, the hydrodynamic coefficient of the screen model will change with the change of flow velocity. Therefore, the Aarsnes formula is suitable for a limited range of flow velocity.

References

[1] Tauti, M. A relation between experiments on model and full scale of fishing net. Nippon Suisan Gakkaishi,171-177(1934).

[2] Tauti, M. The force acting on the plane net in motion through the water. Nippon Suisan Gakkaishi, 1-4(1934).

[3] Zhou Yingqi, Xu Liuxiong. Mechanics of fishing gear., Science Press,53-78(2018).

[4] Baranow, F.I., Theory and assessment of fishing gear. Fish Industry Press, Moscow, (1948).

[5] Aarsnes J V, Rudi H, Løland G., Current forces on cage, net deflection, Engineering for offshore fish farming. Proceedings of a conference organised by the Institution of Civil Engineers. Glasgow, UK, 137-152(1990).

[6] Løland G., Current Force On and Flow Through Fish Farm. Division of Marine Hydrodynamics, The Norwegian Institute of Technology, (1991).

[7] Kristiansen T, Faltinsen O M. Modelling of current loads on aquaculture net cages., Journal of Fluids and Structures, 34:218-235(2012).

[8] Dong Shuchuang, Hu Fuxiang, KUMAZAWA TAISEI, SIODE DAISUK, Tokai Tadashi., Hydrodynamic characteristics of netting panel used for aquaculture net cages in uniform current. Nippon Suisan Gakkaishi82(3), 282–289(2016).

[9] Li Yucheng, Gui Fukun. Experimental study and seleceion of the drag coefficient of knotted and knotless fishing net [J]. China Offshore Platform, 20(6):11-17(2005).

[10]Gui Fukun. Hydrodynamic behavious of deep-water gravity cageD]. Dalian University of Technology, 74-75(2006).

[11]Zhan, J.M., Jia, X.P., Li, Y., Sun, M.G., Guo, G.X., Hu, Y.Z., Analytical and experimental investigation of drag on nets of fish cages. Aquacultural Engineering35,91-101(2006).

[12]GD14-2019, Guidelines for the Inspection of Marine Fishery Facilities,33-34(2019).

[13]Rudi H, Løland G, Furunes L. Model tests with net enclosures. Forces on and flow through single nets and cage systems, Technical Report, (1988).

[14]Cheng Zhou, Liuxiong Xu, Fuxiang Hu, Xiaoyu Qu., Hydrodynamic characteristics of knotless nylon netting normal to free stream and effect of inclination, Ocean Engineering,89-97(2015).

[15]Chen lu. Experomental study on the hydrodynamic coefficients of plane nettings and numerical simulation, Shanghai Ocean University, 20-24(2015).

Attitude Reference Transfer Method for Moving Base Based on Correlated Frames

Lifa Nong, Wei Wu*, Xingshu Wang
College of Advanced Interdisciplinary Studies, National University of Defense Technology,
Changsha, Hunan, China 410073
* Corresponding author: 253050799@qq.com

ABSTRACT

When the monocular visual measurement method is adopted, the initial alignment precision of the moving base tends to be easily interrupted by the surrounding environment. In order to solve this problem, in this study, the authors suggest an initial attitude reference transfer method for moving base based on correlated frames of images. This method utilizes the advantage of inertial reference, namely, high precision of relative attitude measurement within a short time period, transforms the multi-frame sequence images of the planar target to the coordinate system of the initial moment, and averagely reduces noise interruption through more than one image, so that the attitude-calculating precision could be promoted. Simulation research shows that: The attitude reference method based on correlated frames, which is put forward in this study, could effectively reduce random noise interruptions. Within the rotation scope of -1° to 1°, the precision of attitude alignment angles in three directions is better than that of 45".

Keywords: Attitude Reference Transfer; Monocular Visualization; Planar Target; Inertial Reference; Correlated Frames of Attitude

1. INTRODUCTION

Moving combat units on large-sized naval vessels, such as shipboard aircraft, amphibious tanks, and small-sized landing craft, are all fitted with inertial reference devices. Before such combat units are put into use, these devices need to undergo initial alignment [1]. Methods of initial alignment for part of the inertial reference devices under the conditions of sea movement masterly include Inertia/Satellite Navigation Self-aligned Approach [2], Transfer Alignment Method [3], and Camera Measurement Method [4]. Through the camera and collaboration target, the Camera Measurement Method connects the object to be measured and the measuring reference, and after that, the three-dimensional attitude, position, etc. of the object to be measured with respect to the measuring reference are transmitted. This method has attracted extensive attention, as it has the advantages of high precision, non-contact features, real-time measurement, dynamic measurement, simple application, etc. [5-7].

The most suitable choice for collaboration target is the planar target, because it has the advantages of simple structure and flexible mounting, and also because there are always limited spaces on naval vessels. When the camera's parameters are given, the relative attitude between the camera and the plane collaboration target could be deemed as the solution for the PnP (Perspective-n-Points) Problem [8]. An orthographic iterative algorithm is currently one of the most effective methods to solve the PnP Problem [9]. In consideration of the wobbling conditions of naval vessels, as well as the features of the monocular visualization system, simple structure, and small computation burden, Tan X.L. [10] suggested a monocular visual measurement method of rapid alignment for shipboard aircraft, which could efficiently avoid problems like lever arm effect, large misalignment angle, or the problem that the alignment precision of horizontal attitude angle would be on the low side when the time period for initial alignment is shortened. This problem is masterly caused by the fact that monocular visual measurement is not sensitive enough to the information of the depth of focus, and therefore the imaging process of the collaboration target is often easily interrupted by random noises, especially the extraction errors of feature points [11]. For this reason, in order to raise the anti-interruption capacity of monocular visual measurement, the relative attitude between the camera and the collaboration target is often estimated through the adoption of sequence images [12-13].

Under the influence of complicated environments of naval vessels, such as the influence of sea waves and winds, moving combat units under the "non-restrained" status may move slightly at any moment, which makes it difficult for the camera to collect multi-frame images in the same position. In order to solve this problem, this study suggests an initial attitude reference transfer method for moving base based on correlated frames of images. This method sets the attitude

International Conference on Computer Graphics, Artificial Intelligence, and Data Processing (ICCAID 2022)
edited by Ruishi Liang, Jing Wang, Proc. of SPIE Vol. 12604, 1260414
© 2023 SPIE · 0277-786X · doi: 10.1117/12.2674641

measurement model for the planar target as the foundation and, through the utilization of inertial reference to accurately measure the changing level of the camera's attitude, gathers up and correlates the multi-frame images of the planar target in the same position, which can effectively reduce the influence of measurement noises and realize the high-precision initial attitude alignment of moving combat units. Finally, the feasibility of this method is assessed through simulation analysis.

2. MEASUREMENT MODEL OF THE ATTITUDE OF THE PLANAR TARGET

2.1 Basic principles

As far as actual measurement is concerned, the camera is fixed on the inertial local reference, while the plane collaboration target is stuck to the deck or basement of the naval vessels, as shown in Figure 1. The master reference coordinate system of naval vessels (*m* system) is defined as $O_m - X_m Y_m Z_m$; the local reference coordinate system of the moving combat units (*s* system) is defined as $O_s - X_s Y_s Z_s$; the optical center of the camera is defined as the original heart, while the optical axis as the Z_c axis. The camera coordinate system (*c* system) is defined as $O_c - X_c Y_c Z_c$; the planar target coordinate system (*w* system) is defined as $O_w - X_w Y_w Z_w$. As fixed mounting is utilized, the mounting angle $\boldsymbol{R}_s^c$ between the camera coordinate system and the local reference coordinate system of the moving combat units can be calibrated. The mounting angle $\boldsymbol{R}_w^m$ between the planar target coordinate system and the master reference coordinate system of the naval vessels is obtained by inertial vector matching measurement [14].

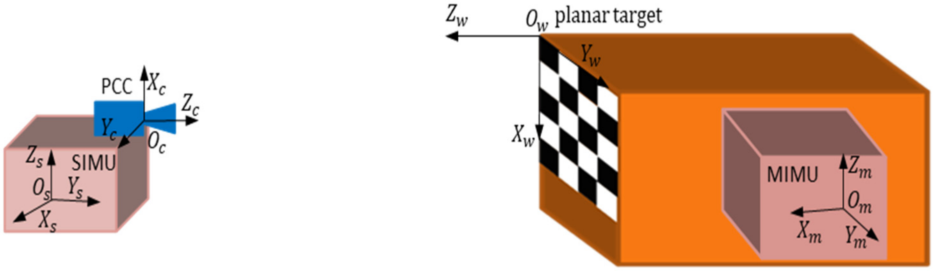


Figure 1 Camera and inertial reference coordinate system

According to the basic principles of measurement with a camera, if a camera is used to continually produce images of the plane collaboration target, features points of the target images could be extracted and, through the change of image points of the feature points, we would be able to estimate the attitude relationship between the camera and the plane collaboration target. The imaging of plane collaboration target is close-shot imaging. When the theorem of pin-hole imaging is satisfied, an arbitrary mark point $P_j\left(j=1,2,\cdots,\mathrm{d},\mathrm{d}\geq 4\right)$ in the planar target coordinate system $\left(X_w,Y_w,Z_w\right)$ is transformed into the image pixel coordinate system $\left(u,v\right)$ as corresponding image point p_j . The transformation relationship is

$$\begin{bmatrix} u \\ v \\ 1 \end{bmatrix} = \frac{1}{Z_c} \begin{bmatrix} F_x & 0 & u_0 & 0 \\ 0 & F_y & v_0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} \boldsymbol{R}_w^c & \boldsymbol{T}_w^c \\ 0^{\mathrm{T}} & 1 \end{bmatrix} \begin{bmatrix} X_w \\ Y_w \\ Z_w \\ 1 \end{bmatrix} \tag{1}$$

$\left(u_0,v_0\right)$ is the principal point of the image; $\left(F_x,F_y\right)$ denotes equivalent focal length of the camera in the horizontal and vertical directions; $\boldsymbol{R}_w^c$ and $\boldsymbol{T}_w^c$ are respectively the rotation matrix and translation vector of the *w* system and *c* system.

When the camera's parameters, aberration coefficients, and coordinates are given, the spin matrix evaluated by Equation (1) could be set as the initial value, while the orthographic iterative algorithm [9] would be utilized for optimization solution, which is capable of finishing the attitude estimation of single-frame planar target images.

2.2 The correlated equations of sequence images of the planar target

In order to enhance the capacity for reducing random noises in monocular visual measurement, this study refers to the thoughts of the attitude of correlated frames [15] and adopts multi-frame correlated images for estimation of the attitude of the planar target. First, ensure that the plane collaboration target would be still. The camera and composite of local reference would be rotated at small angles, while the planar target images shot by the camera in sliding movement, and the attitude information of measurement of the local reference could be collected at the same time; second, correlate two neighboring planar target images through the attitude variation figured out with accurate measurement $\boldsymbol{R}_{S_k}^{S_{k-1}}$ of the local reference, so that the planar target images, which were mutually independent to each other in the past, could be correlated one after one, as shown in Figure 2; finally, more than one images which were correlated should be included and averaged as the final attitude of the planar target. The equation of planar target attitude subject to multi-frame correlation would be deduced in detail in the following.

From the hand-eye calibration equation $\boldsymbol{AX} = \boldsymbol{XB}$, the rotation relationship between the camera and local reference should be

$$\boldsymbol{R}_{C_k}^{C_{k-1}} \boldsymbol{R}_S^C = \boldsymbol{R}_S^C \boldsymbol{R}_{S_k}^{S_{k-1}} \tag{2}$$

In this equation, $\boldsymbol{R}_{S_k}^{S_{k-1}} = \left(\boldsymbol{R}_{S_{k-1}}^i\right)^{\mathrm{T}} \boldsymbol{R}_{S_k}^i$ is the attitude variation of the camera for measurement of the local reference from the time points of frame *k*-1 to frame *k*; $\boldsymbol{R}_{C_k}^{C_{k-1}} = \left(\boldsymbol{R}_{C_{k-1}}^w\right)^{\mathrm{T}} \boldsymbol{R}_{C_k}^w$ is the attitude variation from the time points of frame *k*-1 to frame *k*. Upon consolidation of Equation (2), the planar target attitude of correlated neighboring images could be figured out as follows

$$\boldsymbol{R}_{C_k}^w = \boldsymbol{R}_{C_{k-1}}^w \boldsymbol{R}_S^c \left(\boldsymbol{R}_{S_{k-1}}^i\right)^{\mathrm{T}} \boldsymbol{R}_{S_k}^i \left(\boldsymbol{R}_S^c\right)^{-1} \tag{3}$$

When the planar target attitude of a single frame is given, the fact that the short-time measurement precision of local reference is high could be utilized to figure out the attitude relationship between neighboring image frames, and the equation of the multi-frames correlated target attitude can be derived as

$$\bar{\boldsymbol{R}}_{C_k}^w = \frac{1}{k}\left(\boldsymbol{R}_{C_k}^w + \boldsymbol{R}_{C_{k-1}}^w \boldsymbol{R}_S^c \boldsymbol{R}_{S_k}^{S_{k-1}} \left(\boldsymbol{R}_S^c\right)^{-1} + \cdots + \boldsymbol{R}_{C_1}^w \boldsymbol{R}_S^c \boldsymbol{R}_{S_k}^{S_1} \left(\boldsymbol{R}_S^c\right)^{-1}\right) \tag{4}$$

In this equation, *k* denotes the quantity of correlated planar target image frames; $\bar{\boldsymbol{R}}_{C_k}^w$ denotes that, after *k* image frames are correlated, they would be averaged as the ultimate planar target attitude of frame *k*. From equation (4), it can be seen that the ultimate planar target attitude of frame *k* has been jointly solved and determined by the *k* frames of planar target images in the frame sequence. To obtain the *k*+1 frame, the correlated frames sequence needs to be updated, and the *k*+1 frame is needed in the frame sequence and the target frame at the end of frame sequence has to be removed, and so forth.

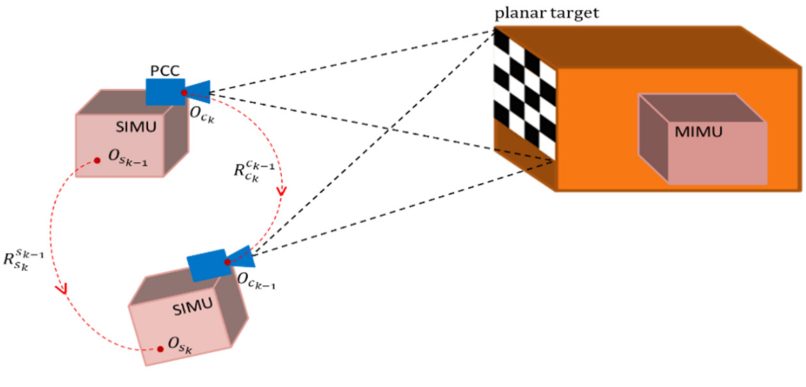


Figure 2 Sketch diagram of the correlated frames of the planar target

3. INITIAL ATTITUDE REFERENCE TRANSFER SCHEME FOR MOVING BASE

When the mounting angles between the camera and local reference, between the planar target and master reference, the reference relationship between the camera and planar target can be accurately measured through the attitude equation of the planar target based on multi-frame correlation, while the attitude reference transfer between the local reference coordinate system and the master reference coordinate system could be realized. The attitude transfer equation could be indicated as

$$\boldsymbol{R}_S^n = \boldsymbol{R}_m^n \boldsymbol{R}_W^m \bar{\boldsymbol{R}}_{C_k}^W \boldsymbol{R}_S^c \tag{5}$$

In this equation, n denotes the coordinate system of navigation; $\boldsymbol{R}_m^s = \left(\boldsymbol{R}_W^m \bar{\boldsymbol{R}}_{C_k}^W \boldsymbol{R}_S^c\right)^{\mathrm{T}}$ denotes the attitude relationship between the master reference and local reference; $\boldsymbol{R}_m^n$ denotes the attitude information of measurement of the master reference.

Under the environment of movement of naval vessels, sensitive angular movement information of local reference includes the absolute attitude of the master reference, as well as the relative attitude between the local reference and the master reference. As camera measurement could only obtain the relative attitude between the camera and planar target, the attitude information measurement of the local reference could not be utilized directly. Therefore,this study puts forward a method that could improve the correlated frames of attitude under the conditions of moving base. In this study, the attitude information of the master reference and local reference is adopted for estimation of the camera’s attitude variation, namely

$$\begin{cases} \boldsymbol{R}_{S_k}^n = \boldsymbol{R}_{m_k}^n \boldsymbol{R}_{S_k}^{m_k} \\ \boldsymbol{R}_{S_{k-1}}^n = \boldsymbol{R}_{m_k}^n{-1} \boldsymbol{R}_{S_{k-1}}^{m_{k-1}} \end{cases} \tag{6}$$

As the alignment of local reference still has not been finished, $\tilde{\boldsymbol{R}}_{S_k}^n$ would be used to stand for $\boldsymbol{R}_{S_k}^n$. Equation (6) will be further consolidated

$$\Delta\tilde{\boldsymbol{R}}_S^m\left(k-1\,|\,k\right) = \boldsymbol{R}_{S_k}^{m_k}\left(\boldsymbol{R}_{S_{k-1}}^{m_{k-1}}\right)^{\mathrm{T}} = \left(\boldsymbol{R}_{m_k}^n\right)^{\mathrm{T}} \tilde{\boldsymbol{R}}_{S_k}^n \left(\tilde{\boldsymbol{R}}_{S_{k-1}}^n\right)^{\mathrm{T}} \boldsymbol{R}_{m_{k-1}}^n \tag{7}$$

In this equation, $\boldsymbol{R}_{m_k}^n$ denotes the attitude value of measurement of the master reference in the time of frame k; $\tilde{\boldsymbol{R}}_{S_k}^n$ denotes the result of navigation calculation based on the information of position, velocity, and attitude further transmitted by the master reference through the utilization of data collected from the spinning top and accelerometer for local reference in the time of frame k. In view of the fact that the mounting angles of the master reference and local reference are very small, while the time intervals between different frames are very short, that $\tilde{\boldsymbol{R}}_{S_k}^n\left(\tilde{\boldsymbol{R}}_{S_{k-1}}^n\right)^{\mathrm{T}}$ could satisfy the requirements of small angle and approximation. $\Delta\tilde{\boldsymbol{R}}_S^m\left(k-1\,|\,k\right)$ could be deemed as the attitude variation $\boldsymbol{R}_{C_k}^{C_{k-1}}$ of the camera from the time of frame k-1 to frame k. After the consolidation of Equations (4) and (7), the attitude equation of planar target of multi-frame correlation under the moving base environment could be figured out

$$\bar{\boldsymbol{R}}_{C_k}^W = \frac{1}{k}\left(\boldsymbol{R}_{C_k}^W + \boldsymbol{R}_{C_{k-1}}^W \Delta\tilde{\boldsymbol{R}}_S^m\left(k-1\,|\,k\right) + \cdots + \boldsymbol{R}_{C_1}^W \Delta\tilde{\boldsymbol{R}}_S^m\left(1\,|\,k\right)\right) \tag{8}$$

To sum up, the procedures of the initial attitude transfer scheme for moving base based on correlated frames are shown in Figure 3.

Step 1: Calibrate the camera's parameters, distortion parameters, mounting angles of the camera and local reference, as well as mounting angles of the planar target and master reference;

Step 2: Collect concurrently the planar target images shot by the camera in sliding movement, as well as the attitude information of measurement of the master reference and the local reference;

Step 3: Correct the distortions of the sequence images of the planar target;

Step 4: Extract the coordinates of feature points of the images of the planar target, and adopt the orthographic iteration method to estimate the attitude of each image of the planar target;

Step 5: Adopt the attitude equation of the planar target featuring multi-frame linkage to realize the correlation of the sequence images and figure out the ultimate attitude of the planar target;

Step 6: Adopt the attitude transfer equation to realize the attitude transfer between the local reference coordinate system and the master reference coordinate system.

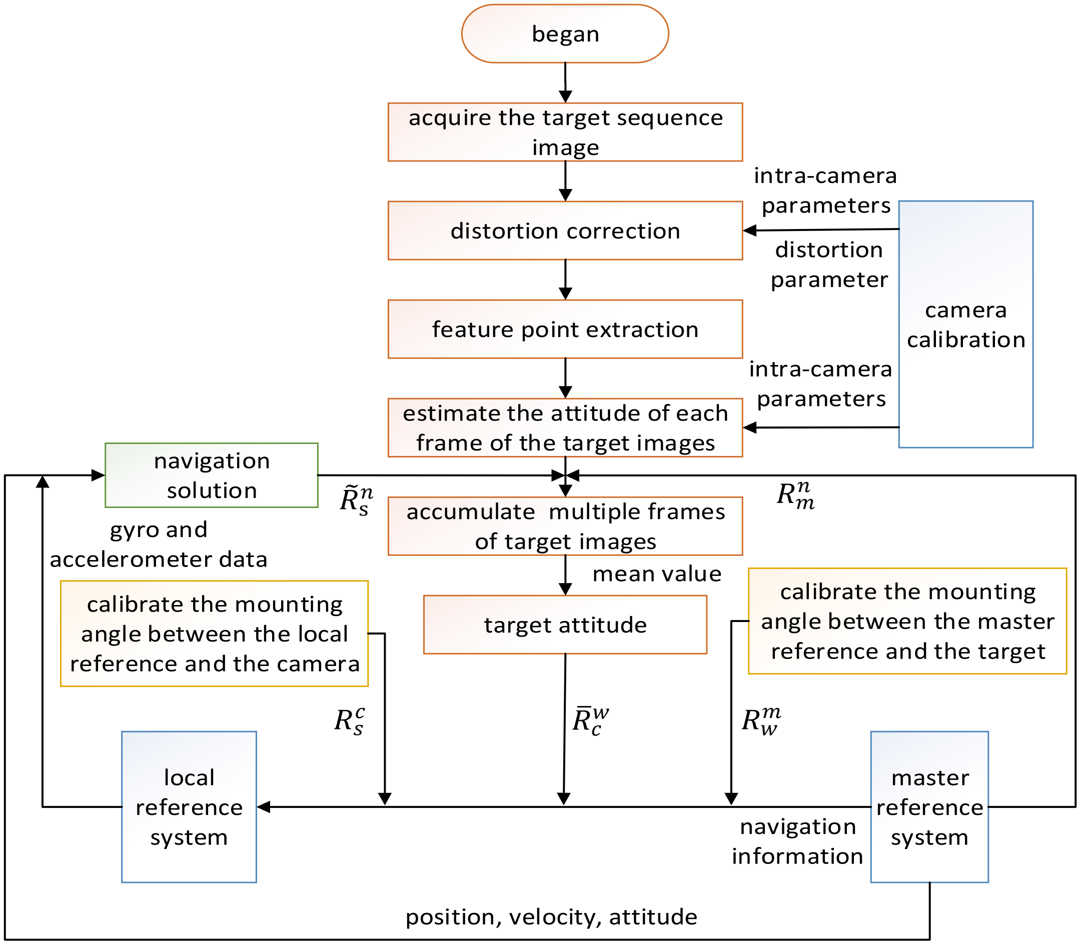

began
acquire the target sequence image
distortion correction
feature point extraction
estimate the attitude of each frame of the target images
intra-camera parameters
distortion parameter
camera calibration
intra-camera parameters
navigation solution
$\tilde{R}_s^n$
accumulate multiple frames of target images
R_m^n
gyro and accelerometer data
mean value
calibrate the mounting angle between the local reference and the camera
target attitude
calibrate the mounting angle between the master reference and the target
R_s^c
$\bar{R}_c^w$
R_w^m
local reference system
master reference system
navigation information
position, velocity, attitude

Figure 3 Attitude transfer procedures of moving base

4. EMULATION ANALYSIS

In order to validate the initial attitude reference transfer method for moving base based on correlated frames of images which can be referred as Correlated Frames (CF) method for short, the influences of the number of correlated frames, error in image point extraction, and error in mounting angle ($\boldsymbol{R}_S^c$ and $\boldsymbol{R}_W^m$) on the precision of attitude alignment transfer would respectively be taken into consideration. As far as emulation is concerned, the error setting of the spinning top and accelerometer of the master reference and local reference is shown in Table 1; the setting of measurement parameters and conditions of the perspective models of the selection center of imaging models, non-camera distortion, calibrated error of the camera, and attitude of the planar target, is shown in Table 2. The setting of emulation experiment data is as follows:

(1) Configure attitude variation of the master reference and local reference, which has been caused by factors such as waves on the sea, and satisfy the sinusoidal variation of amplitude being 1° and frequency being 2π;

(2) Randomly generate the mounting angle $\boldsymbol{R}_S^{c_{\text{real}}}$ of the camera and local reference, as well as the mounting angle $\boldsymbol{R}_W^{m_{\text{real}}}$ of the planar target and master reference;

(3) Randomly generate 60 moving matrices of orthographic attitude;

(4) Set the initial value $\boldsymbol{R}_{S_1}^{i_{\text{real}}}$, and generate the corresponding local reference attitude matrix $\boldsymbol{R}_{S_k}^{i_{\text{real}}}$ according to equation (3) when the camera shoots image frame *k*;

(5) Set the translation vector of the camera and planar target as (0,0,2000) mm. When the coordinates of mark points $\left(X_w, Y_w, Z_w\right)$ are given, the coordinates of real image points corresponding to the mark points of the collaboration target generated in each movement could be figured out according to Equation (1).

This study adopts the Root-mean-square Error (RMS) between the estimated value $\boldsymbol{R}_{s_{\text{est}}}^{n}$ and the actual value $\boldsymbol{R}_{s_{\text{real}}}^{n}$ as the indicator of precision assessment.

Table.1 Parameters for gyro error and accelerometer error

Error parameters	**Gyro constant bias**	**Gyro random walk**	**Accelerometer constant bias**	**Accelerometer random walk**
	$\left(^\circ/\text{h}\right)$	$\left(^\circ/\sqrt{\text{h}}\right)$	$\left(\mu g\right)$	$\left(\mu g/\sqrt{\text{Hz}}\right)$
MIMU	0.003	0.001	5	5
SIMU	0.01	0.001	50	50

Table.2 Parameter conditions for attitude measurement of a planar target

Simulation parameter	**Condition setting**
Feature point coordinate /mm	(-100,100,0), (-100,-100,0), (100,100,0), (100,-100,0)
Production error of cooperative target /mm	0.01
Pixel dimension /μm	3.45×3.45
Focal length of camera /mm	25
principal point coordinates of camera /pixels	(1224,1024)

4.1 Influence of the number of correlated frames on the precision of attitude alignment

Errors to some extent exist in the extraction of image points of the planar target, and the calibration of mounting angle. In order to better analyze the relationship between the number of correlated frames and attitude alignment precision, the authors added Gaussian white noise with the average value of 0 and the standard deviation of 0.05 pixels to the coordinates of image points, and added Gaussian amplifier errors with the average value of 0 and the standard deviation of 15", and then the emulation results are shown in Figure 4.

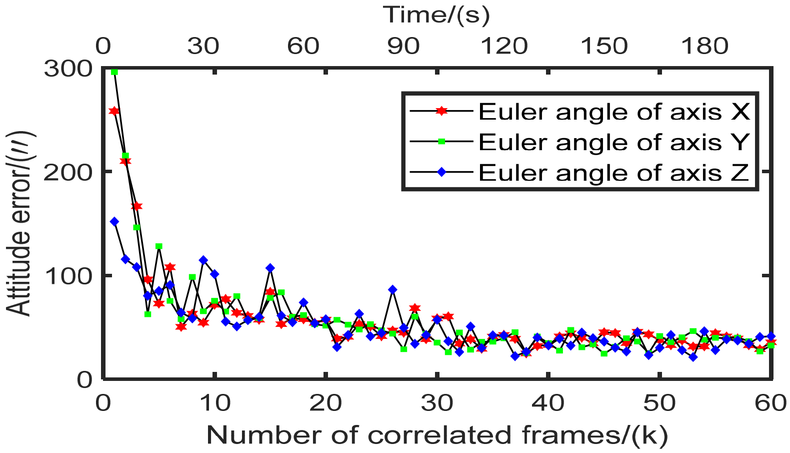
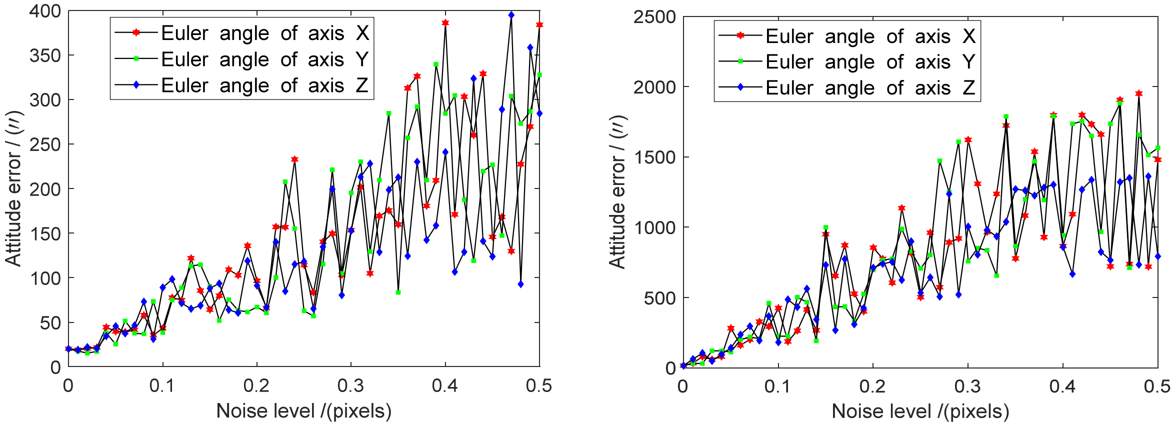


Figure 4 The relationship between errors in attitude alignment and the number of correlated frames (k)

From Figure, when error in image point extraction and error in mounting angle are certain, with the increase of the number of correlated frames k, alignment errors of the Eulerian angles of the X, Y, and Z axes will gradually be reduced. When $k \geq 35$, errors in attitude alignment tend to become stabilized. For this reason, $k = 35$ could serve as the optimal quantity of correlated frames and be used in the subsequent emulation analysis. It could be seen from the results that the method suggested in this study could promote the alignment precision of the Eulerian angles of the X and Y axes (horizontal attitude angle) to a level approximately equal to that of the Eulerian angles of the Z axis, while it only takes approximately 100s for the alignment results of attitude to become stabilized.

4.2 Influence of errors in the extraction of image points on the precision of attitude alignment

When the errors in the mounting angle are set as Gaussian amplifier errors with the average value of 0 and standard deviation of 15". Add Gaussian white noise with the mean value of 0 and standard deviation of 0~0.5pixels to the image point coordinates of 60 target images one by one. The interval of standard deviation amplitude variation is 0.01pixel. When $k=35$, the statistical results of errors would be shown in Figure 5. As can be seen from the figure, the attitude alignment measurement accuracy of CF method is better than that of single-frame method. When the extraction error of image points is less than 0.1pixels, the attitude alignment accuracy of CF method is better than 50 ". Compared with single-frame method, the alignment accuracy of horizontal attitude angle and Z-axis Euler Angle are increased by 5 times and 3 times, respectively. When the error of image point extraction is too large, the accuracy of attitude alignment will be poor and unstable, especially in the single-frame method.



CF method

single-frame method

Figure 5 The relationship between the errors in attitude alignment and errors in the extraction of image points

4.3 Influence of errors in mounting angle on the precision of attitude alignment

Errors in the extraction of image points during emulation are set as Gaussian white noise with the average value of 0 and standard deviation of 0.05 pixels, while the changing scope of errors in the mounting angle is 0~0.05°. When *k*=35, the results would be shown in Figure 6. From this figure, it can be seen that as the errors in mounting enlarge, the errors in attitude alignment would gradually become larger, and the relationship between these two types of errors is of a positive nature; when the errors of mounting angle are smaller than 0.008°, the precision of attitude alignment could be mastertained within 60".

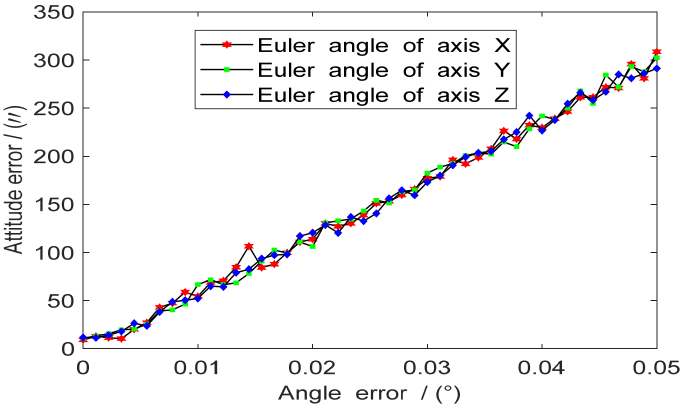


Figure 6 The relationship between errors in attitude alignment and mounting angle errors

According to the aforesaid analysis, set the errors of mounting angle as Gaussian amplifier errors with the average value of 0 and the standard deviation of 15", set the errors in the extraction of image points as Gaussian white noise with the average value of 0 and the standard deviation of 0.05 pixels, and set the number of correlated frames as 35, the emulation results of attitude alignment of the composite of the camera and local reference under different rotation amplitudes could be figured out, which are shown in Table 3. From this Table, it can be seen that, within the rotation scope of -1° to 1°, the method described in this study could realize attitude transfer of high precision, while the alignment precision of attitude angles of three directions is better than 45".

Table.3 Root mean square error of attitude alignment under different rotation ranges

Rotational range	RMS_X / "	RMS_Y / "	RMS_Z / "
±0.1°	44.98	37.55	31.38
±0.2°	31.05	40.05	44.90
±0.3°	43.93	38.42	44.82
±0.4°	42.51	31.64	32.74
±0.5°	32.31	34.52	43.32
±0.6°	33.81	32.26	33.31
±0.7°	41.39	43.49	37.68
±0.8°	39.11	39.85	43.66
±0.9°	34.53	33.44	40.70
±1.0°	32.66	33.26	39.12

5. CONCLUSION

In order to satisfy the demand for high-precision alignment of the initial attitude of moving base under the aligning environment, this study suggests an initial attitude reference transfer method for moving base based on correlated frames. This method utilizes the advantages of cameras, such as high precision, non-contact features, and rapid response, to estimate the three-dimensional attitude between the camera and plane collaboration target, realizes rapid alignment of attitude angles between the master reference and local reference, and solves the problem that the alignment accuracy of currently-available monocular visual measurement could easily be interrupted by random noises. The results of value-emulation experiments indicate the correctness and effectiveness of the attitude reference transfer method mentioned in this study. The alignment accuracies of Euler angles in three directions are all better than 45", while the alignment time is approximately 100s; compared with single-frame method, the alignment accuracies of horizontal attitude angle and Euler angle (Z Axis) are respectively five and three times higher. This show that this method brings about robustness against noises to some extent, which would provide a theoretical basis and support to the verification of semi-real object (emulation) and verification of actual equipment in subsequent experiments.

REFERENCES

[1] Y. Xu," Defense capability of aircraft carrier formation," Weapons knowledge, 241-248 (2010).

[2] W. An, J. Xu, H. He and P. Jiang, "A Method of Deflection of the Vertical Measurement Based on Attitude Difference Compensation," in IEEE Sensors Journal, 13125-13136(2021).

[3] Y. Wang, J. Xu and B. Yang, "A New Polar Rapid Transfer Alignment Method Based on Grid Frame for Shipborne SINS," in IEEE Sensors Journal, 16150-16163(2022).

[4] W. Liu, X. Yang and J. Zhang, "A Robust Target Detection Algorithm Using MEMS Inertial Sensors for Shipboard Video System," 2020 27th Saint Petersburg International Conference on Integrated Navigation Systems (ICINS), 1-5(2020).

[5] Q. Guo, Y. Quan, and Z. Zhu, "Workpiece Posture Measurement and Intelligent Robot Grasping Based on Monocular Vision," in Eighth International Conference on Measuring Technology & Mechatronics Automation, 919-922(2016).

[6] Y. Huang, M. Zhu, Z. Zheng and K. H. Low, "Linear Velocity-Free Visual Servoing Control for Unmanned Helicopter Landing on a Ship With Visibility Constraint," in IEEE Transactions on Systems, Man, and Cybernetics: Systems, 2979-2993(2022).

[7] G. Zhang, M. Kontitsis, N. Filipe, P. Tsiotras, and P. A. Vela, "Cooperative Relative Navigation for Space Rendezvous and Proximity Operations using Controlled Active Vision, " Journal of Field Robotics, 205-228(2016).

[8] M. A. Fischler, R. C. Bolles, "Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography, " Communications of the ACM, 381–395(1981).

[9] J. Wang, W. Huang, J. Zhang, Y. Zhang, G. Li, and M. Zhu, "Motion Parameters of Space Validator Relay Videometrics Method," Guangxue Xuebao/Acta Optica Sinica(2021).

[10] X.L. Tan, "Research on fast transfer method of shipboard attitude reference based on camera measurement, " National University of Defense Technology(2016).

[11] P.Y. Qu and W. Hou, " Attitude Accuracy Analysis of PnP Problem Based on Error Propagation Theory," Optical Precision Engineering, 479-487(2019).

[12] L. Wen, Z. Yingjun and Y. Xuefeng, "A Feature-based Method for Shipboard Video Stabilization," 2019 IEEE 2nd International Conference on Electronic Information and Communication Technology (ICEICT), 315-322(2019).

[13] J. H. White and R. W. Beard, "An iterative pose estimation algorithm based on epipolar geometry with application to multi-target tracking," in IEEE/CAA Journal of Automatica Sinica, 942-953(2020).

[14] X. Ma, S. Qin, X. Wang, W. Wei, J. Zheng, and P. Yao, "Hull Structure Monitoring Using Inertial Measurement Units," IEEE Sensors Journal, 2676-2681(2017).

[15] L. Ma, D. Zhan, G. Jiang, S. Fu, and S. Qin, "Attitude-correlated frames approach for a star sensor to improve attitude accuracy under highly dynamic conditions, " Applied Optics, 7559-7566(2015).

Ship damage submarine cable accident investigation —Trajectory analysis based on computer chart operation of VTS monitoring records

Aimin Wei · [1], Hui Wang [1]*, Xin He [1], and Donglou Li [1]

[1]Maritime School of Hainan Vocational University of Science and Technology, Haikou, China

[*] Corresponding author: juliawanghuia@163.com

Abstract: At 2:37:34 seconds on April 14, 2017, a large area of the Qing Zone suffered a power outage. During the troubleshooting of the power Supply Bureau, it was found that there was a suspicious self-unloading vessel, "HXX 338," anchored in the anchor forbidden area of the submarine cable and weighed anchor. Hainan Power Grid sued the suspicious ship and won the lawsuit, but the ship owner insisted on the ship not damage the submarine cable. The power grid entrusts the author as the maritime expert to make a professional judgment concerning the suspicion of this accident. The author makes use of navigation common sense: after anchoring, the position of the ship must be within the radius of the arc with the anchor position as the center of the circle, the length of the chain thrown out plus the sum of the distance between bridge and anchor chain hole; put in another way, if the ship's position is known after anchoring, the anchor position must be in a circle with the ship's position as the center of the circle, the length of the chain thrown out plus the distance between bridge and the bow's chain hole and as the radius. The trajectory of the ship was analyzed, and the suspicious ship was identified as the ship resulting in the accident. Investigate and analyze the facts of the ship in question moving in the prohibited anchor zone and hooked up submarine cables, find out the causes and distinguish the responsibilities. This article has explored an effective analysis for maritime investigation.

Key words: anchoring prohibited zone; submarine cable; VTS ship track monitoring diagram; Channel dredging; high probability criteria

1. Introduction

On April 14, 2017, a large area of Qinglan Zone suffered a power outage. During the troubleshooting of the power Supply Bureau, it was found that there was a suspicious self-unloading vessel "HXX 338" anchored in the anchor forbidden area of submarine cable and weighed anchor. It was suspected that hanging up the submarine cable caused the power failure. Hainan Power Grid sued the suspicious ship and won the lawsuit, while the suspected ship owner believed that (1) During the anchoring period, the anchor position of "HXX 338" was always outside the anchoring forbidden area, and there was no anchor dredging; (2) During the tidal surge, the hull of the ship did pass in and out of the prohibited anchor zone without changing its anchor position, but the water depth was greater than the draft of the ship, and the ship was not grounded, so the bottom of the ship could not have touched the cable under the seabed; (3) Even if the "HXX 338" is dredged anchor and the anchor in the water enters the prohibited anchor zone, it is impossible to damage the submarine cable 1.5 meters deep under the seabed because the anchor claw is only about 0.6 meters deep in the earth. The shipowner appealed.

Power grid entrusts the author as maritime expert, make professional judgment with respect to the suspicious point of this incident. The author makes use of navigation common sense: after anchoring, the position of the ship must be within the radius of the arc with anchor- position as the center of the circle, the length of the chain thrown out plus the sum of the distance between bridge and anchor chain hole. In other words, if the ship's position is known after anchoring, the anchor position must be in a circle with the ship's position as the center of the circle, the length of the chain thrown out plus the distance between bridge and the bow's chain hole and as the radius. The trajectory of the ship was analyzed, and the suspicious ship was identified as the ship causing the accident. Investigate and analyze the facts of the ship in question moved in the prohibited anchor zone and hooked up submarine cables, find out the causes and distinguish the responsibilities.

International Conference on Computer Graphics, Artificial Intelligence, and Data Processing (ICCAID 2022)
edited by Ruishi Liang, Jing Wang, Proc. of SPIE Vol. 12604, 1260415

1.1 Particular of MV.HXX 33

Year built: 2016 Classification: ZC

Overall length: 104.8.00 m Length of ship: 82.00 m

Overall beam: 18.00 m Depth molded: 6.00 m

Design draft: 4.55 m Gross tonnage: 2936 tons

Deadweight: 4356 tons Speed: 12 knots

Main engines: 70013,500 BHP Propellers: Twin screw controllable

Rudder: double rudder Anchors: 2x spade-anchors; weigh 1.929 kg;

Anchor rod: 1.5 meters, Length of anchor claw: 0.9 meters;

Chain: Port and starboard 8 shackles with 25 meters each shackle

1.2 The accident

At 2:37 and 34 seconds on April 14, 2017, a large area of Qingan area suffered a power outage. During the troubleshooting, the power Supply Bureau found that the sand ship “HXX338” dropped anchor and weighed anchor in the anchor prohibited area where the submarine cable of Qinglan Port was located. The power supply Bureau suspected that the ship anchor hooked the 35 kV cable, which caused the submarine cable to break, thus causing the power outage accident.

Hainan Power Grid Company sued the owner of the ship at Haikou Maritime Court. The court held that tiff if a ship shall break the submarine cable, three conditions should be met at the same time: one is to anchor in the prohibited area; second, the anchor should hook the cable; third, ship movement. Haikou Court confirmed that the “HXX338” hooked submarine cable caused losses, and ordered the owner of “HXX338” to compensate for the maintenance cost of Hainan power grid power facilities of RMB 2,027,700 yuan.

The ship-owner of “HXX338” refuses to accept the judgment of the maritime court of the first instance and appeals to Hainan High People’s Court. The owner’s expert opinion is that (1) the position recorded in the logbook of HXX338 at 1940 on 13 April 2017 is “19 ˚33’30″ N; 110 ˚ 49’ 42 E, that is, the anchorage has been outside the prohibited anchoring area, there is no anchor dredging; (2) During the tidal surge, “HXX338” did pass in and out of the anchoring prohibited zone under the condition that the anchor position did not change, but the ship did not run aground, and the bottom of the ship could not scrape the cable under the seabed; (3) even if “HXX338” has gone through anchor dredging and entered the prohibited anchor zone, due to the length of the anchor claw is 90 cm and the angle of the anchor claw is 42 degrees, it is presumed that the depth of the anchor claw can only reach 0.6 meters, and it is impossible to hook the cable 2.1 meters below the seabed. Considering that the evidence for the damage caused by “HXX338” hooking submarine cables was insufficient, the ship owner appealed to the Hainan High People’s Court, requesting the court to cancel the first-instance judgment.

Hainan Power Grid Company hereby entrusts the author to conduct the maritime investigation. In this case, issue expert opinions and attend the trial to express opinions.

2. Maritime investigation

The crux of the case is one. Whether the anchor position of the vessel is within the submarine cable anchorage zone; the second is whether the anchor dropped by the ship can hook the cable. To better clarify the case, the author referred to the following documents: VTS monitoring records from Qianlan Maritime Bureau of the first instance; expert opinion on the appeal provided by the appellant (defendant of the first instance); supplementary comments on the appeal provided by the appellant (defendant of the first instance); Port Guide to Qianlan Port; Tide tables at Qianlan Harbor; record of first inquiry; Photocopy of ship’s log book; Chart of the ship.

2.1 Determine whether HXX338's anchorage is in the prohibited anchorage area as recorded in the logbook?

(1) Is the position recorded in the log of HXX338 the ship's firm anchorage at 1940?

The "HXX338" ship log is extremely simple, and it only records the end time and location of Anchorage. The owner asset the position recorded in the log book was "19 ˚33 30″ N; 110 ˚ 49 48 E" is the anchor position.

From the general sense of navigation, after anchoring, the position of the ship must be within the arc with the anchor position as the center of the circle, the sum of the distance between the length of the chain thrown and the bridge to the bow anchor chain hole as the radius (for HXX 338:82-20+25×3 =137 meters). Computer mapping was carried out based on the anchoring position of the appealing party and the screenshot of the ship's track chart recorded in the VTS monitoring records of the Maritime Safety Administration. It was apparent that all the positions of the ship from 1940 to 0205 were outside the circle, with the anchoring position as the center and the length of the cast chain plus the length of the ship as the radius. So 33˚60″N; 110˚49-48 E" is not the stable anchor position after 1940. See **Figure 1.**

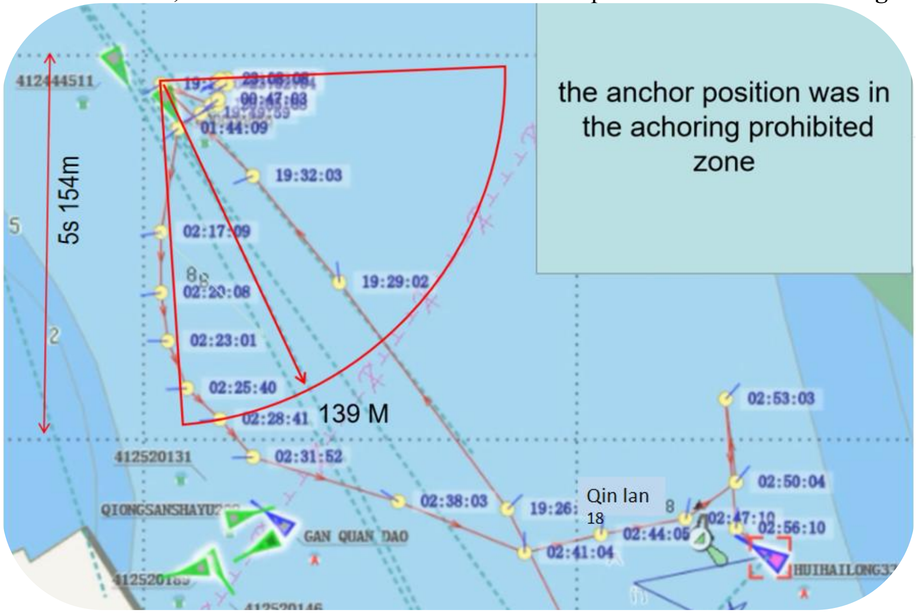


Figure 1. VTS monitoring record chart against log anchorage

(2) Since the position data recorded in the log book of "HXX338" was confirmed not to be the stable anchorage position after 1940, what was the true anchor position of "HXX338" after 1940?

According to the first trial defense statement, we finally confirmed that HXX338 drew out the port anchor and three shackles into the water. According to the VTS track record of Qinglan Maritime Bureau, the position of the ship in 1935 was 19˚33' 34.5″N; 110˚49-36.7 E. This position is also the farthest position from the anchor position due to the influence of wind currents after the ship is anchored. The VTS monitoring signal of Qinglan Port comes from the AIS signal of Mulan Tau Lighthouse (the highest lighthouse in Asia). The ship position signal collected by VTS comes from the ship GPS signal sent by ship AIS, and the GPS antenna of the HXX338 bridge is located 20 meters forward from the ship's stern.

Regarding navigation common sense, the anchor position must be in a circle with the anchor position as the center of the circle, the length of the chain thrown, and the distance between the bridge and the chain hole as the radius. HXX338 anchor chain throw length is three shackles, the ship length of all is 82 meters, and the bridge is 20 meters forward aft. Therefore, the maximum distance between the ship position and anchor position is 75 + (82-20) = 137 m. The anchorage position must be within the 137-meter radius of the 1935 position as the center of the circle and close to the track line. According to the chart work on the screenshot of the maritime Administration's monitoring record, we found that the arc was in the anchoring prohibited area. Therefore, it can be confirmed that the anchor position of ship "HXX338" after 1940 was anchored is within the anchoring prohibited zone and is at least several meters deep into the anchoring prohibited zone, as shown in attached **Figure 2.**

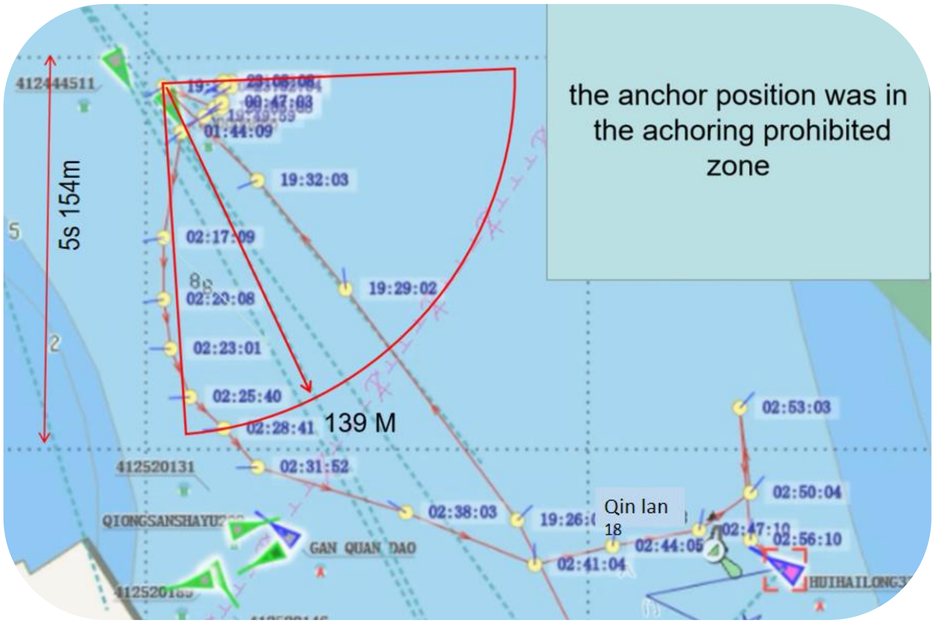


Figure 2. The anchor position was within the anchoring prohibited zone

(3) HXX338 logbook records that the ship dropped port anchor out three shackles into the water, but why did it not dredge anchor in the strong wind and rapid current?

The ship's logbook records a northeasterly wind of 6/7, gale relative to the ship's tonnage, at the time of anchoring. Find out the tidal current condition of Qinglan port: the high tide near the wharf flows to the northwest at 1.3 knots, and the low tide flows to the south at 2.5 knots. When the current exceeds 1.2 knots, it is generally considered to be "rapids" in ship handling. The ship is equipped with a 1920 kg spade anchor. The length of the chain released is only three shackles. The actual length of the chain lying on the seabed is about 2.5 shackles, and the grip provided by the ship's anchor and chain is very limited.

According to the data provided in the ship maneuvering textbook issued by MSA, in a water depth less than 30 meters, under the condition of wind force less than level 7, the recommended length of single anchor mooring mode should be 5-6 shackles according to statistics to ensure safety. "HXX338" uses a single anchor with three anchor chains to enter the water in the case of strong wind and rapid currents, which only reaches about half of the recommended length of the anchor throwing chain. In this case, an anchor dredging accident is very easy to happen. However, from the VTS track monitoring record chart of the Maritime Administration, it can be seen that after 1940, the ship did not incur "anchor dredging". It is presumed that the ship "HXX338" was most likely anchored to solid objects underwater.

(4) Where is HXX338's stable anchor position after 1940?

The ship's logbook record: HXX338 started to heave anchors from 0205 and 0250 anchors out of the water, costing a total of 45 minutes. According to the requirements of the *Code for the Construction and Classification of Steel Seagoing Ships* issued by the China Classification Society, the anchor equipment of the ship should be able to pull up the anchor chain at a speed of not less than 9 meters per minute in a water depth of 82.5 meters. It should be within 10 minutes for the ship to heave up three shackles of anchor chains under normal circumstances. It took several times longer than normal for the ship to heave up the three shackles of anchor chains, indicating that there was an abnormal event happened on the ship, but the ship's logbook did not record anything about it. But in the record of the maritime traffic accident investigation provided by the Maritime Bureau, the captain stated, "when 1 or 2 shackles of the anchor chains were left, anchor chain could not get down into the chain chamber…"

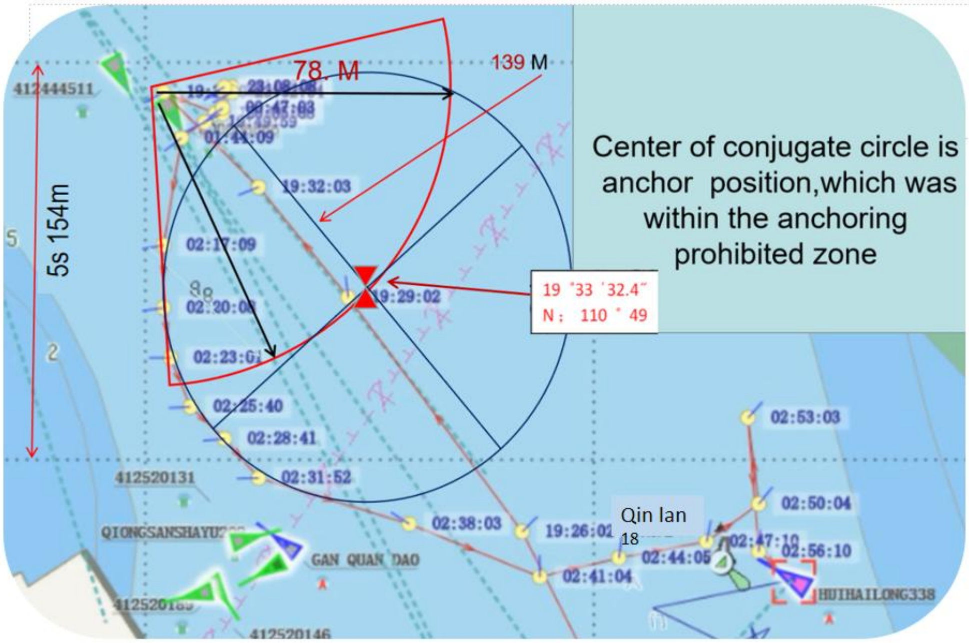


Figure 3. The Center of the conjugate circle is the anchor position

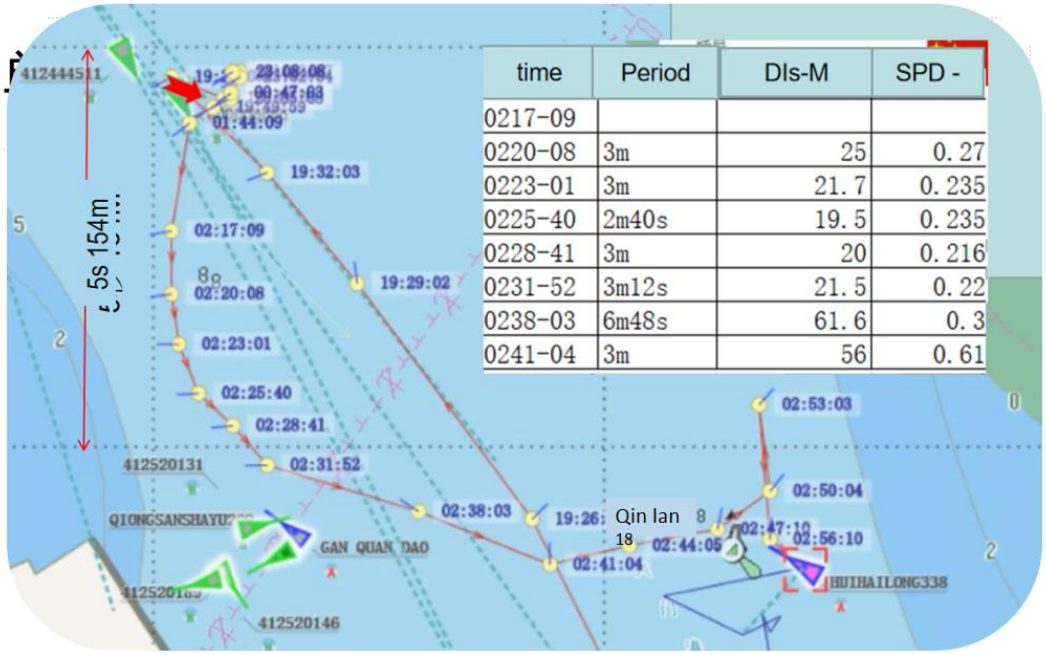
“The anchor chain could not get down into the chain chamber” will inevitably lead to the top of the anchor heaving. It can be seen from the track chart of “HXX338” provided by VTS of Maritime Administration: from 0217 to 0229, the track of HXX338 forms a very regular arc, and the ship’s moving speed is stable, as shown in **Figure 3.** The condition for the occurrence of the above phenomenon must be “stopped heaving anchor,” and the ship moves in a circle around the anchor position.

The conjugate circle center corresponding to the regular arc in this section is the real “anchor position” after 1940. with computer graphics, we can measure its position: “19°33 32.4″N; 110°49 39 E”. This was the anchor position since 1940 (see **Figure 3**).

From the above analysis, it can be determined that the ship “HXX338” was anchored in the prohibited anchoring area, most likely hooked to submarine cable; it was difficult to heave up anchor when lifting, the main engine was used, and the submarine cable was damaged by forced pulling, which led to tripping and power failure.

(5) After analysis, there is an obvious causal relationship between the trajectory tendency of HXX338 and the damage to the submarine cable.

We trace the ship’s entire motion track to a reasonable and coherent explanation, which can be shown as follow: after the ship stopped mooring around 2017, the ship tended to move along the tangent line of the arc under the action of strong wind and centrifugal force. At this time, the ship’s movement exerted great tension on the anchor chain. However, because the anchor caught a solid object underwater, which failed to be dragged, so the anchor chain was strained. Under the action of anchor chain elasticity and gravity, the ship was forced to drag to the anchor chain direction. Therefore, at 02:25:40, the ship’s position slightly deviated from the tangent direction, and at 02:28:41, the ship’s position retrenched into the arc. After that, from 02:28:41 to 02:31:52, the ship’s track appeared to pull and rebound again. After the ship track deviates from the arc, the ship starts to accelerate along the tangent direction of the arc. After the track bounces back slightly, it continues to accelerate along the tangent direction. At this time, the “underwater solid object” caught by the anchor begins to be dragged under the combined force of the strong airflow pressure and anchor heave force. It can be seen from the ship’s moving track that the ship’s moving speed is obviously faster and faster, and the object hooked is gradually dragged away from the bottom until 02:38:03, when the ship’s moving speed is greatly accelerated again, and the blocking force caused by the hooked has been eliminated. Speed changes are shown in the attached table (Figure 4).



time	Period	DIs-M	SPD -
0217-09			
0220-08	3m	25	0. 27
0223-01	3m	21. 7	0. 235
0225-40	2m40s	19. 5	0. 235
0228-41	3m	20	0. 216
0231-52	3m12s	21. 5	0. 22
0238-03	6m48s	61. 6	0. 3
0241-04	3m	56	0. 61

Figure. 4. Speed varied during anchor heaving

At 02:37:34, the power supply bureau of the 35-kV Qingjia line tripped power outage, resulting in a large area of Qinglan power outage. It was found to be caused by damage to submarine cables in the prohibited-anchoring zone. A study of the MSA's ship tack monitoring records shows that, although other vessels passed by during this period, no other vessel was found to have anchored, weighed anchor, or engaged in other underwater towing activities of a similar nature in the waters except HXX338. Therefore, it can be concluded that since the Ship HXX 338 was anchored in the no-anchor zone in violation of regulations, it was most likely hooked with cables. When HXX338 was lifting the anchor, it was difficult for the ship to heave up the anchor, and the main engine was used to forcibly pull the submarine cable, which was damaged, resulting in a large area of power failure.

2.2 Can MV. HXX338 cast anchor hook submarine cable?

Whether the anchor flukes can catch the cable depends on two key data points: the depth of the submarine cable and the depth of the flukes into the earth.

(1) Documents of Hainan Electric Power Company record that there are three 35-kV submarine river crossing submarine cables from Qinglan Port to Dongjiao, each of which is 870 meters long and 12 centimeters in diameter, and both ends are fixed on the terminal frame on the shore. The actual depth of the submarine cable construction is 2.1 meters. Since the 1990 s, Qinglan Port has made several channel renovation projects. The 5000-ton channel expansion project of Qinglan Port was started in July 2006 and completed in November 2009. The channel was dredged from the historical of 6-7 meters to 7.6 meters, and the water depth was maintained at 7.6 meters. Later, it was used as a Wenchang Satellite launch base, and logistics support port of Sansha City for further maintenance, and the water depth of the chart reached more than 8 meters. After several dredging, the channel depth deepens more than 1 meter, which is bound to lead to the shallow buried of underwater cables.

(2) The biting depth of the anchor is closely related to the bottom sediment. Under the condition of soft mud bottoms in rivers and lakes, it is not uncommon for the entire anchor body to sink several meters into the bottom. The bottom nature in the forbidden anchorage area of Qinglan Port is "mud." "The length of the flukes is 0.9 meters, and the calculated depth of the flukes is only 0.6 meters into the earth, so it is impossible to hook the cable below 1.5 meters…" It's just a dull geometric calculation, which is erroneous.

The above two points are to indicate that the anchor of “HXX 338” may be hooked to the submarine cable when anchored in the prohibited area.

At that very time, Hainan Electric Power Design and Research Institute provided a Qinglan-Dong Jiao 35 kV submarine cable construction completion drawing. The engineering coordinates where the submarine cable is located in the completed drawing were converted into geographical coordinates and marked on the screenshot of the VTS ship track monitoring record chart provided by MSA. It was observed that the anchor of HXX338 happened to cross the position of the submarine cable in the process of anchor throwing and lifting. The VTS monitoring record chart of Qinglan Maritime Safety Bureau shows that no other ship, except HXX338, dropped anchor, pulled anchor, or pulled anchor in the prohibited area at the time of the incident. Combined with the large area power failure when the anchor was off the bottom, it clearly shows that there is a causal relationship between the anchor throwing and lifting operation of the wheel and the cable being pulled off (see Figure 5).

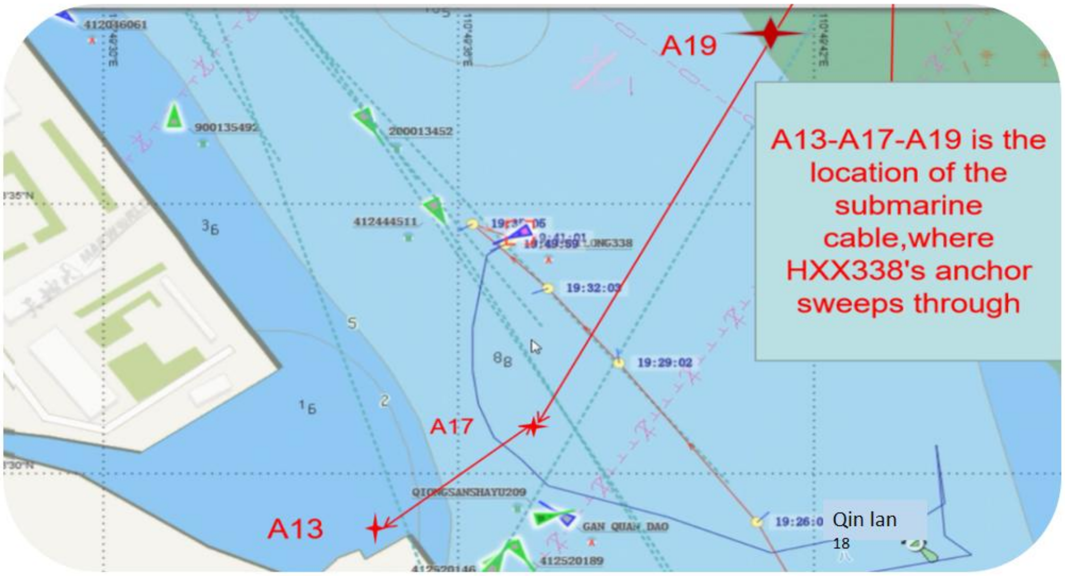


Figure 5. The location of the submarine cable

3. Investigation conclusion of this case analysis

During the period from 1940 to 0205 on April 14, 2017, the anchor position of “HXX338” shall be in the anchoring forbidden zone. In the process of dropping anchor, the anchor passes through the position of submarine cable; the ship hooked submarine cables in strong currents, making it difficult to heave the anchor. As a result, ships started their main engines, pulled anchors, and dragged submarine cables, which broke the Qinglan river crossing cable and caused a massive blackout. The facts of the first instance are clear, and the conclusion is correct.

Due to the 35-kV Qinglan line tripping power outage, a large area of Qinglan power outage was found to be caused by damage to submarine cables in the prohibited-anchoring zone. The study of the MSA’s ship tack monitoring records shows that although other ships passed by during this period, no other vessel was found to have anchored, weighed anchor, or engaged in other underwater towing activities of a similar nature in the waters except HXX338. Hainan High People’s Court, based on the standard of the high probability of civil litigation, accepted this analysis. The final judgment was delivered on 20 May 2019; the appeal was dismissed, and the judge upheld it.

4.Conclusion

This case adopts computer chart operation, combined with a VTS track monitoring chart and cable completion chart, to investigate and analyze the fact that the ship involved in the accident was moving in the prohibited anchor area and hooked up the submarine cable, find out the reason, distinguish the responsibility, and explore an effective analysis for maritime investigation.

Acknowledgment

This work was supported by the Application Curriculum Reform Program of Hainan Vocational University of Science and Technology, HKKG2018-4. & Hainan Province Philosophy and Social Science Planning Project: A Study on the International Ship Registration System under the Background of Free Trade Port (Grant No. HNSK(ZX)21-90).

References

[1] Hong biguang . Ships manoeuvring Dalian: Dalian Maritime University Press, 2008

[2] Ministry of Land and Resources Submarine Cable Pipeline Protection Regulations, Ministry of Land and Resources Order No. 24 2004

[3] Interpretation of the Supreme Court on the Application of the Civil Procedure Law of the People's Republic of China, 2015

[4] Wu Zhaolin. Maritime investigation and analysis. Dalian: Dalian Maritime University Press, 2016

[5] electronic chart https://www.chart.gov.cn/ Chart/ Customer/ Digital Map Products. Aspx

[6] Hainan Province Submarine Cable Protection Joint Conference System Notice; Qiongfum Office (2011) No. 129

FEM-SPH Adaptive Coupling Calculation Model for Small-caliber Bullet Penetrating Ceramic Composite Target

Wei Xia[1], Xin Jia[1*], Zheng xiang Huang[1], Yu Wang[1], Qing le Liu[1], Tao Zhang[1]
[1] College of Mechanical Engineering, Nanjing University of Science and Technology, Nanjing 210094, Jiangsu, China
2 63850 troop, Baicheng, 137001, Jilin, China
[*] jiaxin@mail.njust.edu.cn

Abstract

In order to accurately simulate the dynamic response process of ceramic composite target under the penetration of 7.62mm steel core projectile in ballistic test., this paper selects FEM (Finite Element Method), SPH (Smoothed Particle Hydrodynamics), SPH-FEM fixed coupling algorithm, SPH-FEM adaptive coupling algorithm as the simulation scheme, and uses LS-DYNA finite element software to analyze the penetration of small-caliber bullets. The research shows that: by analyzing the dynamic response process of ceramic failure and the residual velocity of the projectile, the FEM-SPH adaptive coupling algorithm can simulate the failure modes of ceramics at each stage, and quantify the simulation accuracy with the residual velocity of the projectile. The error of the adaptive coupling algorithm is 1.17%. The FEM-SPH adaptive coupling simulation model is reasonably applied to simulate the impact process of the bullet penetrating the ceramic target.

Key words: small caliber bullet; ceramic composite armor; FEM-SPH adaptive coupling

1. INTRODUCTION

At present, the simulation of the penetration of small-caliber bullets into ceramics usually adopts FEM modeling, SPH [1] modeling· FEM-SPH fixed coupling algorithm [2, 3], etc.; for FEM simulation , the processing algorithm of the failed unit is to delete the failed unit directly, which fails to dynamically simulate the splashing phenomenon of ceramic fragments in the ballistic test of the ceramic target plate and does not consider the impact of the splashed ceramic fragments on the bullet, so the simulation results are quite different from the actual ones. big. For the simulation modeling of SPH (both projectile ceramics are SPH), when the contact between SPHs is carried out, the penalty function related to SPH erosion is difficult to define and the contact interface is not clear, which leads to the phenomenon of unstable stress propagation. The dynamic modeling of the target system has gradually developed from FEM and SPH modeling to FEM-SPH coupled simulation modeling. The FEM-SPH coupling algorithm can be divided into two categories [3,4]: the first is the fixed coupling algorithm, which defines the initial area of the FEM and SPH in the target plate, and the position remains unchanged during the penetration response process, in which the coupling method is point Coupling one by one; the second is an adaptive coupling algorithm. At the initial moment, the simulation model is all modeled by finite element. In the subsequent simulation process, the large deformation element of the material or the damaged and failed element is converted into SPH particles, and through the SPH particles Continue to penetrate.

In this paper, LS-DYNA finite element difference software is used to simulate the penetration of 7.62mm steel core projectile into ceramic composite target. By changing the traditional simulation method of bullets and target plates, the FEM-SPH adaptive coupling simulation method can obtain the dynamic damage response process of ceramic composite plates and the residual velocity of the projectile, which significantly improves the simulation accuracy of small caliber bullets penetrating ceramic composite targets. The results of this paper can provide some reference for the structural design of small caliber bullets penetrating composite targets.

2. SIMULATION MODEL ESTABLISHMENT

The main structure of the ceramic composite armor of the finite element model consists of three parts, as shown in Figure 1. The penetration of the ceramic composite target plate with a steel core projectile with a diameter of 7.62 mm was used, and the average mesh size was 0.5 mm. In order to reduce the amount of computation, a quarter model is established based on the symmetry of the overall model.

International Conference on Computer Graphics, Artificial Intelligence, and Data Processing (ICCAID 2022)
edited by Ruishi Liang, Jing Wang, Proc. of SPIE Vol. 12604, 1260416
© 2023 SPIE · 0277-786X · doi: 10.1117/12.2674616

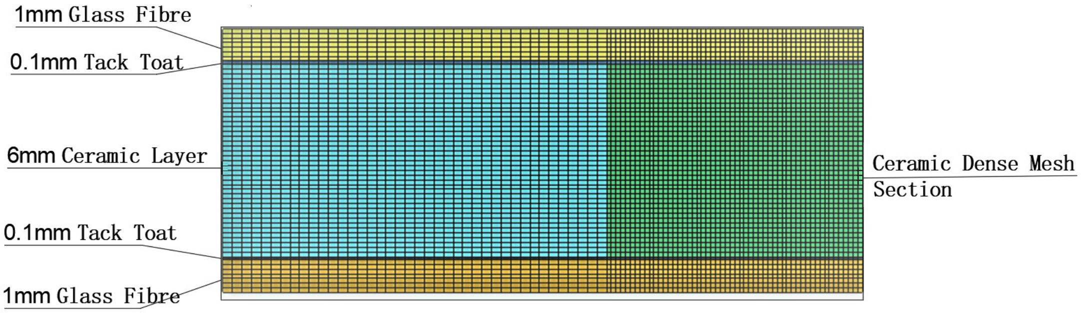


Figure 1. Mesh division diagram of ceramic composite target plate

Symmetry constraints are imposed on the symmetry plane of the finite element model, and non-reflection boundaries are imposed on the composite target boundary [5]. The minimum mesh size of the composite target plate is 0.3mm, and the densified area is 20mm. In order to ensure the calculation accuracy and improve the calculation efficiency, the target plate is divided into denser meshes at the distance from the penetration center, and the influence of stress and strain in the farther area is relatively small. The grid division is gradually sparse, and the sparse grid size is 1.5mm, as shown in Figure 1.

3. SIMULATION MODEL OF CERAMIC COMPOSITE ARMOR-BULLET SYSTEM

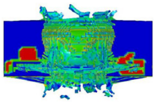
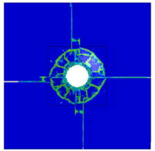
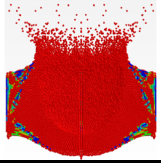
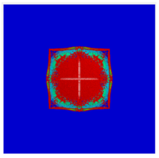
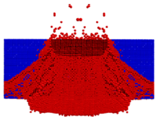
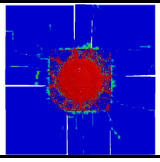
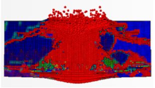
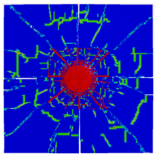
In this paper, various methods of simulation modeling are carried out for 7.62mm steel core projectile penetrating ceramic composite armor. The specific simulation scheme See Table 1.

Table 1 Simulation scheme of ceramic composite armor

Modeling Method	Projectile	Ceramics	Glass Fiber
1	FEM	FEM	FEM
2	FEM	SPH-FEM fixed coupling	FEM
3	SPH	xed coupling	FEM
4	FEM	SPH-FEM adaptive coupling	FEM

4. SIMULATION RESULTS AND ANALYSIS

Table 2. Damage nephogram of ceramic plate corresponding to different simulation schemes

Simulation Scheme	Sectional View	Back View
1		
2		
3		
4		

As shown in Table 2, in simulation scheme 1, both the projectile and the ceramic plate are modeled by FEM. Radial cracks and annular cracks appear on the back plate of the target plate, but the damage area of the overall target plate is small. This is because the FEM modeling adopts the deletion algorithm for the ceramic elements that fail in the penetration process, and fails to simulate the reverse impact of ceramic fragments on the projectile in the actual impact process. And the ceramic profile does not dynamically reflect the critical ceramic cone failure in ceramics.

In simulation scheme 4, the projectile body is modeled by FEM, and the ceramic plate is adaptively coupled with SPH and FEM, which can simulate the failure stage of the ceramic. That is, when the projectile penetrates the ceramic plate, stress waves are first generated in the penetration direction, and at the same time, tensile waves are generated on the back side. The stress will exceed the tensile strength of the tensile ceramic itself, resulting in tensile cracks on the surface, which will continue to expand and intersect with the front cracks (circumferential cracks and conical cracks) and continue to expand. The ceramic plate is further damaged, and the final failure form is ceramic. Cone [6]. This is more consistent with the simulation case.

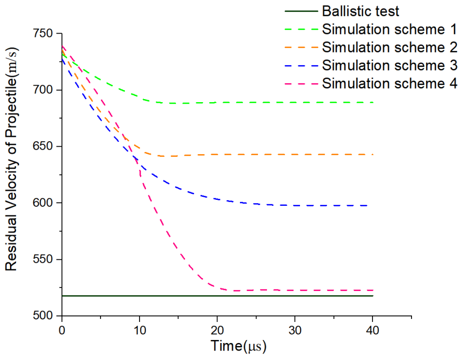


Fig. 2 Time history diagram of projectile velocity

Table 3 Projectile residual velocity and simulation error

Working Condition	Ballistic Test	Simulation Scheme 1	Simulation Scheme 2	Simulation Scheme 3	Simulation Scheme 4
Residual Velocity (m/s)	512	698.5	649	595	518
Error（%）	0	36.43	26.76	16.21	1.17

With reference to the residual velocity of the ballistic test, the residual velocity of the projectile is used as the quantitative index of the simulation results. It can be seen from Fig. 2 and Table 3 that the residual velocity of the projectile in the simulation scheme 1 is the largest, which is 698.5m/s, and the error is 36.43%. It can be seen that the simulation scheme 1 is inaccurate for the bullet penetrating the ceramic plate. Compared with the simulation scheme 1, the simulation scheme 2 can simulate the dynamic penetration process of the projectile and the ceramic plate. it can be seen that the macroscopic crack on the ceramic surface is not simulated. Compared with the first two simulation schemes, the simulation scheme 3 has a significant reduction in error. From Table 3, it can be seen that the propagation of damage is unstable, which eventually leads to inaccurate residual velocity of the projectileThe residual velocity of the projectile in simulation scheme 4 is the minimum, which is 518m/s, and the error is 1.17%, which meets the basic error tolerance

5. CONCLUSION

- In this paper, the modeling method of FEM-SPH coupling calculation model for 7.62 steel core projectile penetrating ceramic composite armor is studied. Based on the different simulation modes in the simulation process of projectile penetrating ceramic, the ceramic failure modes and residual velocity of the projectile in different simulation schemes are analyzed. The FEM-SPH adaptive coupling simulation method is applied to accurately simulate the impact response process of gun projectile ceramic composite target. The simulation accuracy of steel core projectile penetrating ceramic composite armor is significantly improved.
- The research results of this paper are of general significance and can be applied to the simulation analysis of ballistic impact of other steel cores, which is helpful to improve the accuracy and efficiency of ballistic impact simulation analysis.

REFERENCES

[1] Zhu J analysis on impact resistance and disaster reduction and prevention effect of debris flow dam based on sph-fem [D], 2021

[2] Yu Q, Wang T S. SPH analysis method for large amplitude sloshing of variable mass liquid in liquid filled spacecraft [J]. Acta Astronautica Sinica, 2021, 42 (1)

[3] Liu S, Zhang W G, FEM-SPH coupling calculation model of armor piercing incendiary projectile penetrating ceramic composite armor and glass composite armor, explosion and impact 2021,41(1)

[4] Hu D An, Han X, et al Research progress of smooth particle method and its coupling with finite element method [j] Chinese Journal of mechanics, 2013, 45 (5): 639– 652 DOI:10.6052/0459-1879-13-092

[5] Zaera R, Sanchez-Saez S, Perez-Castellanos JL, et al. Modelling of the adhesive layer in mixed ceramic/metal armours subjected to impact [J]. Composites: Part A, 2000, 31: 823–833

[6] R. ZAERA*t and V. SA¨NCHEZ-GA¨ LVEZAnalytical Modeling of noraml and oblioueballistic impact on ceramic/metal lightweight armours 2002

Adaptive analysis of rapid prediction of coal quality information based on image analysis

Ming Liu *, Hua Zhang
CHN Energy Xinjiang Chemical Co.,Ltd. , Urumqi 831404, Xinjiang, China
* Corresponding author: 793471497@qq.com

Abstract

Machine vision technology has shown great potential for development and application in the coal heat utilization and coal chemical production process It is important to carry out image analysis-based stability adaptation analysis of coal heat utilization and coal chemical equipment. To further expand its application, this work reviewed the characteristics of different vision sensor systems and image processing algorithms in the process of image acquisition and processing, conducted applicability analysis for rapid prediction of coal quality, presented the current problems and future development trends of image analysis technology in the coal heat utilization and coal chemical production process. It provided theoretical guidance and reference for promoting automation, intelligence, and mechanization in the coal industry.

Keywords: machine vision; image processing; coal quality rapid prediction

1. Introduction

Coal is the main energy source for industrial production in China, and coal thermal utilization and coal chemical industry play an important supporting role in national economic development[1]. At present, China's energy consumption structure is still dominated by coal, and it is difficult to change in the short term, and the resulting problems of coal resources development and utilization will also exist for a long time[2]. According to the strategy of comprehensive utilization of coal resources and environmental protection, coal beneficiation is the key. Coal beneficiation, as the source technology of the clean coal industry, can effectively remove the ash and sulfur from coal. Therefore, in order to realize the sustainable development of the coal industry and improve the comprehensive utilization of coal resources, clean coal technology must be vigorously developed[3-5].

The traditional gangue sorting process refers to the use of conveying devices to transport raw coal from the well to the well, after screening, and then individual gangue and coal larger than 100mm will be transported to the artificial gangue inspection workshop. The method is mainly based on the workers' own experience, through the gangue and coal appearance of brightness, color and shape of the difference to their identification, judgment and sorting. This coal sorting method is not only inefficient, but also labour-intensive. The gangue sorting rate is largely affected by human factors such as the quality of production personnel, physical quality, responsibility and management level, which ultimately cannot guarantee the quality of coal production. Therefore research and development of modern gangue automatic sorting system is imminent.

With the development of artificial intelligence technology, computer vision and image processing technology are gradually applied to the coal gangue sorting link. As the coal gangue sorting operation is mainly done by using the difference of glossiness and greyscale of coal and gangue, so the reflected light and wavelength of coal and gangue are different under the light irradiation, and accordingly the peak of greyscale distribution of both are also different. Based on this principle, some scholars have compared the standard ash distribution maps of gangue and coal with their actual ash distribution maps to distinguish between coal or gangue based on the distribution and size of the peaks. Research based on this approach has been widely reported, and the differences are mainly in the image processing algorithms. For example, Sun et al.[6] compared focusing functions in terms of accuracy, noise immunity, number and width of peaks, and analyzed the advantages and disadvantages of each focusing function. Hobson, DM et al.[7] introduced texture, a visual feature represented by the distribution of greyscale of pixels and their surrounding spatial neighborhoods, and used it to distinguish between coal and gangue; Debi Prasad Tripathy et al.[8] proposed three symbiotic matrix-based extension methods. methods and combined with a classification model of neural networks for gangue sorting. Although existing studies in the field of coal thermal utilization and coal chemical industry can effectively identify coal gangue[9], prediction of coal quality information such as particle size, density composition and ash content has not yet been

International Conference on Computer Graphics, Artificial Intelligence, and Data Processing (ICCAID 2022)
edited by Ruishi Liang, Jing Wang, Proc. of SPIE Vol. 12604, 1260417
© 2023 SPIE · 0277-786X · doi: 10.1117/12.2674986

achieved. Therefore, how to achieve rapid prediction of coal quality information is a key issue in the development of breakthroughs in the automation of coal production processes.

In this paper, we compared and analyzed the characteristics of different visual sensor systems and image processing algorithms, carried out the adaptation analysis of coal quality information rapid prediction methods such as intelligent coal transport detection, coal bin coal level detection and coal gangue sorting, and proposed the problems and development directions of coal quality information rapid prediction technology based on image analysis.

2. Image acquisition technology

Vision sensors are an important way of capturing image information by machine vision technology, which can be divided into two-dimensional (2D) image sensors and three-dimensional (3D) image sensors according to the number of dimensions of the captured image. For the selection of vision sensors, several factors such as the coal chemical production process, operation object and working environment should be taken into consideration to fully utilize the advantages of sensors while meeting the working requirements[10]. Based on the number and characteristics of vision sensors, the current mainstream vision technologies include monocular vision, binocular stereo vision, multi-vision, and panoramic vision.

3. Photocatalytic dye degradation

Image processing is a technique that uses a computer to analyze an image to achieve the desired result and is the core aspect of a machine vision system. Its main components include image digitization, image enhancement and restoration, image segmentation and image recognition[16]. Among them, image segmentation is the main problem in machine vision research. The so-called image segmentation refers to the division of an image into several disjoint regions based on features such as grayscale, color, spatial texture, geometry, etc., so that these features show consistency or similarity within the same region and distinct differences between different regions[17]. Existing studies have reached a consensus on the general laws of image segmentation. Threshold segmentation is an algorithm that determines a certain threshold value for image segmentation based on the distribution characteristics of the image grayscale values. Let the original grayscale image be f(x, y), select a grayscale value T as the threshold by some criterion, and compare the size of each pixel value with T; the pixel value greater than or equal to T is one class, change its pixel value to 1; the pixel value less than T is another class, change its pixel value to 0. The process of turning a grayscale image into a binary image g(x, y) is also called image binarization, as shown in Eq. (1) shows.

$$g(x,y)=\begin{cases}1 & f(x,y)\geq T\\ 0 & f(x,y)\leq T\end{cases} \tag{1}$$

Table 1. Characteristics and main problems of different vision technologies.

No.	Vision technology	Feature	Major problems
1	Monocular vision technology [11]	The structure is simple, the algorithm is mature and the computational effort is small.	The working conditions are more idealized and the practical application of the measurement is not good.
2	Binocular stereo vision technology [12]	The imaging effect is no longer affected by changes in lighting and can use the rich three-dimensional spatial information to identify the target object positioning more accurately easy to operate and can obtain a parallax map.	The actual working process requires precise knowledge of the spatial location relationship between the two cameras, and the 3D information of the scene environment requires two cameras to take two images of the same scene from different angles at the same time and perform complex matching to recover the 3D information of the visual scene more accurately.
3	Multi-eye vision technology [13]	It can solve the problem of matching polysemy in a binocular stereo-vision system and improve the matching accuracy.	The trinocular vision system must reasonably place the relative positions of the three cameras, and its structural configuration is more cumbersome than that of the binocular vision system, and the matching algorithm is more complex and less real-time.
4	Multispectral imaging technology [14-15]	Removing a large amount of useless information, a small amount of data, fast processing speed; practical, less affected by environmental conditions; conducive to the development of convenient, practical, and fast monitoring devices.	One spectral band corresponds to one lens and filter. If the accuracy of the spectrum is improved, the number of lenses will increase and the cost will be superimposed accordingly.

Various threshold segmentation algorithms are classified according to the different ways of threshold selection, such as the maximum inter-class variance method and the maximum entropy-based threshold segmentation method.

Maximum interclass variance method let the resolution of the image be $M \times N$ and the probability of occurrence of gray levels in the image be:

$$p_i = \frac{n_i}{M \times N}\text{，} i\text{=0，1，2，…，}L\text{-1} \tag{2}$$

where L is the total number of gray levels in the image and n_i is the number of gray level occurrences.

All pixels are divided into two classes according to a certain threshold T. Let low grayscale be the target region and high grayscale be the background region, and the probability of distribution in the pixel images of the two classes is:

$$p_o = \sum_{i=0}^{T} p_i \;\; p_B = \sum_{i=T+1}^{L-1} p_i \tag{3}$$

The mean value of the pixel values in both categories is:

$$\mu_o = \frac{1}{p_o}\sum_{i=0}^{T} i \times p_i \;\; \mu_B = \frac{1}{p_B}\sum_{i=T+1}^{L-1} i \times p_i \tag{4}$$

The mean value of the overall gray scale is:

$$\mu = p_o \times \mu_o \times p_B \times \mu_B \tag{5}$$

The variance between the two categories is:

$$\mu_o^2 = \frac{1}{p_o}\sum_{i=0}^{T} pi(i-\mu_o)^2 \;\; \sigma_B^2 = \frac{1}{p_B}\sum_{i=T+1}^{L-1} p_i(i-\mu_B)^2 \tag{6}$$

The total intraclass variance is:

$$\sigma_{in}^2 = p_o \cdot \sigma_o^2 + p_B \cdot \sigma_B^2 \tag{7}$$

The interclass variance between the two classes is:

$$\sigma_b^2 = p_o \times (\mu_o - \mu)^2 + p_B \cdot (\mu_B - \mu)^2 \tag{8}$$

The threshold *T* that minimizes the intra-class variance or maximizes the inter-class variance, or minimizes the ratio of intra-and inter-class variance, is the optimal threshold.

Entropy is a measure of uncertainty in information theory, a measure of the amount of information contained in the data, and when entropy takes the maximum value, it indicates that the amount of information obtained is the maximum.

To perform image thresholding, the image is divided into two classes, and entropy can be used as the criterion for classification: the maximum amount of information can be obtained from the image when the sum of the average entropy of the two classes is the maximum, and the threshold used for classification at this time is the optimal threshold.

For digital images, the entropy of two regions, target, and background, when taking the threshold value of *T* is:

$$H_o(T) = -\sum_{i=0}^{T} \frac{p_i}{p_o} \log \frac{p_i}{p_o} \quad H_B(T) = -\sum_{i=T+1}^{L-1} \frac{p_i}{p_B} \log \frac{p_i}{p_B} \tag{9}$$

The entropy function used for the evaluation is:

$$J(T) = H_o(T) + H_B(T) \tag{10}$$

When the entropy function takes the maximum value corresponding to *T* is the optimal threshold value sought.

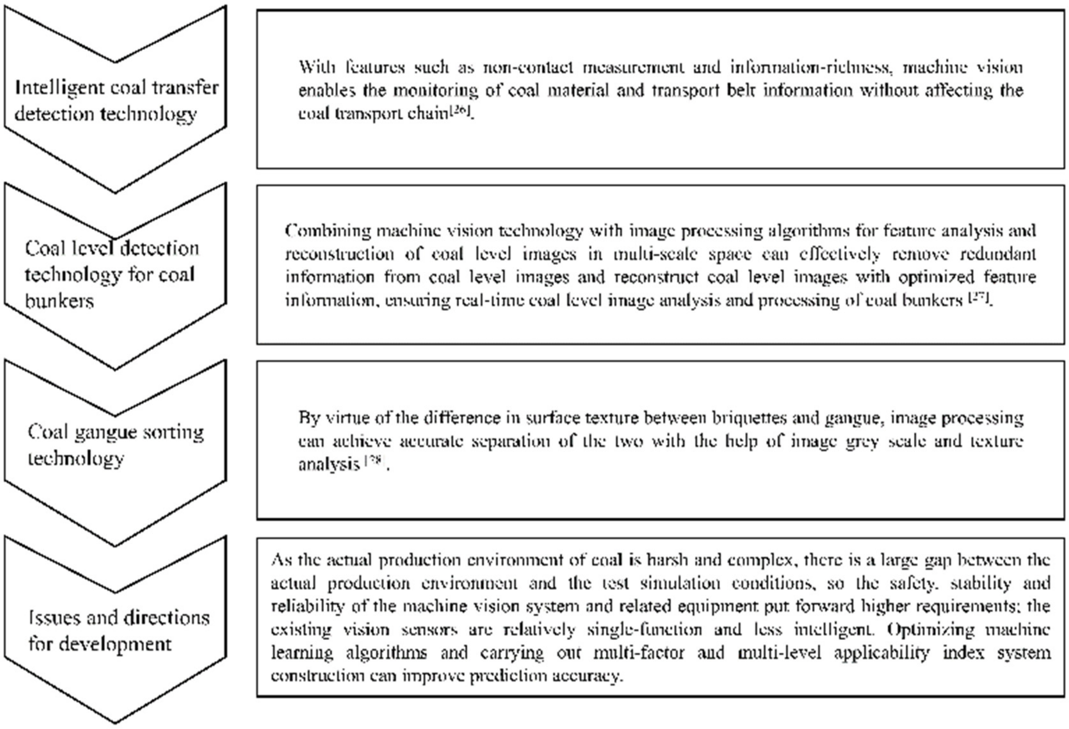
Image segmentation methods also include region-based image segmentation methods, edge detection-based image segmentation methods, and wavelet analysis and wavelet transform-based image segmentation methods, and the technical characteristics and applicability of different image segmentation methods are analyzed in Table 2.

Table 2. Technical advantages, disadvantages, and applicability analysis of different image segmentation methods

Algorithm name	Technical advantages	Shortcomings	Scope of application
Thresholding[18]	Direct use of the grey-scale properties of the image is computationally simple, computationally efficient, and fast.	Sensitive to noise, insignificant for grey scale differences and overlapping segmentation of different target grey scale values, so it needs to be combined with other methods.	Suitable for segmentation of targets and backgrounds with large differences[19].
Region[20]	Effectively overcome the shortcomings of small continuous image segmentation space existing in other methods, with better region features.	Tending to cause over-segmentation of images, combining edge detection with region segmentation can get good segmentation results.	Suitable for segmenting images with region structure[21].
Edge[22]	Search detection is fast and works well for edge detection.	Can't get a better region structure, contradiction between noise immunity and detection accuracy during edge detection. Accuracy increases at the expense of noise immunity and vice versa.	There is low noise interference, but the nature of the regions varies greatly from one to another[23].
Wavelet transform[24]	A local domain transforms in the air and frequency domains, thus effectively extracting information from a signal and performing multi-scale analysis of a function or signal through operational functions such as scaling and translation, solving many problems that cannot be solved by the Fourier transform.	As it operates in the frequency domain, it is insensitive to noise; a suitable filter needs to be selected.	Used for edge detection, multi-scale edges can be extracted and the type of edge can be distinguished by the calculation and estimation of the singularity of the image[25].

4. Analysis of the adaptability of rapid prediction technology for coal quality information

Figure 1 analyses the adaptability of coal quality information rapid prediction for intelligent coal transmission detection, coal bin coal level detection and coal gangue sorting, and forms the problems and development directions of coal quality information rapid prediction technology based on image analysis.


Intelligent coal transfer detection technology
Coal level detection technology for coal bunkers
Coal gangue sorting technology
Issues and directions for development

With features such as non-contact measurement and information-richness, machine vision enables the monitoring of coal material and transport belt information without affecting the coal transport chain[26].

Combining machine vision technology with image processing algorithms for feature analysis and reconstruction of coal level images in multi-scale space can effectively remove redundant information from coal level images and reconstruct coal level images with optimized feature information, ensuring real-time coal level image analysis and processing of coal bunkers [27].

By virtue of the difference in surface texture between briquettes and gangue, image processing can achieve accurate separation of the two with the help of image grey scale and texture analysis [28].

As the actual production environment of coal is harsh and complex, there is a large gap between the actual production environment and the test simulation conditions, so the safety. stability and reliability of the machine vision system and related equipment put forward higher requirements; the existing vision sensors are relatively single-function and less intelligent. Optimizing machine learning algorithms and carrying out multi-factor and multi-level applicability index system construction can improve prediction accuracy.

Figure 1. Rapid detection process of coal quality information based on image analysis.

5. Conclusion

This paper compares and analyses the characteristics of image processing techniques and algorithms based on machine vision technology. The study provides an important reference for the selection and optimization of methods for the rapid detection and prediction of media information. The technical process of rapid coal quality information detection and prediction has been formed by carrying out the analysis of the adaptability of coal quality information rapid prediction method, and the future research trends are discussed based on the above analysis.

1. The use of image recognition and machine vision and other means to build optimized algorithms can improve the safety, stability and reliability of the coal quality information rapid prediction system and related equipment.

2. Propose a rapid real-time detection technology solution for different coal quality raw materials, select image information acquisition hardware devices with high applicability and stability, realize dynamic monitoring at key control points, while optimized machine learning algorithms can obtain prediction models with high reliability and accuracy, and promote the application of intelligent equipment technology in coal quality information analysis.

3. Combining the different characteristics of coal quality and technical needs, the construction of a multifactor and multi-level applicability index system is carried out to form a suitability evaluation method that meets the requirements of technology, economy, and social aspects, which can promote the thermal utilization of coal and coal chemical industry to improve quality and increase efficiency.

References

[1]. Shi, J., Huang, W., Han, H., & Xu, C. (2020). Review on treatment technology of salt wastewater in coal chemical industry of China. Desalination, 493, 114640.

[2]. Shi, J., Huang, W., Han, H., & Xu, C. (2021). Pollution control of wastewater from the coal chemical industry in China: Environmental management policy and technical standards. Renewable and Sustainable Energy Reviews, 143, 110883.

[3]. Zhang, B., Wang, S., Wang, D., Wang, Q., Yang, X., & Tong, R. (2022). Air quality changes in China 2013–2020: Effectiveness of clean coal technology policies. Journal of Cleaner Production, 366, 132961.

[4]. Melikoglu, M. (2018). Clean coal technologies: A global to local review for Turkey. Energy Strategy Reviews, 22, 313-319.

[5]. Zhao, Y., Cui, Z., Wu, L., & Gao, W. (2019). The green behavioral effect of clean coal technology on China's power generation industry. Science of The Total Environment, 675, 286-294.

[6]. Wang, T., Chen, B., Zhang, Z., Li, H., & Zhang, M. (2022). Applications of machine vision in agricultural robot navigation: A review. Computers and Electronics in Agriculture, 198, 107085.

[7]. He, W., Liu, T., Han, Y., Ming, W., Du, J., Liu, Y., ... & Cao, C. (2022). A review: The detection of cancer cells in histopathology based on machine vision. Computers in Biology and Medicine, 105636.

[8]. Hashmi, A. W., Mali, H. S., Meena, A., Khilji, I. A., & Hashmi, M. F. (2021). Machine vision for the measurement of machining parameters: A review. Materials Today: Proceedings.

[9]. Yin, H., Yi, W., & Hu, D. (2022). Computer vision and machine learning applied in the mushroom industry: A critical review. Computers and Electronics in Agriculture, 198, 107015.

[10]. Pundir, M., & Sandhu, J. K. (2021). A systematic review of quality of service in wireless sensor networks using machine learning: Recent trend and future vision. Journal of Network and Computer Applications, 188, 103084.

[11]. Shi, Z., Xu, Z., & Wang, T. (2022). A method for detecting pedestrian height and distance based on monocular vision technology. Measurement, 111418.

[12]. Guo, X. D., Wang, Z. B., Zhu, W., He, G., Deng, H. B., Lv, C. X., & Zhang, Z. H. (2022). Research on DSO vision positioning technology based on binocular stereo panoramic vision system. Defence Technology, 18(4), 593-603.

[13]. Jaddoa, M. A., Gonzalez, L., Cuthbertson, H., & Al-Jumaily, A. (2021). Multiview eye localisation to measure cattle body temperature based on automated thermal image processing and computer vision. Infrared Physics & Technology, 119, 103932.

[14]. Zhu, R. Z., Feng, H. G., & Xu, F. (2023). Deep learning-based multimode fiber imaging in multispectral and multipolarimetric channels. Optics and Lasers in Engineering, 161, 107386.

[15]. Turukmane, A. V., Alhebaishi, N., Alshareef, A. M., Mirza, O. M., Bhardwaj, A., & Singh, B. (2022). Multispectral image analysis for monitoring by IoT based wireless communication using secure locations protocol and classification by deep learning techniques. Optik, 271, 170122.

[16]. Chen, W., Wang, W., Wang, K., Li, Z., Li, H., & Liu, S. (2020). Lane departure warning systems and lane line detection methods based on image processing and semantic segmentation: A review. Journal of traffic and transportation engineering (English edition), 7(6), 748-774.

[17]. Qureshi, I., Yan, J., Abbas, Q., Shaheed, K., Riaz, A. B., Wahid, A., ... & Szczuko, P. (2022). Medical Image Segmentation Using Deep Semantic-based Methods: A Review of Techniques, Applications and Emerging Trends. Information Fusion.

[18]. Tamal, M. (2020). Intensity threshold based solid tumour segmentation method for Positron Emission Tomography (PET) images: A review. Heliyon, 6(10), e05267.

[19]. Poletti, E., Zappelli, F., Ruggeri, A., & Grisan, E. (2012). A review of thresholding strategies applied to human chromosome segmentation. Computer methods and programs in biomedicine, 108(2), 679-688.

[20]. Cufi, X., Munoz, X., Freixenet, J., & Marti, J. (2003). A review of image segmentation techniques integrating region and boundary information. Advances in imaging and electron physics, 120, 1-39.

[21]. Yuan, M., Jing, Y., Armstrong, R. T., & Mostaghimi, P. (2022). Prediction of local diffusion coefficient based on images of fractured coal cores. Journal of Natural Gas Science and Engineering, 100, 104427.

[22]. Lu, Y., Duanmu, L., Zhai, Z. J., & Wang, Z. (2022). Application and improvement of Canny edge-detection algorithm for exterior wall hollowing detection using infrared thermal images. Energy and Buildings, 274, 112421.

[23]. Elmi, S., & Elmi, Z. (2022). A robust edge detection technique based on Matching Pursuit algorithm for natural and medical images. Biomedical Engineering Advances, 4, 100052.

[24]. Deng, F., Li, H., Wang, R., Yue, H., Zhao, Z., & Duan, Y. (2021). An improved peak detection algorithm in mass spectra combining wavelet transform and image segmentation. International Journal of Mass Spectrometry, 465, 116601.

[25]. Wang, C., Shi, A. Y., Wang, X., Wu, F. M., Huang, F. C., & Xu, L. Z. (2014). A novel multi-scale segmentation algorithm for high resolution remote sensing images based on wavelet transform and improved JSEG algorithm. Optik, 125(19), 5588-5595.

[26]. Karimpouli, S., Tahmasebi, P., & Ramandi, H. L. (2020). A review of experimental and numerical modeling of digital coalbed methane: Imaging, segmentation, fracture modeling and permeability prediction. International Journal of Coal Geology, 228, 103552.

[27]. Jiang, X., Mao, S., Li, M., Liu, H., Zhang, H., Fang, S., ... & Zhang, C. (2022). MFPA-Net: An efficient deep learning network for automatic ground fissures extraction in UAV images of the coal mining area. International Journal of Applied Earth Observation and Geoinformation, 114, 103039.

[28]. Lu, S., Li, M., Ma, Y., Wang, S., & Zhao, W. (2022). Permeability changes in mining-damaged coal: A review of mathematical models. Journal of Natural Gas Science and Engineering, 104739.

Research on the Analysis Method of Civil Aircraft Operational Safety Data

Jie Yu[1*]

[1] Shanghai Aircraft Design and Research Institute, Commercial Aircraft Corporation of China, Ltd,
Shanghai, 201210, China

[*]Corresponding author's e-mail: yujie1@comac.cc

Abstract

In the civil aircraft certification process, theoretical or empirical data provided by equipment suppliers are used to demonstrate compliance with the safety requirements related to CCAR 25.1309. During the aircraft operation phase, operational safety monitoring is required to continuously evaluate the operational data to confirm or correct the theoretical data given in the pre-certification system safety assessment process and to indicate the actual safety level of the aircraft. Through the operational safety data analysis method, the actual level of the aircraft is compared and analyzed with the deviation of the pre-certification safety target value, and corrective measures are taken to improve the aircraft operational safety level if necessary.

Keywords: Data Analysis, Monitoring, Operational Safety, Civil Aircraft

1. Introduction

The safety design of civil aircraft is a crucial factor in determining the life of an aircraft model [1], and with the rapid development of civil aircraft, the means to ensure operational safety is becoming more and more important [2]. Only by ensuring civil aircraft operational safety and overall aircraft airworthiness can we continuously improve aircraft operational efficiency [3], reduce aircraft cost consumption, and thus continuously improve the competitive strength of aircraft manufacturers [4]. In the civil aircraft certification process, the conformity to the safety requirements related to CCAR 25.1309 provisions is mainly based on theoretical data or empirical data (e.g., FMEA) provided by equipment suppliers [5]. In the aircraft operation phase, the condition monitoring system of the aircraft is the key to ensure the safe and stable operation of civil aircraft [6], and the data analysis method is the core of the operational safety monitoring work and the key to processing various types of information [7]. The deviation of the actual level of the aircraft from the pre-certification safety target value is compared and analyzed, and corrective measures need to be taken to improve the level of aircraft operational safety if necessary [8]. This paper is the first one in China to use operational safety data analysis method to continuously evaluate operational data from the main manufacturer's point of view, to confirm or correct the theoretical data given in the pre-certification system safety assessment process to indicate the actual safety level of the aircraft [9], and to verify the reasonableness of the pre-certification safety design, and its innovation and application value have been verified in actual work.

2. Analysis Steps

2.1. Analysis process

The safety data analysis method in the operation phase is to evaluate the safety level of the type aircraft during operation by processing and analyzing the operation data and event report data, combining with the design data, and giving the warning level, including the following steps.

- Operational qualitative data analysis. For the FTA related to catastrophic (Class I) and hazardous (Class II) failure states, analyze the correspondence between the operation phase event reports and the design FTA bottom events, and update and improve the FTA logical structure.

- Operation quantitative data analysis. Calculate the updated FTA bottom event occurrence probability using the data from the operation phase[10].

- Recalculate the probability of functional failure state. Recalculate the probability of each functional failure state based on the updated FTA logical structure and the probability of bottom events.

International Conference on Computer Graphics, Artificial Intelligence, and Data Processing (ICCAID 2022)
edited by Ruishi Liang, Jing Wang, Proc. of SPIE Vol. 12604, 1260418
© 2023 SPIE · 0277-786X · doi: 10.1117/12.2674693

- Determine the alarm level. For the recalculated functional failure state probability, compare it with the target probability and determine the alarm level.

2.2. Data entry

The data required for operational safety data analysis include：

- Aircraft operation data. Including: basic aircraft operation information, APU usage data, ETOPS situation, and engine usage data;

- Event reports during the operation phase. including: incident reports, usage difficulties, operational interruptions, supplier equipment inspection reports;

- FTA data in the design phase.

Collect all the above data from the whole fleet history of this civil aircraft, and conduct screening and preliminary analysis, including merging duplicate data, eliminating data with missing information, etc., and analyze and extract operational safety related parameters such as problem description [11], ATA involved, equipment part number, equipment name, number of installed aircraft, total operational flight time, etc. Through data collection and processing in the operation phase, the above data is converted into the basic data required for operation safety data analysis.

3. Qualitative Data Analysis

3.1. Basic requirements for qualitative data analysis

The operational qualitative data analysis work serves as the basis for quantitative assessment of top events (functional failure states) by analyzing the correspondence between the event reports of the operational phase and the bottom events in the design fault tree. In this process, the following differences need to be addressed.

- The FTA involves more specific information, such as the failure rate of components, than the operational event reports.

- The event in question should not have occurred during the operational phase (e.g. catastrophic, hazardous failure states) or did not occur during the reporting examination cycle content.

Therefore, the data analysis work carried out in the operation phase requires a combination and evaluation of the theoretical analysis results generated during the design process and the data on problems occurring during operation to obtain the level of safety in the operation phase.

3.2. Qualitative data analysis methodology

For larger (Class III) and smaller fault states, their probability of occurrence during the operational phase shall be calculated by the actual event assessment during the operational phase.

For catastrophic (Class I) and hazardous (Class II) fault states, they should be evaluated again using their FTA support material. the FTA contains the logical relationships between the underlying events and the underlying events that constitute the logical model of the fault state, called the FC model. The main methods to assist in analyzing the logical relationships of the FC model and modifying and refining the FC model through operational data include:

- If the impact of a run event may be comparable to the impact of an underlying event or part of a branch in the FTA not considered in the FC model, the event should be added to the FC model.

- If the logical relationships in the FC model do not match the logical relationships reflected in the operational data, the logical relationships in the FC model shall be adjusted to match the logical relationships in operations.

- If the operation event represents a part (such as a branch or subtree) rather than an underlying event in the logical relationship of the entire FC model, the operation event will replace that part.

The updated FC model is obtained after combining the operational data correction, called *FC**. For events in the FC that are not observed during the examination period, they are retained in the *FC** model.

4. Quantitative Data Analysis

4.1 Sampling requirement

After the FC^* model is built, the updated bottom event probability needs to be evaluated. In general, the sampling period is chosen to be within one year of flight data for a given aircraft type. If the sample size of data within one year is not sufficient to support the estimation of bottom event probability, the sampling period can be extended appropriately. Factors such as external environment, age of the aircraft, and senior equipment should also be taken into account when sampling.

4.2 Quantitative data analysis method

It is assumed that the failure of electronic components conforms to the exponential distribution characteristics. The calculation in statistical time T is given in equation (1):

$$T = \sum_i N_i T_i \tag{1}$$

T, statistical duration of aircraft operating time during the sampling period, FH;

N_i, the number of parts installed in the *i*-th aircraft, in units;

T_i, total flight hours of the *i*-th aircraft, FH.

The failure rate under the "optimistic" estimate is given in equation (2):

$$\lambda_{low} = \begin{cases} min(C_l(1,T), \lambda_{SSA}), & r = 0 \\ C_l(r,T), & r \geq 1 \end{cases} \tag{2}$$

λ_{SSA}, state failure rate under system security assessment;

r, the number of times a failure state is counted during the sampling period, in units of one.

The failure rate λ_{high} under the "pessimistic" estimate is given in equation (3):

$$\lambda_{high} = C_h(r,T) \tag{3}$$

The statistics $C_l(r,T), C_h(r,T)$ in the "optimistic" and "pessimistic" estimated failure rates are calculated in equations (4) and (5):

$$C_l(r,T) = \frac{\chi^2(1-\alpha,2r)}{2T} \tag{4}$$

$$C_h(r,T) = \frac{\chi^2(\alpha,2(r+1))}{2T} \tag{5}$$

α, Confidence level, between 0 and 1.

4.3 Calculation method of functional failure state occurrence probability

Based on the bottom event occurrence probability that has been evaluated, the probability of occurrence of the functional failure state is recalculated using the FC^* model based on the FTA calculation method. The probability of the functional failure state is $P_{low}(FC^*)$ using the failure rate calculation formula under optimistic estimation, and the probability of the functional failure state is $P_{high}(FC^*)$using the failure rate calculation formula under pessimistic estimation.

5. Alarm Analysis

5.1 Alarm level and response requirements

The probabilities of catastrophic (Class I) and hazardous (Class II) functional failure states, as assessed by operational data, are compared with the design values to determine the alert levels and corresponding countermeasures.

The severity of the determined probability of functional failure states varies at different levels of operational maturity. The assessment of alert levels and countermeasures depends on the operational maturity of the aircraft type. Aircraft operational maturity generally goes through three stages, as shown in Table 1.

Table 1.Definition of Operational Maturity

Maturity	Description
Maturity Stage I	The total flight time of the fleet is less than 200,000 flight hours after the aircraft is put into operation.
Maturity Stage II	The total flight time of the fleet is between 200,000 and 800,000 flight hours.
Maturity Stage III	The aircraft is in operation with a total fleet flight time greater than 800,000 flight hours.

The alarm level expresses the severity of the failure and the required countermeasures, as detailed in Table 2.

Table 2. Alarm Levels and Response Measures

Alarm level	Related Requirements
1	Further observation is required.
2	Develop a survey plan and if needed design changes should be developed in a timely manner.
3	Immediate action is required. For example, design changes, operational restrictions, and corresponding maintenance work.

5.2 Functional failure status alarm analysis

For Class I and II failure states, the operational safety alert levels and corresponding countermeasures are determined by comparing their operational data evaluation results with the design indicators. The design indicators are described as follows.

- Safety indicator $P_S(FC)$: the safety indicator defined in AC25.1309, such as the probability of occurrence of a Class I functional failure state of 10^{-9} times per flight hour and a Class II functional failure state of 10^{-7} times per flight hour;
- Design indicator $P_D(FC)$: the probability of occurrence of functional failure states obtained by design and analysis, usually $P_D(FC) < P_S(FC)$.

The functional failure state alarm judgment matrix in Table 3 is used to determine the operational security alarm level of the functional failure state. The symbol # indicates the comparative relationship between the two, for example: A#B, indicating that A is similar to B or A<B. When the design index $P_D(FC)$ and safety index $P_S(FC)$ of an event trigger the alarm level inconsistently, the higher alarm level should be selected.

Table 3. Function Failure Status Alarm Determination Matrix

Situation	Alarm level		
	Maturity Stage I	Maturity Stage II	Maturity Stage III
$P_{low}(FC^*)\#P_{high}(FC^*)\#P_s(FC)$	0	0	0
$P_{low}(FC^*)\#P_s(FC) < P_{high}(FC^*)$	1	2	2
$P_s(FC) < P_{low}(FC^*)\#P_{high}(FC^*)$	2	3	3
$P_{low}(FC^*)\#P_{high}(FC^*)\#P_D(FC)$	0	0	0
$P_{low}(FC^*)\#P_D(FC) < P_{high}(FC^*)$	0	1	1
$P_D(FC) < P_{low}(FC^*)\#P_{high}(FC^*)$	1	2	2

6. Data Feedback

The results of the operational safety monitoring and evaluation should be fed back to the competent safety unit, and corrective measures should be taken when necessary to improve the level of aircraft operational safety, to ensure the operational safety of the aircraft, and to update the safety design documents when necessary. The data feedback required for the operational safety assessment includes：

- Completion of the updated FC^* model.
- The probability of occurrence of functional failure states recalculated using operational data.
- Alarm levels and their responses.
- Relevant raw data and the setting parameters and assumptions of the calculation process.

7. Summary

This paper is the first time in China to use the operational data evaluation method to compare and analyze the deviation of the actual safety level of the aircraft with the pre-certification safety target value, and to take corrective measures to improve the operational safety level of the aircraft and ensure the operational safety of the aircraft if necessary according to the warning level. The method has been verified and applied in the actual type work, providing a good theoretical basis and technical reserve for the development of domestic large aircraft to establish the method and process of civil aircraft safety design verification, and solving the problem that the rationality of the preliminary aircraft safety design cannot be verified, which has important significance and good application prospects in the field of safety analysis.

References

[1] Kang D. (2017) Application of big data mining analysis in aero-engine condition monitoring and fault diagnosis. Journal of Xi'an Aviation College,35(5):42-46.

[2] Chen C. (2016) Research on model aircraft condition monitoring system. Science & Education Guide,11:129-130.

[3] Bi F. (2014) Civil aviation engine maintenance and management based on condition monitoring technology. Science and Wealth,7:278.

[4] Wu Y, Li J, Wu X. (2017) Deep Belief Network-based Engine Condition Monitoring for Civil Aviation. Computerized measurement and control,25(7):28-31.

[5] Lin LG. (2019) Research and system development of reliability monitoring technology for a certain type of aircraft. Science & Technology Vision,12:64-65.

[6] Li YP. (2005) Research and system development of reliability monitoring technology for a certain type of aircraft. Journal of the Civil Aviation Institute of China,06:18-21.

[7] Guo YW, Jia YX. (2018) Research on Reliability Monitoring Technology for Civil Aircraft Airworthiness Certification Test Flight Phase. Aviation Maintenance and Engineering,03:45-48.

[8] Zhang P. (2017) Civil aviation aircraft flight management system reliability monitoring simulation study. Computer Simulation,34(08):105-109.

[9] Xie Y.(2016)Research on the characteristics parameters of civil aircraft prediction and health management technology and system[J].Science & Technology Vision,20:287-289.

[10] Yan S, Xue H. (2014) Aircraft Area Safety Analysis Process Optimization and Implementation Methodology. Aviation Science and Technology,04:36-41.

[11] Li T, Ye B, Ding X, Wang D, Chen Z. (2022) Exploring the application of regional safety analysis in aircraft final assembly process. Aerospace Manufacturing Technology, Z1:92-97.

A Method for Extracting Correlative Features of Power Grid Big Data Based on Improved Deep Learning

Xinyan Wang[1], Ying Zhu[1*], Yongjie Ning [2], Jiacheng Du[1], Jingli Jia[1],
[1]State Grid Henan Information and Telecommunication Company, Zhengzhou, Henan 450052;
[2]State Grid Puyang Power Supply Company, Puyang 457000;
Corresponding author: tammychen2022@163.com

ABSTRACT

With the continuous maturity of big data, artificial intelligence, Internet of Things and other technologies, the rapid development of smart grid has been helped, but at the same time, the increasing line loss power has also attracted widespread attention. In the process of building a smart grid, each link of the grid operation generates a large number of multi-source heterogeneous data, including line loss data and line loss cause related data, which constitutes the big line loss data. First of all, considering the mining efficiency in big data, FP growth algorithm in association rule learning is selected to search the frequent item set of line loss features. Support, confidence and lift are used as evaluation indicators to analyze the association relationship between the causes of line loss; Secondly, a line loss prediction model based on deep learning is established. By eliminating the influence of line loss characteristics in turn, the correlation contribution of line loss causes to line loss is calculated to quantify the line loss caused by line loss causes. After verification, the depth confidence network and BP depth neural network as the prediction model of the depth learning method are superior to the shallow artificial neural network model in the prediction effect, and the prediction accuracy means the reliability of the contribution calculation. Finally, combined with the above two aspects of analysis, the causes of line loss in the substation area are comprehensively evaluated, and guidance suggestions are given to assist power enterprises in decision-making.

Key words: improve deep learning; Power grid big data; Relevance; feature extraction

1. INTRODUCTION

The research of distribution network is still in the development stage, which is caused by objective constraints such as the complex internal structure of distribution network and relatively backward construction progress. With the steady development of society, the growth rate of electricity consumption is slowing down, and the investment in the power supply side has gradually reached saturation. The main direction of capital inflow has gradually shifted from the backbone network to the distribution network side. With the support and guidance of a series of new policies issued by the state, the State Grid has increasingly invested in the construction of distribution networks, and the theoretical research and practice related to distribution networks have also ushered in new opportunities.

In recent years, with the rise of big data, artificial intelligence, cloud computing, Internet of Things and other concepts, the power system is gradually moving towards intelligence and informatization, and the traditional power grid is also transforming to smart grid. In this process, a large number of data closely related to line loss are constantly generated, forming the big data of line loss. By means of deep learning, machine learning, data mining and other methods, this paper takes the big data of power grid line loss as an example to reveal the huge value behind the data related to the cause of line loss and line loss data. If we simply locate the cause of line loss without quantifying the results, it is difficult to find the key points. If we simply quantize the results of line loss analysis without locating the causes, it is difficult to grasp the direction of loss reduction. In order to combine the two, we should not only explore the diversified causes of line loss, but also explore the relationship between the causes of line loss and line loss. Therefore, it is particularly important to carry out the analysis and prediction of the correlation feature extraction of distribution network big data, and then solve the problem of line loss assessment and governance.

2. EXTRACTION METHOD OF POWER GRID BIG DATA CORRELATION FEATURES

2.1 Cause and characteristics of big data of power grid line loss

In addition to the fixed and inevitable power loss, other losses are generally caused by abnormal conditions. The

International Conference on Computer Graphics, Artificial Intelligence, and Data Processing (ICCAID 2022)
edited by Ruishi Liang, Jing Wang, Proc. of SPIE Vol. 12604, 1260419
© 2023 SPIE · 0277-786X · doi: 10.1117/12.2674980

abnormal manifestations of line loss constitute the cause of line loss. In this paper, the cause of line loss is divided into file anomaly, meter anomaly, acquisition anomaly and station operation anomaly. Among them, meter and acquisition exceptions are essentially equipment failures. The location of the failure is mainly in the metering module, acquisition module and communication module of the equipment. However, from the previous occurrence of exceptions, the types of equipment failures are complex and diverse, and the relative occurrence frequency is high. Therefore, the equipment failures are further analyzed in detail. File exceptions mainly include inconsistent records of the same information in different systems, and omissions and mistakes due to negligence. The abnormal operation of the station area is mainly due to the unreasonable planning and scheduling of the station area, which aggravates the degree of loss. The problem needs to be solved from the level of station area management. The cause of line loss is an abstract summary of specific anomaly forms with some commonalities. Each cause of line loss has a variety of more detailed characteristics. Through mining the characteristics of line loss, accurate positioning of line loss is achieved, possible consequences are analyzed, and anomalies are evaluated.

The consequences caused by different line loss characteristics are reflected in the amount of abnormal line loss. This paper refers to the Manual for Handling the Line Loss Anomaly in the Same Period in the Substation Area compiled by the Marketing Department of State Grid Corporation of China. According to the magnitude of line loss, the substation area can be divided into high loss substation area, negative loss substation area and non computable line loss substation area. The line loss characteristics are divided into subordinate cause categories, and the possible losses caused by each characteristic are summarized. Among them, there are mainly two cases of long-term high loss and sudden high loss in the high loss station area; There are three cases of long-term negative loss, small negative loss and sudden negative loss in the negative loss station area; If the line loss is not calculable, it means that the power supply is zero or null, and the power consumption is null. If the characteristics are related to the acquisition and communication modules, it is a lack of acquisition, and the statistical line loss may become larger, but the user has paid for the power consumed by himself. That is to say, from an economic perspective, this part of the power is not lost, and this part of the loss is also called a false communication loss.

2.2 Large data characteristic data set of power grid line loss

According to the actual situation of line loss in a certain area and the practical experience of local experts and staff, the characteristic judgment rules of specific line loss causes are formulated, so as to extract the existing line loss characteristics from the line loss cause feature library, and further collect the line loss feature data set for subsequent analysis.

After the judgment rules of line loss characteristics are formulated, the calculation and statistics can be carried out according to the original meter data and archive data, and then the line loss characteristics and their data sets can be extracted. This is the data basis for the subsequent analysis of the correlation relationship of line loss characteristics. The data set takes the daily abnormal characteristics of the distribution network substation area as the minimum unit to collect data, in which the line loss characteristic data of file exceptions, meter exceptions and acquisition exceptions are collected according to the number of times per day in the substation area, and the line loss characteristic data of abnormal operation in the substation area is collected according to the length of time per day. The characteristic fields and numerical units of the dataset are shown in Table 1.

Table 1 Collection of Line Loss Characteristic Data Set

Line loss characteristics	Numerical unit	Characteristic field	Line loss characteristics	Numerical unit	Characteristic field
Inconsistent relationship between Taiwan households	Day/time	A1	Phase failure	Day/time	B9
Inconsistent comprehensive magnification	Day/time	A2	Current transformer fault	Day/time	B10
Zero meter	Day/time	A3	Power metering fault	Day/time	B11
Overcapacity of meter	Day/time	B1	Acquisition fluctuation	Day/time	C1
Reverse the meter	Day/time	B2	Acquisition missing	Day/time	C2
The meter stops	Day/time	B3	Power factor overrun	Days/Hours	D1
Loss of voltage	Day/time	B4	Three phase unbalance	Days/Hours	D2

Loss of current	Day/time	B5	overload	Days/Hours	D3
Measurement inaccuracy	Day/time	B6	heavy load	Days/Hours	D4
Wiring error	Day/time	B7	Light load	Days/Hours	D5
Open phase	Day/time	B8			

3. ASSOCIATION RULES BASED ON IMPROVED DEEP LEARNING

3.1 Improve the construction of deep learning association rules

To improve deep learning association rules, unsupervised learning is used to mine knowledge. The purpose is to find strong association rules in the data set according to some evaluation indicators. The key to improving deep learning association rules is to mine frequent itemsets. Itemsets are subsets of several item (event) sets (or full sets, which are special subsets but not empty sets). When the support of an itemset is greater than the threshold set according to experience, it is called frequent itemsets. There are two main methods for mining frequent itemsets: Apriori algorithm and FP Growth algorithm.

The basic idea of Apriori is that if an item set belongs to a frequent item set, then all non empty subsets of the item set are called frequent item sets. On the contrary, if an item set does not belong to a frequent item set, then all supersets of the item set do not belong to a frequent item set. In the algorithm, non frequent itemsets are no longer searched for frequent itemsets with higher number of items by means of "pruning". The basic steps of the algorithm are:

(1) For a given dataset, first search for one frequent itemset, set the support threshold ε_1 for one itemset, and retain the one itemset whose support is greater than the threshold. Only the remaining one itemset is eligible for subsequent calculation.

(2) On the basis of the reserved $k-1(k>1)$ itemsets in the previous step, connect the candidate frequent $k-1$ itemsets, set the support threshold of k itemsets to ε_k, and search for qualified frequent k itemsets.

(3) Repeat step 2 to search until no frequent $k+1$ itemsets can be found, that is, if the support of all candidate $k+1$ itemsets does not reach the threshold ε_{k+1}, then all frequent itemsets will be output and the algorithm ends.

The Apriori algorithm scans the data set every time it judges whether the candidate frequent item set meets the support threshold. When the data volume is too large or the candidate item set is too large, it will take a lot of time to traverse, resulting in low efficiency. The FP growth algorithm only needs to traverse the database twice to complete the search of frequent item sets. The first time is to count all the elements that appear, and the second time is to search for the item sets that meet the frequent conditions. The greater the amount of data, the more obvious the speed advantage of FP growth search.

FP growth algorithm stores data in a prefix tree structure called FP tree. The root node of the FP tree is an empty set, while other nodes record a single element and its occurrence times. The element with more times is closer to the root. All elements on the path from each leaf node to the root node on the tree appear at the same time.

In the process of mining frequent itemsets, the conditional pattern base (CPB) plays an important role. CPB refers to the set of all prefix paths that exist at the end of the target leaf node of the FP tree. CPB extracts from the FP tree, builds a conditional FP tree with all CPBs for each frequent itemset, and finds a new FP tree composed of corresponding CPBs, which is a set composed of frequent itemsets. This process is recursive until the algorithm ends the condition. In general, the process of FP growth algorithm is divided into two steps: building the FP tree and extracting CPB to continue building the FP tree, and mining frequent itemsets from the FP tree.

3.2 Frequent Item Set Evaluation Indicators

There are three common frequent item set evaluation indicators, namely, support, confidence and promotion.

Support ($Support$) indicates the ratio of the number of simultaneous occurrences of all events of the $\{X_1,\ldots,X_k\}(1\le k\le n)$ transaction in the data set to the total data sample. If the events occur very frequently at the same time, it indicates that there may be a correlation in the $\{X_1,\ldots,X_k\}(1 \le k \le n)$.

$$Support(X_1,\ldots,X_k) = P(X_1,\ldots,X_k) \tag{1}$$

The confidence level ($Confidence$) represents the probability that all events in the transaction $\{X_{q+1},\ldots,X_k\}$ will occur at the same time when a frequent item set $\{X_1,\ldots,X_q\}(q < k)$ occurs, that is, the conditional probability in probability. If the confidence level is too low, it indicates that the occurrence of itemset $\{X_1,\ldots,X_q\}$ has little relationship with the occurrence of itemset $\{X_{q+1},\ldots,X_k\}$, which further indicates that the correlation between them is weak.

$$Confidence(X_{q+1},\ldots,X_k \left| X_1,\ldots,X_q\right.) = \frac{P(X_1,\ldots,X_k)}{P(X_1,\ldots,X_q)} \tag{2}$$

The lifting degree ($Lift$) is the ratio of the probability of simultaneous occurrence of $\{X_{q+1},\ldots,X_k\}$ and the probability of $\{X_{q+1},\ldots,X_k\}$ in the case of $\{X_1,\ldots,X_q\}$, reflecting the correlation between $\{X_1,\ldots,X_q\}$ and $\{X_{q+1},\ldots,X_k\}$. The higher the lift degree is, the higher the positive correlation is. The lower the lift degree is, the higher the negative correlation is. If the lift degree is equal to 1, there is no correlation.

$$Lift(X_{q+1},\ldots,X_k \Leftarrow X_1,\ldots,X_{k-1}) = \frac{P(X_1,\ldots,X_k)}{P(X_1,\ldots,X_q)P(X_{q+1},\ldots,X_k)} \tag{3}$$

3.3 Large data prediction process of power grid based on deep learning

Before analyzing the relationship between the cause of line loss and line loss, it is necessary to build a line loss prediction model based on deep learning. The data set of the new principal component feature $T_1, T_2, \ldots, T_w$ optimized by PCA is used as the input, and multiple hidden layers are set in the middle to conduct in-depth learning on the massive line loss feature data. For any group of training samples $(t_1, t_2, \ldots, t_w)$, the input layer starts to learn layer by layer through each hidden layer until the output layer. If the output layer results do not meet the expectations, the error is propagated layer by layer in reverse, and the weights and offsets of neurons are constantly adjusted until the requirements are met, The final output line loss of the model is Loses. The process of forecasting the line loss of the distribution network substation area through the in-depth learning method is shown in Figure 1, which is divided into the following steps:

(1) Acquisition of sample data related to line loss prediction in distribution network substation area: including line loss characteristic data and line loss volume data. Then, the non-technical part of the line loss is calculated by statistical line loss and technical line loss, and the line loss data is normalized. Finally, the training set data and test set data are divided;

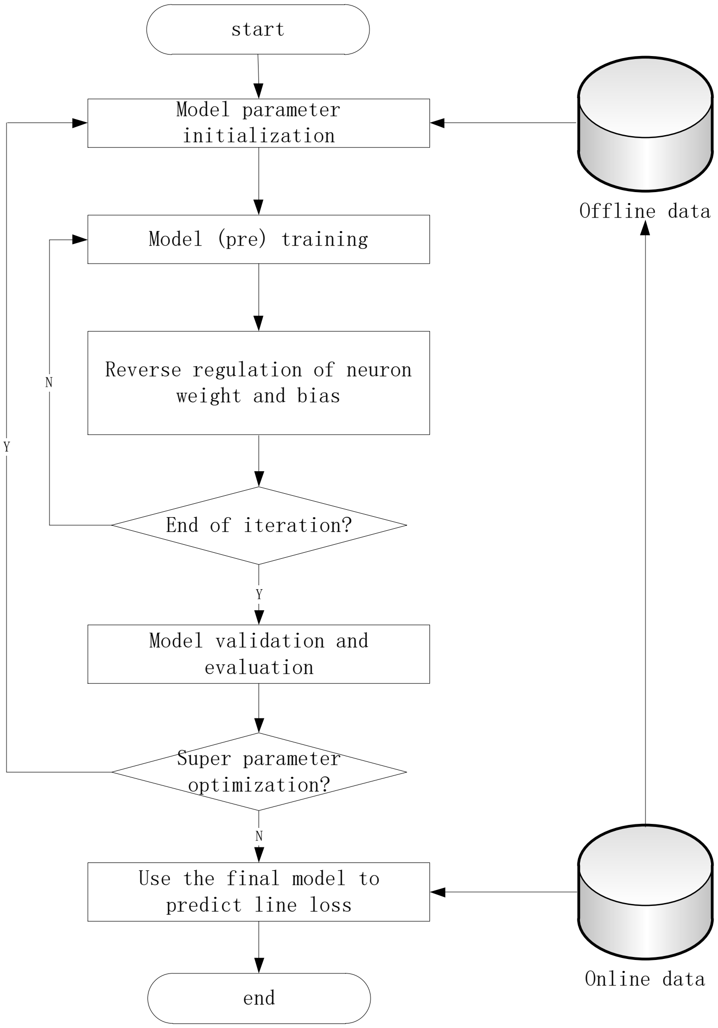

start
Model parameter initialization
Model (pre) training
Reverse regulation of neuron weight and bias
End of iteration?
N
Y
Model validation and evaluation
Super parameter optimization?
Y
N
Use the final model to predict line loss
end
Offline data
Online data

Figure 1 Power grid big data prediction process based on deep learning

(2) Initialization of line loss prediction model: including the selection of activation function of hidden layer and output layer, the selection of loss function, and the determination of super parameters such as the number of hidden layers, the number of neurons in each layer, learning rate, momentum, and training batch. Where, if the DBN-nn model is used, the number of RBM layers and the number of regression layers need to be specified respectively;

(3) Line loss prediction model (pre) training: take line loss characteristic index data as input, train layer by layer from the input layer to the output layer of the model, and update the neuron state of each layer;

(4) Reverse parameter adjustment of line loss prediction model: the non-technical line loss value is used as the model output, the root mean square loss function is used, and Adam optimization algorithm is used as the gradient descent algorithm of the model. The weight and bias of neurons are constantly adjusted to make them close to the optimal parameters. Repeat steps (3) and (4) until the model meets the required accuracy or reaches the set number of iterations, stop training and save the model parameters;

(5) Verification and evaluation of line loss prediction model: input the labeled test set data into the completed training line loss prediction model, calculate the predicted line loss value, compare it with the real line loss value, and evaluate the model by calculating the loss function;

(6) Line loss prediction model super parameter re optimization: set the number of iterations for super parameter optimization. If the model evaluation results do not meet the requirements, go back to step (2) to reinitialize the super parameters, otherwise go to step (7);

(7) Online use of line loss prediction model: input online line loss data, and update the line loss prediction results to the database in real time. At the same time, the newly generated data can be added to the offline data set to train the model to increase the richness of model samples and training accuracy.

3.4 Large data prediction process of power grid based on deep learning

When the line loss prediction model training is completed, prepare to input a new batch of line loss characteristic data. Before input, each group of data shall be subject to exception elimination processing, that is to say, the daily abnormal characteristics (non-zero values) in the station area shall be returned to zero in turn, and then the adjusted line loss characteristic data in the station area shall be input into the line loss prediction model in turn, and the predicted line loss value $\mathrm{Losses}_{zero}^{(i)}$ after the elimination of abnormal effects shall be output. Then compare with the true value of line loss $\mathrm{Losses}_{real}^{(i)}$, and calculate the associated contribution of the first line loss feature to the line loss (set as AC). It can be seen that if an anomaly does not occur, the predicted value does not change from the true value, and the associated contribution of the anomaly feature is 0.

$$AC_i = \frac{\mathrm{Losses}_{real}^{(i)} - \mathrm{Losses}_{zero}^{(i)}}{\mathrm{Losses}_{real}^{(i)}} \tag{4}$$

In order to reflect and understand the meaning of this indicator more intuitively, the contribution degree is divided into hundreds,

$$AC_i(\%) = \frac{AC_i}{\sum_{k=1}^{N} AC_k} \times 100\% \tag{5}$$

Obviously, the accuracy of the line loss prediction model plays a key role in the reliability of the correlation contribution, which is reflected in the prediction accuracy of $\mathrm{Losses}_{zero}^{(i)}$. Based on reliable correlation contribution, it is necessary to focus on the anomaly characteristics with high contribution, take measures in advance to prevent the occurrence of anomaly or minimize the loss caused by the occurrence of anomaly, and pay attention to the occurrence of other associated anomalies.

4. CASE ANALYSIS

4.1 Basic information

A total of 1012 multi anomaly stations with line loss characteristics occurring more than 5 times on May 1, 2021 in a region were selected. After preliminary feature screening, 13 features were retained. According to the above, the correlation analysis of line loss causes and the correlation analysis of line loss causes and line loss have been completed. In the past research, scholars have conducted line loss analysis on more than one aspect. This paper will combine the analysis of these two aspects to conduct a more comprehensive analysis and develop more targeted loss reduction strategies.

4.2 Overall design of power grid line loss big data association model

The overall design of the line loss correlation model is shown in Figure 2:

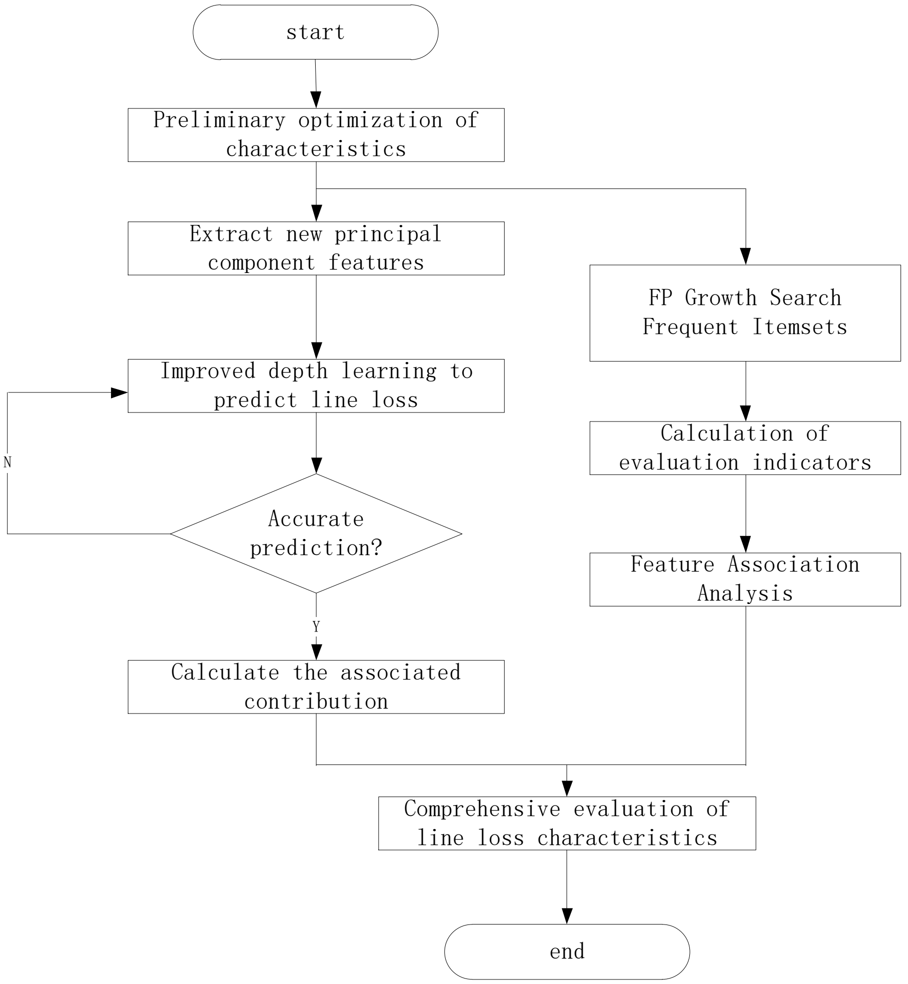

start
Preliminary optimization of characteristics
Extract new principal component features
FP Growth Search Frequent Itemsets
Improved depth learning to predict line loss
N
Calculation of evaluation indicators
Accurate prediction?
Feature Association Analysis
Y
Calculate the associated contribution
Comprehensive evaluation of line loss characteristics
end

Figure 2 Overall design of line loss correlation model

After the initial optimization of removing similar features and irrelevant features from line loss features, line loss feature data will be used for two operations. On the one hand, FP Growth algorithm, which is more efficient in calculation, is used to mine frequent itemsets of line loss features, and the correlation of line loss features is analyzed according to the calculation results of evaluation indicators. On the other hand, the line loss feature data is input into the PCA model for dimension reduction and optimization, and new principal component features are extracted. Furthermore, the new feature data set is input to the line loss prediction model based on deep learning. In this paper, BPDNN model and DBN-DNN model in the deep learning method are selected for the prediction process. For the trained line loss prediction model, predict the data that eliminate the influence of abnormal characteristics, and then calculate the correlation contribution of the line loss characteristics to the line loss, and find out the main causes of the line loss. Finally, the line loss characteristics are comprehensively evaluated based on the results of the two aspects.

4.3 Comprehensive evaluation scheme

The comprehensive assessment scheme of substation area line loss of distribution network is shown in Figure 3:

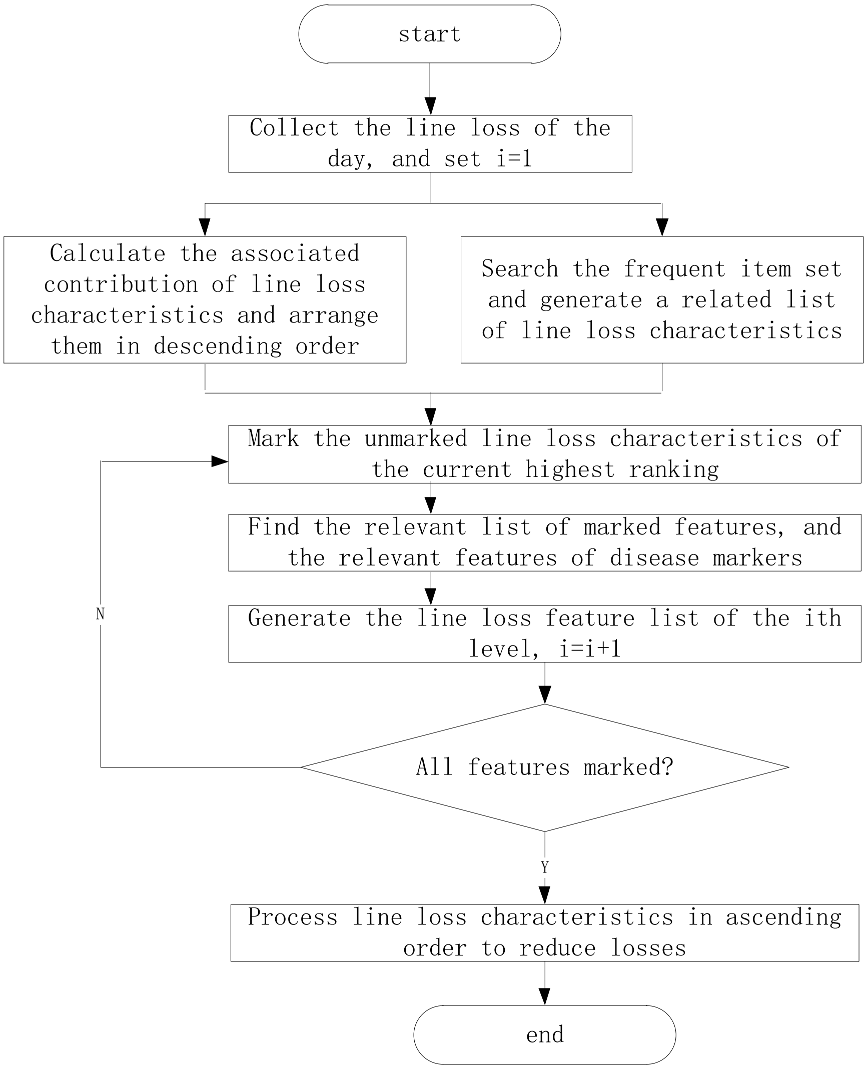

start
Collect the line loss of the day, and set i=1
Calculate the associated contribution of line loss characteristics and arrange them in descending order
Search the frequent item set and generate a related list of line loss characteristics
Mark the unmarked line loss characteristics of the current highest ranking
Find the relevant list of marked features, and the relevant features of disease markers
Generate the line loss feature list of the ith level, i=i+1
N
All features marked?
Y
Process line loss characteristics in ascending order to reduce losses
end

Figure 3 Comprehensive Assessment Scheme for Line Loss in Radio Area of Distribution Network

For the occurrence of abnormal line loss features in all substation areas of the distribution network every day, the frequent item set is searched, and the relevant list of line loss features is formed according to the associated evaluation indicators. For the daily situation of a single station, calculate the associated contribution of the line loss characteristics that have occurred, and rank the contribution in descending order. Start from the line loss feature with the first contribution ranking, mark the feature, and then find the relevant list of the feature on the current day. If the features occurring on the current day correspond to the features in the list, this feature should also be marked to get the first level of the line loss feature list to be solved. Next, continue to search for unmarked line loss features in the order of contribution, repeat the above feature marking steps until all the features occurred are marked, and finally conduct loss reduction guidance according to the order of the hierarchy, giving priority to solving high contribution and its related line loss features.

4.4 Results and Analysis

The correlation coefficients are calculated for different characteristics, and the results are shown in the thermodynamic diagram in Figure 4. The line loss characteristics corresponding to each field number in the figure are shown in Table 1

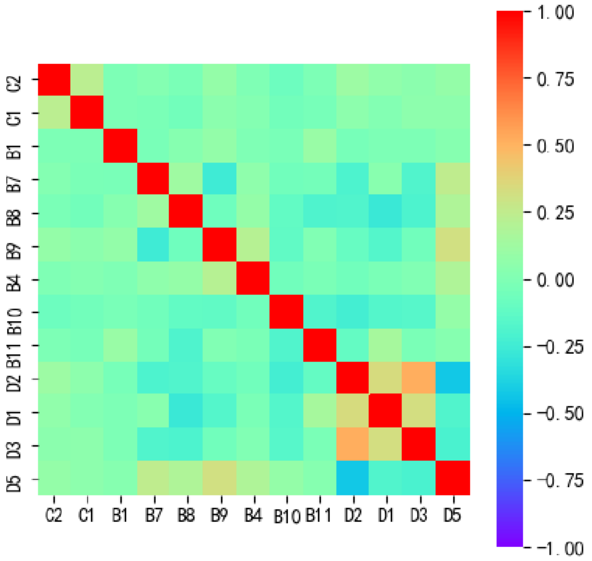


Figure 4 Thermal Diagram of Line Loss Characteristic Correlation Coefficient

From all stations, a multi abnormal station area with eight line loss characteristics is selected as the analysis object. The process of comprehensive line loss assessment for this station area is as follows:

(1) Calculate the associated contribution of line loss features and arrange them in descending order. The eight line loss characteristics and their associated contributions are C2 (14.3%), D1 (14.1%), C1 (13.9%), D2 (13.9%), B7 (12.7%), D3 (12.1%), B9 (10.1%) and B4 (9%) respectively;

(2) According to the calculation result of correlation coefficient between line loss features, set the minimum correlation threshold value to 0.1, and then generate a correlation list of line loss features, as shown in Table 2:

Table 2 List of Line Loss Characteristics

Line loss characteristics	Related List	Line loss characteristics	Related List
C2	C1 D2	B10	B8 B9 B11 D1 D2 D3
C1	C2	B11	B8 B10 D1 D2
B1	B11	D2	C2 B7 B8 B9 B10 D1 D3 D5
B7	B8 B9 D2 D3 D5	D1	B8 B9 B10 D1 D2 D5
B8	B7 B10 B11 D1 D2 D3 D5	D3	B7 B8 B10 D1 D2 D5
B9	B4 B7 B10 D1 D2 D5	D5	B4 B7 B8 B9 D1 D2 D3
B4	B9 D5		

(3) The line loss characteristics are processed according to the priority order of the hierarchy, and the processing process is shown in Table 3.

(3.1) Start with the feature C2 with the highest contribution, and find out the relevant features C1 and D2 according to Table 2. Therefore, C2, C1 and D2 are the first level, with the highest priority of processing, and will be marked as the processed feature after processing;

(3.2) Starting from D1, which has never been marked and has the highest contribution at present, relevant features B7, B9 and D3 can be found to form the second level, and these features can be processed and marked;

(3.3) Similarly, find feature B4 and deal with it. So far, all features have been marked, and the evaluation of the cause of line loss in the platform area has ended.

Table 3 Processing Process of Line Loss Characteristics in Substation Area

Class	Unhandled characteristics	Processing characteristics	Processed Features
$i=0$	C2 D1 C1 D2 B7 D3 B9 B4	/	/
$i=1$	D1 B7 D3 B9 B4	C2 C1 D2	C2 C1 D2
$i=2$	B4	D1 B7 D3 B9	C2 C1 D2 D1 B7 D3 B9
$i=3$	/	B4	C2 C1 D2 D1 B7 D3 B9 B4

5. CONCLUSION

This paper introduces the line loss correlation analysis based on association rules and deep learning, including the line loss cause correlation analysis based on association rules and the line loss prediction based on deep learning. On the one hand, association rules are used to mine frequent itemsets of line loss features. Aiming at the inefficiency of Apriori algorithm, which needs to query the database many times, this paper proposes an association analysis method based on FP growth. It only needs to query the data twice to complete the search for frequent itemsets, which greatly improves the efficiency. Then calculate the evaluation index of frequent itemsets, and analyze the correlation between line loss features. On the other hand, BPDNN and DBN-DNN in the depth learning method are used to predict the line loss, and the line loss prediction process based on depth learning is designed. The prediction model is obtained through training, and the impact of abnormal line loss features is eliminated in turn before prediction, so as to calculate the associated contribution of line loss features to line loss. Finally, combined with the correlation analysis of the two aspects, the comprehensive evaluation of the correlation relationship of line loss is carried out, and the guidance and suggestions for loss reduction are given.

REFERENCES

[1] Yang Jinggang, Deng Min, Ma Yong, et al. PRPD data feature extraction method based on depth learning [J] Electric measurement and instrument, 2020, v.57; No.728(03):104-109+120.

[2] Chunjin, Zhang Xinchang, Guo Haijing, et al Population spatial sampling method based on multi-source information and deep learning feature extraction [J] Surveying and Mapping Bulletin, 2021 (8): 6

[3] Wang Sufang, Xie Fang Research on tractor electrical fault diagnosis method based on deep learning theory and big data [J] Research on Agricultural Mechanization, 2021, 43 (6): 5

[4] Deng Xiong, Wang Hongchun Face detection algorithm based on depth learning and feature fusion [J] Computer Application, 2020, 40 (4): 7

[5] Zhu Yanmin, Xu Ailan, Sun Qiang New progress in air quality prediction methods based on depth learning [J] China Environmental Monitoring, 2020, 36 (3): 9

[6] Tian Sheng, Long Anyang Point cloud classification method based on graph convolution and multi-layer feature fusion [J] Advances in Laser and Optoelectronics, 2023, 60 (14)

[7] Du Peng, Ding Shifei DGA domain name detection method based on mixed word vector depth learning model [J] Computer Research and Development, 2020

[8] Ji Ziheng, Wang Bin Research progress of sketch retrieval methods based on depth learning [J] Computer Engineering and Science, 2021, 43 (12): 16

[9] Yan Yan, Cong Yiming, Adnan Mahmood, et al Statistical release and privacy protection method of location big data based on deep learning [J] Journal of Communication, 2022, 43 (1): 203-216

[10] Song Yong, Hou Bingnan, Cai Zhiping Network intrusion detection method based on deep learning feature extraction [J] Journal of Huazhong University of Science and Technology: Natural Science Edition, 2021, 49 (2): 6

[11] Wang Hui, Ouyang Xiu, Liu Qinghua, et al Structural Feature Detection Method of Ground Penetrating Radar 2D Profile Image Based on Depth Learning [J] Journal of Electronics and Information, 2022, 44 (4): 11

[12] Zhang Shengwen, Zhou Xi, Li Bincheng, et al Extraction method of part machining feature information based on image depth learning [J] China Mechanical Engineering, 2022, 33 (03): 348-355

[13] Wei Yulong, Lv Penghui, Lin Tao, et al Temperature and humidity control method of painting air conditioner based on big data in-depth learning [J] Electroplating and coating, 2020, 39 (6): 5

[14] Yan Xingyu, Gu Hanming, Luo Hongmei, Yan Youping Intelligent identification of seismic facies based on improved depth learning method [J] Petroleum Geophysical Exploration, 2020 (6): 10

[15] Jiang Chen, Wang Yuan, Hu Junhua, et al Power entity information recognition method based on deep learning [J] Power grid technology, 2021, 45 (6): 9

[16] Li Hongnan, Zhang Wensheng, Fu Xing Wind resistance vulnerability assessment of transmission tower structure based on big data in-depth learning [J] Journal of Civil Engineering, 2022 (009): 055

Application of accident tree method and Analytic Hierarchy Process in emergency vehicle traffic analysis

Zhaona Lu [1], Yan Wang [1*], Chenyu Hu [1], Tong Xu [1], Chen Chen [1], Kejia Ma [1]
(1. School of Automotive Engineering, Nantong Institute of Technology, Nantong 226000, China)
* Wang Yan: wangfetter@163.com

ABSTRACT

In order to determine the factors affecting emergency vehicle traffic and find out the mitigating measures, the accident tree method was used to qualitatively analyze the four main factors and 20 sub-factors affecting emergency vehicle traffic. Combined with analytic hierarchy process for quantitative analysis, using MATLAB to determine the weight of factors, the aim is to reduce the impact of subjective factors. The results show that among the four main factors, the weight value of traffic factor and road factor is as high as 0.834, which is the primary focus to determine the solution measures. Among the 19 sub-factors, the comprehensive weight sum of traffic accidents caused by intersection signal priority equipment, social vehicles not allowed, few lanes, complex terrain and emergency vehicles passing through the intersection is 0.543, which is more than half of the total weight. Therefore, it is necessary to solve the problem targeted.

Key words: emergency vehicle passage; Accident tree method; Analytic hierarchy Process (AHP); MATLAB

1. INTRODUCTION

With the continuous expansion of urban scale and the continuous growth of car ownership, the accident rate is also increasing. In order to solve the emergency, emergency vehicles need to arrive at the scene of the accident in the shortest time. However, as a part of the traffic flow, the emergency vehicle is affected by external factors, so it is often difficult to reach the scene at the fastest speed to solve the accident, resulting in more serious loss of life and property. According to the statistics, in the rush hour, the blocked rate of fire trucks is as high as 95%, and the road is less than one kilometer needs dozens of minutes. The factors affecting the traffic of emergency vehicles need to be analyzed and improved in the future work.

At present, the research on emergency vehicle traffic mainly focuses on the analysis and summary of the factors affecting emergency vehicle traffic and puts forward the corresponding solutions. Bi Xudong[1-2]Et al. proposed to establish a route selection model for emergency vehicles and to use the improved Dijkstra algorithm to solve it, so as to select the optimal path and reduce the passage time of emergency vehicles Fang Lei, He Jianmin[3]The site The site selection planning of emergency system with comprehensive analytic hierarchy process (AHP) and objective planning method was proposed by others, aiming to achieve reasonable coverage of emergency services and reduce travel distance. Wang Gaofei, Liu Song[4-5] used The analytic hierarchy process (AHP) was used to evaluate and study urban road capacity, from which road factors affecting emergency vehicle traffic could be analyzed. Li Xiaowei[6-7] Through the quantitative analysis of urban road traffic efficiency and its influencing factors, the authors summarized the traffic factors affecting emergency vehicles. In this paper, through literature reference, case analysis and expert research, the accident tree method is adopted to make a qualitative analysis of the factors affecting the emergency vehicle passage, and the analytic hierarchy process is adopted to determine the main factors affecting the emergency vehicle passage, and the corresponding solutions are found.

2. ACCIDENT TREE METHOD AND ITS APPLICATION

Through the social survey of 200 drivers of social vehicles, non-motor vehicles and pedestrians, the main reasons for the delay of emergency vehicles are summarized statistically as shown in the following table.

International Conference on Computer Graphics, Artificial Intelligence, and Data Processing (ICCAID 2022)
edited by Ruishi Liang, Jing Wang, Proc. of SPIE Vol. 12604, 126041A
© 2023 SPIE · 0277-786X · doi: 10.1117/12.2674555

Table 1 Emergency vehicle delay cause statistics

Cause of delay	The number of	Proportion (%)	Cause of delay	The number of	Proportion (%)
Heavy traffic	51	25.5%	The emergency vehicle violated the traffic signal and had an accident	22	11%
Severe weather impact	11	5.5%	Non-motor vehicles occupy motor lanes	32	16%
Social vehicles don't move	49	24.5%	Social vehicles are illegally driven	35	17.5%

The fault tree analysis method originated from fault tree analysis, which uses logical reasoning to identify and evaluate the risk of various systems. It can not only analyze the direct causes of accidents, but also reveal the potential causes of accidents in depth. Based on the relevant literature and the above actual investigation, the factors affecting the passage of emergency vehicles were analyzed, and the accident tree was drawn from the four main factors and 19 sub-factors, as shown in Figure 1.

Taking the delay of emergency vehicle rescue as the final event, the delay of emergencyvehicle has the germination stage, the occurrence stage of delay inducement and the development stage of delay. The delayis mainly caused by the low level of drivers, the risk of human factors, terrain and climate restrictions, traffic congestion and backward equipment. The embryonic stage of potential risk of delay mainly refers to the influence of human factors, such as unclear or wrong rescue position provided by the rescued, unauthorized change of waiting place, driver not familiar with the terrain, pedestrians affecting traffic travel and other delays in the rescue process, resulting in vehicle delay. The main causes of delay are terrain restrictions, climate conditions and traffic environment restrictions. The reason for terrain restrictions is that complex mountain sections and rural roads put forward higher requirements on drivers' driving ability. Weather conditions are limited because rain, snow, fog and other bad weather can obscure drivers' vision and make roads slippery, preventing them from improving speed. Traffic environmental restrictions include large number of route intersections, small number of lanes, uneven road surface, road construction, road congestion and traffic control. The development stage of the delay includes the insufficient emergency vehicles can not be dispatched for rescue, the accident can not be located in time on the way, and the social vehicles do not give way. The delay caused by the occupation of emergency lanes and congestion, and the lack of complete and advanced traffic equipment such as traffic signal priority equipment to ensure the priority of emergency vehicles.

Table 2 Symbols of each node in the accident tree

symbol	meaning	symbol	meaning
T	Emergency vehicle delay	X7	There was an accident at the intersection with an emergency vehicle
A1	Potential risk of delay (bud)	X8	Social vehicles don't move
A2	Delay inducement effect	X9	Social vehicles occupy emergency lanes
A3	Occurrence of delay (development)	X10	The number of route intersections is large
B1	The driver drove poorly	X11	Few lanes
B2	Topographic climate limitation	X12	Uneven road surface
B3	Traffic congestion	X13	Construction of road
B4	Backward equipment of all kinds	X14	Lack of intersection signal priority equipment
X1	The rescue position is not clear or wrong	X15	Section road congestion
X2	The rescued changed the waiting place without permission	X16	Road section Traffic control
X3	The driver was unfamiliar with the terrain	X17	Rain, snow and fog inclement weather
X4	Non-motor vehicles and pedestrians affect vehicle traffic	X18	Complex terrain
X5	The emergency vehicle broke down	X19	The roads are not well lit at night
X6	Emergency vehicles are not adequately equipped		

Proc. of SPIE Vol. 12604 126041A-327

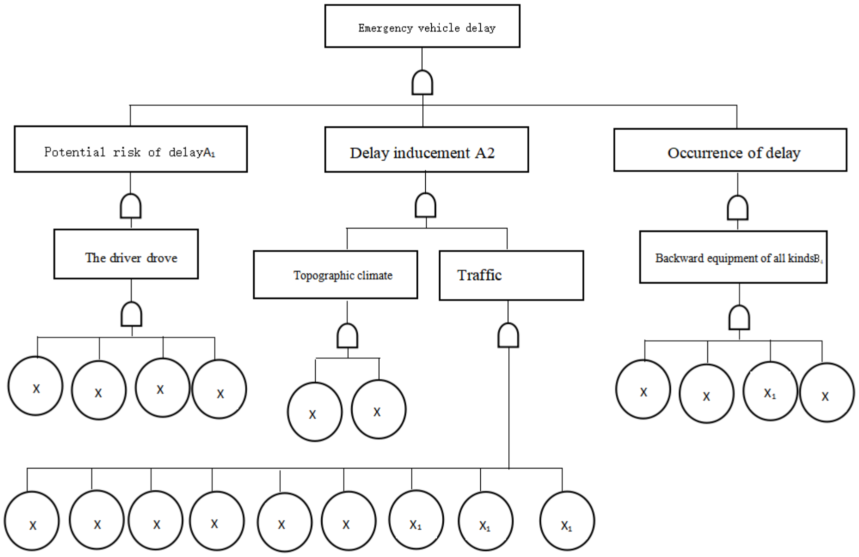


FIG. 1 Emergency vehicle delay time delay accident tree

3. ANALYTIC HIERARCHY PROCESS

3.1 Establishment of hierarchical structure model

Analytic Hierarchy Process (AHP) is a decision making method that decomposes elements related to decision into objective, criterion, scheme and other levels, and analyzes them on this basis.

Create an indicator system based on the accident tree. With the emergency vehicle traffic delay as the target layer and the human factor, vehicle factor, road factor and environmental factor as the criterion layer, the factors with high correlation degree at the bottom of the accident tree are integrated to obtain each element of the index layer, as shown in Table 3.

Table 3 Target layer, criterion layer and index layer elements of the index system

Target layer	Criterion layer element	Index layer elements
Emergency vehicle travel time delay T	Human Factors A1	Rescue position is not clear or wrong B11 The rescued changed the waiting place without permission Driver is not familiar with terrain B13 Non-motor vehicles and pedestrians affect vehicle passage B14
	Vehicle factor A2	Emergency vehicle has broken down B21 Emergency vehicles are underequipped B22 Emergency vehicles passing through the intersection of traffic accident B23 Social vehicles do not give way to B24 Social vehicles occupy emergency lane B25
	Road factor A3	The number of route intersections is more B31 Fewer lanes B32 Uneven road surface B33 Road construction B34 Lack of intersection signal priority equipment B35 Section road congestion B36 Section Traffic control B37
	Environmental factors A4	Rain, snow and fog inclement weather B41 The terrain is difficult Lack of light on the road B43 at night

3.2 Calculation of single rank of emergency vehicles' passing time delay hierarchy

The value of each element in the evaluation matrix reflects the understanding of the relative importance of the two elements of the same level and the upper elements, which will have a great impact on the decision outcome. In the specific scale judgment, it needs to rely on professional knowledge, which can be determined by a certain number of experts or through directional survey of relevant people. The judgment generally adopts 1-9 and reciprocal scale, 1 represents equally important, 3 represents slightly important, 5 represents significantly important, 7 represents strongly important, and 9 represents absolutely important. Among them, 2, 4, 6 and 8 are the intermediate values of the two adjacent judgments. By comparing the four factors of the middle layer in pairs, the judgment matrix of the middle layer is constructed by combining the 1-9 scale and the reciprocal scale. As shown in Table 3, is the maximum eigenvalue of the matrix, and CI is the consistency index. When CI=0, it is the complete consistency; when CI is close to 0, it is the satisfactory consistency; when CI is larger, the consistency is worse. λ_{max} RI is the average random consistency index, and the larger the matrix order is, the possibility of random deviation of consistency will also increase.CR is the random consistency ratio. When CR<0.1, the judgment matrix passes the consistency test.

Table 4 Intermediate layer judgment matrix and weight results

T	A1	A2	A3	A4	The weight	One time inspection
A1	1	0.2	1/7	1	0.075	$\lambda_{max} = 4.083$ CI = 0.028 CR = 0.031 < 0.1 Satisfy consistency
A2	5	1	1/3	3	0.269	
A3	7	3	1	5	0.565	
A4	1	1/3	0.2	1	0.091	

Through the analysis of the data in Table 4, it can be seen that road factors are the main factors affecting the passage of emergency vehicles, followed by vehicle factors, and the weight value of the two is as high as 0.834. It is necessary to strengthen the reasonable improvement of traffic environment management. Prevent emergency vehicles from being unable to travel fast on the road due to traffic congestion during the rescue process, resulting in delays and failure to arrive at the scene in time for rescue. The reasonable improvement of traffic management can also reduce the delay of social vehicles and the occurrence of traffic accidents. Road construction should also be strengthened to avoid accidents caused by backward traffic and road facilities. Strengthen the ideological construction of motor vehicle drivers, non-motor vehicles and pedestrians, abide by laws and regulations, consciously maintain traffic order, avoid emergency situations, and open up a "road of life" for the rescued. At the same time, attention is paid to the level training of emergency vehicle drivers, which can cope with various poor terrain and climate conditions, and strive for the "golden rescue time".

3.3 Calculation of the total rank of emergency vehicles' passing time delay hierarchy

The calculation of comprehensive weight includes: The first step, using Matlab to calculate the single-layer weight from the bottom layer (B layer) to the middle layer (A layer).Second, multiply it with the weight of the middle layer (layer A) over the top layer (layer T).Finally, the resulting value is the comprehensive weight value and passes the consistency test. The calculation process of total hierarchical sorting is shown in Table 5.

Table 5 Bottom judgment matrix and comprehensive weight results

A1	B11	B12	B13	B14	The weight	Combined weight of weight	One time inspection
B11	1	1/3	0.2	0.143	1/3	0.048	$\lambda_{\max} = 4.074$ CI = 0.025 CR = 0.028 < 0.1 Satisfy consistency
B12	3	1	1/3	0.2	1/3	0.090	
B13	5	3	1	0.2	1	0.184	
B14	7	5	5	1	3	0.500	

A2	B21	B22	B23	B24	B25	The weight	Combined weight of weight	One time inspection
B21	1	1/3	0.2	0.143	1/3	0.048	0.013	$\lambda_{\max} = 5.245$ CI = 0.061 CR = 0.055 < 0.1 Satisfy consistency
B22	3	1	1/3	0.2	1/3	0.090	0.024	
B23	5	3	1	0.2	1	0.184	0.050	
B24	7	5	5	1	3	0.500	0.135	
B25	3	3	1	1/3	1	0.178	0.048	

A3	B31	B32	B33	B34	B35	B36	B37	The weight	Combined weight of weight	One time inspection
B31	1	1/3	3	1	1/7	0.2	3	0.080	0.045	$\lambda_{\max} = 7.598$ CI = 0.1 CR = 0.073 < 0.1 Satisfy consistency
B32	3	1	1	3	0.2	1/3	5	0.115	0.065	
B33	1/3	1	1	3	0.2	1/3	3	0.084	0.047	
B34	1	1/3	1/3	1	1/7	0.2	1	0.043	0.024	
B35	7	5	5	7	1	3	9	0.416	0.235	
B36	5	3	3	5	1/3	1	7	0.230	0.130	
B37	1/3	0.2	1/3	1	1/9	1/7	1	0.032	0.018	

A4	B41	B42	B43	The weight	Combined weight of weight	One time inspection
B41	1	0.2	1/3	0.106	0.010	$\lambda_{\max} = 3.039$ CI = 0.019 CR = 0.037 < 0.1 Satisfy consistency
B42	5	1	3	0.633	0.058	
B43	3	1/3	1	0.261	0.024	

Through the analysis of the data in the above table, it can be seen that the comprehensive weight of B36 (Traffic congestion), B35 (lack of intersection signal priority equipment) and B24 (social vehicles do not allow traffic) exceeds 0.1, which is a factor that needs to be focused on. The main reason for the blocked passage of emergency vehicles is that there are too many social vehicles in front of them and they do not avoid the emergency vehicles, which leads to the emergency vehicles being stuck on the road. The long waiting time for the signal light and the lack of complete signal control system to ensure the priority passage of emergency vehicles force the passage time to be extended. The timing of signal lights should be adjusted reasonably and emergency vehicles should be given the right of way. Because of the large number of people around the road, small vendors privately occupy the road, resulting in emergency vehicles have to slow down. At the same time, if the intersection with complex road conditions and large traffic flow is not properly managed, it will also cause serious congestion and affect the passage of emergency vehicles. Therefore, strong supervision and guidance should be carried out for every intersection with a large traffic flow, and timely rectification should be made for the phenomenon of road streetization. Otherwise, it will not only affect the traffic of vehicles, but also produce great safety risks. Drivers of social vehicles should also have the awareness and action to avoid emergency vehicles and consider the lives of others. The weights of B14(non-motor vehicles and pedestrians affect the passage of vehicles), B25(emergency vehicles pass through the intersection of traffic accidents), B25(social vehicles occupy the emergency road), B31(a large number of route intersections), B33(uneven road surface) and B42(complex terrain) are about 0.04~0.06.The news about social vehicles occupying the emergency passage, causing emergency vehicles to be unable to reach the destination quickly and resulting in tragedy is frequently broadcast. People should have a sense of responsibility, resolutely maintain the life passage of emergency vehicles, for the good of themselves and others. Drivers of emergency vehicles should also have regular physical examinations to ensure their health and prevent discomfort during travel that may endanger their lives and the lives of others. The weight values of B22 (insufficient emergency vehicle), B34 (road construction) and B43 (insufficient road light at night) are between 0.02 and 0.04, among which there are road hardware facilities involved, so it is necessary to improve the quality of the road surface and upgrade the road. On motor vehicles and non-motor vehicles and pedestrians illegal behavior should also be strictly punished, emphasize the importance of traffic safety, so that people form awareness. The emergency vehicles themselves should also be inspected and maintained regularly.

4. CONCLUSION

Through the accident tree method, the paper qualitatively analyzes the 4 main factors and 20 sub-factors affecting the emergency vehicle traffic, which is intuitive and clear. By importing the accident tree to form a hierarchical structure model for quantitative analysis, it is concluded that traffic factors and road factors are the main factors affecting the passage time of emergency vehicles by using MATLAB calculation. Among the sub-factors, there is a lack of intersection signal priority equipment, social vehicles are not allowed to allow, and the number of lanes is small. The sum of comprehensive weights of traffic accidents caused by complex terrain and emergency vehicles passing through intersections is 0.543. Therefore, corresponding solutions should be developed to reduce the impact of these factors. In this paper, the accident tree method and analytic hierarchy process are used in the emergency vehicle traffic analysis. Firstly, the influencing factors are determined from a qualitative perspective, and then the weight of the influencing factors is calculated from a quantitative perspective. This method is intuitive and scientific, and is worth popularizing and applying in other fields.

REFERENCES

[1] Bi Xudong. Research on Key Issues of Urban Emergency Vehicle Priority [D]. Southwest Jiaotong University,2014

[2] Zhao Jianyou, Xiao Yu, ZHU Xinyuan, Zhao Yang. Emergency Vehicle Routing Optimization Method considering Demand Urgency [J]. Journal of Harbin Institute of Technology,2022,54(09):27-34.

[3] Fang Lei, He Jianmin. Location Planning model of emergency system based on AHP and objective planning [J]. Systems Engineering Theory and Practice, 2003, (12):116-120.

[4] [Wang Gaofei, Liu Song. Research on Urban road capacity evaluation based on Analytic Hierarchy Process [J]. Henan Science and Technology,20,39(31):128-130.

[5] Yang Zheming, Liu Junmeng, Zhao Cong, Wang Ling. Impact of lane occupation on urban road capacity [J]. Journal of North China University of Science and Technology (Natural Science Edition),2017,39(01):75-79.

[6] Li Xiaowei. Quantitative Analysis of Urban road traffic efficiency and its Influencing factors. Beijing Jiaotong University,2012.

[7] Study on the influence of lane occupation on urban road capacity based on mathematical model [J]. Public Standardization,2020(18):53-55.

Prediction of High-speed Train Wheel-rail Relationship based on Maximum Information Coefficient and LSTM

Yulong Cui[a], Wei Dong[a], Yanhong Shi[a], Yanghui Niu[b]*, Shuwei Shen[b]
[a]CRRC Qingdao sifang, Qingdao, China
[b]Southwest Jiaotong University, Chengdu, China
*Corresponding author: niuyanghui2019@163.com

ABSTRACT

Aiming at the prediction of wheel-rail relationship of high-speed trains, a prediction method combining maximum information coefficient (MIC) and LSTM was proposed. Firstly, the dynamic model of high-speed train was established by SIMPACK and the original data set was obtained. Secondly, the maximum information coefficient is used to preprocess the data. Then, a prediction model of wheel-rail relationship based on long short-term memory neural network was built. Adam optimizer was used to optimize the learning rate and network structure. Finally, the optimized long short-term memory neural network is used to predict the wheel-rail relationship. The prediction results of wheel-rail relationship show that when the number of model iterations is 45, and the number of hidden layers is 110, the average absolute error percentage of model prediction is the smallest, and the value is 0.0247. Under these conditions, the predicted result is very close to the real value, that is, the data pretreated by the maximum information coefficient can make the model accurately predict the change trend of wheel-rail relationship, which can provide support for further research.

Keywords: maximum information coefficient, long short-term memory neural network, wheel-rail relationship, dynamic model

1. INTRODUCTION

Rail transit has become the first choice for people to travel because of its green, safe, convenient and comfortable characteristics. The wheel-rail relationship is an inevitable basic problem in the railway field, which not only directly affects the safety of rail vehicle operation and passenger ride comfort, but also has an important impact on operating costs. The research on the wheel-rail relationship is not only related to the application technology, but also involves the basic theoretical issues, which is an important support to ensure the safety, efficient operation and technological innovation of high-speed railway. Therefore, more and more attention has been paid to the research of wheel-rail relationship.

In recent years, there have been many studies on rail personalized grinding, rail wear evolution, wheel-rail contact relationship, wheel-rail contact force and so on, which laid a foundation for further exploring the coupling relationship of wheel-rail interaction. Taking the equivalent conicity as the evaluation index and combining with the line profile detection, Zeng [1] analyzed the relationship between the equivalent conicity and the fault phenomena such as car shaking and shaking. Huang [2] explored the influence of abnormal wear of wheel tread on wheel-rail contact behavior and dynamic behavior by means of dynamic simulation. Based on the principle of trace method, Xu [3] analyzed the wheel-rail contact geometry and mechanical properties under different gauge and rail cant parameters. The above research qualitatively measures the wheel-rail contact relationship from the perspective of principle and evaluation index, and lacks quantitative description from the perspective of data.

In addition, some scholars quantitatively evaluate the wheel-rail relationship from the perspective of wheel-rail force. Ahmed [4] studied the dynamic interaction between wheel and rail in the presence of unsupported sleepers by an effective numerical method based on the central finite difference theory. Magdy [5] used ANN technology to model the wheel-rail vertical force of simplified standard urban rail train, but it has certain requirements for simplified degrees of freedom, and there is a problem of slow model response; Pang [6] used the improved large artificial neural network to predict the wheel-rail force. The above research lacks the screening of model input, and there is a large redundancy in the input, which leads to low accuracy of subsequent prediction, and relies on a large number of experimental data, which brings certain challenges to the acquisition of data and the inspection and elimination of abnormal data.

In order to solve the problem that the actual wheel-rail force measurement is not easy, the current prediction technology only combines the traditional neural network prediction accuracy is limited, the dynamic model calculation speed is slow,

International Conference on Computer Graphics, Artificial Intelligence, and Data Processing (ICCAID 2022)
edited by Ruishi Liang, Jing Wang, Proc. of SPIE Vol. 12604, 126041B
© 2023 SPIE · 0277-786X · doi: 10.1117/12.2674513

and it is not easy to spread on site. This paper proposes a method of train wheel-rail relationship prediction based on the combination of maximum information coefficient and LSTM. Firstly, according to the dynamic model, the input parameters of working conditions and the output parameters of wheel-rail force under different working conditions are obtained as the data set of model training. The normalized and maximum information coefficient algorithm is used to extract the characteristic parameters highly correlated with the predicted target, and the effective data set is obtained. For the effective data set, the LSTM algorithm is used to construct the wheel-rail force prediction model, and the high-precision prediction of wheel-rail force is realized.

2. DYNAMIC MODEL OF HIGH-SPEED TRAIN

The high-speed vehicle is modeled as a four-axis multi-rigid body system, including a car body, two bogie frames, four wheelsets and eight axle boxes. Since the high-frequency vibration response caused by the elasticity of each component is not the focus of this paper, they are all regarded as rigid bodies and the elastic deformation of each component is ignored. Each bogie frame is supported on two wheelsets by an axle box, a coil spring system, and a vertical shock absorber. The car body is supported on the two bogie frames by the air spring system, two vertical shock absorbers, two lateral shock absorbers, traction bars and anti-roll bars. The vehicle system modeling has a total of 50 degrees of freedom: the body, bogie and wheelset have six degrees of freedom in vertical, transverse, longitudinal, nod, shake and roll directions respectively. The axle box body has only one nodding degree of freedom. See Table 1 for specific degrees of freedom settings.

Table 1. Freedom of vehicle dynamics model

Part name	Longitudinal	Lateral	Vertical	Nod	Shake	Roll
Car body	x_c	y_c	z_c	β_c	ψ_c	ϕ_c
Bogie frame	x_t	y_t	z_t	β_t	ψ_t	ϕ_t
Wheelset	x_w	y_w	z_w	β_w	ψ_w	ϕ_w
Axle box	—	—	—	β_a	—	—

Considering that the track structure has a great influence on wheelset0 excitation, the current representative dynamic models of ballastless track include long-pillow embedded ballastless track model, elastic supporting block ballastless track model and layout ballastless track model. In this paper, the long-pillow buried ballastless track model is selected as the research object, and the rail vibration differential equation is established based on the Timoshenko beam model.

The wheel-rail spatial contact geometry is solved by trace method, the wheel-rail normal strive is solved by Hertz nonlinear elastic contact theory, and the wheel-rail creep force is solved by Kaller linear theory, and then nonlinear correction is carried out.

3. MIC-LSTM PREDICTION ALGORITHM

3.1 Maximum information coefficient

Maximum information coefficient (MIC) is a method proposed by Reshef [7-9] to quantitatively measure the degree of linear or nonlinear correlation between two variables. Its core idea is to partition the scatter graph composed of two variables by using the grid to realize data partition blocks and encapsulate the relationship between variables, so as to discover the deep relationship between variables. Compared with other traditional statistical correlation coefficients, MIC has two obvious advantages: universality and uniformity. Universality means that it can capture all kinds of functional relations, not limited to some specific functional relations. Uniformity means that for different relationship types, when the same noise level is added, the values of the maximum information coefficients between the calculated variables are similar. For a given ordered data set, input variable X and output variable y are divided a times on X-axis and b times on Y-axis, then mutual information $I(D,a,b)$ can be expressed as:

$$I(D,a,b)=\sum_{a,b}p(x,y)log(\frac{p(x,y)}{p(x)p(y)}) \tag{1}$$

where $p(x)$ and $p(y)$ are the independent probability distribution, $p(x, y)$ the joint probability density function, and approximate estimation by the ratio of the points in the grid to the total points.

By changing the position of partition interval points, different mutual information values can be obtained, and the maximum mutual information value is assumed to be, which can be used for comparison of different dimensions through normalization processing of mutual information values, then the maximum information coefficient is defined as:

$$MIC(D) = \max_{a*b<n^{0.6}}\{\frac{maxI(D, a, b)}{log(min\{a, b\})}\} \tag{2}$$

where n is the length of data, MIC values between 0 and 1.

3.2 Long short-term memory neural network

In 1997, Hochreater and Schmidhuber [10-12] proposed the network structure of LSTM. LSTM model is a special RNN cycle structure, which can solve the problems of gradient explosion and gradient disappearance encountered in RNN model, so it is widely used in the fields related to time series data. The inherent characteristics of LSTM make it an ideal choice for anomaly detection tasks involving time series and nonlinear data flow [13]. The main characteristic of LSTM is to control the information state of every moment in the neural network through "gate" structure.

The internal data operations of the unit model of LSTM are shown in Figure 1.

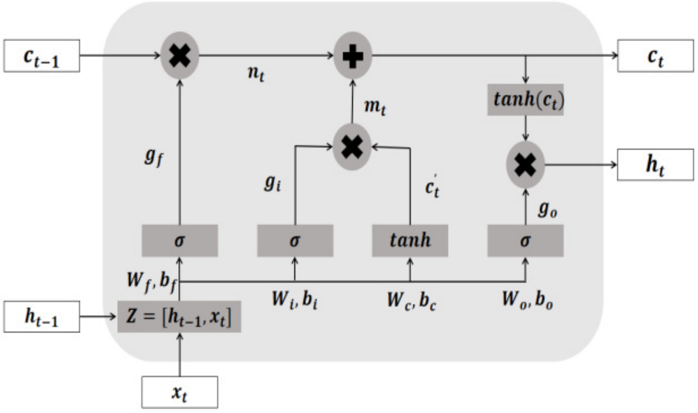


Figure 1. Unit structure diagram of LSTM

Based on the conventional recurrent neural network, LSTM introduces forgetting gate, input gate and output gate units to control the state iteration of LSTM units, and controls the memory information, the acquisition of input information and the transmission of output information through the gate coefficient to improve the efficiency and stability of the classification network.

The difference between LSTM and general recurrent neural network mainly lies in the calculation of hidden state of recursive network. The LSTM unit state at time t is determined by two parts: discarding useless information of the unit and reserving some useful information. The state of hidden layer at this time is obtained by the useful information selected by the output gate.

$$h_t = o_t \tanh(C_t) \tag{3}$$

$$C_t = f_t C_{t-1} + i_t \widetilde{C_t} \tag{4}$$

where f_t、i_t、o_t is the gating coefficient of forgetting gate, input gate and output gate, $\widetilde{C_t}$ the update vector at the time of input unit state t.

The calculation formulas are respectively:

$$f_t = \sigma\left(W_f \times [h_{t-1}, x_t] + b_f\right) \tag{5}$$

$$i_t = \sigma(W_i \times [h_{t-1}, x_t] + b_i) \tag{6}$$

Proc. of SPIE Vol. 12604 126041B-364

$$o_t = \sigma(W_o \times [h_{t-1}, x_t] + b_o) \tag{7}$$

$$\widetilde{C_t} = tanh(W_c \times [h_{t-1}, x_t] + b_c) \tag{8}$$

where W_f、W_i、W_o is the weight matrix of each gated node, W_c the memory weights, b_f、b_i、b_o the bias of each gated node, b_c the memory cell bias, σ the Sigmod function.

LSTM realizes its special long short-term data memory function through the three gates of input gate, forgetting gate and output gate, making the model more suitable for learning the characteristics of related and continuous data [14].

3.3 MIC-LSTM prediction method

For the prediction of wheel-rail relationship of high-speed trains, the implementation process of the network prediction model is shown in Figure 2. The realization of MIC-LSTM network prediction model can be divided into three stages: the first stage is data preprocessing. Through dynamic model simulation, the original data set is obtained, which is normalized, and its time-domain characteristics are extracted from the processed sample data. The second stage is the screening of input variables, which mainly includes the selection of parameters that have great influence on the results by using the screening method of maximum information coefficient. The third stage is the construction of prediction model. The LSTM network parameters were initialized, and the parameters were optimized by combining the evaluation indexes to construct the LSTM network prediction model.

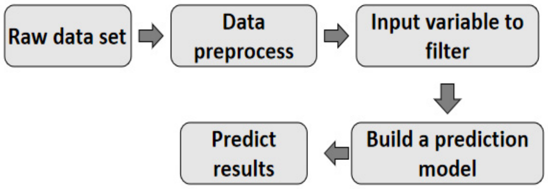


Figure 2. Frame diagram of MIC-LSTM prediction model

The average absolute error percentage (MAPE) was taken as the evaluation index of the prediction model [15] to evaluate the closeness between the predicted value and the real value.

$$MAPE = \frac{1}{N}\sum_{i=1}^{N}\left|\frac{y_i - \widetilde{y_i}}{y_i}\right| \tag{9}$$

Where y_i is real value, $\widetilde{y}_1$ the predictive value.

4. ALGORITHM VALIDATION

In this paper, input parameters and output parameters of high-speed train dynamics model data are selected for training. Partial output parameters of the dynamic model are shown in Table 2.

Table 2. Partial output parameters of dynamic model

Vertical vibration acceleration(g)	Lateral vibration acceleration(g)
-0.0214037	-0.0687349
-0.0289523	-0.144862
-0.0341701	-0.154441
-0.0379032	-0.0918801
-0.0410805	0.0195403
…	…

Taking the working condition parameters of the dynamic model as the original data, the sections are divided according to the interval of 5% of the sampling frequency, and the time-domain indexes in each section are extracted, mainly including parameters such as peak value, variance, mean value, kurtosis, skewness, waveform factor and pulse factor. For the problem that the dimensions of many kinds of parameters are not unified, the normalization [16] method is used to facilitate the training of subsequent models. The normalization method adopts formula (10). Combined with the prediction and alarm ability of the model for wheel-rail force, 98.5% of the maximum value in the section is taken as the target output for the corresponding output wheel rail force, which can improve the alarm ability of the model.

$$x_i = \frac{x - x_{min}}{x_{max} - x_{min}} \tag{10}$$

where：x_i is the normalized value, x、x_{max}、$x_{min}$are the maximum value and the minimum value respectively.

By calculating the *MIC* values between the time-domain characteristics and the output wheel rail force, the results are shown in Table3. Take the parameters with large MIC value and control the number of parameters to within 10. Determine the optimal number of input parameters with the evaluation index of the final model to obtain an effective data set with low redundancy and strong coupling relationship, which not only overcomes the problem of high latitude and nonlinearity between state parameters, but also ensures that the effective information will not be discarded.

Table 3. Time domain characteristics of vertical acceleration and MIC value of output force

Time Domain Features	Wheel-rail vertical force	Wheel-rail lateral force
Maximum	0.7287	0.5174
Minimum	0.8884	0.5250
Average	0.6019	0.4213
Peak-Peak	0.7401	0.5142
Effective value	0.7430	0.5162
…	…	……

Taking the vertical force related data set as an example, the calculation results of transverse acceleration and vertical acceleration are compared longitudinally. The effective data set of vertical force can be obtained by taking out the data with the largest correlation with vertical force and the largest mic value. The processed data is divided into two parts: training set and test set. The training set is used to input the training model in LSTM model, so as to optimize the network structure and internal parameters; The test set is used to verify the performance of the optimized model.

Set the learning rate of LSTM network as 0.005, the maximum number of iterations as 50, and the optimizer as Adam optimizer. In order to study the influence of LSTM network parameters on prediction accuracy, the number of iterations and the number of hidden network layers are selected for testing, and the average absolute error percentage (MAPE) is used as the evaluation index. Take the single-layer LSTM network as the training model, change the number of iterations, and analyze the impact of different iterations on the prediction accuracy of the model, as shown in Table 4.

Table 4. Comparison of evaluation indexes of different iterations

Index	Number of iterations			
	35	40	45	50
MAPE	0.0266	0.0263	0.0247	0.0250

It can be seen from table 4 that as the number of iterations increases, the value of the evaluation index MAPE first decreases and then increases. When the number of iterations is 45, the prediction accuracy of single-layer LSTM network training is better.

The number of fixed iterations is 45, and the number of hidden layers of LSTM network is gradually increased. The evaluation indicators are shown in Table 5.

Proc. of SPIE Vol. 12604 126041E-56

Table 5. Comparison of evaluation indexes of different hidden layers

Index	Hidden layers			
	90	100	110	120
MAPE	0.0260	0.0252	0.0249	0.0252

It can be seen from table 5 that when the number of hidden layers is 110, the value of MAPE is small. Therefore, the number of hidden layers of LSTM network is 110 and the number of iterations is 45.

Taking the prediction of vertical force as the goal, the relevant data sets are screened, and the calculation results of transverse acceleration and vertical acceleration are compared longitudinally. The effective data sets of vertical force can be obtained by taking out the first six data items with the greatest correlation with vertical force and the largest mic value, which are respectively: vertical acceleration index: minimum value and rectified average value; Lateral acceleration index: peak-peak value, peak value, variance and standard deviation. A total of 800 sample points in the parameters are taken out, of which the first 80% are divided into training sets and the last 20% are divided into test sets; Input the above optimized parameters into the MIC-LSTM prediction model to predict the wheel rail force, and the prediction results and relative prediction errors can be obtained, as shown in figure 3 and figure 4.

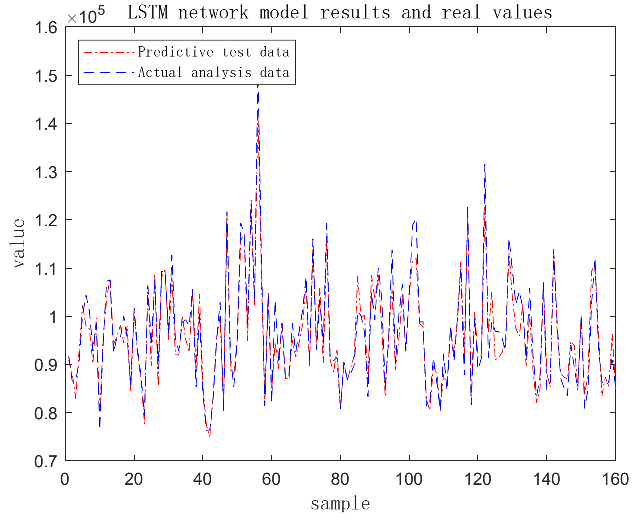


Figure 3. Comparison of predicted results and real values

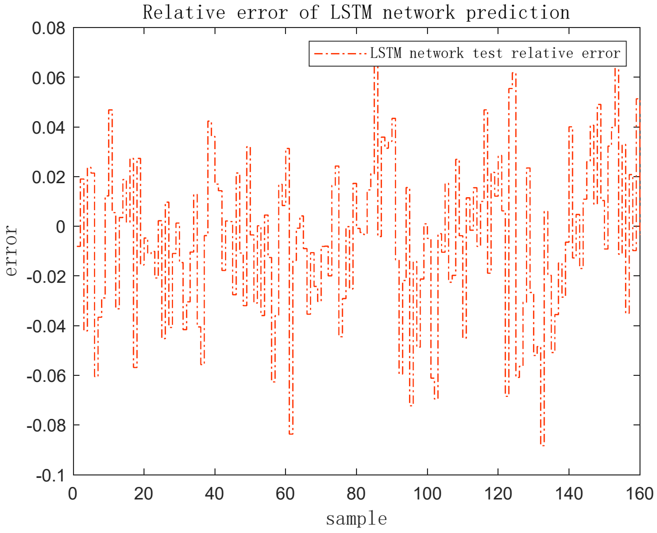


Figure 4. The relative error between predicted value and true value

5. CONCLUSION

This paper applies the maximum information coefficient to the screening of input variables, removes the state parameters that have little correlation with the target variables, uses the LSTM network prediction model to predict the target variables at the future time, and combined with the evaluation index, it is concluded that the input variable gauge pretreated by the maximum information coefficient has better prediction results. The model improves the prediction accuracy to a certain extent. This method can be further applied to the prediction of wheel rail relationship of high-speed train, predict the state of wheel rail relationship in the future, and provide data closer to the real value for the early fault early warning of wheel rail relationship of high-speed train.

REFERENCES

[1] ZENG J, GAN F, LUO G B. Wheel-rail relationship detection and equivalent taper management of rail vehicles [J]. Modern Urban Transit, 2021(6):29-34 (in Chinese)

[2] HUANG Y B, ZHONG H, WANG W J, et al. Analysis of the influence of abnormal wear of wheel tread on wheel-rail relationship [J]. Journal of Sichuan University(Engineering Science Edition), 2014,46(S1):198-202 (in Chinese)

[3] Xu J M, Zheng Z G, Lai J, et al. Influence of track parameters on wheel-rail contact behavior of high-speed turnout [J]. Journal of Southwest JiaoTong University, 2022,57(05):990-999 (in Chinese)

[4] JIAN J Z, AKW Ahmed, Subhash Rakheja & Amir Khajepour (2010) Development of a vehicle-track model assembly and numeric method for simulation of wheel-rail dynamic interaction due to unsupported sleepers, Vehicle System Dynamics, 48:12, 1535-1552, DOI: 10.1080/00423110903540751.

[5] El-Sibaie M. Computer Model Developed to Predict Rail Passenger Car Response to Track Geometry [R].US: Department of Transportation, Federal Railroad Administration, 2000.

[6] PANG X M. Wheel-rail force prediction based on artificial neural network[D]. Journal of Nanjing University of Science and Technology, 2012 (in Chinese)

[7] WEN T, DONG D, CHEN Q, et al. Maximal Information Coefficient-Based Two-Stage Feature Selection Method for Railway Condition Monitoring[J]. Transactions on Intelligent Transportation Systems, 2019,20(7):2681-2690

[8] ZHAO L, GONG J X, HUANG D R, et al. Fault Feature Selection Method of Gearbox Based on Fisher Score and Maximum Information Coefficient[J]. Control and Decision ,2021,36(9):2234-2240 (in Chinese)

[9] LIU G Q, WANG X Q, WEI D, et al. Feature Selection Method for Software Defect Number Prediction Based on Maximum Information Coefficient[J]. Telecommunications Science,2021,37(5):133-147 (in Chinese)

[10] ZHUANG Y X, LI Q, YANG B, et al. An End-to-End Approach for Bearing Fault Diagnosis Based on LSTM[J]. Noise and Vibration Control,2019,39(6):187-193 (in Chinese)

[11] LEI J H, LIU C, JIANG D X. Fault Diagnosis of Wind Turbine Based on Long Short-term Memory Networks [J]. Renewable Energy, 2019, 133: 422-432.

[12] Hochreiter S, Schmidhuber J. Long Short-term Memory [J]. Neural Computatio-n,1997, 9 (8): 1735-1780.

[13] Hundman K, Constantinou V, Laporte C, et al. Detecting Spacecraft Anomalies Using LSTMs and Nonparametric Dynamic Thresholding [C] the 24th ACM SIGKDD International Conference. ACM, 2018.

[14] SHAO Z T, XU D R, XU W L, et al. Radar Active Jamming Recognition Based on LSTM and Residual Network [J/OL]. Systems Engineering and Electronics (in Chinese)

[15] WANG H N, HAN H A, GUO Y X, et al. Short-Term Load Forecasting of Power System Based on Artificial Neural Network and Mutual Information Theory [J/OL]. China Measurement and Test:1-8 (in Chinese)

[16] ZHOU Z, ZHOU F, et al. Application Research on Bearing Degradation Prediction Based on E2E Deep VAE-LSTM [J/OL]. Application Research of Computers:1-8 (in Chinese)

Analysis of Statistical Data of the Local-oriented Freight Volume of Inland Water Transport in China

Fuyou Huang*, Bin Chen
Institute of Transportation Development Strategy & Planning of Sichuan Province, Chengdu, China
* Corresponding author: fuyou.huang@hotmail.com

ABSTRACT

In this paper, we propose the concept of local-oriented freight volume of inland water transport, which can reflect the inland waterway transportation level in a province or a city more truly and objectively. Based on the characteristics of China's inland water transport, the local-oriented freight transportation can be broken down into two parts: short-distance transportation and long-distance transportation. The freight volume of short-distance transportation can be captured in existing statistical reports, the freight volume of long-distance transportation can be calculated through the basic data of report system for ships entering and leaving ports. We take Sichuan province as an example, and investigate the correlation between local-oriented freight volume and GDP. The results show that the correlation between local-oriented freight volume and GDP is higher than that between published freight volume and GDP, this indicates that the local-oriented freight volume and its statistical method proposed in this paper is feasible and applicable.

Keywords: inland water transport, local-oriented freight volume, statistical method, correlation analysis, regional economy

1. INTRODUCTION

Generally, freight volume is considered as a barometer of economics, it can reflect the economic performance in a region, such as economic growth and industrial structure, and there is a certain quantitative relationship between freight volume and economic indicators.

Currently, the majority of social and economic indicators such as population and GDP are counted by administrative region in China. However, like highway transportation, the freight volume of inland water transport in a province or city, which is published by National Statistics, is the freight volume of the ships registered in the administrative region(called local ships), which including the freight volume generated by the ships in both in the administrative region and other regions. Meanwhile, the freight volume of nonlocal ships in the administrative region is not included in the published data. Therefore, the current published freight volume of water transport cannot reflect the waterway transportation performance level in a province or city truly and objectively, it may lead to decision-making bias about planning and policy for water transportation administration department.

In the existing literature on freight volume of water transport, most of them only optimized and improved statistical methods according to the statistical principle of local ships. For example, based on the statistical principle of local ships and the actual situation of Hunan province in China, Chen proposed a new calculation method of the freight volume blending monthly fluctuation coefficient and proportion [1]. Song carried out freight structural analysis and prediction of waterway transportation in the upper reaches of the Yangtze river, but the freight volume of watery transport is still obtained by the traditional statistical method [2]. In fact, Xie and Luo pointed out the shortcomings of existing statistical methods in 2009, they confirmed that it is very necessary to investigate the local-oriented freight volume of water transport in the administrative region, but they did not establish a systematic statistical method [3]. In the field of highway transportation, there have been some studies about the local-oriented freight volume [4-6], which can provide reference for us.

In this paper, we summarize the existing statistical systems and methods of the freight volume of watery transport, and propose a statistical approach of the local-oriented freight volume of watery transport. Then, Sichuan province is taken as an example to conduct empirical analysis, the feasibility and accuracy of the established statistical approach is also verified in this paper.

International Conference on Computer Graphics, Artificial Intelligence, and Data Processing (ICCAID 2022)
edited by Ruishi Liang, Jing Wang, Proc. of SPIE Vol. 12604, 126041C
© 2023 SPIE · 0277-786X · doi: 10.1117/12.2674844

2. STATISTICAL METHOD

In a province or city in China, the published freight volume and freight turnover volume of water transport come from the comprehensive statistical survey system of transportation, which is issued jointly by the National Statistics and Ministry of Transport. The comprehensive statistical survey system of transportation contains many basic transportation data of local ships by full sample survey. In particular, using simple analysis and calculation, the freight volume and freight turnover volume of every kind of goods generated by local ships in every city, can be obtained from the comprehensive statistical survey system of transportation.

In addition, a statistical survey system of highway and waterway transportation enterprises is another system related to the statistics of waterway transportation, which contains the throughput volume of every kind of goods of every important port or wharf. But the origin-destination data of goods can not be got from the statistical survey system of highway and waterway transportation enterprises.

Fortunately, the Maritime Safety Administration of the Ministry of Transport began to implement and promote the reporting system for ships entering and leaving ports in 2017, the basic origin-destination data of goods of every important port can be got from this system.

Considering the characteristics of China’s inland water transport and the existing statistical resources, the local-oriented freight transportation is divided into two parts: short-distance transportation and long-distance transportation. Thus, the local-oriented freight volume F is

$$F = F_s + F_l \tag{1}$$

where F_s is the local-oriented freight volume of water transport in short distance, and F_l is the local-oriented freight volume of water transport in long distance. In a similar way, the local-oriented freight turnover volume T is

$$T = T_s + T_l \tag{2}$$

where T_s is the local-oriented freight turnover volume of water transport in short distance, and T is the local-oriented freight turnover volume of water transport in long distance.

In general, the goods transported in short distance are mainly sand and stones and ores. The transportation activities are done within the administrative district, and are carried out by local ships. Thus, in short-distance transportation, the corresponding local-oriented freight volume and freight turnover volume of water transport are obtained by the comprehensive statistical survey system of transportation.

The long-distance transportation is considered to be transportation across cities, all the origin-destination data of goods of every port can be obtained from the reporting system for ships entering and leaving ports. Thus, in long-distance transportation, the corresponding local-oriented freight volume and freight turnover volume of water transport are calculated from basic data.

In some cities, there are multiple ports or docks, and there is a small exchange of goods between ports or docks. The transportation activities are done within the cities, and the corresponding freight volume and freight turnover volume of water transport should not be included in the category of long-distance transportation. Thus, in inland water transport of a province or city, we have

$$F_l = \sum_{i=1}^{a} A_i + \sum_{j=1}^{a} B_j - 2\sum_{i=1}^{a}\sum_{j=1}^{a} C_{ij} \tag{3}$$

$$T_l = \sum_{i=1}^{a} D_i + \sum_{j=1}^{a} E_j - 2\sum_{i=1}^{a}\sum_{j=1}^{a} H_{ij} \tag{4}$$

where a is the number of ports or docks of a province or city, A_i is the quantity of goods arriving at the *i*th port or dock, B_j is the quantity of goods shipped from the *j*th port or dock, C_{ij} is the quantity of goods transferred from the *i*th port or dock to the *j*th port or dock, D_i is the turnover volume of goods arriving at the *i*th port or dock, E_i is the turnover volume of goods shipped from the *j*th port or dock, H_{ij} is the turnover volume of goods transferred from the *i*th port or dock to the *j*th port or dock.

3. EXAMPLE ANALYSIS

In this section, we take the inland water transport performance in Sichuan province of China in 2019 as an example. We first summarize the characteristics of inland water transport in Sichuan province, and then analyze the local-oriented freight volume and freight turnover volume of inland water transport of every city. In addition, compared with the correlation between published freight volume and GDP, we examine the correlation between local-oriented freight volume and GDP.

In 2019, inland water transport occurs in 13 cities in Sichuan province. Details are as follows.

There is only short-distance transportation in 10 cities (Zigong, Panzhihua, Guangyuan, Suining, Neijiang, Nanchong, Guangan, Dazhou, Ziyang and Liangshan), almost all ships engaged in short-distance transportation are registered locally. Thus, the local-oriented freight volume and freight turnover volume of water transport of the 10 cities are consistent with the published data.

In Leshan city, most of the freight volume and freight turnover volume of water transport are generated by the local ships which carry sand and gravel within the city, this corresponding freight volume and freight turnover volume of water transport can be also obtained easily from the comprehensive statistical survey system of transportation. Besides, there is a small amount of goods shipped to other cities or come from other cities, and the corresponding part of freight volume and freight turnover volume of water transport, can be calculated from the basic data of the reporting system for ships entering and leaving ports. That is to say, the local-oriented freight volume and freight turnover volume of water transport of Leshan city are composed of two parts.

As sand mining is prohibited in the trunk line of Yangtze river, all the sand and gravel needed in Luzhou and Yinbin cities come from other cities or provinces. Meanwhile, Luzhou port and Yibin port are the main ports in Sichuan province, almost all shipments from other provinces are loaded and unloaded in the two ports. Thus, all the local-oriented freight volume and freight turnover volume of water transport of the two cities are calculated from the basic data of the reporting system for ships entering and leaving ports.

The local-oriented freight volume and freight turnover volume of water transport of the 13 cities in Sichuan province, are shown in Table 1 and Table 2. In Table 1 and Table 2, LOFV is the local-oriented freight volume, LOFTV is the local-oriented freight turnover volume, SDT is short-distance transportation, LDT is long-distance transportation, TTT is ten thousand tons, and TTTK is ten thousand ton-kilometers.

Table 1. The local-oriented freight volume of water transport in 13 cities.

City	LOFV in SDT (TTT)	LOFV in LDT (TTT)	Total LOFV (TTT)
Zigong	107	0	107
Panzhihua	28	0	28
Luzhou	0	951	951
Guangyuan	695	0	695
Suining	74	0	74
Neijiang	189	0	189
Leshan	228	75	303
Nanchong	1478	0	1478
Yibin	0	909	909
Guangan	487	0	487
Dazhou	340	0	340
Ziyang	270	0	270
Liangshan	54	0	54

Table 2. The local-oriented freight turnover volume of water transport in 13 cities.

City	LOFTV in SDT (TTTK)	LOFTV in LDT (TTTK)	Total LOFTV (TTTK)
Zigong	1583	0	1583
Panzhihua	1465	0	1465
Luzhou	0	1658394	1658394
Guangyuan	2202	0	2202
Suining	373	0	373
Neijiang	612	0	612
Leshan	13754	40724	54478
Nanchong	40887	0	40887
Yibin	0	1654511	1654511
Guangan	40752	0	40752
Dazhou	9200	0	9200
Ziyang	1800	0	1800
Liangshan	3794	0	3794

Table 3 shows the published freight volume and freight turnover volume of water transport, and GDP of 13 cities in 2009. In Table 3, MY is million yuan, PFV is published freight volume, PFTV is published freight turnover volume. After carrying out correlation analysis, we find that the correlation between published freight volume and GDP is 0.46, and the correlation between local-oriented freight volume and GDP is 0.55. That is to say, compared with the correlation between published freight volume and GDP, the correlation between local-oriented freight volume and GDP is increased by 9 percentage points. Similarly, the correlation between published freight turnover volume and GDP is 0.41, and the correlation between local-oriented freight turnover volume and GDP is 0.60. Compared with the correlation between published freight volume and GDP, the correlation between local-oriented freight volume and GDP is increased by 19 percentage points. Theses indicates that the local-oriented freight volume and freight turnover volume can reflect the development of regional economy better.

Table 3. The published data in 13 cities.

City	GDP (MY)	PFV (TTT)	PFTV (TTTK)
Zigong	1428490	107	1583
Panzhihua	1010130	28	1465
Luzhou	2081260	1986	2279846
Guangyuan	941850	695	2202
Suining	1345730	74	373
Neijiang	1433300	189	612
Leshan	1863310	303	82692
Nanchong	2322220	1478	40887
Yibin	2601890	617	590328
Guangan	1250440	784	40752
Dazhou	2041490	340	9200
Ziyang	777800	270	1800
Liangshan	1676300	54	3794

In addition, we found that the increase of correlation is caused mainly by the change of freight volume and freight turnover volume of water transport in Luzhou, Yibin and Leshan. In Luzhou city, there are many local shipping companies, and cargo transportation services are offered by them throughout Sichuan province and other regions along the Yangtze river, this leads to that the freight volume of local ships in other cities is much larger than the freight volume of nonlocal ships in Luzhou city. Accordingly, the published freight volume of water transport is much larger than the local-oriented freight volume in Luzhou city. In Yibin city, there are few local shipping companies, many goods are transported by nonlocal shipping companies, this leads to that the published freight volume of water transport is less than the local-oriented freight volume. Similar to Yibin city, the published freight volume of water transport is also less than the local-oriented freight volume in Leshan city.

4. CONCLUSIONS

The freight volume and freight turnover volume of water transport are important basis to develop policies and programs of water transport. However, the published data are the freight volume and freight turnover volume of local ships, which cannot reflect the real demand and operations of a region in water transport.

In this paper, the concept of local-oriented freight volume and freight turnover volume of inland water transport is proposed, and a practical statistical and extrapolating method is given based on the existing statistical resources. Then, we take the inland water transport performance in Sichuan province in 2019 as an example, and examine the correlation between local-oriented freight volume and GDP, as well as the correlation between local-oriented freight turnover volume and GDP. Compared with the correlation between the published freight volume and GDP, the correlation between local-oriented freight volume and GDP is increased by 9 percentage points. Compared with the correlation between the published freight turnover volume and GDP, the correlation between local-oriented freight turnover volume and GDP is increased by 19 percentage points. Theses indicates that the local-oriented freight volume and freight turnover volume can reflect the development of regional economy and the demand of waterway transportation more truly and objectively.

ACKNOWLEDGMENT

This work was supported by Sichuan Science and Technology Program (Grant No. 2022JDRC0063) and the Science and Technology Project of Transportation Department of Sichuan Province (Grant No. 2021-D-04).

REFERENCES

[1] Chen, J., “The research on calculation method of highway and waterway transportation volume and application in Hunan”, Changsha University of Science & Technology, 2015.

[2] Song, K., “Structural analysis and prediction of waterway transportation goods in the upper reaches of the Yangtze river”, Chongqing Jiaotong University, 2019.

[3] Xie Y. and Luo, R., “Problems and countermeasures of highway and waterway transportation statistics”, China Transportation Review, 10, 69–71 (2009).

[4] Xiong, D., Huang, F. and Zhong, M., “Statistical analysis and research on freight volume of regional highways in Sichuan province”, Transport Energy Conservation & Environmental Protection, 16(79), 66–70 (2020).

[5] Duan, L., Luan, Q. and Zhao, X., “Improved statistical method measuring freight traffic volume in regional road network”, Journal of Transportation Systems Engineering and Information Technology, 20(4), 28–33 (2020).

[6] Yan, S., Xiao, R. and Yang, M., “Statistical approach for the region-oriented volume of freight transport on highway”, Journal of Traffic and Transportation Engineering, 20(6), 109–113 (2020).

Operation and maintenance interactive system base on artificial intelligence and big data information system

Shanshan Wang[1*], Tongle Liang[2]
[1]Guangdong Industry Polytechnic, Guangzhou 510300, China
[2]Guangdong Vocational College of Post and Telecom, Guangzhou 510630, China

Youth Innovation Project of Guangdong Province (2019GKQNCX054)

*Corresponding author: shanshanwang202@126.com

ABSTRACT

In order to better meet the development needs of interconnection and further improve the core operation capability of hospitals, the research of operation and maintenance interactive system based on artificial intelligence and big data information system is proposed. In the process of building the hospital information operation and maintenance platform, there are still some problems such as insufficient support in technology and operation and maintenance resources. It is necessary to effectively optimize the operation and maintenance system with limited human resources to ensure the stable operation and development of the hospital information system. Under this background, this paper analyzes the problems existing in hospital information management and the current situation of operation and maintenance of information technology, and discusses the mode construction of operation and maintenance platform based on big data artificial intelligence architecture. In the construction of hospital information operation and maintenance platform, we should focus on improving the service level and meeting the needs of patients, so as to continuously enhance the security, integration and interactivity of the system platform, thus promoting the sustainable development of China's medical reform.

Keywords: artificial intelligence; Big data; System operation and maintenance; Pattern construction; operational capability

1. INTRODUCTION

With the increasing types of hospital information, the difficulty of information data processing is also increasing. It is necessary to strengthen the introduction of advanced technical means and methods to comprehensively and accurately analyze hospital information data, so as to effectively improve the overall management level of the hospital[1]. At present, the results of medical data analysis in many hospital information systems can't meet the needs of hospital management, and there are some phenomena such as nonstandard data operation[2-3]. At the same time, there are different standards of data transmission and storage among different regions and hospitals, which leads to the ineffective circulation and sharing of hospital data across regions and hospitals. In addition, there are many monitoring equipment, medical laboratory equipment and radiation equipment in the hospital, and their daily operation will produce a large amount of data information[4]. However, due to the inability to effectively integrate the systems, patients have repeated diagnosis and treatment in different degrees[5]. This wastes a lot of medical resources, but also aggravates the contradiction between doctors and patients. Therefore, hospitals need to integrate various practical problems existing at this stage and strengthen the application of advanced technologies such as big data and artificial intelligence[6-7].

On the basis of this research, this paper proposes an idea of building an operation and maintenance platform architecture integrating big data analysis and artificial intelligence algorithms, providing more convenient and accurate operation and maintenance early warning and capacity analysis for hospital informatization operation and maintenance workers, and thus providing quality assurance for hospital informatization business. By using the AIOps operation and maintenance platform, the operation and maintenance personnel can be released from the repetitive and mechanical work, reduce the workload of the operation and maintenance personnel, and let the operation and maintenance personnel undertake more important work. In order to obtain more standard and accurate non personal empirical operation and maintenance decisions, the hospital's information service construction and operation will be better, and the availability and reliability of business will be better guaranteed.

International Conference on Computer Graphics, Artificial Intelligence, and Data Processing (ICCAID 2022)
edited by Ruishi Liang, Jing Wang, Proc. of SPIE Vol. 12604, 126041D
© 2023 SPIE · 0277-786X · doi: 10.1117/12.2674679

2. OPERATION AND MAINTENANCE PLATFORM MODE BASED ON ARTIFICIAL INTELLIGENCE BIG DATA ARCHITECTURE

Hospital information business system has the characteristics of wide coverage, multiple applications and strong coupling among various systems[8]. When the operation and maintenance personnel analyze the operation and maintenance data, there are usually problems that can't effectively evaluate the deep-seated problems and performance of the system[9-10]. Strengthening the application of Alops operation and maintenance mode of big data artificial intelligence has become an important direction for the development of hospital operation and maintenance of information technology[11]. By combining all kinds of data generated in the operation and maintenance process with machine learning and algorithm analysis, to a certain extent, the operation and maintenance work can overcome the excessive dependence on professional knowledge and experience, break through the manpower limitations existing in the information department, and effectively solve the contradiction between input and performance[12]. The hospital intelligent operation and maintenance platform needs to be constructed from the existing information operation and maintenance data of the hospital. It can effectively contact all kinds of network and equipment performance logs and other data originally closed in multiple sets of isolated information systems through data connection, and make use of advanced technologies such as big data technology to analyze and manage them, so as to transform them into all kinds of valuable data[13]. In the actual construction, the four-layer architecture and five display layers can be adopted to realize the construction of hospital operation and maintenance of information technology platform. Among them, the four levels of operation and maintenance platform construction need to ensure the correctness of their sequence and the strong correlation between levels [14]. Through the data storage layer, the data and logs of all information operation and maintenance equipment in the hospital are screened and analyzed in a near real-time and real-time manner. Operation and maintenance platform architecture is shown in Figure 1, Tables 1 and 2[15-16].

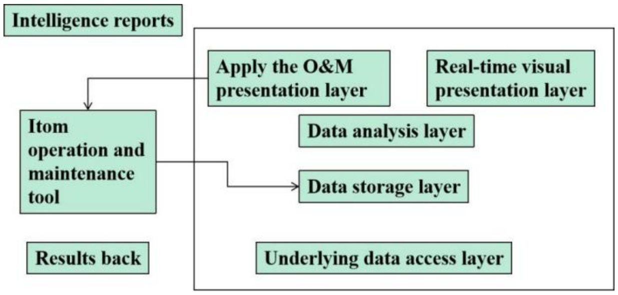

Intelligence reports
Apply the O&M presentation layer
Real-time visual presentation layer
Itom operation and maintenance tool
Data analysis layer
Data storage layer
Results back
Underlying data access layer

Figure 1. Platform architecture

Table 1. Bottom data access layer of operation and maintenance platform architecture

Bottom data access layer	network equipment	Terminal security software	database	Business system	safety equipment
	middleware	Computer room monitoring	VMs	server	

Table2. Data analysis layer and data storage layer of table operation and maintenance platform architecture

Data analysis layer	data modeling	machine learning	Business analysis	Deep learning
Data cache layer	Big data cache			

In the operation and maintenance platform, the bottom data interface and data storage layer mainly collect and store a large amount of daily data generated by various basic equipment resources, and provide data for specific operation and maintenance algorithm analysis[17]. Among them, some work, such as server log screening, is done manually, and network data collection is done by machines and equipment. This enables the operation and maintenance personnel to be effectively liberated from the heavy work that relied on experience analysis in the past, optimizes and innovates the information operation and maintenance system, and applies the surplus human resources to more important positions. In addition, the continuous standardization and complexity of operation and maintenance can further improve the quality and efficiency of hospital information construction.

As the core layer of hospital information operation and maintenance platform, the data analysis layer needs to strengthen the application of big data screening and artificial intelligence analysis, and at the same time, the data analysis layer needs to provide valuable reference data[18]. At present, the hospital operation and maintenance work needs to provide effective decision-making and treatment schemes through the application of operation and maintenance performance layer. At the same time, this layer can reasonably plan the future development of operation and maintenance work in combination with business demands, including shrinking and expanding forecasts, etc[19]. If more physical resources need to be allocated due to the shortage of human resources in the hospital information department, the platform artificial intelligence analysis and prediction model can be used to judge the rationality of the resource demand put forward by the application developer. In addition, this layer can also transmit valuable operation and maintenance reports obtained by intelligent analysis to the data storage layer, providing effective data for machine learning, thus realizing the closed-loop process of machine learning.

The value of the whole operation and maintenance platform is mainly reflected by the real-time visual display layer[20]. In practice, all the non-information workers in the hospital can know and master the operation and maintenance data through this layer, and at the same time provide a reliable basis for strengthening the information construction in the later period of the hospital, and further clarify the information development direction of the hospital.

3. ARTIFICAL INTELLIGENCE FOR IT OPERATIONS (AIOPS) BASED ON AI BIG DATA ARCHITECTURE

Now, the proposed AIops operation and maintenance mode provides an important development direction for hospital informatization operation and maintenance. Through the analysis of various data generated in operation and maintenance, machine learning and algorithm analysis are combined. To a certain extent, the operation and maintenance work can get rid of the dependence on human experience and knowledge, liberate the manpower constraints of the information department, and effectively solve the contradiction between input and performance. Therefore, the operation and maintenance architecture (AIops) based on big data AI must be the development trend of hospital IT operation and maintenance. The construction of AIOps intelligent IT operation and maintenance platform must be based on the existing information operation and maintenance data of the hospital, connect various equipment performance log data and network data and other access data enclosed in multiple sets of isolated information systems, and then transform them into various valuable information through the management and analysis of big data to help the hospital achieve monitoring and early warning of information operation and maintenance. As shown in Figure 2, the hospital AIOps informatization operation and maintenance platform is built on five presentation layers through a four tier architecture. The four levels of operation and maintenance platform construction are highly related and in sequence, and the latter needs the previous level as the basis. The underlying data access is the foundation. All the hospital's IT operation and maintenance equipment data and logs are concentrated in the data storage layer in a real-time or near real-time manner for filtering and analysis.

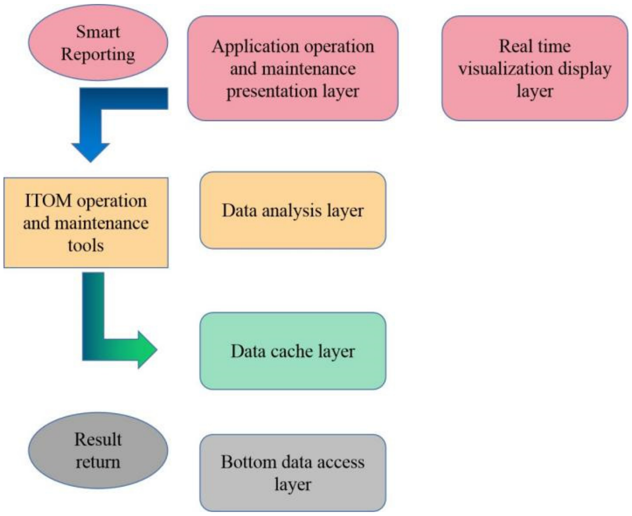

Smart Reporting
Application operation and maintenance presentation layer
Real time visualization display layer
ITOM operation and maintenance tools
Data analysis layer
Data cache layer
Result return
Bottom data access layer

Figure 2. AIops Operation and Maintenance Platform Architecture

Considering that the hospital informatization construction is mostly outsourced to a third-party software company for development in the system software development stage, while after the hardware resources and system are online, the operation and maintenance of the system is conducted in a hybrid way of independent and outsourcing. Based on the above scenario of hospital information construction, the role division of AIops operation and maintenance has the following three roles: traditional operation and maintenance engineer; Data analyst; O&M development engineer. Therefore, the key to the successful implementation of AIOps operation and maintenance platform is to handle the relationship between the above three roles. Based on the characteristics of the information construction in the hospital, we should adjust measures to local conditions, fully integrate with the system business, make targeted configuration for the role division, operation and maintenance relationship and operation and maintenance technology of the operation and maintenance personnel, and organically combine the operation and maintenance department with the third-party software development company through the data and conclusions obtained from the platform, so as to promote the development of core information business.

4. CONCLUSION

This paper proposes an idea of building an operation and maintenance platform architecture that integrates big data analysis and artificial intelligence algorithms. Operation and maintenance platform can release operation and maintenance personnel from repetitive and mechanical work, reduce the workload of operation and maintenance personnel, and let the operation and maintenance personnel take on more important work. In order to obtain more standard and accurate non personal empirical operation and maintenance decisions, the hospital's information service construction and operation will be better, and the availability and reliability of business will be better guaranteed.

REFERENCES

[1] Xiao, B. , & Wang, W. . (2021). Intelligent network operation and maintenance system based on big data. Journal of Physics: Conference Series, 1744(3), 032033-.

[2] Zhang, N. , Zhang, W. , & Shang, Y. . (2021). Research on integrated energy system of power grid based on artificial intelligence algorithm of machine learning. IOP Conference Series: Earth and Environmental Science, 714(4), 042035 (7pp).

[3] Wang, H. . (2021). Design of power line safety operation and maintenance monitoring system based on cloud computing. International Journal of Information and Communication Technology, 19(3), 242-.

[4] Li, C. , & Cui, J. . (2021). Intelligent sports training system based on artificial intelligence and big data. Mobile Information Systems, 2021(1), 1-11.

[5] Zhang, S. , Lu, C. , Jiang, S. , Lu, S. , & Xiong, N. N. . (2020). An unmanned intelligent transportation scheduling system for open-pit mine vehicles based on 5g and big data. IEEE Access, PP (99), 1-1.

[6] Yang, Y. . (2020). Research on brush face payment system based on internet artificial intelligence. Journal of Intelligent & Fuzzy Systems, 38(1), 21-28.

[7] Rfa, D. , Sr, A. , Mm, B. , Cglb, C. , & Nd, C. . (2020). Smart society and artificial intelligence: big data scheduling and the global standard method applied to smart maintenance. Engineering, 6(7), 835-846.

[8] Li, J. , & Wang, T. . (2021). Research on the application of artificial intelligence technology in intelligent operation and maintenance of industrial equipment and system. Journal of Physics Conference Series, 1992(3), 032090.

[9] Sun, Z. . (2021). Research on informatics system and practice prospects based on artificial intelligence mathematical algorithm. Journal of Physics: Conference Series, 1865(4), 042076-.

[10] Zhao, S. , & Wang, H. . (2021). Enabling data-driven condition monitoring of power electronic systems with artificial intelligence: concepts, tools, and developments. IEEE Power Electronics Magazine, 8(1), 18-27.

[11] Si, C. , & Shi, W. . (2021). Establishment and improvement of financial decision support system using artificial intelligence and big data. Journal of Physics: Conference Series, 1992(3), 032082-.

[12] Chen, K. , Zu, Y. , & Cui, Y. . (2020). Design and implementation of bilingual digital reader based on artificial intelligence and big data technology. Journal of Computational Methods in Sciences and Engineering, 20(2), 1-19.

[13] Ren, Q. . (2021). Application analysis of artificial intelligence technology in computer information security. Journal of Physics: Conference Series, 1744(4), 042221 (7pp).

[14] Choi, W. . (2020). A study on the intelligent disaster management system based on artificial intelligence. Korean Society of Hazard Mitigation, 20(1), 127-140.

[15] Liu, Q. , & Huang, Z. . (2020). Research on intelligent prevention and control of covid-19 in china's urban rail transit based on artificial intelligence and big data. Journal of Intelligent and Fuzzy Systems, 39(21), 1-6.

[16] Huang, W. , Ren, J. , Yang, T. , & Huang, Y. . (2021). Research on urban modern architectural art based on artificial intelligence and gis image recognition system. Arabian Journal of Geosciences, 14(10), 1-13.

[17] Lv, X. , & Li, M. . (2021). Application and research of the intelligent management system based on internet of things technology in the era of big data. Mobile Information Systems, 2021(16), 1-6.

[18] Yao, J. , & Liu, J. . (2021). Research on computer network technology system based on artificial intelligence technology. Journal of Physics: Conference Series, 1802(4), 042028 (6pp).

[19] Li, S. . (2020). Structure optimization of e-commerce platform based on artificial intelligence and blockchain technology. Wireless Communications and Mobile Computing, 2020(12), 1-8.

[20] Chen, P. . (2020). Design of travel itinerary planning system based on artificial intelligence. Journal of Physics: Conference Series, 1533(3), 032078 (6pp).

Risk Continuity Analysis of Transmission Line Channel Based on Historical Hidden Danger Data

Fei Wang*, Lingqi Kong
Zhiyang Innovation Technology Co., Ltd., Zibo, 255000, China
* Corresponding author: wangfeichn@163.com

ABSTRACT

This paper proposes a continuous analysis method of transmission line channel based on historical early warning data. It calculates the feature data, the real-time alarm data and the corresponding feature data, and passes the threshold value R_0. Determine whether it is a continuous alarm. In this paper, through the extraction of sample data extraction and calculation of Pearson moment correlation coefficient, solved the fireworks, foreign body alarm for low frequency, seasonal periodic, conventional single comparison and mechanical have good dimension reduction analysis cannot effectively identify the problem, for subsequent application scenarios such as alarm level intelligent annotation, AI image recognition model suspected false alarm and omission of sample identification model support, and improve the intelligent level of transmission line shipment inspection.

Keywords: transmission line channel; historical early warning data; hidden danger continuity analysis; data processing; data analysis; intelligent transportation and inspection

1. INTRODUCTION

As an important national infrastructure, the safety and stability of transmission line are related to people's production and life. In the process of long-distance electric energy transmission, the transmission line carries high loads and works in the uncertain natural environment. Therefore, the investigation of hidden dangers in the channel environment is one of the most important transportation and inspection items, and it is easy to encounter different reasons and different types of hidden dangers and threats. Hidden perils in transmission channel from the initial artificial patrol, to through the image acquisition equipment timing capture and return images for artificial patrol, now using artificial intelligence technology to capture the image of hidden trouble automatic recognition, the image acquisition equipment for terminal recognition, drone inspection, video monitoring, 5G transmission and collection networking, Internet of things technology application, greatly improve the inspection efficiency and security.

Because the visual inspection of transmission line channels is widely used, the automatic identification of visual information has been realized and the alarm objects appear in the image, such as machinery, fireworks, foreign objects, etc. In addition to the basic statistical analysis report, can be based on the alarm data for data mining, such as continuous alarm of real-time alarm data judgment, alarm level based on the results of intelligent annotation, image recognition model suspected false alarm and omission sample identification, but the application of the above scenarios need continuous alarm identification technical support, fireworks, foreign body alarm for low frequency, seasonal periodic, regular single comparison and mechanical have good applicable dimension analysis can not effectively identify the fireworks, foreign body alarm [1].

Because the continuous alarm of transmission line channel is discrete data, it is challenging to find a data processing method to efficiently and accurately. In the relevant technical field, the Pearson product moment correlation coefficient is used to process the data for continuous data and waveform comparison. However, it is difficult to apply the Pearson product moment correlation coefficient to the discrete data such as continuous alarm of visual pyrotechnic foreign matter. This paper proposes a feasible and accurate continuous alarm determination method of fireworks and foreign matter, to provide data support for the intelligent maintenance scene of transmission lines.

International Conference on Computer Graphics, Artificial Intelligence, and Data Processing (ICCAID 2022)
edited by Ruishi Liang, Jing Wang, Proc. of SPIE Vol. 12604, 126041E
© 2023 SPIE · 0277-786X · doi: 10.1117/12.2674541

2. THE CONTINUOUS ALARM SAMPLE DATA WAS PREPROCESSED AND ACQUIRED THE FEATURE DATA

2.1 Data preprocessing

This paper puts forward a kind of transmission line channel hidden danger continuity analysis method based on historical warning data, through the feature extraction of sample data and calculation of Pearson product moment correlation coefficient, solve the fireworks, foreign body alarm for low frequency, seasonal periodic, regular single comparison and have good applicable to mechanical dimension reduction analysis can not effectively identify the problem. The continuous alarm sample data of a visual inspection equipment of the transmission line are preprocessed and the characteristic data are obtained. In this paper, it is determined that fireworks and foreign body alarm in the equipment image is identified by AI image recognition model, because the author in other works, this paper does not say too much; to confirm whether the alarm is continuous alarm, compared with mechanical alarm, because of fireworks and foreign matter occurred less times, randomness in time and region, at present, existing methods, have not found effective machine learning means to solve [2].

In the collected historical hidden danger data, there are a small number of data with two different hidden danger results. This kind of data is because the inspector found a false alarm when confirming the results of the AI image analysis model recognition, and the inspector will produce a new hidden danger data, and will be marked in the corresponding field. For the data with two different results, the results confirmed by the inspector shall prevail, and the identification result data of the AI image analysis model corresponding to the data is excluded. In addition, a small number of AI image analysis models are not accurately identify, and the hidden danger data found by inspectors remains normally [3].

Secondly, most of the collected historical hidden dangers data are regularly collected by the transmission channel visualization remote inspection and capture equipment. After the analysis of the image recognition model, it is confirmed that there are certain hidden dangers [4]. However, due to the inconsistent image acquisition intervals set by different image acquisition devices, such as 30 minutes, 15 minutes, 10 minutes, 5 minutes, etc., and the inconsistent capture interval of the same device in different times because of artificial adjustment, automatic adjustment of intelligent alarm strategy and other reasons.

In this regard, the weight of data is needed. First, the hidden danger data is sampled for 30 minutes; in the second, it sets different weights for the hidden trouble data of 30 minutes, setting the hidden danger data to 1,15 minutes, 0.33, the interval of 5 minutes, 0.17, and the hidden danger data generated by other non-image capture devices as 1.

In this paper in data analysis, the two different processing method multiple comparison verification, two different data cleaning method and no obvious advantages, for this reason, this paper selected the first way, namely according to the time interval sampling way, its advantage is compared to the second method after data cleaning data will reduce about 27%, and do not need additional weight information, in big data analysis can reduce the computation, improve the real-time. Preprocess the continuous alarm sample data and real-time alarm data, keeping only the time field.

2.2 Obtain feature data

Build a two-dimensional array with a value of all 0, with the number of sampling points per day, and its behavior is 12 or a multiple of 12 N * 12, corresponding to 1~N * 12 months, where N is the natural number. The data required for the model construction of this paper needs to be limited to 12 months, because the fireworks and foreign body alarms have a strong periodicity, and they need to cover a complete natural cycle year. For example, in the wheat harvest season, dry season, there will be more fireworks, and there will be more foreign bodies in the windy season. In addition, similar continuous months should be a multiple of 12 months, such as from July 1 last year to June 30 this year, we should not only have the data of the past three months, which cannot reflect the periodicity.

Perverse continuous alarm sample data, assign to the array, assignment principle is: according to each data belongs to determine the corresponding line position, then consider the time of the corresponding sampling point, determine the column in the 2 dimensional array position, determine the position value after 1, pass after the completion of the characteristic data [5].

3. CALCULATE THE CORRELATION COEFFICIENT R

After processing the real-time alarm data of fireworks and foreign matter, the corresponding data in the aforementioned characteristic data can calculate the Pearson product moment correlation coefficient to obtain the correlation coefficient R. For real-time data, the feature sequence is obtained according to the number of sampling points per day. The acquisition method is to construct a full 0 value sequence with the number of elements and the same as the number of sampling points per day [6]. For each alarm data, the position value is added by 1 according to the sampling point corresponding to its time. The correlation coefficient of the resulting feature sequence and the data corresponding to the feature data is calculated to obtain the correlation coefficient R [7].

$$R = \begin{cases} 0\ , The\ X\ elements\ are\ all\ zero\ or\ the\ Y\ elements\ are\ all\ zero \\ \dfrac{\sum_{i=1}^{n}(X_i-\bar{X})(Y_i-\bar{Y})}{\sqrt{\sum_{i=1}^{n}(X_i-\bar{X})^2}\sqrt{\sum_{i=1}^{n}(Y_i-\bar{Y})^2}}\ , Others \\ 1\ , The\ X\ set\ is\ equal\ to\ the\ Y\ set \end{cases} \tag{1}$$

In formula (1), set X is the feature sequence corresponding to real-time data, and set Y is the row sequence corresponding to real-time data month corresponding to continuous alarm sample; the advantage of this design is that in order to judge the input data, this paper improves the correlation coefficient of Pearson product moment, adding that the elements of X or Y are 0 and X and Y are equal [8].

4. CONTINUOUS ALARM ANALYSIS OF THE HIDDEN DANGER OF THE TRANSMISSION LINE CHANNEL

Take the obtained Pearson product moment correlation coefficient R and compare R with the threshold R_0. If $R > R_0$ is satisfied, the real-time alarm data belongs to continuous alarm; otherwise, it does not belong to continuous alarm. The continuous alarm sample data in this paper refers to the alarm data of the same equipment with a time span of 2 years. The amount of time span has an impact on the determination accuracy of this paper. According to the experiment, the span time of 2~5 years is more appropriate [9]. Considering the difficulty of data acquisition, it is more appropriate to use 2 years of data. If the time span is too long, the error will be introduced in the previous data due to the geomorphological changes, and if the span is short, the randomness will be relatively large. Therefore, the time span of the sample data determined in this paper is selected as 2 years [10].

The threshold R_0 refers to the general evaluation standard of Pearson product moment correlation coefficient, whose full value range corresponding meanings are: 0.8-1.0 very strong correlation, 0.6-0.8 strong correlation, 0.4-0.6 moderate correlation, 0.2-0.4 weak correlation, 0.0-0.2 very weak correlation or no correlation. The threshold value of R_0 takes a value of 0.6.

5. THE EXAMPLE ANALYSIS

5.1 The paper model is applied to determine the fireworks continuous alarm

A equipment ID of 99000843025795 identifies the fireworks alarm in the image taken by 2022-9-9. Query the historical fireworks alarm data of 2020-9-9~2022-9-8. There are 187 marked continuous alarm sample data, the time span is 2020-9-16 09:54:17~2022-8-29 14:12:31. This paper model is used to determine whether the 6 real-time fireworks alarm data of the device 2022-9-9 belong to the continuous alarm. In the above data, the continuous alarm data includes 25 fields including alarm self-increase ID, time, alarm content, and image storage ID; the device image acquisition interval is 60 minutes; the threshold R_0=0.6 is the general evaluation standard of Pearson product moment correlation coefficient, which is constant in this model.

Table 1 Example of the preprocessed alarm data

Sequence number	time	Sequence number	time
1	2020-9-16 09:54:17	6	2020-11-23 15:03:40
2	2020-9-16 10:27:19	7	2020-11-23 15:40:41
3	2020-9-16 10:59:23	…	…
4	2020-10-18 11:33:42	186	2022-8-1 18:28:36
5	2020-10-18 12:18:55	187	2022-8-29 14:12:31

A method of hazard continuity analysis of transmission line channel based on historical warning data is proposed as follows. Preprocess the continuous alarm data and the real-time alarm data, and retain only the time attribute. Some continuous alarm sample data as shown in the table 1.

Table 2 Real-time alarm data

Sequence number	time	Sequence number	time
1	2022-9-9 17:02:05	4	2022-9-9 19:32:04
2	2022-9-9 18:32:30	5	2022-9-9 20:02:03
3	2022-9-9 19:02:04	6	2022-9-9 20:32:03

Real-time alarm data as shown in the table 2. A two-dimensional array is constructed with the number of sampling points per day 24h/0.5h=48, the number of sampling points per day, and its behavior is 12, corresponding to 1~12 months, so a two-dimensional array data [48][12] is constructed, and all the values are initialized as 0. Traversing the alarm data, assign values to the array. The assignment principle is as follows: first, determine the corresponding row position in the two-dimensional array according to the month when each data belongs to, then consider the sampling point corresponding to the time, and determine the column position in the two-dimensional array. After determining, add the position value by 1, and the characteristic data is obtained after the traversal: [[0,0,0,0,0,0,0,1,0,0,0,2,0,0,0,0,0,0,0,1, 0,0,0,3,0,1,0,1,0,1,0,1,0,0,0,1,0,0,0,0,0,0,0,0,0,0,0,0],[...],[...],[...],[...],[...],[...],[...],[0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,1,0,1,0,3, 0,1,0,0,0,0,0,1,0,1,0,1,0,1,0,0,1,5,1,3,3,0,0,0,0,0,0],[...],[...],[...,0,1,0,1,0,0,0,0,0,1,0,2,0,0,0,0,0,0,0]]. Parts of the data after the assignment are shown above. Processing of the real-time alarm data yields the feature sequence [0,0,0,0,0,0,0,0, 0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,1,0,0,1,1,1,1,1,0,0,0,0,0,0]. This paper can calculate the correlation coefficient of the data corresponding to the obtained feature data, that is, the September data [0,0,0,0,0,0,0,0,0,0,0,0,0,0, 0,0,0,1,0,1,0,3,0,1,0,0,0,0,0,1,0,1,0,1,0,1,0,0,1,5,1,3,3,0,0,0,0,0,0], and obtain the correlation coefficient R=0.6867. The obtained Pearson product-moment correlation coefficient R=0.6867> R_0=0.6, so this model determines that the real-time alarm data belongs to the continuous alarm.

In this paper, after the model triggers the continuous alarm, it is pushed to the maintenance personnel. According to the image of the maintenance personnel, it is the continuous fireworks alarm caused by the fire of the dry corn stalk of the transmission line. After urgent treatment, the accident of short circuit of the transmission line caused by solid particles is avoided.

5.2 Apply the paper model to determine the continuous alarm of floating objects

A equipment ID of 99000845000770 identified foreign body alarm in the image taken in 2022-9-9, queried the historical foreign body alarm data of 2020-9-9~2022-9-8, among which 91 marked continuous alarm sample data, the time span is 2020-10-3 17:24:12~2022-8-21 10:17:23. Based on the above data, judge whether the 7 real-time foreign body alarm data of the device 2022-9-9 belong to the continuous alarm.

Using the same processing steps as in 5.1, the Pearson product-moment correlation coefficient R=0.6567 was obtained to determine a continuous alarm of foreign matter. In this paper, after the model triggers the continuous alarm, it is pushed to the maintenance personnel. After the image of the maintenance personnel confirms, it is a continuous foreign body alarm caused by a piece of black plastic cloth on the transmission line. After triggering the continuous alarm, the inspection personnel were arranged to remove foreign objects to avoid the accident of short circuit of the transmission line caused by suspended matter.

This paper combined with application, based on the specific sample data characteristics extraction method and calculation of Pearson product moment correlation coefficient, unsupervised learning, based on unsupervised learning, realize the automatic continuous alarm judgment of real-time alarm data, from two real-time alarm data time distribution, and maintenance personnel confirmation, it is a continuous alarm. Good solve the fireworks, foreign body alarm because of low frequency, seasonal periodic, cannot effectively identify continuous alarm problem, for the subsequent application scenarios such as alarm level intelligent annotation, AI image recognition model suspected false alarm and omission sample identification provides a model support, and improve the intelligent level of the transmission line shipment inspection.

Proc. of SPIE Vol. 12604 1260401-382

6. CONCLUSION

This paper proposes a continuous analysis method of hidden dangers of transmission line channels based on historical early warning data, which belongs to the field of intelligent operation and inspection of transmission lines. Based on the annotated sample data of continuous alarm of transmission line channels, we judge whether the real-time alarm data belongs to continuous alarm or not. The idea of this paper is to find the final solution only after the special agreement of the data. A complete model is provided, in which the Pearson correlation coefficient is one of the bright spots, and the described serialization and feature extraction are very difficult to think of. The discrete data were weighted serialized to obtain the feature data, and only based on such data has the possibility of performing the analysis using the Pearson's coefficient. At the same time, the present paper defines the calculation of Pearson's product moment correlation coefficient optimization, in order to meet the sequence similarity calculation, and solved the low because of fireworks, foreign class occurrence frequency, in some cases characteristic data line elements are 0 and some cases real-time data features and characteristic data corresponding line sequence set, unable to calculate the correlation coefficient. Definition optimization makes the calculated results fit into practical significance. In this paper, the improved operation method of Pearson product moment correlation coefficient is selected to solve the difficulty of the algorithm with similar conventional comparison waveform but not corresponding position. It can determine the continuous alarm of real-time alarm data of fireworks and foreign matter, which provides technical and model support for subsequent application scenarios such as intelligent annotation of alarm level, suspected false alarm of image identification model and missed sample annotation. Based on the specific sample data feature extraction method and calculation of Pearson moment correlation coefficient, because of the special data reduction means and feature extraction method, make the discrete data Pearson moment correlation coefficient similarity judgment into feasible scheme, based on the unsupervised learning based on the model in the paper, solve the fireworks, foreign body alarm because of the low frequency of occurrence, seasonal periodic, unable to effectively identify the problem of continuous alarm.

The method in this paper can get the expected results very well. The corresponding model system has been deployed for use in several regions, but there are also some shortcomings. First, in terms of hidden danger data sources, in some areas are not preserved for a long time, or the messy data sources, difficult to efficiently fuse or sample, for such problems, need to do long-term effective collection, sorting and integration of data sources, such as adding effective time mark, marking good collection method, manual confirmation, etc.; second, incomplete collection of the hidden danger data, because most of the hidden danger data is the images collected by the image capture equipment installed on the transmission channel tower, results after identification by the AI image analysis model, but in some areas, some poles and towers are not equipped with image acquisition equipment, or the uneven installation density of the image acquisition equipment, leading to the difference between the analysis and the actual inspection, for this kind of problem. The continuous improvement of data sources and more perfect hidden danger data sampling strategy. This is also one of the key work of the subsequent iteration and upgrade of this model.

REFERENCES

[1] Zhang, Y. , Wang, W. , Huang, P. , & Jiang, Z. . (2019). Monocular vision-based sense and avoid of uav using nonlinear model predictive control. Robotica, 37(9), 1-13.

[2] J Xiong, Liu Z, Chen S, et al (2020). Visual detection of green mangoes by an unmanned aerial vehicle in orchards based on a deep learning method. Biosystems Engineering, 194:261-272.

[3] Zhiyang Innovation Technology Co., Ltd., Zhan Xingang, Wang Fei, etc. An identification method for visual alarm of high incidence of transmission line channel: ZL201910807542.2 [P].2020-05-08.

[4] Mederos-Henry, F. , Hermans, S. , & Huynen, I. . (2019). Microwave characterization of metal-decorated carbon nanopowders using a single transmission line. Journal of Nanomaterials, 2019, 1-11.

[5] Naing, T. H. , Janudom, S. , Rachpech, V. , Mahathaninwonga, N. , & Thiwong, S. . (2019). Corrosion behavior of galvanized steel for porcelain insulator's pin in hvac transmission line. Key Engineering Materials, 803, 45-49.

[6] Wang, J. , & Zhang, Y. . (2021). Traveling wave propagation characteristic-based lcc-mmc hybrid hvdc transmission line fault location method. IEEE Transactions on Power Delivery, PP(99), 1-1.

[7] Zhiyang Innovation Technology Co., Ltd., Yang Jun, Wang Fei, etc. A alarm method of transmission line channel: ZL201911012721.3 [P].2021-09-07.

[8] Yamane T, Chun P J (2020). Crack Detection from a Concrete Surface Image Based on Semantic Segmentation Using Deep Learning. Journal of Advanced Concrete Technology, 18(9):493-504.

[9] Zhao Y, Li Y, Shi W, et al (2019). Mutual Coupling Reduction Between Patch Antenna and Microstrip Transmission Line by Using Defected Isolation Wall. Applied Computational Electromagnetics Society journal, 34(1):100-106.

[10] Niitsu, K. , Nakanishi, T. , Murakami, S. , Matsunaga, M. , Kobayashi, A. , & NM Karim, et al. (2019). A 65-nm cmos fully integrated analysis platform using an on-chip vector network analyzer and a transmission-line-based detection window for analyzing circulating tumor cell and exosome. IEEE transactions on biomedical circuits and systems, 13(2), 470-479.

A data fusion-based method for detecting difficult yoga postures

Kaiyue Wang*
Shandong Management University, Jinan City, Shandong, China 250357
*Corresponding author: w_kyue@163.com

ABSTRACT

The current traditional human pose detection methods mainly acquire the pose data collected by sensors and realize the pose discrimination by edge computing decision unit, which leads to poor detection effect due to the lack of fusion of sensor data. In this regard, a high difficulty yoga body pose detection method based on data fusion is proposed. Complementary filtering, Kalman filtering and adaptive Kalman filtering methods are used to fuse sensor data for pose prediction, and a skeleton extraction model is constructed to build a multilayer perceptron network architecture to achieve the detection of difficult yoga asana poses. In the experiments, the proposed human posture detection method is validated. The analysis of the experimental results shows that the proposed method has a high comprehensive evaluation index and excellent detection performance when used for the detection of difficult yoga asana postures.

Keywords: data fusion; difficult yoga asana poses; human pose detection; Kalman filter algorithm

1. INTRODUCTION

Human pose detection, a current research hotspot in computer vision, has a wide range of applications in life, such as in video surveillance to ensure security in the public domain and in human-machine interaction to enhance the fluency between humans and machines[1]. Human pose detection is an algorithmic process that uses convolutional neural networks to detect key nodes of the human body in pictures or videos and then connects these key points. The complete human node information is obtained by connecting the different key points of the human body. Before applying deep learning algorithms to human pose detection, many algorithms used graphical structure-based algorithms to deal with the human pose detection problem. These methods are mainly based on local detectors, which use the principle that the model is built by the intrinsic connection between the key points of the human body. Although the recognition accuracy is improved, it is often susceptible to uncertainties such as shooting angle and illumination. In addition to this it is also susceptible to manual annotation and other factors such as edge features, directional gradient histograms, such annotation requires a lot of labor and material resources. Human pose estimation can usually be divided into two kinds of two-dimensional human pose and three-dimensional human pose. Since most of the images that need to be processed in real life are two-dimensional images, we generally understand human pose detection as two-dimensional human pose detection[2] . This paper also reviews human pose detection algorithms for two-dimensional conditions. According to the different application scenarios of different algorithms, human pose detection is divided into two cases: single-person and multi-person. The multi-person pose detection algorithm differs from the single-person pose detection algorithm in that the multi-person pose detection algorithm requires the detection of the key points of the human body in the picture and the accurate division of the key points of each person. Generally speaking, there are two approaches: top-down and bottom-up. The top-down approach includes both human detection and individual human keypoint detection, i.e., the human body is first detected in the image by the target detection algorithm, and then the keypoints of each human body are detected based on the detected human body. The bottom-up approach, on the other hand, contains two parts: key point detection and clustering combination of the detected key points, in which the key points of all people in the image are first detected, and then the detected key points are clustered and analyzed to further combine into different individuals. Yoga asanas are relatively static, and the movements require a certain degree of human skeletal positioning, so in computer vision, the detection of human skeletal points can make a preliminary judgment on the correctness of the movements. Regarding human skeletal point detection, Toshev et al. proposed Deep Pose as an early algorithm for human skeletal joint point detection. He transformed the human skeletal joint point estimation problem from the original image processing and template matching problem to CNN image feature extraction and key point coordinate regression, and used some regression criteria to estimate the occluded or non-appearing human joint nodes. However, the robustness of this method is poor, and the human body has complex and variable movements, thus leading to the weak applicability of this method. Human skeletal images can avoid the effect of illumination variation and the interference of external scene and background environment, so the combination of skeletal images and RGB images is used to identify yoga movements as well as criteria thus improving the robustness of the model[3] .

International Conference on Computer Graphics, Artificial Intelligence, and Data Processing (ICCAID 2022)
edited by Ruishi Liang, Jing Wang, Proc. of SPIE Vol. 12604, 126041F
© 2023 SPIE · 0277-786X · doi: 10.1117/12.2674850

2. DATA FUSION-BASED PREDICTION OF HIGHLY DIFFICULT YOGA ASANAS

Sensor data fusion refers to the multi-level, multi-faceted and multi-level processing of data from multiple sensors to produce new meaningful information that is not available from any single sensor. In this paper, data fusion of measurement data from two sensors is performed using complementary filtering, Kalman filtering and adaptive Kalman filtering methods, respectively, through several experiments to obtain more accurate data[4] .

The Kalman filtering algorithm is an optimal autoregressive data processing algorithm that provides optimal solutions in most application scenarios. Its most important feature is that it can estimate the state quantities of the system in real time and improve the performance of the filter by correcting the a priori estimates of the state quantities based on new observations using a recursive algorithm[5] .

The basic idea of the Kalman filter algorithm is to use a recursive form of calculation, that is, based on the estimated value of the previous moment x_{k-1} and the measured value of the current moment $z(t)$, to find the estimated value of the current moment $x_{k|k}$ with the principle of optimal estimation of the minimum mean squared error, the algorithm needs to establish two equations of system and measurement, the specific equation expression is shown below.

$$\begin{cases} x_k = A_k x_{k-1} + B_k u_k + w_k \\ z_k = H_k x_k + v_k \end{cases} \tag{1}$$

Where, x_k is the state vector of the system at time k, A_k is the system state transfer matrix, B_k is the input transition matrix of the system, u_k is the control quantity of the system at time k, w_k is the system noise, z_k is the observation quantity of the system at time k, H_k is the measurement-to-observation transition matrix of the system, and v_k is the observation noise[6] . Assuming that w_k and v_k are mutually independent, normally distributed white noise and the covariance matrices are Q_k and R_k , respectively, the following expression can be obtained.

$$\begin{cases} w_k \sim N(0, Q_k) \\ v_k \sim N(0, R_k) \end{cases} \tag{2}$$

It is under the condition that equation (2) holds that the Kalman filter algorithm uses the system equations to obtain the a priori estimate of the state vector. The correction of the prior estimate is then implemented using the measurement equation to obtain the a posteriori estimate of the state vector. Thus the Kalman filtering algorithm contains two parts, the time update equation and the measurement update equation, in each cycle. The expression of the time update equation is as follows.

$$x_{k|k} = A_k * x_{k-1} + B_k * u_k \tag{3}$$

$$P_{k|k-1} = A_k * P_{k-1} + Q_k \tag{4}$$

Eq. (3) is mainly used to obtain the a priori estimate of the current state $x_{k|k}$, while Eq. (4) is used to compute the a priori covariance $P_{k|k-1}$, both of which are prepared for the measurement update equation. The expression of the measurement update equation is shown below.

$$K_k = P_{k|k-1} * H_k * (H_k * P_{k|k-1} + R_k) \tag{5}$$

$$x_{k|k} = x_{k|k-1} + K_k * z_k \tag{6}$$

$$P_{k|k} = (1 - K_k * H_k) P_{k|k-1} \tag{7}$$

Eq. (5) calculates the Kalman gain K_k based on the prior covariance $P_{k|k-1}$, while Eq. (6) compensates the prior state estimate $x_{k|k-1}$ using the Kalman gain K_k and the measured value z_k to obtain the posterior state estimate $x_{k|k}$, and

Proc. of SPIE Vol. 12604 126041F-2

finally the covariance $P_{k|k}$ of the posterior state estimate $x_{k|k}$ is calculated by Eq. (7) for the time update equation[7] at the next moment.

Through the above steps, the prediction of difficult yoga postures by Kalman filtering algorithm can be realized, and the effective fusion of the predicted data can be completed to provide data support for the subsequent detection of difficult yoga postures.

3. BONE EXTRACTION MODEL DESIGN

The bone extraction model is to convert the ordinary RGB person image into a skeletal pose image. Some methods in deep learning can acquire the skeletal joint point images directly on the ordinary 2D camera captured images of people, so the skeletal images of human body can be extracted without any additional equipment. The bone extraction model is obtained based on the OpenPose model design. The model first detects the joint points of the person in the image, then clusters the detected joint points, and finally connects the joint points of the human body together[8]. The process of extracting skeletal data of yoga movements using OpenPose model is as follows.

Firstly, the input image is convolved through the first ten layers of VGG19 to generate the corresponding convolutional feature maps, then the generated convolutional feature maps are fed into the multi-order network for predicting the heat map of key points and describing the orientation of the connections of the joints, and finally the bipartite graph maximum weight matching algorithm is used to assemble the key points to obtain the human skeleton[9].

The output of the multi-order network in the Openpose model is obtained from $S^t = (S_1, S_2, ..., S_J)$ after generating the heat map of the body's joints; and from $L^t = (L_1, L_2, ..., L_C)$ after describing the connection direction of the joints. Where J is the number of joints in the human body; C is the number of associated regions. The associated area is the arm, leg, etc. S_J It refers to the heat map corresponding to the Jth joint point, which can be considered as a probability value; L_C is the direction corresponding to the Cth associated region .[10]

In predicting the joint point heat map, for the Jth joint point of each person, let its position be $x_j \in R^2$, then the true position is a two-dimensional Gaussian distribution centered at x_j and denoted by S_j^* . The true position corresponding to the Jth joint point is $S_j^*(p) = \max S_j^*(p)$, and P denotes a single position, i.e., the true position of a person's joint point is taken as the maximum value by pixel point. In predicting the direction of the connection of the joint points, for the cth association region, which can also be understood as the region connecting the joint points j_1 and j_2 , the true direction is denoted by L_c^* . If the position P is on this association region, otherwise it is a zero vector. The specific calculation formula is as follows.

$$L_c^*(p) = v = \frac{x_{j2} - x_{j1}}{\left\| x_{j2} - x_{j1} \right\|} \tag{8}$$

where v is actually a unit vector of j_1 pointing to j_2 . As long as P satisfies that it is on the line segment $j_1\ j_2$ or within a threshold distance from the line segment $j_1\ j_2$, it is considered that P is on that associated region. Finally, for all locations in a certain association region, each pixel point is processed using the averaging process with the following processing expression.

$$L_c^*(p) = \frac{1}{n_{c(p)}} \sum L_c^*(p) \tag{9}$$

where $n_{c(p)}$ represents the number of non-zero vectors at position p, i.e., only non-zero vectors are involved in the mean calculation. Once the model obtains the heat map of the nodes, a series of candidate points are identified for each node section using non-maximal value suppression. The combination of these candidate points with each other can generate a large number of possible association regions, so it is necessary to define the weights of the combination between two key points j_1 and j_2 . The specific calculation formula is shown below.

$$E = \int_{u=0}^{u=1} L_c(p(u)) \cdot \frac{d_{j2}}{d_{j1}} \tag{10}$$

where d_{j1} and d_{j2} denote the coordinates of j_1 and j_2 , respectively. Intuitively, if the direction of the points on the line segment coincides with the direction of the line segment, the greater the E, then the more likely it is that these two joints form an association region. So in this way it is possible to remove the association region[11] that is not there.

Through the above steps, we can convert the image of a person into a skeletal pose image, detect different joints of the person, and construct a skeletal node extraction model.

4. HUMAN POSTURE DETECTION ALGORITHM DESIGN

The human posture detection algorithm is mainly used to determine and classify the recognized human skeleton point data. 18 and 25 human skeleton point output formats are provided by Openpose, the difference between the two is mainly in the number of foot key points, because the foot key point information is not high for the final classification result image, but the increase of key points will greatly affect the recognition speed[12]. Therefore, the format of 18 key points is chosen here, which can increase the recognition speed as much as possible while ensuring the accuracy.

After obtaining the output of human skeletal points, a multilayer perceptron will be chosen to classify the output. Multilayer perceptron is developed from perceptron and is characterized by having multiple intermediate hidden layers relative to the perceptron, hence also known as deep neural network[13]. The architecture for classifying human skeletal point data using a multilayer perceptron is shown in Figure 1.

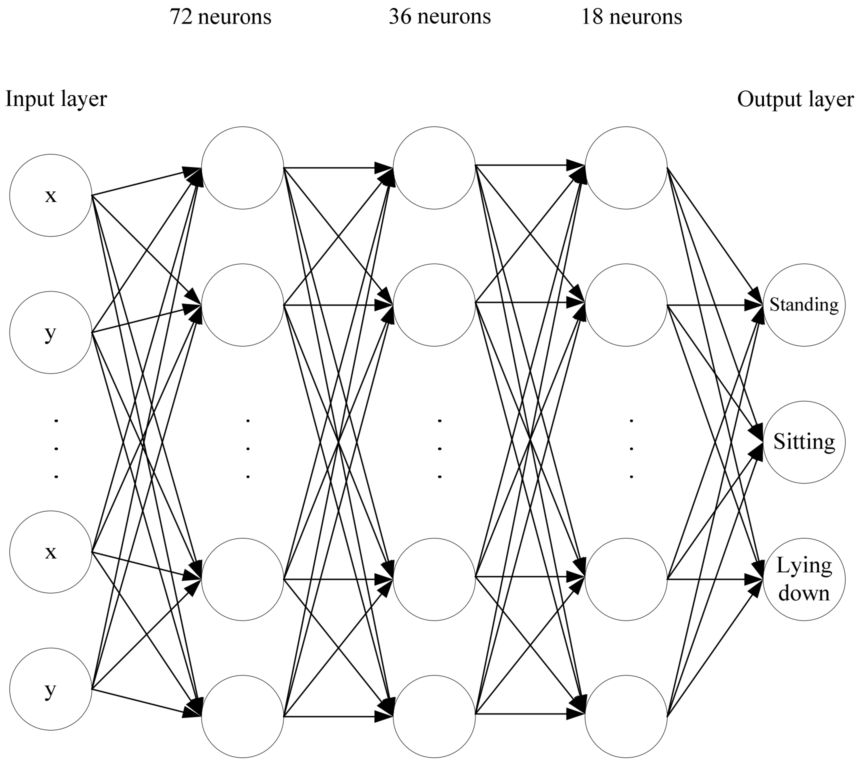


Figure 1. Multilayer perceptron network architecture for human posture classification

From Figure 1, it can be seen that since the inputs are the coordinates of the key points of the human skeleton, i.e., 18 (x, y) coordinates, there are 36 inputs to the multilayer perceptron. The outputs are the probabilities of 3 postures: standing, sitting, and lying. There are 3 hidden layers in between, where layer 1, layer 2, and layer 3 have 72, 36, and 18 neurons, respectively. Each input neuron is connected to the next layer neuron with a weight value w , and a bias value b. The expression of the relationship between input and output is as follows.

$$y_j^{h+1} = \sum_i w_j^{h+1} x_j^{h+1} + b_j^{h+1} \tag{11}$$

Where, y_j^{h+1} represents the output of the jth neuron at layer h+1, x_j^{h+1} represents the output of the jth neuron prior at layer h+1, w_j^{h+1} represents the weight value of the connection x and y, and b_j^{h+1} represents the corresponding bias value.

A ReLU function is used between two adjacent layers to perform a nonlinear transformation of the results of the previous layer. The multi-layer perceptron network is trained with back propagation to correct the parameters of the hidden layer. In the backpropagation process, the stochastic gradient descent (SGD) method[14] will be used in order to make the model training faster and increase the accuracy. First of all, the loss function needs to be defined, and the specific function expression is as follows.

$$J(\omega) = \frac{\sum_{j=1}^{m} [h_w(x_i) - y_i]^2}{2} \tag{12}$$

where m is the number of samples to be subjected to gradient descent, y_i is the i-th sample out of m samples, and $h_w(x_i)$ is the corresponding neuron output. ω is the parameter between two neurons. The objective is to minimize the value of the loss function $J(\omega)$, so to correct ω in the direction of its negative gradient, the required new parameter ω' is expressed as follows.

$$\omega' = \omega - \frac{\partial J(\omega)}{\omega_j} \tag{13}$$

However, the multilayer perceptron at this point still has some shortcomings. Firstly, it is difficult to choose the initial learning rate of the network, and secondly, the learning rate of the network is limited by the pre-defined adjustment rules and the same learning rate for each parameter. Therefore, in this case, the network is optimized using an adaptive gradient scheme. Its formula for making corrections to the network parameters is as follows.

$$g_i = \frac{J(\omega_{t-1})}{\omega} \tag{14}$$

$$w_{t-1} = \omega_t - \alpha \tag{15}$$

Where: g_i denotes the gradient at time step t, ω_t denotes the parameter at time step t; α is the learning rate of the network. From Equation (13), the parameter ω_t has an opposite relationship with the gradient g_i , and the change of g_i in the denominator will affect the learning rate α . For the frequently changing parameter g_i will accumulate in the denominator, thus making the parameter ω_t updated more and more slowly, while the sparsely changing parameter g will make the parameter ω_t updated faster, so the use of adaptive gradient allows each parameter to adjust the learning rate by itself according to the different situations. And while performing gradient descent, the problem of gradient disappearance or gradient explosion can occur. At this time, using the batch normalization method to process the input values of each hidden layer after linear activation can effectively improve the training speed of the model, speed up the convergence process, and also enhance the final detection effect[15].

The above steps will complete the design of the difficult yoga asana pose detection algorithm, combining the contents of this section with the above mentioned bone extraction model and pose prediction related contents, so that the design of the difficult yoga asana pose detection method based on data fusion is completed.

5. TESTING AND ANALYSIS

5.1 Test PREPARATION

In order to prove that the proposed data fusion-based difficult yoga asana pose detection method outperforms the traditional human pose detection method in terms of detection accuracy, an experimental test session is constructed to verify the actual detection effect of this yoga asana pose detection method after the theoretical part of the design is completed. In this experiment, two traditional human pose detection methods are selected for comparison, namely the deep learning-based human pose detection method and the multimodal information fusion-based human pose detection method.

In the experiment, pictures of yoga poses were collected from 400 volunteers according to the MOOC's course schedule. Four of these sets of poses were used so far, and each picture was labeled with a movement category and score level. It should be noted that the subjects in the training set and the test set were different in order to eliminate individual differences. The yoga poses dataset was created to facilitate yoga poses judging and to verify the accuracy of single and joint models for yoga poses recognition and scoring. The CPU model of the desktop computer used in the experiment is Intel Core i5 -4460 m with 3. 2GHz, GPU model is GXT 1060, memory is 16GB, operating system is Windows 10 , Python version number is 3.6, and Pytorch framework is used.

In order to verify the effectiveness of the data fusion-based method for detecting difficult yoga postures, this experiment chose to validate four yoga postures, Mountain Pose, Walking Stick Pose, Phantom Chair Pose and Supine Rising Leg Pose, and to illustrate the evaluation index of each posture. In order to make the yoga posture database more reliable, all the data in this paper were obtained from the volunteer pictures of MOOC and online teaching. The pictures that met the requirements were selected and stored in the database according to different categories. The evaluation criteria were based on the mean scores of the three teachers' ratings. The collected dataset is shown in Table 1, where rating refers to the rating level of yoga poses, which is divided into three levels: excellent, good and medium; quantity refers to the number of images present in the dataset. Finally 4,800 valid images were collected in the yoga posture dataset, and 400 images for each yoga posture action under each rating. These images were finally used as the dataset for the three detection methods in this experiment.

Table 1 Yoga posture data set

Yoga postures	Rate	Number
Mountain Pose	Excellent	400
	Good	400
	Medium	400
Walking Stick Pose	Excellent	400
	Good	400
	Medium	400
Phantom Chair Pose	Excellent	400
	Good	400
	Medium	400
Supine Pose	Excellent	400
	Good	400
	Medium	400

In order to verify the effect of the pose detection model, the training dataset is fed into the skeleton extraction model to extract the corresponding skeletal pose. At this time, the training data is divided into RGB data and skeletal data. The RGB data and skeletal data are data processed in the same proportion to generate multimodal data. The amount of data is the same as RGB data and skeletal data. These three types of data are fed into the joint model in the same way for training, and the training process uses the same data enhancement method, the same learning rate, and the same training method. Finally, the trained model is validated with test data separately using the trained model.

5.2 ANALYSIS OF TEST RESULTS

The evaluation index chosen in this paper is the detection accuracy of the human pose detection algorithm, and the specific measure is the experimentally set comprehensive evaluation index S. S is derived from the determination of the yoga pose category and the determination of the score. The S metric for each detection model is calculated as follows.

$$S = \frac{1}{n}\sum_{i} A_i W_i \tag{16}$$

Where, n represents the yoga category, A_i represents the category accuracy of the i-th category of yoga postures, and W_i represents the score accuracy of the i-th category of yoga postures. The combined evaluation values of different algorithms were obtained according to the above formula, and the experimental results shown in Figure 2 were obtained.

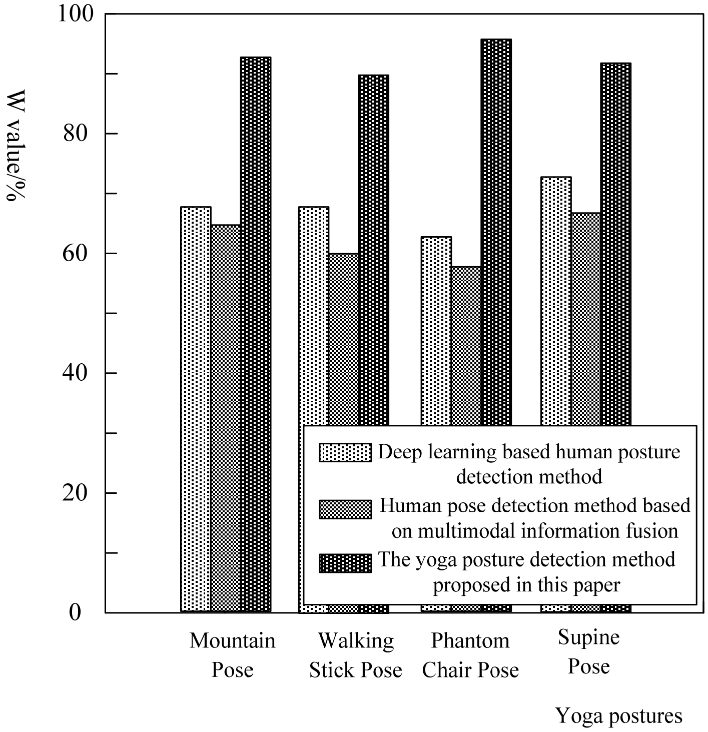


Figure 2. Comprehensive index comparison

According to the above experimental results, it can be seen that although skeletal data has a greater advantage in judging action categories, some joint points may not be collected when extracting bones, so the final recognition effect of using skeletal data alone does not reach the optimum. The human posture detection method based on deep learning has a greater advantage in distinguishing action categories, but when evaluating action scores, it cannot effectively evaluate the action scores due to the interference of clothing environment; while the yoga posture detection model proposed in this paper shows a greater advantage in combining the advantages of RGB data and skeletal data, and at the same time using one model to complete the recognition task In this paper, we can quickly and accurately classify and score yoga poses.

6. CONCLUDING REMARKS

The data fusion-based method of high difficulty yoga asana pose detection proposed in this paper can effectively improve the model accuracy by predicting the yoga asana pose and constructing a skeleton extraction model, while fusing RGB data as well as skeleton data with high robustness and real-time performance. In future research work, it is also necessary to use scene information for human motion recognition, fuse the results of object recognition and scene analysis, and generate contextual information of the surrounding environment, which can be used to guide human behavior recognition and analysis.

REFERENCES

[1] Wang B, He W, Yang Z, et al. An Unsupervised Sentiment Classification Method Based on Multi-Level Fuzzy Computing and Multi-Criteria Fusion [J]. IEEE Access, 2020, 8:145422-145434.

[2] Yp A, Wl A, Xlb C, et al. Integrated fusion framework based on semicoupled sparse tensor factorization for spatio-temporal-spectral fusion of remote sensing images [J]. Information Fusion, 2021, 65:21-36.

[3] Wang H, Guo H, Zhang K, et al. Automatic sleep staging method of EEG signal based on transfer learning and fusion network [J]. Neurocomputing, 2022, 488:183-193.

[4] Zhu Z, Arezki Y, Cai N, et al. Data Fusion-based Method for the Assessment of Minimum Zone for Aspheric Optics [J]. Computer-Aided Design and Applications, 2020, 18(2):309-327.

[5] Yao G, Yin Y, Li Y, et al. High-precision and wide-wavelength range FBG demodulation method based on spectrum correction and data fusion. [J]. Optics express, 2021, 29(16):24846-24860.

[6] Liu S, Wu J, Zhang X, et al. Research on data classification and feature fusion method of cancer nuclei image based on deep learning [J]. International Journal of Imaging Systems and Technology, 2022, 32(3):969-981.

[7] Song X, Zheng B, Tan Y, et al. Dynamic Measurement Method of Near-Bit Borehole Trajectory Parameters Based on Data Fusion [J]. Petroleum Drilling Techniques, 2022, 50(1):38-44.

[8] A Y M, Envelope J H A P , B J C , et al. A novel grid generation method based on multi-resolution data fusion for 2D shallow water models [J]. Journal of Hydro-environment Research, 2022, 45:29-38.

[9] Envelope S D A, B J S, A J C F, et al. Optimized in-vehicle multi person human body pose detection [J]. Procedia Computer Science, 2022, 204:479-487.

[10] Cai W Y, Guo J H, Zhang M Y, et al. GBDT-Based Fall Detection with Comprehensive Data from Posture Sensor and Human Skeleton Extraction [J]. Journal of Healthcare Engineering, 2020, 2020(9):1-15.

[11] Khanian M, Golpayegni S, Rostami M. A new multi-attractor model for the human posture stability system aimed to follow self-organized dynamics - ScienceDirect[J]. Biocybernetics and Biomedical Engineering, 2020, 40(1):162-172.

[12] Qu J, Wu C , Li Q , et al. Human Fall Detection Algorithm Design Based on Sensor Fusion and Multi-threshold Comprehensive Judgment [J]. Sensors and materials, 2020, 32(4):1209-1221.

[13] Han J, Song W, Gozho A , et al. LoRa-Based Smart IoT Application for Smart City: an Example of Human Posture Detection [J]. Wireless Communications and Mobile Computing, 2020, 2020(2):1-15.

[14] Rocha-Ibarra E, Oros-Flores M I, Almanza-Ojeda D L, et al. Kinect Validation of Ergonomics in Human Pick and Place Activities Through Lateral Automatic Posture Detection [J]. IEEE Access, 2021, 9:109067-109079.

[15] Lv Z, Dong Z, Liu X, et al. Intelligent Multi-Source Heterogeneous Structured Data Characteristics Simulation[J]. Computer Simulation,2022(6)451-455,501.

The improved clustering algorithm is used to analyze the data anomalies in the network environment

Xiaojia Lin
Fujian Business College, Fuzhou, Fujian, 35000, China
Qwe1570414@163.com

ABSTRACT

In this paper, an improved clustering algorithm is proposed and a heterogeneous model based on this model is developed. A new data extraction technology, such as data classification, network platform anomaly detection, distributed maximum frequent sequence extraction, comparison and mining of maximum frequent sequence data, is adopted. Through the comparison experiment, it is found that the algorithm can better reflect the correlation with the corresponding abnormal data types, and can better reflect the actual use of the algorithm.

Keywords: improved clustering algorithm; Network platform; Data mining; Abnormal Data

1. INTRODUCTION

With the popularization of computer technology, the development of the Internet platform presents a trend of "global integration". The huge Internet system supports the spread and dissemination of the Internet, which brings great convenience for the intrusion of the Internet. In view of the increasingly complex network platform and the increasingly changeable network attack mode, how to identify and obtain these abnormal information more quickly and more accurately, so as to reduce the damage caused by abnormal information to the operation of network platform? In this context, scholars have carried out in-depth discussions, analyzed various kinds of data, and given a variety of different mining ways. For example, the commonly used data mining technology is to carry out data mining through K-means clustering algorithm. This method can be used to effectively classify unlabeled data, reduce the time consuming of labeling, and improve the speed of mining [2]. However, this method has certain limitations. Although it has good practical value, it must be classified first, and it is highly sensitive to noise, so the conditions for its appearance and use become more and more complex, resulting in an increasing bit error rate, which can no longer adapt to the security operation requirements of network system [3]. Therefore, on the basis of the analysis of the clustering algorithm, combined with the improvement of the clustering algorithm, the algorithm is deeply discussed.

2. DESIGN OF ABNORMAL DATA MINING METHOD FOR NETWORK PLATFORM BASED ON IMPROVED CLUSTERING ALGORITHM

With the popularization of modern computer technology, the development of the Internet presents a global trend. The huge Internet system brings great convenience for people's information exchange and transmission, but also creates a lot of favorable environment for people who illegally invade the Internet. In view of the increasingly complex network platform and the increasingly changeable network attack mode, how can we identify and obtain these abnormal information more quickly and more accurately, so as to reduce the damage caused by abnormal information to the operation of network platform? In this context, scholars have carried out in-depth discussions, analyzed various kinds of data, and given a variety of different mining ways. For example, the commonly used data mining technology is to carry out data mining through K-means clustering algorithm. This method can effectively cluster unlabeled data, reduce the time consuming of annotation, and improve the speed of mining [2]. However, this method has certain limitations. Although it has good practical value, it must be classified first, and its sensitivity is very high. Therefore, the conditions of its emergence and use are becoming more and more complex, leading to its increasing bit error rate, and it cannot adapt to the security operation requirements of the network system [3]. Therefore, on the basis of the analysis of the clustering algorithm, combined with the improvement of the clustering algorithm, the algorithm is deeply discussed.

2.1 Network platform running data classification based on improved clustering algorithm

In order to accurately discover abnormal information in the network environment, it is necessary to determine the types of data that can be mined. Only under specific types can the correlation and consistency between the final mining and the

International Conference on Computer Graphics, Artificial Intelligence, and Data Processing (ICCAID 2022)
edited by Ruishi Liang, Jing Wang, Proc. of SPIE Vol. 12604, 126041G
© 2023 SPIE · 0277-786X · doi: 10.1117/12.2674851

corresponding data types be guaranteed, so as to obtain the expected mining results. A new clustering method is used to classify abnormal data. In other words, under the operating environment of the network platform, several nodes are randomly selected as the center point of the cluster. As described in Equation (1), a single data and the spacing of the center point are calculated:

$$j = \sqrt{\sum_{i=1}^{n}(t_1 - t_2)^2} \tag{1}$$

In Formula (1), j represents the distance between a specific network platform data and a central point. t1 and t2 represent two eigenvalues, and i represents the number of a parameter. According to formula (1) above, the classified clusters are updated. The next step is data mining based on the key cluster, and on this basis, the spacing between each node and the core node is analyzed. The longer the distance, the more likely this type of data is to be abnormal data, and the closer the distance, the less likely it is to be abnormal data, so as to achieve the classification of the operation data on the network.

2.2 Detecting network platform anomalies

After the classification method of running data of network platform is determined, the actual running environment of network platform can be detected. This process can be selected from MapReduce, where the specific exception detection and parallel processing flow is shown in Figure 1.

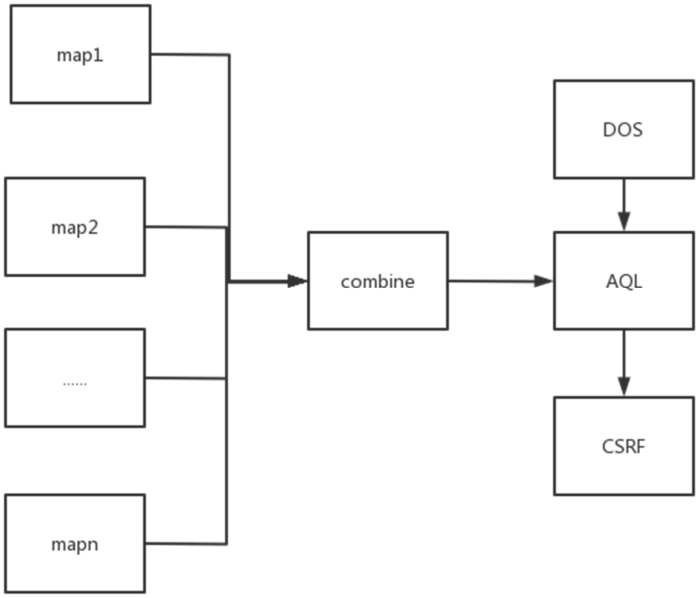


Figure 1: Schematic diagram of network platform anomaly detection and parallelization process

When detecting the anomalies of the network platform, the map function can be used to determine the abnormal type of the data. The operation data classification method of the network platform designed is introduced into the map, each data that needs to be detected for anomalies is allocated, and y mappings are sent in < key, value>. In the operating environment of the network platform, the distance between each anomaly detection data and k is calculated from the three-dimensional perspective, and the minimum distance is determined. In the detection process, if abnormal data is found, the alarm processing function can be used to analyze the operation records and rules of the network platform, and determine whether the system has triggered the rule determination. If so, the response will be fed back to the network platform client. If not, check whether any rules are processed. If yes, go back to the previous step. If it is not triggered, the alert is determined to be repeatable, if it is repeatable, it is updated, if not, it generates a new alert and a flag is added to the alert, thus completing the anomaly detection of the entire network platform.

2.3 Maximum frequent sequence extraction of distributed network platform

After mastering the abnormal resources on the network platform, the text of the web page is recorded, the captured content is preprocessed, and the URL of the server where the platform belongs is saved, and it is matched with the time series. The algorithm trims the data stream at the network level and reduces the amount of computation in the current data stream extraction process. In this process, the historical data of the platform terminal is used as a reference, the user's interest is analyzed through the access times of different programs, and some historical data in the text is deleted and screened according to the user's interest consumability of various information, and the recent access behavior is determined by the user's weight, and the following formula (2) is obtained: Through the weighted analysis of the model, the following formula (2) is obtained.

$$\mathrm{W} = \frac{1}{1+\beta} \tag{2}$$

Formula (2): W represents the user's action data screened by the web platform (belongs to the past); Decay factors representing the final data; tc represents a series of present moments; t' represents a series of representations of the present moment. Since the damping factors of the network terminals of each site are very different, we must choose a larger damping factor to make the decision. According to the above methods, we complete the extraction process of the maximum distribution frequency. The first step is to capture the behavior of the data set and perform the initial screening. The second step is to convert the data of the data stream into the time scale that can be directly calculated, and then use the network spider to obtain the maximum time scale, and tc in the expression given above; The third step is to transform IP and corresponding data, and then preprocess the data, and then build the WASD database according to the characteristics of the dialogue; The fourth step is: create a data table, fill in a data table, and then according to the user access process to obtain the access order of each data; The fifth step, according to the above calculation formula, calculate the user access interface weight and user interest relationship; In the sixth stage, the end users of the platform are screened by referring to the MPFs method, and then sent to the database. The other data that has been screened is input to the terminal, and then the maximum frequency array is generated by using the timing sequence of the information, and the sequence value of the data is taken as the final extraction.

2.4 Comparison and mining of maximum frequency sequences

On this basis, the comparison of the maximum frequency sequence is divided into two steps by comparing and exploring the maximum frequency. The first step is to use the maximum frequency sequence to obtain users' daily life, simulate the hacked test data, and compare the array and the gene sequence according to the collected gene sequence [10][10]. The method can compare the data concatenation with the normal working mode of the user to find out whether there is any unusual phenomenon in the running state of the network terminal. The second step is based on the previous step. It can be said that after the comparison of the previous step, the data obtained is relatively similar to the length of the original array sequence. Therefore, we can refer to the direct dialogue method of biology and use the vector routine method to calculate the Haiming spacing of the sequence accessed by the user. Can carry out the range calculation. Then, starting from the dynamic data plan, by studying the optimal session sequence, we can combine the required support decision with the required calculation range, and judge whether the decision behavior can stably interact with the information of the end user according to the required distance. NW algorithm is used to compare the similarity between user sequences, and then the data in the interval is compared with the data in this region, and the data in this region is compared for further mining of future data, and it is compared with other data to provide sufficient basis for future data mining work.

On this basis, we also introduce the Combine and Reduce functions. Through mining different isomerism, we make statistics of different isomerism, and then import different types of isomerism information into the same Reduce respectively. By analyzing the data, the types of abnormal data in the data warehouse are obtained, and the remaining data of the data warehouse are obtained, and the data are processed accordingly, so as to obtain all the data of the whole network.

3. COMPARISON EXPERIMENT

In order to test whether the proposed method can discover the abnormal information of network platform in reality, and the mining results are correct and efficient. In the test, we selected KDDCup99 as the test object, which included five commonly used attack types, including Normal, DOS, U2R and other five different attack types, and divided them into I, II, III, IV and V. Before the test, the test data must be preprocessed to ensure that both methods can get the best results and make the test get better results. For the discrete attributes in the test data group, they are converted into continuous features, and then normalized and normalized, and the following formula is obtained:

$$X_v = \frac{X_u - AVG}{STAD} \tag{3}$$

In Formula (3), represents the experimental data set after standardization and normalization; Xij stands for unprocessed experimental data set; AVG represents the average value of the experimental data set; STAD represents the mean absolute deviation in a set of experimental data. From the above selected data sets, they were divided into experimental test group data and experimental training group data in a ratio of 1:2, and they were divided into five groups according to their attack types for experiment. Among them, each group has 1000 normal data and random abnormal data. Table 1 shows the specific situation of the five experimental data groups.

Table 1: The specific situation of the five experimental data groups

Setoftests	AttackType	Numberofrecords	Normalnumberofrecords	Numberofexceptionrecords
Test1	Type 1	1125 pieces	1000pieces	125pieces
Test2	Type 2	1105 pieces	1000pieces	105pieces
Test3	Type 3	1098 pieces	1000pieces	98pieces
Test4	Type 4	1115 pieces	1000pieces	115pieces
Test5	Type 5	1108 pieces	1000pieces	108pieces

On the basis of the specific conditions of the experimental data group in Table 1, the mining methods of the experimental group and the control group were respectively used to mine the five test sets for the convenience of comparison.

Table 2 Comparison results of recognition rate between experimental group and control group

Datasource	Recognitionrateofexperimentalgroup	Recognitionrateofcontrolgroup
SJY-01	92.65%	65.26%
SJY-02	98.56%	48.65%
SJY-03	94.26%	82.36%
SJY-04	98.62%	45.28%
SJY-05	97.56%	44.27%

As can be seen from Table 2, the recognition rate of the experimental group is significantly higher than that of the control group, and will not be affected by interference data in the data source. The correlation between the mining results and the corresponding attack types is taken as the evaluation index, and its calculation formula is as follows:

$$\sigma = \frac{t}{n \cdot K} \tag{4}$$

In the formula, the correlation between the mining results and the corresponding attack types; This paper describes the sum of the maximum distribution frequency of each node in the mining by proper clustering and classification. Represents the total number of data in the test data set; Indicates the type of the data to be attacked. On this basis, the correlation between the experimental group and the control group in various test sets is obtained respectively. The higher the correlation, the higher the degree of matching between the excavation results and the corresponding attack data, and the higher the use value of the excavation results. The smaller the correlation is, the worse the matching degree of its corresponding attack data is, and the less the use value of its excavation results is. By combining the above discussion, the conclusion of the experiment is drawn.

From the two charts and the previous analysis, it can be seen that among the data of the five test sets, the data mining of the experimental group has a correlation of 0.80, which indicates that the data mining of the experimental group has a good correlation with the corresponding attack types, and has a high mining application value. The correlation between the five test sets of the control group was 0.70~0.40, which significantly reduced the mining effect of the experimental group, indicating that the mining results of the control group were less consistent with the corresponding aggressive data, and its mining value was also lower. The practice shows that the improved clustering algorithm can better discover the abnormal information in the real network environment, so as to lay the foundation for the formulation of the system operation and management strategy.

4. CONCLUSION

The superiority of this new data mining technique in practice is proved by comparison experiment. The mining algorithm described in this paper can be used to mine all kinds of attack data accurately and improve the stability performance of the system. Because in the future system, because the working conditions of the system are more and more complex, the intrusion mode of the system will be more and more, and the abnormal information of the system will also change. Based on the above problems, this paper will conduct more exploration and exploration in the future work, in order to improve the performance of clustering algorithm.

REFERENCES

[1] Wang Han, Zhang Feng, Xue Huifeng. Mining and Reconstruction of industrial water intake anomaly Data based on Wavelets-Support Vector Machine [J]. Computer Applications and Software, 201,38(05):61-68+81.

[2] Kuang Hua, He Xin, He Mi, et al. Detection of Abnormal Voltage Data in Distribution Network Based on Bidirectional Long Term and Short Term Memory Neural Network [J]. Science Technology and Engineering, 201,21(24):10291-10297.

[3] Liu Shengwa, Zhou Yajie, Gao Xiang, et al. Design and implementation of downhole anomaly warning platform based on Big data technology [J]. Internet of Things Technology,20,10(03):67-69.

[4] XU Changyun. Research on Library Big Data Analysis and Collaborative anomaly Detection Platform under the Internet [J]. Automation Technology and Application,20,39(07):133-136.

[5] Ouyang Bulgarelli. Risk Identification from Abnormal Financial Data of P2P online lending Platform -- A Case study of T Company [J]. Chinese Business Theory,2019(17):162-163.

[6] WANG Songqing. Application of data encryption and abnormal data Self-destruct Technology in network information security [J]. Electronic Technology and Software Engineering,2021(14):252-253.

[7] Wan Lei, Chen Cheng, Huang Wenjie, et al. Power Anomaly Detection Method Based on BRB and LSTM Networks [J]. Electric Power Construction, 201,42(08):38-45.

[8] Xu Lei, Wang Jianxin. Anomaly Network data Mining Algorithm based on Fuzzy Neural network [J]. Computer Science,2019,46(4):73-76. (in Chinese)

[9] Yang Wei. Malicious Intrusion data mining Method of Cloud Computing Platform Ship Communication network [J]. Ship Science and Technology,20,42(20):104-106.

[10] Wu Junjie. Research on data mining method of abnormal nodes in high-load grating sensor network [J]. Laser Journal,2019(2):68-72. (in Chinese)

[11] Denning D E. An intrusion-detection model[J]. IEEE Transactions on software engineering, 1987 (2): 222-232.

[12] Chandola V, Banerjee A, Kumar V. Anomaly detection: A survey[J]. ACM computing surveys (CSUR), 2009, 41(3): 15.

[13] MacDonald N. Information Security is Becoming a Big Data Analytic Problem[J]. Gartner, (23 March 2012), DOI= http://www. gartner. com/id, 2012, 1960615.

[14] Pablo Salazar. OpenSOC: An Open Commitment to Security [EB/OL]. https://blogs.cisco.com/security/opensoc-an-open-commitment-to-security, 2014.

[15] Huawei. CIS Network Security Intelligent System [EB/OL]. http://e.huawei.com/cn/products/enterprise-networking/security/bigdata-apt/cis, 2017.

[16] Huang Yuan-fei, Ji-yong, JIN Li-ping. Discussion on the Situation of Network Information Security and Related Hot Issues [J]. Department of Telecommunications Science, 2009, 25 (02) : 16-20.

[17] Design and implementation of Network Abnormal behavior detection System based on Big Data technology [D]. Jiangxi University of Finance and Economics,2016.

[18] Liu Jing, Gu Lize, Niu Xinxin, Yang Yixian. Research on Network anomaly detection based on single classification Support vector Machine and active learning Journal of Communications,2015,36(11):136-146. (in Chinese)

[19] Qian Yekui, Chen Ming, Ye Lixin, Liu Fengrong, Zhu Shaowei, Zhang Han. Whole-network anomaly detection method based on Multi-scale Principal Component Analysis [J]. Journal of Software,2012,23(02):361-377.

[20] Kayacik H G, Zincir-Heywood A N, Heywood M I. Selecting features for intrusion detection: A feature relevance analysis on KDD 99 intrusion detection datasets[C]. Proceedings of the third annual conference on privacy, security and trust. 2005.

A Pneumonia Classification Accuracy Enhancement Scheme Based on Data Augmentation, Data Preprocessing and Transfer Learning

Tianfu Mao [1,a], Teng Li [2,b], Yu Zhang [2,c]
[1] Computer Science and Technology, Dongguan open University, Guangdong Dongguan, China, 523000
[2] College of Computer and Information Science, Southwest University, Chongqing, China, 400715
[a]tellyouwhat@qq.com
[b]liteng0264@email.swu.edu.cn
[c]zhangyu@swu.edu.cn

Abstract.

As an infectious disease, pneumonia can cause great harm to human health. If pneumonia can be detected and treated early, its harm will be greatly reduced. Previously, hospitals relied on specialized doctors to diagnose diseases, but with advances in computer technology, deep learning is widely used in the medical field. In recent years, many excellent pneumonia classification methods have been proposed. They can judge whether a patient is infected with pneumonia based on their chest X-ray image, which effectively solves the shortage of professional doctors. In this paper, a convolutional neural network was proposed for pneumonia classification, and the pneumonia classification model was trained based on 1211 real chest X-ray image provided by Third Military Medical University. Experimental results on the test set show that the convolutional neural network proposed in this paper is not dominant, and its classification accuracy is only 72.0%, which is lower than the other three pneumonia classification models compared. Therefore, this paper integrates data enhancement, data preprocessing, transfer learning technology, then proposes a pneumonia classification accuracy enhancement scheme. This scheme improves the classification accuracy of the proposed pneumonia classification model from 72.0% to 82.0%, and the classification effect exceeds the other three pneumonia classification models compared.

Keyword: Pneumonia Classification; Classification Accuracy; Data Augmentation; Transfer Learning

1. Introduction

Pneumonia refers to inflammation of the lungs caused by bacteria, viruses or other factors. It is a common but dangerous disease. The early symptoms of pneumonia are similar to the common cold. If the patients receive treatment as soon as possible and the immune system is not damaged, they will recover quickly. If the patient is not diagnosed and treated in time, it is easy to aggravate the disease. In addition, pneumonia is also contagious, with the elderly, young children, and people with low immunity more likely to be infected. According to the statistics from World Health Organization, pneumonia kills more children each year than measles, dysentery, and malaria combined. In developing countries, tens of thousands of children die from pneumonia every year, mainly because of lack of medical resources. A professional doctor can accurately diagnose pneumonia, but the training cost of doctors is too high. Therefore, hospitals urgently need computer-aided diagnosis technology to identify pneumonia quickly and accurately to reduce the mortality rate of pneumonia patients. Pneumonia can be diagnosed by chest X-ray, lung CT, chest ultrasound, and chest magnetic resonance imaging. Chest X-ray is one of the best ways to detect pneumonia due to its lower side effects, shorter time-consuming, higher-quality images, and cheaper equipment.

Deep learning, as the main branch of machine learning, has shown great potential in natural language processing and computer vision. In the deep learning field, Convolutional Neural Network(CNN) has achieved remarkable success in image-based tasks such as classification, detection, and so on. Unlike machine learning algorithms, CNN can learn more abstract features from large-scale image datasets. Although deep learning-based computer-aided diagnosis techniques cannot completely replace well-trained doctors, they can complement clinical decision-making, and in some medical fields, CNN models even have expert-level diagnostic capabilities. In view of the above advantages, this paper uses CNN to complete the pneumonia classification task.

In this paper, a pneumonia classification accuracy enhancement scheme-model feature enhancement scheme is proposed, which can improve the effect of CNN on pneumonia classification tasks. In the proposed model feature enhancement

International Conference on Computer Graphics, Artificial Intelligence, and Data Processing (ICCAID 2022)
edited by Ruishi Liang, Jing Wang, Proc. of SPIE Vol. 12604, 126041H
© 2023 SPIE · 0277-786X · doi: 10.1117/12.2674701

Proc. of SPIE Vol. 12604 126041H-398

scheme, there are three parts, which are the data enhancement part, data preprocessing part, and fine-tuning of the convolutional neural network part. In the data enhancement part, random rotation, offset, scaling and other operations are used, the purpose is to increase the samples in the dataset and improve the generalization ability of the pneumonia classification model. In the data preprocessing part, techniques such as Gaussian filtering and histogram equalization are used to filter image noise, so that the contrast of the chest X-ray image is stronger and the lung area is highlighted. In the fine-tuning part, transfer learning technology is used. After the data enhancement part, the dataset input to the network has been expanded many times, so transfer learning techniques are needed to improve the fitting ability of the pneumonia classification model. To verify the model's performance, this paper tests it on the test set. After using the model feature enhancement scheme, the experimental results show that the accuracy, F1 score, recall rate, and other metrics of the pneumonia classification model are improved.

The other contents of this paper are as follows: Section 2 gives the related work of other authors in diagnosing pneumonia or other diseases, including their methods, datasets, and achievements. Section 3 introduces the datasets used in this paper and the evaluation metrics used in the experiments. The model feature enhancement scheme proposed in this paper is given in section 4. In section 5, experimental results and analysis are given. Finally, Section 6 concludes the paper.

2. Related Works

With the rapid development of deep learning technology, it has been widely used to assist disease diagnosis. Grewal et al. [1] used deep learning technology to detect intracerebral hemorrhage in brain CT images. Salido et al. [2] used deep learning technology to detect whether patients had skin cancer in dermoscopy images. Ragab et al. [3] combined deep learning technology and a Support Vector Machine to detect breast cancer. Gao et al. [4] used deep learning technology to detect brain tumors in brain CT images. This section focuses on the achievements of deep learning applications in pneumonia classification.

Jain et al. [5] proposed two CNN models, and selected four classic classification models - VGG16, VGG19, ResNet50, and Inception-V3, and made them trained on the public chest X-ray image dataset provided by Kaggle. The results on the test set show that the accuracy of the six models is 85.26%, 92.31%, 87.28%, 88.46%, 77.56%, and 70.99%, respectively, and the second network they proposed achieves the best results. Compared with another proposed network, it has one more convolution layer to extract features, and also adds dropout after the fully connected layer, and reduces the learning rate. In addition to its excellent performance in the accuracy index, it also has a recall rate of 98% and an F1 score of 94%. Compared with the other five models, the over-fitting situation is also the least, so it can help radiologists to diagnose pneumonia efficiently and accurately.

Rajpurkar et al. [6] proposed the CheXNet algorithm. It is a convolutional neural network with 121 layers. The authors trained it on the ChestX-ray14 dataset, which is the largest public chest X-ray image dataset with more than one hundred thousand frontal chest X-ray images covering 14 diseases. The final test results showed that CheXNet exceeded the average level of radiologists in F1 score, and had a good predictive ability for 14 lung diseases.

Due to the global outbreak of new coronary pneumonia in 2020, radiologists have become a scarce resource in hospitals, so it is necessary to use deep learning technology to assist doctors in diagnosing new coronary pneumonia quickly and efficiently. In this context, Toraman et al. [7] proposed Capsnet, which can detect COVID-19. Capsnet is not a traditional convolutional neural network stacked layer by layer, but Capsule Networks. The classification accuracy of Capsnet is as high as 97.24% on the binary classification (new coronary pneumonia and normal) task, and the classification accuracy on the multi-classification (new coronary pneumonia, normal, and pneumonia) task also reaches 84.22%. Therefore, Capsnet can assist doctors in diagnosing new coronary pneumonia and relieve the pressure of diagnosis in hospitals.

Hashmi et al. [8] proposed an efficient pneumonia detection model using a weighted classifier-based method. The method integrates a series of excellent deep learning models for weighted prediction and outperforms all individual models. In addition, it also used transfer learning to fine-tune the model for higher classification accuracy, and data augmentation techniques were used to balance the proportions of different classes within the original dataset. Finally, the model achieved a classification accuracy of 98.43% on the test set.

Yu et al. [9] proposed CGNet, which can classify pneumonia and normal from chest X-ray images. CGNet consists of three parts: feature extraction, graph-based feature reconstruction, and classifier. CGNet first trains other excellent convolutional neural networks with transfer learning techniques and takes the learned features as input for the next two parts. Then the graph-based feature reconstruction recombines the features to generate new features. Finally, a shallow convolutional neural network GNet is used to classify c pneumonia. CGNet achieved an accuracy of 98.72%, a sensitivity of 1, and a specificity of 97.95% on the public chest X-ray image dataset.

3. Materials and Methodology

3.1 DataSet

The dataset used in the paper comes from the Third Military Medical University. All chest X-ray images were selected by professional radiologists and manually labeled to ensure the correctness and representativeness of the dataset. To better identify pneumonia, all chest X-ray images were screened a second time, removing low-quality images for quality control. The final dataset has a total of 1211 chest X-ray images, including 611 pneumonia samples and 600 normal samples.

In addition to the private dataset, this paper also downloads the public chest X-ray image dataset with a total of 5856 images on Kaggle.

3.2 Performance Metrics for Classification

In the field of the pneumonia classification task, in addition to accuracy and loss, there are other evaluation metrics, such as recall rate, precision, F1 score, AUC value, and ROC curve. Next, the meaning and calculation formula of these metrics will be introduced:

TP (True Positive): indicates the number of pneumonia samples predicted as pneumonia;

TN (True Negative): indicates that the number of normal samples predicted as normal;

FP (False Positive): indicates that the number of normal samples predicted as pneumonia;

FN (False Negative): indicates that the number of pneumonia samples predicted as normal;

Recall rate: indicates the proportion of correctly predicted pneumonia samples to the total pneumonia samples. The calculation formula is given in (1):

$$Recall = TP/(TP + FN) \tag{1}$$

Precision: indicates the proportion of correctly predicted pneumonia samples to the actual predicted pneumonia samples. The calculation formula is given in (2):

$$Precision = TP/(TP + FP) \tag{2}$$

The ROC curve can show how the relationship between recall and precision changes when changing the discriminative threshold of the model. Suppose there is a pneumonia classification model that outputs a value between 0 and 1 for each chest X-ray image. The user needs to set a threshold for each category. When the output result falls within the threshold range of a certain category, the model would then consider this chest X-ray image to be in that category. Constantly changing this threshold can balance the relationship between precision and recall.

AUC value: indicates the area of ROC curve. The larger the AUC value, the better the model performance.

4. Proposed Methods

4.1 Proposed Convolutional Neural Network

Figure 1 shows the overall architecture of the convolutional neural network proposed in this paper, which consists of two parts: a feature extractor and a classifier. In the feature extractor, each layer uses the output of its previous layer as the input, and the output of this layer is used as the input of the subsequent layer. The classifier is placed on top of the CNN, which is often referred to as the fully connected layer. The classifier needs a one-dimensional feature vector to perform the computation, so the output of the feature extractor needs to be converted into the one-dimensional feature vector required by the classifier. The fully connected process is called flattening, which flattens the output of the convolutional layer into a lengthy feature vector that is used in the classification process of the final Dense layer.

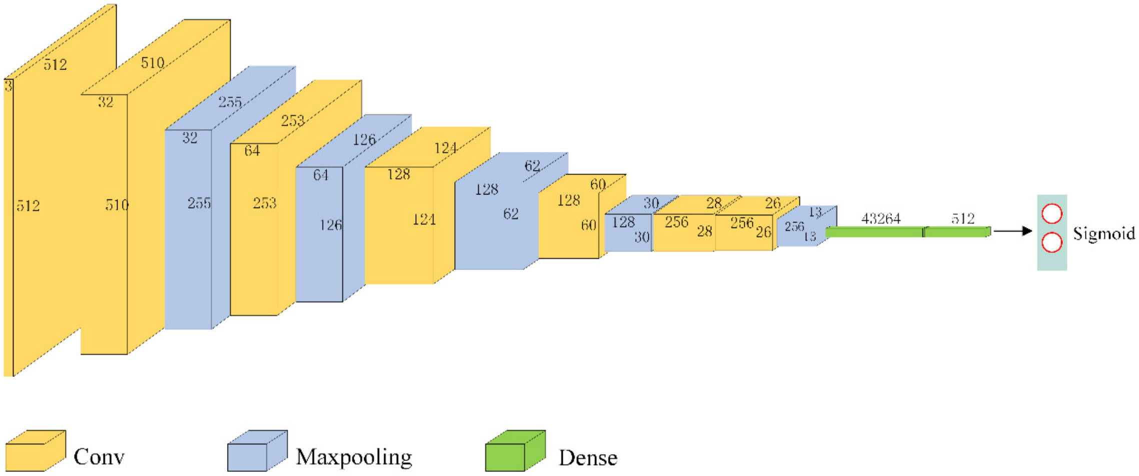
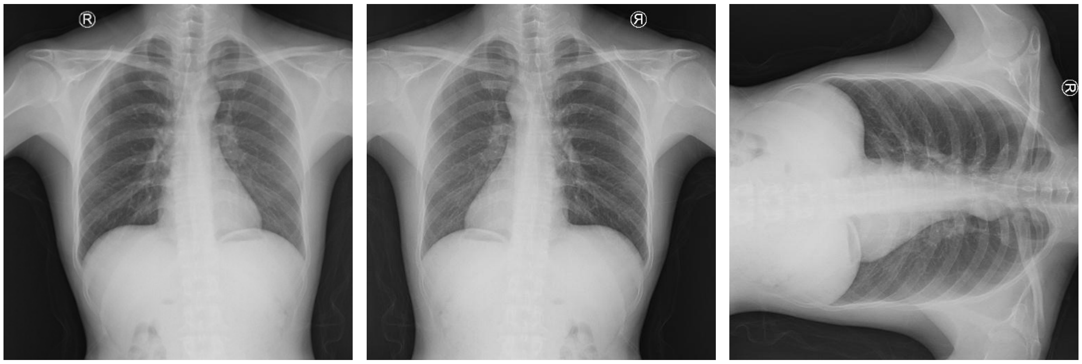


Figure 1. Overall architecture of the convolutional neural network.

4.2 Model Feature Enhancement Scheme

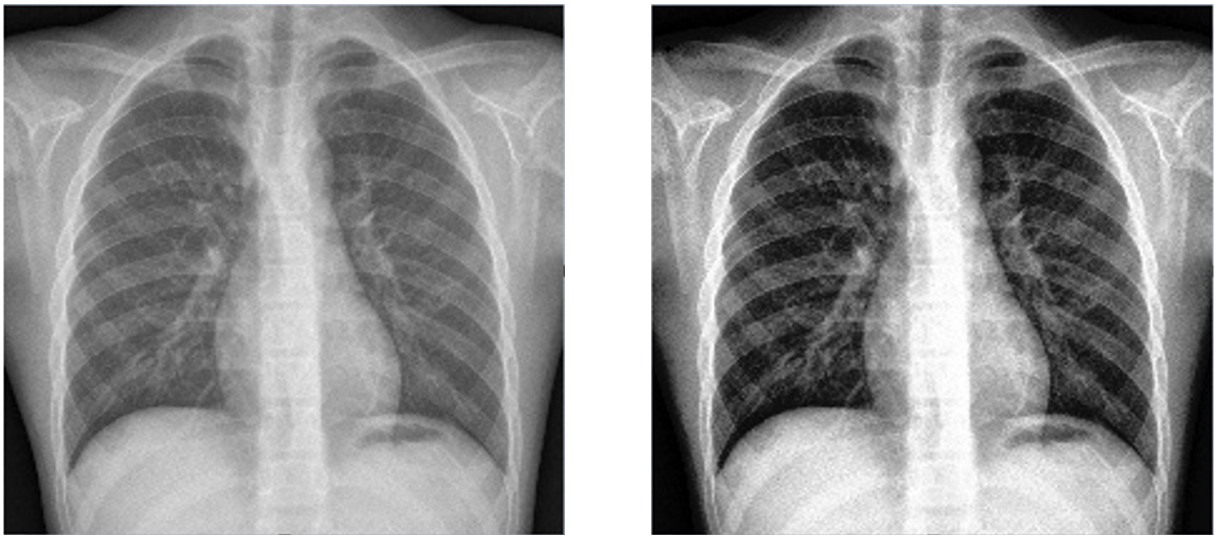
When the number of samples in the dataset is insufficient, the convolutional neural network should not be too complicated, otherwise, there will be serious overfitting problems. Therefore, the convolutional neural network proposed in this paper is relatively simple. However, it is also necessary to properly expand the dataset, which can help the convolutional neural network learn more features and help improve the accuracy of pneumonia classification. Data augmentation will make minor modifications to the existing sample data so that multiple data can be obtained. Common data augmentation operations include flips, translations, rotations, etc. As shown in Figure 2, (a) is the original chest X-ray image, (b) is the flipped chest X-ray image, and (c) is the chest X-ray image rotated by 90 degrees.



(a) original chest X-ray image (b) flipped chest X-ray image (c) rotated chest X-ray image

Figure 2. Data enhancement of the original image.

In addition to data augmentation, some preprocessing operations can be performed on the dataset to further improve the image quality. For example, histogram equalization can enhance the contrast of the image to highlight the inconspicuous parts. It may help the convolutional neural network more easily to classify the category. As Figure 3 shows, (a) is the original chest X-ray image, and (b) is the chest X-ray image after histogram equalization. It can be seen that the chest cavity in (b) is more obvious.



(a) original chest X-ray image

(b) flipped chest X-ray image

Figure 3. Data enhancement of the original image.

Since the dataset has been expanded, and there are many open chest X-ray image datasets on the Internet, transfer learning can also be used. First, the convolutional neural network learns some features on large-scale public datasets, and then performs secondary learning on expanded and preprocessed private datasets, which can further improve the accuracy of pneumonia classification.

In summary, the overall process of the model feature enhancement scheme is shown in Figure 4.

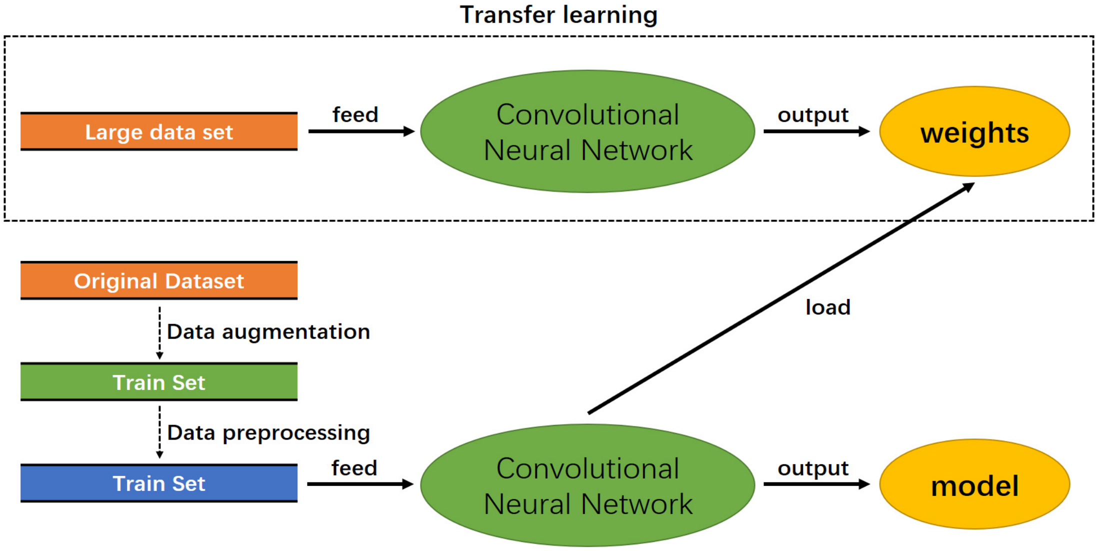


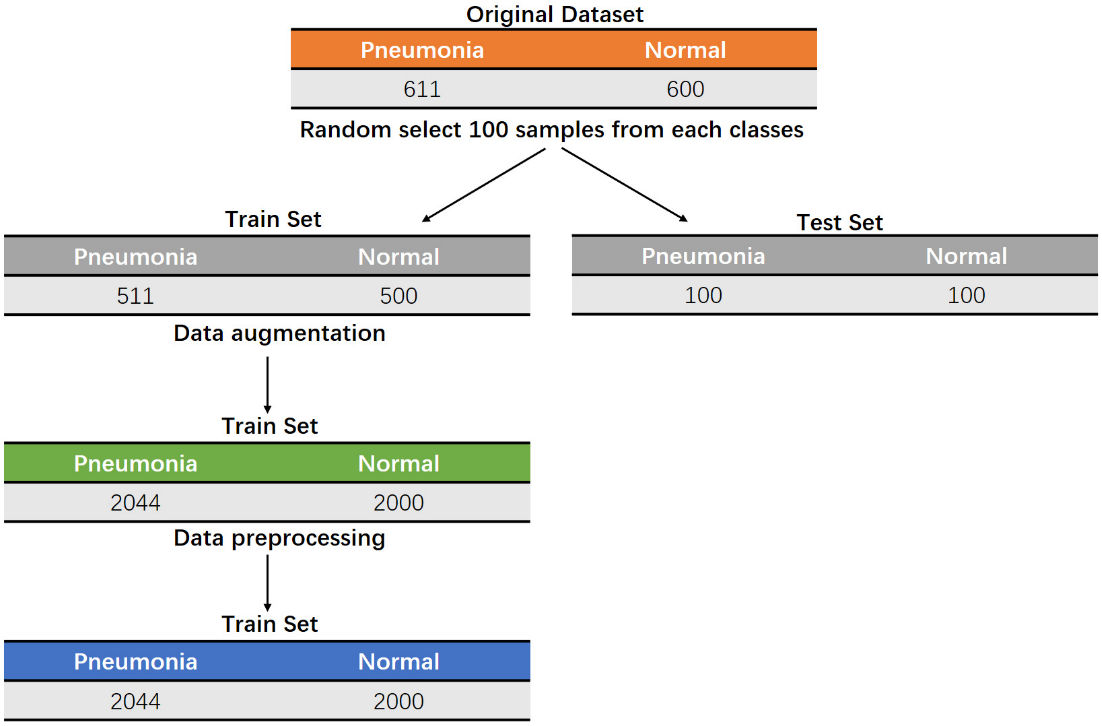
Figure 4. The overall process of the model feature enhancement scheme.

5. Experimental results and discussion

5.1 Experimental material

To build a convolutional neural network, this section uses the Keras framework and tensorflow-gpu backend. All the experiments were performed on a standard server with four 8GB Nvidia GeForce GTX 2080 Super GPUs. The version of cuDNN is 7.4, and the version of CUDA Toolkit is 10.1. The RAM of the server is 64GB, and the operating system of the server is Windows10 Profession 64Bit.

The experiments used four kinds of convolutional neural networks, namely RachnaJainNet[5], HarshSharmaNet[14], OkekeStephenNet[15], and the CNN proposed in Section 4.1. The experiments adopt the private dataset provided by the Third Military Medical University and the public dataset of Kaggle. The private dataset includes a total of 611 pneumonia samples and 600 normal samples, and 100 samples from each of the two classes are randomly selected as the test set. After dividing the test set, there are still 1011 samples left in the private dataset, and these 1011 samples are used as the first training set. Next, data enhancement was performed on 1011 samples, each sample was flipped horizontally, rotated randomly, and added with salt-pepper noise. The expanded training set has a total of 4044 samples, including 2044 pneumonia samples and 2000 normal samples, which are used as the second training set. Finally, the histogram equalization is performed on the expanded dataset, which is used as the third training set. Figure 5 shows the composition of the experimental dataset.



Original Dataset

Pneumonia	Normal
611	600

Random select 100 samples from each classes

Train Set

Pneumonia	Normal
511	500

Test Set

Pneumonia	Normal
100	100

Data augmentation

Train Set

Pneumonia	Normal
2044	2000

Data preprocessing

Train Set

Pneumonia	Normal
2044	2000

Figure 5. Composition of the dataset.

5.2 Experimental scheme

In the experimental part, this paper sets up four groups of contrast experiments to test the effect of the model feature enhancement scheme. The first set of experiments is as follows: the four CNNs are trained on the first training set, then the models with the lowest loss are tested on the test set. The second set of experiments is as follows: our proposed CNN is trained on the second training set, then the model with the lowest loss is tested on the test set, and compared with the first experimental results. The third set of experiments is as follows: our proposed CNN is trained on the third training set, then the models with the lowest loss are tested on the test set, and compared with the first two experimental results. The fourth set of experiments is as follows: first, our proposed CNN is training on a public dataset on Kaggle, then is training on the third training set by transfer learning, and finally. The models with the lowest loss are tested on the test set and compared with the first three experimental results. In each set of experiments, the samples of the training set were randomly divided, 70% samples were used for training and the other 30% samples were used for validation.

In the experimental part, the epoch of each network is set to 100, and the batch size of each network is set to 32. Other hyperparameters are set according to the reference paper. The specific hyperparameters of different networks are shown in Table 1.

Table 1. Hyperparameters of different networks.

Method	Batch size	Optimizer	Learning rate	Epochs	Loss function’
RachnaJainNet	32	Adam	0.001	100	Binary crossentropy
HarshSharmaNet	32	Adam	0.0001	100	Binary crossentropy
OkekeStephenNet	32	Adam	0.0001	100	Binary crossentropy
ProposedNet	32	Adam	0.001	100	Binary crossentropy

5.3 Experiment 1

Four kinds of convolutional neural networks will be repeated the training 10 times on the first training set, and the models with the smallest loss will be saved as the optimal models. The specific Accuracy-Loss curves of each network are shown in Figure 6 to Figure 9.

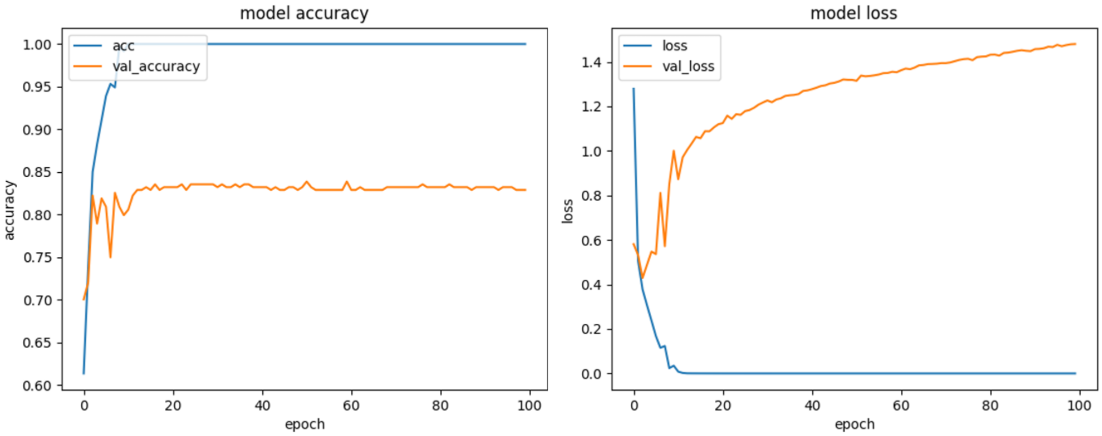


Figure 6. The Accuracy-Loss curve for RachnaJainNet.

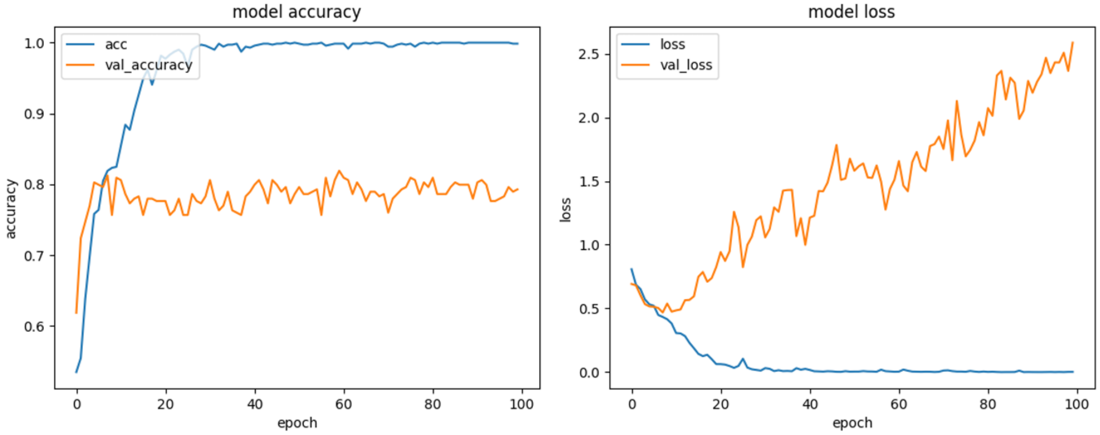


Figure 7. The Accuracy-Loss curve for HarshSharmaNet.

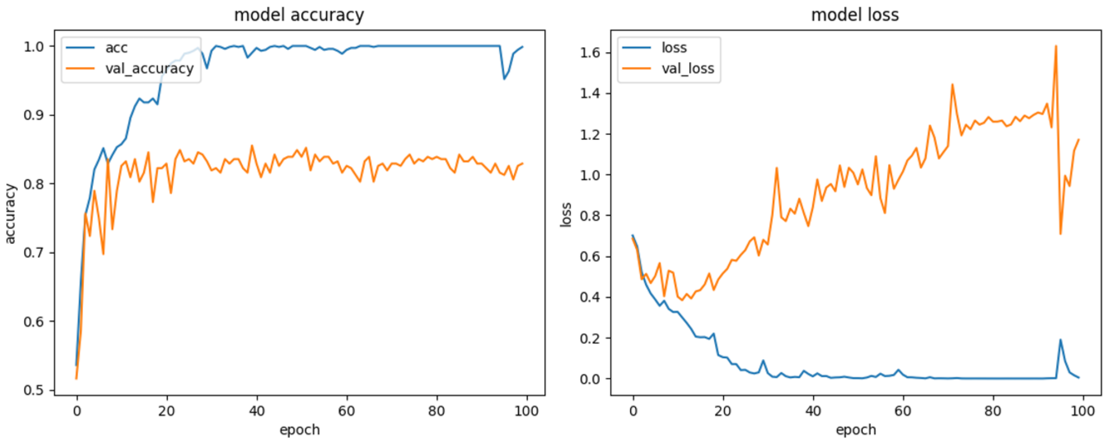


Figure 8. The Accuracy-Loss curve for OkekeStephenNet.

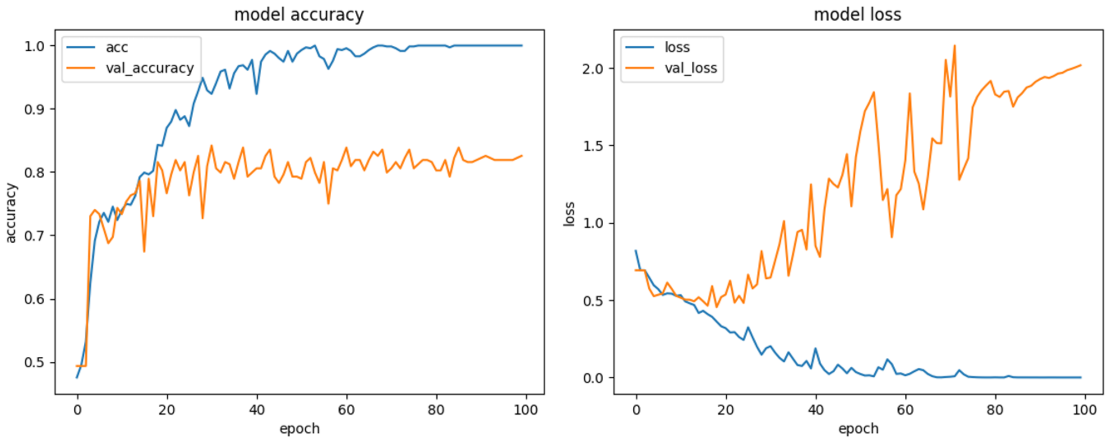


Figure 9. The Accuracy-Loss curve for ProposedNet.

The optimal model of each network is used to test on the test set, and the classification accuracy of each model is obtained. The specific accuracy results are shown in Table 2. From Table 2, it can be seen that the classification effect of ProposedNet is worse than the other three methods, and its performance is not good in Experiment 1.

Table 2. Classification accuracy of different models in Experiment 1.

Method	Accuracy
RachnaJainNet	78.5%
HarshSharmaNet	74.0%
OkekeStephenNet	**79.5%**
ProposedNet	72.0%

5.4 Experiment 2

ProposedNet does not achieve good performance in Experiment 1. After analysis, it may be because there are too few samples in the first training set, which makes ProposedNet unable to learn the main features of chest X-ray images. To this end, in Experiment 2, ProposedNet will be repeated the training 10 times on the second training set, and the model with the smallest loss will be saved as the optimal model. After that, the optimal model will be tested on the test set and compared with the results of Experiment 1. The specific accuracy comparison results are shown in Table 3.

Table 3. Comparison of the results of Experiment 2 with Experiment 1.

Method	Accuracy
RachnaJainNet	78.5%
HarshSharmaNet	74.0%
OkekeStephenNet	**79.5%**
ProposedNet-Experiment 1	72.0%
ProposedNet-Experiment 2	79.0%

From Table 3, it can be seen that the classification accuracy of ProposedNet in Experiment 2 has been greatly improved, which is 7.0% higher than the ProposedNet in Experiment 1. The classification effect of ProposedNet in Experiment 2 exceeds HarshSharmaNet and RachnaJainNet, but it is still worse than OkekeStephenNet.

5.5 Experiment 3

In Experiment 2, the performance of ProposedNet has been greatly improved, but it still has room for improvement. Although the second training set has three times the number of samples compared to the first training set, there are still many chest X-ray images that are not clear enough, so histogram equalization is used to highlight the chest region of the chest X-ray. In Experiment 2, ProposedNet will be repeated the training 10 times on the third training set, and the model with the smallest loss will be saved as the optimal model. After that, the optimal model will be tested on the test set and compared with the results of Experiment 1 and Experiment 2. The specific accuracy comparison results are shown in Table 4.

Table 4. Comparison of the results of Experiment 3 with Experiment 1 and Experiment 2.

Method	Accuracy
RachnaJainNet	78.5%
HarshSharmaNet	74.0%
OkekeStephenNet	79.5%
ProposedNet-Experiment 1	72.0%
ProposedNet-Experiment 2	79.0%
ProposedNet-Experiment 3	**81.5%**

From Table 4, it can be seen that the classification accuracy of ProposedNet in Experiment 3 has been improved, which is 2.5% higher than the ProposedNet in Experiment 2. The classification effect of ProposedNet in Experiment 3 also exceeds OkekeStephenNet. Now ProposedNet has the best performance among the comparison methods.

5.6 Experiment 4

The training set on Experiment 3 has 4044 images, at this time, the network structure of ProposedNet is relatively simple. Therefore, Experiment 4 chooses transfer learning to further improve the performance of ProposedNet.

First, ProposedNet will be repeated the training 10 times on the public dataset of Kaggle, and the network parameters under the minimum loss will be saved. Then, ProposedNet will load the pre-trained network parameters, then it will be repeated the training 10 times on the third training set, and the model with the smallest loss will be saved as the optimal model. Finally, the optimal model will be tested on the test set and compared with the results of Experiment 1, Experiment 2, and Experiment 3. The specific accuracy comparison results are shown in Table 5.

Table 5. Comparison of the results of Experiment 4 with Experiment 1, Experiment 2, and Experiment 3.

Method	Accuracy
RachnaJainNet	78.5%
HarshSharmaNet	74.0%
OkekeStephenNet	79.5%
ProposedNet-Experiment 1	72.0%
ProposedNet-Experiment 2	79.0%
ProposedNet-Experiment 3	81.5%
ProposedNet-Experiment 4	**82.0%**

From Table 5, it can be seen that the classification accuracy of ProposedNet in Experiment 4 has been slightly improved, which is 0.5% higher than the ProposedNet in Experiment 3.

5.7 Discussion

From the results of Experiment 1 to Experiment 4, it can be seen that the proposed model feature enhancement scheme can effectively improve the classification accuracy of ProposedNet, and the accuracy rate is increased from 72.0% to 82.0%. The classification accuracy of all experimental models is shown in Figure 10. From Figure 10, it can be seen that the classification accuracy of ProposedNet has been significantly improved after applying the model feature enhancement scheme.

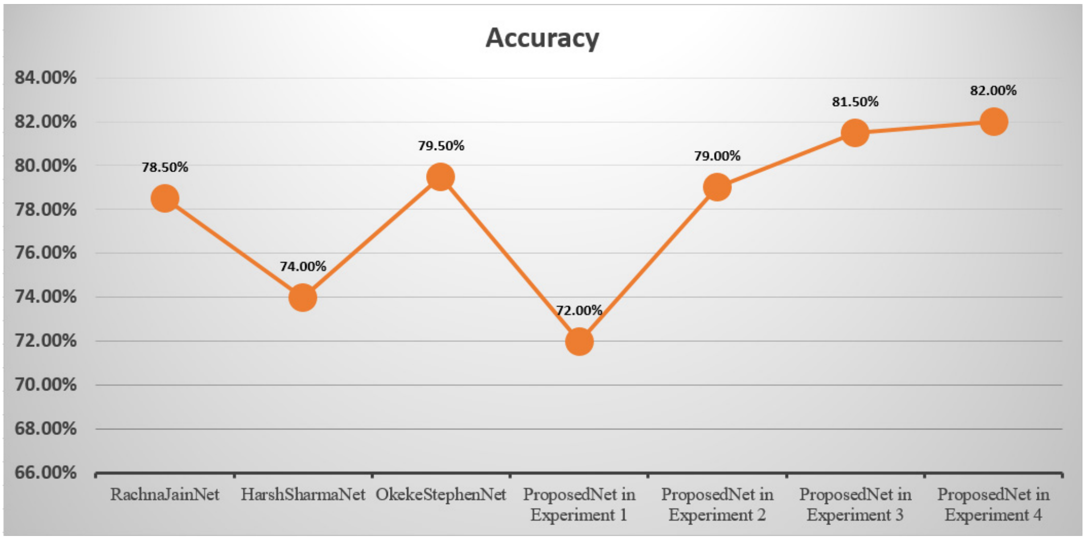


Figure 10. Model-Accuracy curve.

In addition to accuracy, the evaluation metrics of the model include F1 score, recall rate, etc. All metrics of ProposedNet in the four experiments are shown in Table 6. It can be seen that the model feature enhancement scheme can not only improve the classification accuracy but also improve other metrics.

Table 6. Metrics of ProposedNet in each experiment.

Method	Accuracy	Recall rate	Precision	F1 score
ProposedNet-Experiment 1	72.0%	55.0%	83.3%	0.66
ProposedNet-Experiment 2	79.0%	74.0%	82.2%	0.78
ProposedNet-Experiment 3	81.5%	76.0%	**85.4%**	0.80
ProposedNet-Experiment 4	**82.0%**	**79.0%**	84.0%	**0.81**

6. Conclusion

Based on the popular pneumonia classification problem in recent years, this paper designs a model feature enhancement scheme that can improve the accuracy of pneumonia classification. The scheme consists of three parts: data enhancement, data preprocessing, and transfer learning.

The design idea of the model feature enhancement scheme is: first, use data enhancement technology to expand the original dataset to an appropriate number, then use data preprocessing technology to enhance the contrast of the image and highlight its core lesion area. Finally, the convolutional neural network is trained on the large-scale public dataset firstly, then the trained parameters are loaded by transfer learning technology, and the convolutional neural network is second trained on the processed private dataset.

On the dataset provided by the Third Military Medical University and the public dataset of Kaggle, the model feature enhancement scheme improves the classification accuracy of our proposed model from 72.0% to 82.0% and makes the performance of our proposed model better than the reference RachnaJainNet, HarshSharmaNet, and OkekeStephenNet. In addition to Accuracy, other metrics have also been enhanced, with the F1 score from 0.66 to 0.81, and the recall rate from 55.0% to 79.0%. Although the model feature enhancement scheme can improve the performance of our proposed model, it still has shortcomings. Observing the model-accuracy curve in Figure 10, it can be found that with the advancement of the model feature enhancement scheme, the increase in the accuracy of pneumonia classification is gradually decreasing. For our proposed model, the role of data enhancement is greater than data preprocessing and transfer learning. Therefore, the model feature enhancement scheme can indeed improve the performance of the convolutional neural network, but the improvement effect of each part is different.

Although the model feature enhancement scheme proposed in this paper can improve the pneumonia classification accuracy of convolutional neural networks, the scheme still has room for further improvement and research, which we will continue to explore in future research.

Reference

[1] Grewal M, Srivastava M M, Kumar P, et al. Radnet: Radiologist level accuracy using deep learning for hemorrhage detection in ct scans[C]//2018 IEEE 15th International Symposium on Biomedical Imaging (ISBI 2018). IEEE, 2018: 281-284.

[2] Salido J A A, Ruiz C. Using deep learning to detect melanoma in dermoscopy images[J]. Int. J. Mach. Learn. Comput, 2018, 8(1): 61-68.

[3] Ragab D A, Sharkas M, Marshall S, et al. Breast cancer detection using deep convolutional neural networks and support vector machines[J]. PeerJ, 2019, 7: e6201.

[4] Gao X W, Hui R, Tian Z. Classification of CT brain images based on deep learning networks[J]. Computer methods and programs in biomedicine, 2017, 138: 49-56.

[5] Jain R, Nagrath P, Kataria G, et al. Pneumonia detection in chest X-ray images using convolutional neural networks and transfer learning[J]. Measurement, 2020, 165: 108046.

[6] Rajpurkar P, Irvin J, Zhu K, et al. Chexnet: Radiologist-level pneumonia detection on chest x-rays with deep learning[J]. arXiv preprint arXiv:1711.05225, 2017.

[7] Toraman S, Alakus T B, Turkoglu I. Convolutional capsnet: A novel artificial neural network approach to detect COVID-19 disease from X-ray images using capsule networks[J]. Chaos, Solitons & Fractals, 2020, 140: 110122.

[8] Hashmi M F, Katiyar S, Keskar A G, et al. Efficient pneumonia detection in chest xray images using deep transfer learning[J]. Diagnostics, 2020, 10(6): 417.

[9] Yu X, Wang S H, Zhang Y D. CGNet: A graph-knowledge embedded convolutional neural network for detection of pneumonia[J]. Information Processing & Management, 2021, 58(1): 102411.

[10] Liang G, Zheng L. A transfer learning method with deep residual network for pediatric pneumonia diagnosis[J]. Computer methods and programs in biomedicine, 2020, 187: 104964.

[11] Gabruseva T, Poplavskiy D, Kalinin A. Deep learning for automatic pneumonia detection[C]//Proceedings of the IEEE/CVF conference on computer vision and pattern recognition workshops. 2020: 350-351.

[12] Varshni D, Thakral K, Agarwal L, et al. Pneumonia detection using CNN based feature extraction[C]//2019 IEEE international conference on electrical, computer and communication technologies (ICECCT). IEEE, 2019: 1-7.

[13] Selvaraju R R, Cogswell M, Das A, et al. Grad-cam: Visual explanations from deep networks via gradient-based localization[C]//Proceedings of the IEEE international conference on computer vision. 2017: 618-626.

[14] Sharma H, Jain J S, Bansal P, et al. Feature extraction and classification of chest x-ray images using cnn to detect pneumonia[C]//2020 10th International Conference on Cloud Computing, Data Science & Engineering (Confluence). IEEE, 2020: 227-231.

[15] Stephen O, Sain M, Maduh U J, et al. An efficient deep learning approach to pneumonia classification in healthcare[J]. Journal of healthcare engineering, 2019, 2019.

Research on Laboratory Intelligent Security Management Mode Based on Big Data and Cloud Services

Changning Ji[1*]
[1]Chongqing Three Gorges Vocational College, ChongqingWanzhou, China, 404155
*Corresponding author e-mail: jichangning2022@163.com

Abstract

With the popularization of information technology, the proportion of universities in the field of scientific research investment has increased year by year, which is also mainly used for laboratory investment. With the continuous maturity of cloud service technology, intelligent laboratory management has become the main mode of university management, which will gradually replace the traditional manual management. Through various sensing devices, the intelligent laboratory can realize the intelligent management of equipment, personnel, safety and technical data, which is the main embodiment of intelligence. **Research method:** Through the intelligent laboratory management and control platform, managers can directly intelligentize the management system and control the laboratory, which is the necessity of the laboratory intelligent security management mode. Through the intelligent system, managers can realize the standardization, process and informatization of technical data, which will improve the utilization and sharing rate of experimental equipment. This paper mainly refers to the literature survey and empirical research. Through the investigation of literature to obtain information, this paper comprehensively and correctly understand a method to master the problem to be studied. The literature research method is widely used in various disciplines. Through empirical research, this paper proposes a design mode using scientific instruments and equipment, which can be used to observe, record and measure the intelligent management mode of laboratories step by step. **Conclusion:** The development of laboratory intelligent security management mode based on big data and cloud services has become the development trend of university laboratory management, which is of great significance to improve the scientific research level of universities. Based on big data analysis algorithm, this paper constructs an experimental intelligent security management mode.

Keywords: Big data, cloud service, laboratory intelligence, security management

1. Introduction

With the increasing investment in scientific research hardware equipment in colleges, large and high-end instruments and equipment have gradually entered college laboratories, which has greatly helped the scientific research work in colleges. The laboratory is an important place for teaching and scientific research, which is directly related to the completion rate of scientific research achievements in colleges[1-2]. However, with the establishment of large and medium-sized laboratories, the use of laboratory equipment and instruments is becoming more and more frequent, which urgently needs to be able to change the laboratory safety management mode. The current laboratory management mode has many drawbacks, such as low equipment use efficiency, high error rate of manual management, and difficulty in real-time monitoring of equipment status. With the improvement of the scale, quality and quantity of university laboratories, laboratory safety accidents occur frequently, which is also an urgent problem for universities to solve[3]. Through ZigBee, the laboratory can timely monitor various potential safety hazards and give early warning, which rapidly reduces safety accidents. Through intelligent safety management of laboratories, colleges can effectively reduce the work intensity of managers, which will also reduce the probability of laboratory safety accidents. Therefore, the research on intelligent safety management of laboratories is of great significance. First, through big data and cloud service technology, laboratory intelligence can directly serve relevant departments, which can obtain data and equipment information in real time[4-6]. Therefore, laboratory intelligence has laid a foundation for analysis and decision-making, which also ensures the accuracy of test data. Second, through big data and cloud service technology, laboratory intelligence can improve the safety and reliability of test data, which can prevent data distortion. Third, through big data and cloud service technology, laboratory intelligence can improve the scientificity of analysis, judgment and decision-making, which can highlight flexibility and convenience. The key to learning laboratory safety management is mainly divided into the following contents: strengthening the construction of laboratory system, establishing a strong laboratory team, and optimizing the application of intelligent technology. By improving the initiative of the team, universities can stimulate the vitality and vigor of the management team.

International Conference on Computer Graphics, Artificial Intelligence, and Data Processing (ICCAID 2022)
edited by Ruishi Liang, Jing Wang, Proc. of SPIE Vol. 12604, 126041I
© 2023 SPIE · 0277-786X · doi: 10.1117/12.2674938

Li Chunbo (2021) and others have made great achievements in the intelligent control of chemicals in the laboratory, which also makes different types of treatment plans according to different accidents[7]. Through the intelligent laboratory safety management, the safety accident of the experiment has been greatly guaranteed, which also promotes the innovation of a variety of laboratory safety management modes, such as double person double lock, alarm system, emergency fire-fighting equipment, etc. Li Guangrong (2021) has made great achievements in laboratory information management[8]. He has developed a trinity access control security system, which is an innovative model integrating "civil air defense", "material defense" and "technical defense". Based on MIFARE technology, intelligent management can realize non-contact control. Bristol is a college teacher who is engaged in chemistry teaching. He proposed measures to fill in the risk assessment report online to ensure the purpose of the experiment, which can help us understand the degree of toxicity of experimental drugs from the side. At the same time, he has developed a closed loop management system for the whole process of dangerous goods, which can ensure that all links are connected to form a closed loop management. At the same time, through face recognition technology, he has established monitoring and monitoring linkage alarm, which can evaluate the alarm signal based on GIS platform and analysis algorithm. Through sensor readings and on-site video images, managers can master the on-site situation, which improves management efficiency[9-10].

2. Research Basis

2.1. Laboratory intelligent safety monitoring management

First, intelligent smoke detection monitoring. Through the intelligent smoke monitoring system, the laboratory can automatically monitor, which will realize the intelligent identification of smoke concentration in the laboratory, which also becomes an important mode for rapid identification of laboratory fire disasters[11]. Once a fire occurs in the laboratory, the laboratory can automatically send alarm information for the content that has problems through the intelligent smoke detection system[12]. Second, intelligent gas monitoring. Through the intelligent gas monitoring system, the laboratory can automatically identify the concentration of harmful gases, which will effectively help the experiment to achieve automatic identification function. Once the gas leakage occurs in the laboratory, the laboratory can automatically send alarm information for the content that has already had problems through the intelligent gas monitoring system. The third is intelligent power supply monitoring[13-15]. Multiple intelligent power controllers can be installed in the laboratory to remotely view the power status. Through power supply detection, the laboratory can monitor the energy consumption of the equipment in real time, and can also remotely control the power switch of the equipment. Through power supply testing, it can effectively help laboratory managers master the use of laboratory equipment[16]. When the intelligent power controller detects that the device current is abnormal, the laboratory will send an alarm message for the content of the problem. Fourth, intelligent temperature and humidity monitoring[17-18]. When the temperature and humidity of the laboratory are abnormal, the laboratory can automatically send an alarm message for the content of the problem. Fifth, pedestrian intrusion prevention monitoring. By installing the pedestrian monitoring system and the infrared monitoring system for doors and windows, the laboratory can detect illegal intrusion and pedestrian activities in the laboratory within the deployment time. Through pedestrian monitoring, the laboratory can send out alarm information for illegal intrusion[19-20].

2.2. Laboratory intelligent control system

First, intelligent light control. By installing an intelligent light controller, the laboratory can remotely view the light switch status, which will actively control the light switch. Through intelligent light control, the laboratory can turn on and off lights automatically. Second, laboratory intelligent monitoring. The laboratory has installed several high-definition gun cameras and ball cameras, which can realize the monitoring of the laboratory without dead corners. Through the network client or app, the laboratory administrator can remotely view the laboratory monitoring video, which can be accurately controlled and observed remotely.

2.3. Laboratory safety alarm and safety emergency treatment

First, the emergency call button. All laboratories are equipped with emergency call buttons. When students are in danger in the laboratory, the on-site personnel can press the emergency call button of the laboratory to send out a call for help message. Second, intelligent audible and visual alarm. Through the intelligent audible and visual alarm, the laboratory can automatically send a variety of information to the intelligent audible and visual alarm, including fire, harmful gas leakage, etc. Third, establish a laboratory safety emergency response process. When the laboratory environmental safety monitoring equipment detects that the laboratory has fire, harmful gas leakage, abnormal equipment current, excessive laboratory temperature and illegal pedestrian intrusion, the laboratory can automatically upload the information to the central management system. Through the laboratory intelligent monitoring video to confirm the danger information, the management personnel can make the safety emergency treatment at the first time.

3. The methods

3.1. FAS Theory Based on Anomaly Detection

Anomaly detection is a method to identify abnormal sample tasks from a group of normal data. In the field of face anti deception, face representation attack detection (PAD) is described as anomaly detection, which is a method that assumes that living samples have some common attributes. Therefore, this study can assume that the living samples are distributed in the tight sphere of the machine learning feature representation space, while the deception samples are far away from the center of the living sample sphere, as shown in Figure 1.

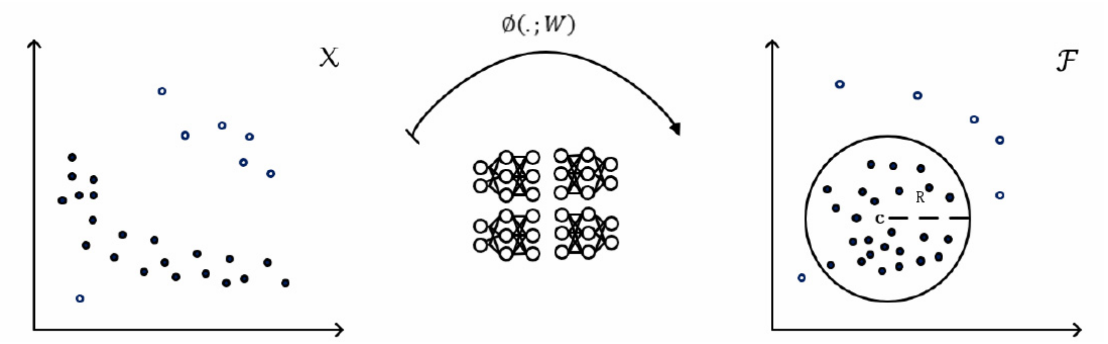


Figure 1. FAS model for anomaly detection

In the FAS method model based on anomaly detection, it is assumed that the input space is $X \in R^p$, the output space is $Z \in R^p$. W is the depth neural network, and the weight is $w = \{w_1, w_2, ..., w_n\}$.Assume that N_t is a living sample $(x_1, x_2, ..., x_{nt} \in X)$, and N_s is a deceptive sample $(y_1, y_2, ..., y_{ns} \in X)$. The FAS objective function for anomaly detection is shown in Formula 1.

$$\begin{aligned} &\arg(W): \\ &\min \frac{1}{n_t} \sum_{i=1}^{n_t} \left\| \phi(x_i, W) - c \right\|^2 \\ &\min \frac{1}{n_s} \sum_{i=1}^{n_s} \left\| \phi(y_i, W) - c \right\|^2 \end{aligned} \tag{1}$$

In the test phase, the score of test sample t is defined as spool, and the spool value is the distance from the calculated sample to the center c point of the moving sample sphere, as shown in Formula 2.

$$s(x) = \left\| \phi(x_i, W) - c \right\| \tag{2}$$

3.2. Spoofing prompt generator

Cheating cue graph can effectively separate real time closed set and cheating open set. Therefore, RGB images can output feature mapping C according to the feature prompt generator. The feature mapping regression loss of living samples is L_1 loss at pixel level, as shown in Formula 3.

$$L_r = \frac{1}{T} \sum\nolimits_{l_i \in live} \left\| C_i \right\|_1 \tag{3}$$

Therefore, this paper constructs a spoofing prompt generator, which is a method based on U-Net architecture. By building multi-scale jump connections, we can generate deception hints, which will be used to judge the difference between live samples and deception samples, as shown in Figure 2.

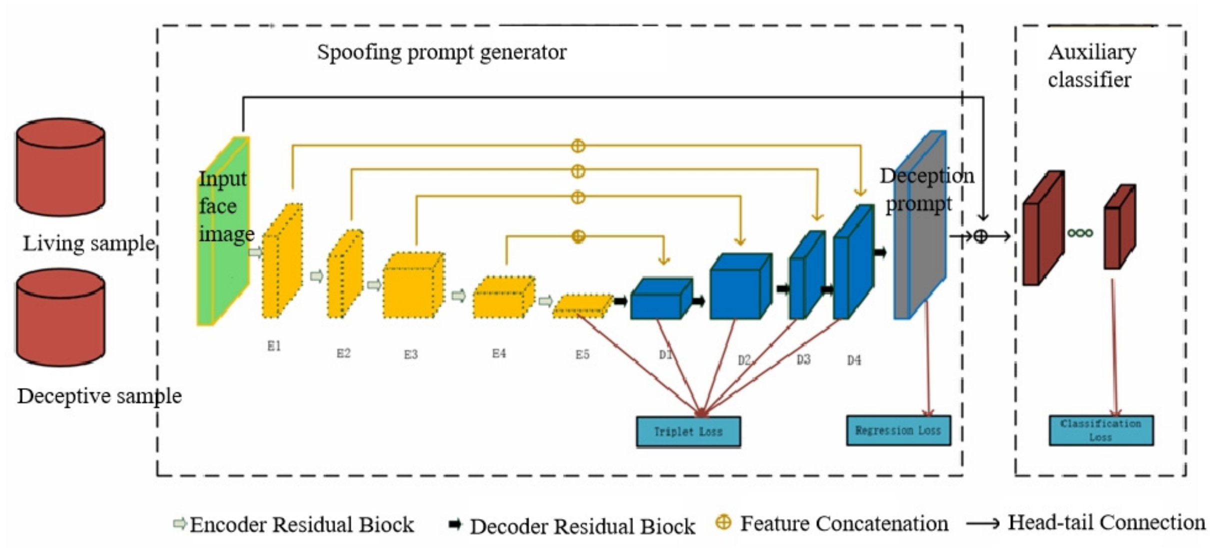


Figure 2. Spoofing prompt generator

3.3. Data fusion technology

Multi sensor data fusion is a new technology for comprehensive processing of multi-source information, which is a way of information fusion based on different information levels, including data level fusion, feature level fusion, decision level fusion, etc. Therefore, we can meet the accuracy requirements of the laboratory safety management system, which mainly includes the following multiple fusion technologies. First, data level fusion. Based on the data collected by sensors, we need to collect data under the same category of sensors, which cannot handle the fusion of heterogeneous data. Second, feature level fusion. This paper can extract the feature vectors of the collected data, which can reflect the attributes of the monitored physical quantities. In image data fusion, we can use edge feature information, which will completely replace all data information. Third, decision level integration. According to the data features of feature level fusion, we can make certain discrimination and classification. In the specific data fusion implementation of laboratory environmental monitoring, we can choose the fusion method according to the characteristics of the application.

4. The verification with intelligent security management mode

4.1. Construction of safety management control platform Based on big data and cloud services

This paper constructs an intelligent laboratory management and control platform, which is a website application. The function of the platform is not only to display information, but also to process related businesses. After analyzing the requirements of existing platforms, the intelligent laboratory management and control platform mainly includes management system and controller. Among them, the intelligent management system mainly includes administrators, teachers and students, which mainly provides users with laboratory information browsing, laboratory opening instructions, equipment browsing, equipment appointment, etc. Therefore, this paper constructs the functional modules of the intelligent laboratory management system, as shown in Figure 3.

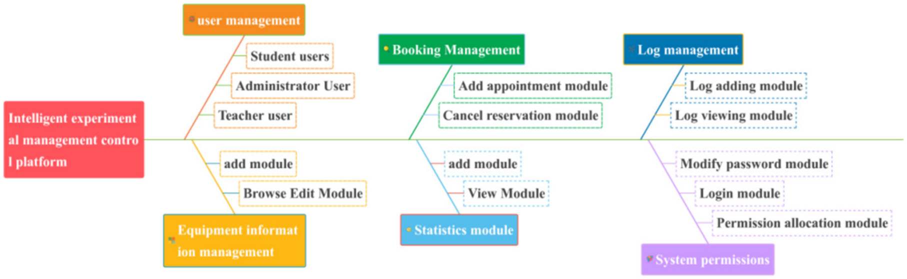


Figure 3. Safety Management Control Platform

4.2. Overall system architecture Based on big data and cloud services

When designing the system interface, this paper mainly considers the usability of the interface and the user's usage habits. Based on CSS, HTML and Bootstrap frameworks, the front-end layout of framework development is more concise and convenient. The overall three-tier architecture of the system. First, the basic function layer, which provides the underlying support for the operation of the system software. Therefore, the system can operate continuously and stably. The basic function layer provides interface operation, logic processing and other functions for the whole system. Second, the data layer, which is mainly responsible for classifying and summarizing the data collected by the SCM through various sensors to the database. At the same time, the mutual transmission of big data and cloud computing data can ensure the normal operation of face recognition. Third, the hardware layer, which mainly consists of the local server, STM32F103RCT6 microcontroller, laser scanner, high-precision full bridge pressure sensor, DS1820, DHT11, etc. The overall architecture design of the system is shown in Figure 4.

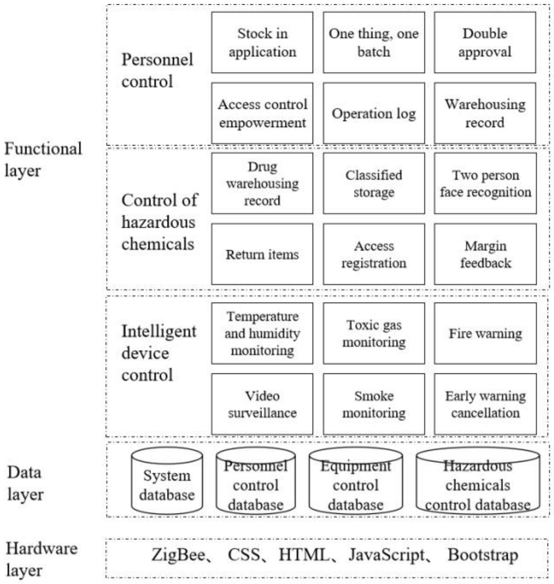


Figure 4. Overall System Architecture

Through Spoofing prompt generator, this paper constructs the overall system architecture, which can better complete the laboratory opening instructions, equipment browsing, equipment reservation and other functions. Based on the FAS model for anomaly detection, we can be used to judge the difference between real-time samples and deceptive samples, which plays an important role in big data screening. Through FAS model, face recognition can completely improve the intelligent security management of the laboratory.

5. Conclusion

The laboratory intelligent management and control platform can meet various needs of laboratory management, and can complete the management, browsing, appointment and other functions of laboratory information. According to the user's permission, the intelligent laboratory can realize the role of management and control, which will make the laboratory management and control work intelligent.

This paper draws the following conclusions: First, this paper analyzes the FAS theory based on anomaly detection and the cheating prompt generator mode that can be processed according to big data. Second, this paper constructs an intelligent, networked, immersive laboratory model, which will realize the safety identification function of the laboratory. Thirdly, this paper constructs the construction of the security management control platform and the overall architecture of the system, realizing the digital and intelligent management design of the laboratory. Shortcomings of this paper: Due to space reasons, this paper did not analyze the overall structure design, system construction, workflow, database design, etc. of the security management control platform. I hope that follow-up scholars can conduct research and design a more intelligent laboratory management platform. Future development prospects: With the popularization of intelligent laboratories, the working mode of mobile terminals will become the main mode of future laboratory work, which will focus more on the design and development of operating systems.

References

[1] Chen Zhuzi, Xu Xiaoqing, Qi Mengwen. Research on the Intelligent Management System of the Central Laboratory of Higher Vocational College [J]. China New Communications, 2022, 24 (15): 67-69

[2] Zhou Fubao, Bai Xiangyu, Chen Xiaoyu, Huang Changzhong, Wang Bingjie. Intelligent management and control of college laboratory safety in the context of "intelligent+" era [J]. Labor Protection, 2022 (08): 10-13

[3] Sun Wenqing, Han Qiang, Zhang Xudong. Intelligent construction and management of computer experimental teaching center in application-oriented universities [J]. Journal of Jilin Normal University of Engineering Technology, 2022,38 (04): 52-55

[4] Qiu Dunguo, Li Fajun, Liao Yong, Wang Jun, Zheng Xiaolin. Security Strategy for Open Sharing of "Mass Entrepreneurship and Innovation" Intelligent Laboratory [J]. Laboratory Research and Exploration, 2022,41 (03): 300-303

[5] Fan Fengxin, Zhou Bing. Design of Intelligent Security Management Scheme for University Laboratory Based on ZigBee Wireless Sensor Network [J]. Network Security Technology and Application, 2022 (01): 77-78

[6] Yang Yan, Zhou Di, Li Fen, Liu Shenpei, Zou Jingrui. Function introduction and operation practice of the intelligent management system of the scientific research public platform of the general hospital [J]. China Medical Biotechnology, 2021,16 (05): 476-480

[7] Zhang Xiuliang. Application and Practice of Intelligent Security Management in University Laboratories [J]. Cyberspace Security, 2021,12 (Z4): 99-102

[8] Lin Peng, Xiang Yunfei, An Ruinan. Inspiration of hydropower intelligent safety management for strengthening the safety construction of university laboratories [J]. Experimental Technology and Management, 2021,38 (06): 7-12+20

[9] Sun Limi, Zhu Li. Research on Innovation Management Mode of University Laboratory under the Background of Mass Entrepreneurship and Innovation Education [J]. China Education Technology Equipment, 2021 (01): 6-7

[10] The New Face of the State Evaluation and the New Journey of Jiangsu Jiaoke [J]. China Highway, 2020 (19): 64-65

[11] Qiu Zige, Chen Wenlong, Jiang Nanping. Safety Management of Life Science Laboratories under the Application of Intelligent Devices [J]. Technology and Market, 2020, 27 (10): 145-146

[12] Lian Jingjing, Pang Xibin, Xu Jin, Wang Haixia, Lu Feng. Application and Practice of Intelligent Management of University Laboratories [J]. Laboratory Research and Exploration, 2020, 39 (07): 255-257+284

[13] Guo Junfang, Zhang Ruoyu, Chen Shujie, Lu Enzi, Zhao Yan, Xu Haohao. The Design of Intelligent Video Monitoring System for University Laboratory Security [J]. Laboratory Research and Exploration, 2020, 39 (05): 297-301

[14] Wang Jingcheng, Yin Gengxin. Discussion on standardized design and decoration of medical laboratories [J]. Journal of Anhui Vocational and Technical College of Health, 2019, 18 (06): 12-13

[15] Liu Weibing. Research on Intelligent Management of University Laboratory Security in the View of the Internet of Things [J]. Computer Products and Circulation, 2019 (12): 162

[16] Ping Honghai. Computer Group Policy Application Security [J]. Electronic Technology and Software Engineering, 2019 (07): 201

[17] Han Shanpeng, Mao Rong. Research on Safety Management of Petroleum Laboratories in Colleges [J]. Laboratory Research and Exploration, 2019, 38 (01): 260-263

[18] Shen Weiwei. Research on Laboratory Security Intelligent Management System [J]. Internet of Things Technology, 2018,8 (12): 61-63

[19] Zhang Weiming. Research on Intelligent Management of University Laboratory Security in the View of the Internet of Things [J]. Microcomputer Application, 2018, 34 (08): 54-56+77

[20] Wang Guoqiang, Wu Min. Composition and Application of Intelligent Laboratory Security System [J]. Experimental Technology and Management, 2018 (04): 151-155

Multi-objective parameter optimization of distributed hydrological models based on data-poor watersheds

Ke Xu[a], Kun Yang[b,c*]
[a]School of Information Science and Technology, Yunnan Normal University, Kunming 650500, China
[b]Faculty of Geography, Yunnan Normal University, Kunming 650500, China.
[c]GIS Technology Engineering Research Centre for West-China Resources and Environment, Ministry Education, Yunnan Normal University, Kunming 650500, China.
[*] Corresponding author: kmdcynu@163.com

ABSTRACT

The distributed hydrological model's high-resolution representation of watershed heterogeneity is especially suitable for hydrological simulation in watersheds with small areas, and the hydrological model needs to use flow observation data for parameter calibration to obtain adapted parameters, but the lack of observation basins is the core problem we face in China, where there are few hydrological stations and they are concentrated in important rivers. This work attempts to optimize the parameters of the distributed hydrological model using limited flow observation data in a small watershed where flow observation is relatively scarce and to explore the value of combining multiple sources of data to optimize the parameters of the distributed hydrological model. Based on the combination of meteorological observations and remote sensing data collected, a distributed hydrological model is constructed based on the WetSpa model for the Jianshan River basin, and a multi-objective genetic algorithm NSGA-II is applied to rate the model to predict the simulated basin runoff process. The results show that the hydrological model optimized by multi-objective parameters has good adaptability in the study area, and the simulation has certain accuracy, which can provide basic support for the simulation of the water environment in the basin and also provide a reference for hydrological simulation.

Keywords: hydrological simulation, multi-objective parameter optimization, WetSpa model, distributed hydrological model.

1 INTRODUCTION

Watershed flow prediction is an important part of surface hydrology, and accurate flow prediction is essential for water management and related infrastructure design for human life, agriculture, industry, environment, and ecosystems. Natural processes of rainfall-runoff processes are widely simplified and generalized into a variety of hydrological models. Hydrological models are important tools for simulating hydrological processes in watersheds and understanding the response mechanisms of hydrological elements, and they are also effective tools for solving hydrological simulation problems[1, 2]. The optimization of hydrological model parameters has become an important research content in the field of hydrology. The essence of hydrological model parameter optimization is to find the best value of each parameter of the hydrological model so that the simulated value of the model is as close to the actual value as possible. The optimization of hydrological model parameters has a crucial influence on the overall performance of the hydrological model and the hydrological forecast results.

However, the lack of information in many watersheds makes obtaining sufficient river flow data for calibration difficult. The limited availability and low quality of observed data reduce the reliability of hydrologic model simulation outputs. Therefore, the development of new schemes to improve the predictive performance of hydrologic models in data-poor watersheds is a challenge in the field of hydrologic research. For river flow prediction in data-poor watersheds only hydrological simulations can be adopted and are widely used in practice[3]. In the past decades, advances in satellite remote sensing technology have made possible the use of hydrologic data such as precipitation, soil moisture, and evapotranspiration. The increased availability of remotely sensed data provides an important opportunity to improve the predictive performance of hydrologic models in watersheds where information is lacking[4-7]. Since remote sensing data has unparalleled advantages in areas where observations are scarce, we can make full use of remote sensing data in hydrologic modeling for data-poor areas. This work attempts to optimize the parameters of a distributed hydrologic model

International Conference on Computer Graphics, Artificial Intelligence, and Data Processing (ICCAID 2022)
edited by Ruishi Liang, Jing Wang, Proc. of SPIE Vol. 12604, 126041J
© 2023 SPIE · 0277-786X · doi: 10.1117/12.2674764

using limited flow observations in a small watershed where flow observations are relatively scarce and to explore the value of combining multiple sources of data to optimize the parameters of a distributed hydrologic model.

2 STUDY AREA

The study area is located in the Jianshan River basin of Chengjiang County, Yuxi City, central Yunnan Province, which is a first-order tributary of Fuxian Lake and is located on the northwestern shore of the lake (Fig. 1). It is located at latitude 24°32'00"-24°37'38' N and longitude 102°47'21"-102°52'02" E. The basin's total area is 35.42 km2, with an elevation of 1,722.0-2347.4 m and a relative height difference of 625.4 m. The three-dimensional climate is obvious. The average annual rainfall in the Jianshan River basin is 1050 mm, and the wet and dry seasons are distinct, with the rainy season from late May to late October, when 75% of the total annual rainfall is received, and the dry season from early November to mid-May, when 25% of the annual rainfall is received. The average annual evaporation is 900 mm, and the soils in the basin are mainly red purple clay and red soil. The area of arable land in the study area is 1174.6 hm2, accounting for 33.2% of the total area; the area of sloping arable land is 347.1 hm2, accounting for 29.6% of the total arable land; the area of forest land is 1 697.5 hm2, accounting for 47.9% of the total area, including 941 hm2 of plantation forest, 339.8 hm2 of shrub forest and 413.4 hm2 of secondary forest.

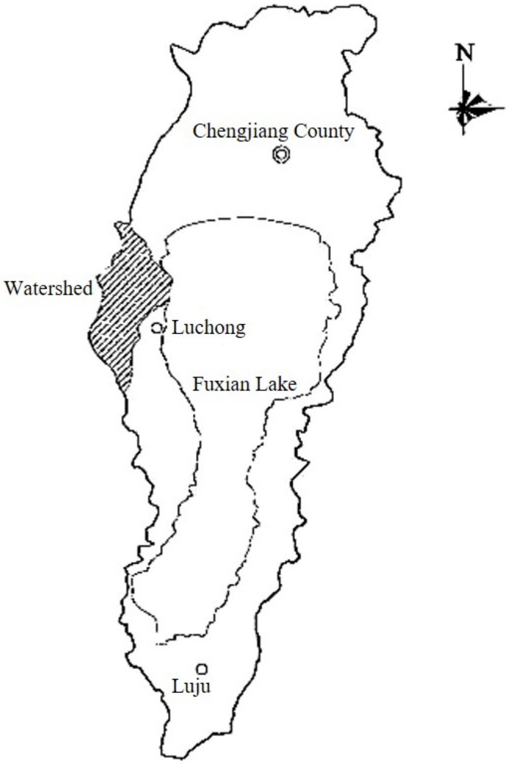


Figure 1.The geographical location of the Jianshan River basin

3 MODELS AND METHODS

3.1 WetSpa Model

The WetSpa model, originally proposed by Wang and Batelaan[8], is a distributed grid-based hydrological model with a time step of days for simulating water and energy exchange between soil, plants, and the atmosphere at the study watershed scale[9]. For each grid cell, the structure is divided vertically into four layers: plant canopy, soil surface, root zone, and groundwater zone, as shown in Figure 2. The hydrological processes considered in the model include precipitation, interception, depression storage, surface runoff, snowmelt, infiltration, evapotranspiration, seepage, and groundwater drainage. The model can predict flood flows and hydrologic profiles can be defined for any number and location in the watershed network and can simulate the spatial distribution of hydrologic variables in the watershed.

Compared with other hydrological models, such as the SWAT model, the WetSpa model is more suitable for the simulation of single-event flood processes, and the calculation period can be chosen in hourly, daily, or monthly time steps, which improves the generality and portability of the model and can dynamically reflect the runoff process at any point in the watershed.

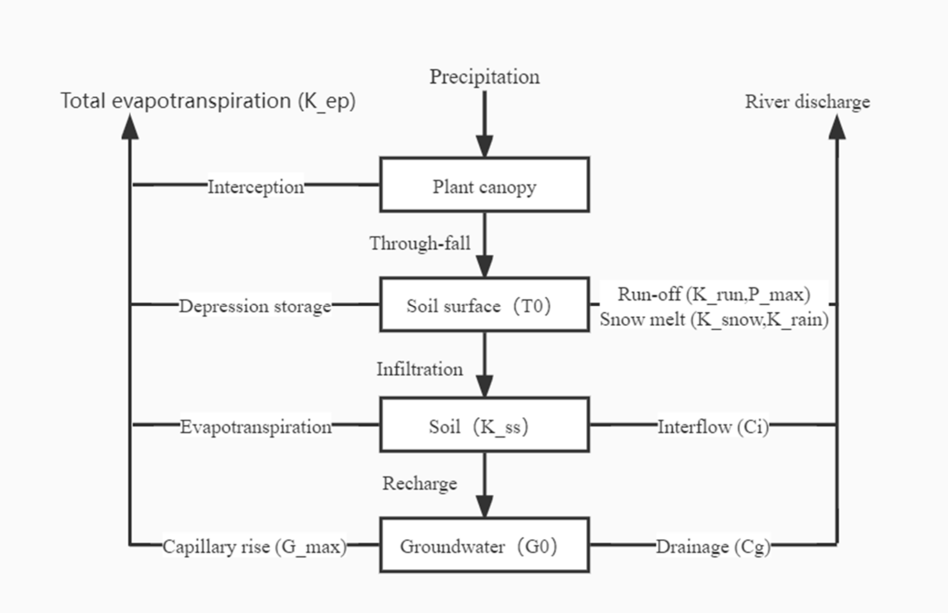


Figure 2. The structure of the Wetspa model with arrows indicating hydrological processes and model parameters in parentheses are shown in Table 1.

3.2 Input Data

The input data for the model include topography, land use, soil type, rainfall, temperature, and evapotranspiration. All the spatial distribution parameters in the model can be derived from three types of data: topography, land use, and soil type. In addition, 11 global parameters applied to each cell or sub-basin need to be prepared before running the Wetspa model (Table 1).

The DEM data of the study area basin were used at 12.5m×12.5 m spatial resolution. Based on the DEM, the Wetspa model can automatically extract the numerical characteristics of the watershed, including determining the flow direction of the cell network, cumulative flow, confluence network, river network chain, river network code, slope, hydraulic radius and boundary delineation of sub-basins, etc., and provide the input of sub-bedding data for the hydrological model. Soil data were obtained from the 1:1 million soil spatial database developed by the Nanjing Institute of Soil Research, Chinese Academy of Sciences. The data were transformed to the same spatial resolution and coordinate system as the DEM data through resampling and coordinate transformation, and reclassified according to the Wetspa model soil classification criteria. The land use data in the study area was obtained from the global surface cover characteristics database provided by USGS, and the data were transformed to the same spatial resolution and the same planar coordinate system as the DEM data through resampling and coordinate transformation, and reclassified according to the Wetspa model land use classification criteria.

The temperature and rainfall observation data used in this study were obtained from the Chengjiang meteorological station and rainfall station near the watershed; evaporation data were obtained from MOD16 remote sensing products[10], extracted from the watershed and interpolated; all data lengths were from 2014-2020. The day-by-day flow data for 2019-2020 used for parameter calibration and model validation were obtained from hydrological station monitoring. The year 2014 was selected as the model warm-up period, the years 2015-2019 were selected for model calibration, and the year 2020 was selected for model validation.

3.3 Multi-objective optimization algorithm NSGA-II

Genetic Algorithm[11] (GA) is a computational model of the biological evolution process that simulates the mechanism of natural selection and genetics of Darwinian biological evolution and is a method to search for the optimal solution by simulating the natural evolution process. In the genetic algorithm, the initial population evolves according to the principle of survival of the fittest and survival of the fittest and generates better and better approximate solutions generation by generation. In each generation, individuals are selected according to their fitness in the problem domain. The genetic operators of natural genetics combine crossover and mutation to produce a population representing a new set of solutions until the optimal solution is produced. Although genetic algorithms can find approximately optimal solutions more efficiently, there are still problems in selecting the best individuals in each generation, and the good individuals may be deleted during the evolutionary process. This led to the proposal of the non-dominated ranking genetic (NSGA) algorithm. NSGA is based on the Pareto dominance relationship.

Multi-objective optimization problems are very different from single-objective optimization problems. When there is only one objective function, one looks for the best solution, which is better than all other solutions, usually the global maximum or minimum, i.e., the global optimal solution. When there are multiple objectives, it is difficult to find a solution that makes all the objective functions optimal at the same time because there are conflicts between the objectives that cannot be compared. Therefore, for multi-objective optimization problems, there is usually a set of solutions that are not comparable between all the objective functions and are characterized by the inability to improve any objective function without weakening at least one other objective function. This solution is called a non-dominated solution or Pareto optimal solution.

NSGA-II[12] is a genetic algorithm based on the NSGA algorithm, which uses genetic methods to evolve the population, a fast non-dominance method to rank individuals, and a crowding distance method to maintain the diversity of the population. The algorithm is based on the following procedures: firstly, the population is initialized, and the first generation of offspring population is obtained by the three basic operations of the genetic algorithm: selection, crossover, and mutation; secondly, from the second generation, the parent population is merged with the offspring population, and the fast non-dominance sorting is performed, and the crowding degree is calculated for the individuals in each non-dominance layer. Finally, a new population of children is generated by the basic operation of the genetic algorithm, and so on until the end of the program is satisfied. This is shown in Figure 3.

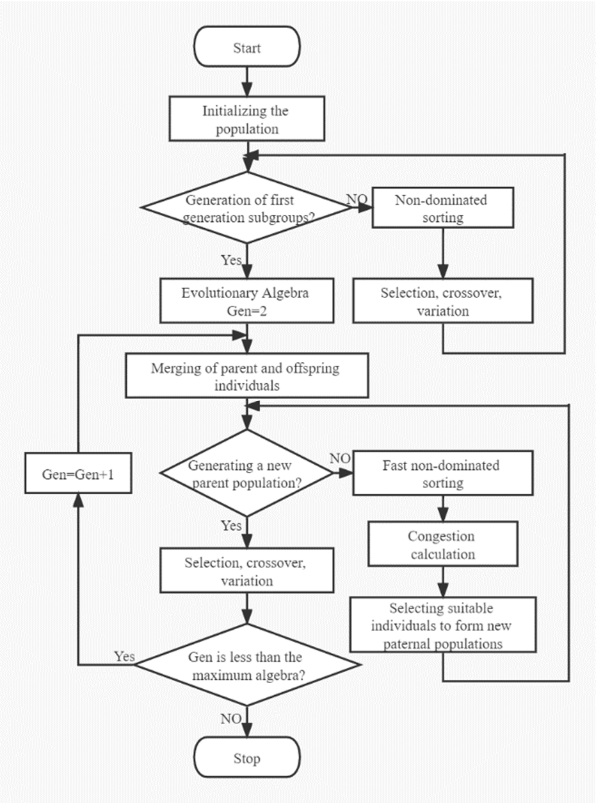

Start
Initializing the population
Generation of first generation subgroups?
NO
Non-dominated sorting
Yes
Evolutionary Algebra Gen=2
Selection, crossover, variation
Merging of parent and offspring individuals
Gen=Gen+1
Generating a new parent population?
NO
Fast non-dominated sorting
Yes
Congestion calculation
Selection, crossover, variation
Selecting suitable individuals to form new paternal populations
Yes
Gen is less than the maximum algebra?
NO
Stop

Figure 3. NSGA-II basic flow

In calculating the non-dominance relationship between individuals in the population, NSGA-II introduces a fast non-dominance ranking algorithm to rank the Pareto ranks of all individuals in the population.

1. find all individuals in the population with i = 0 and deposit them in the current non-dominated set rank1;

2. for each j in the current non-dominated set rank1, traverse the set S_j of individuals it dominates and subtract 1 from N_t for each t in the set S_j, i.e., the number of individuals that dominate the solution of individual t minus 1 (since the individual j that dominates the individual has been deposited in the current non-dominated set), and if N_t-1 = 0, deposit the individual t in another set H.

3. take rank1 as the first level of the set of non-dominated individuals, so that the solved individual in rank1 is optimal. It only dominates individuals and is not dominated by any other individuals. All individuals in this set are assigned the same non-dominated order, and then the above hierarchical operation is continued for the set H, and the corresponding non-dominated order is also assigned, until all individuals are hierarchical, i.e., all are assigned the corresponding non-dominated order.

The advantage of the multi-objective optimization algorithm NSGA-II is that, firstly, the fast non-dominated sorting method is proposed to reduce the computational complexity of the algorithm. Secondly, the crowding degree and crowding degree comparison operators are proposed instead of the fitness sharing strategy which requires a specified sharing radius and is used as the winning criteria in the peer comparison after fast sorting, so that the individuals in the quasi-Pareto domain can be extended to the whole Pareto domain and uniformly distributed, maintaining the diversity of the population. Finally, an elite strategy was introduced to expand the sampling space. By combining the parent population with its resulting offspring population, the next generation population is generated by joint competition, which is helpful to keep the best individuals in the parent generation into the next generation, and by storing all individuals in the population in a hierarchical manner, the best individuals are not lost and the population level is rapidly improved.

3.4 Objective function

The objective function is an important indicator used to evaluate the deviation between simulated and measured values. Different objective functions are used to evaluate different characteristics of the hydrological process, and the selection of the objective function is crucial to the parameter optimization results. In this paper, the Nash efficiency coefficient[13] F1 (emphasizing peak flow simulation) and log-transformed Nash coefficient F2 (focusing on low flow simulation), which are commonly used in domestic and international studies, are selected as the objective functions for model optimization, and their calculation equations are:

$$\text{F1}=1-\frac{\sum_{i=1}^{N}(Q_{Mi}-Q_{Ni})^2}{\sum_{i=1}^{N}(Q_{Mi}-\overline{Q_M})^2}$$

$$\text{F2}=1-\frac{\sum_{i=1}^{N}[ln\,(Q_{Mi})-ln(Q_{Ni})]^2}{\sum_{i=1}^{N}\left[ln(Q_{Mi})-\overline{ln\,(Q_M)}\right]^2}$$

whereQ_N is the simulated runoff volume;Q_M is the observed runoff volume; and $\overline{Q_M}$ is the mean value of the observed runoff volume. The objective of the multi-objective calibration in this study is to maximize F1 and F2.

3.5 Algorithm parameter setting

The NSGA-II algorithm parameters were set as follows: population size 50, iteration number 100, decision variable dimension 11, crossover probability 0.9, variance probability 0.2, and the algorithm termination condition was set to reach the maximum number of iterations.

4 PARAMETER CALIBRATION AND VERIFICATION

The calibration process begins with the identification of feasible parameter values. Model parameter ranges were selected based on watershed characteristics, as described in the documentation and user manual for the WetSpa model[14], and pre-determined feasible parameter ranges are given in Table 1. To generate the initial populations for NSGA-II, this study used Latin Hypercube Sampling[15] (LHS) to explore the full range of feasible parameter values for the model for optimal calibration. Using the LHS to generate 50 sets of parameters as the initial population, the WetSpa model parameters were calibrated against the 2019-2020 day-by-day flow data in the Jianshan River basin, and the multi-objective optimization will result not in a single unique set of parameters, but in a set of parameters consisting of Pareto fronts. The results of the calibration of each parameter are shown in Table 1.

Table 1. Parameters to be calibrated for the WetSpa model: description, symbols, units, preset range of values, and range of values for the Pareto optimal solution obtained by NSGA-II

Description	Parameter	Units	Feasible range	NSGA-II solutions	
				Min	Max
Interflow scaling factor	Ci	–	0–10	8.13	9.95
Groundwater recession coefficient	Cg	d^{-1}	0–0.05	0	0.01
Initial soil moisture factor	K_ss	–	0–2	1.5	1.64
Correction factor for PET	K_ep	–	0–2	1.37	2
Initial groundwater storage	G0	mm	0–500	330.22	500
Groundwater storage scaling factor	G_max	mm	0–2000	1992.49	2000
Base temperature for snowmelt	T0	℃	−1–1	0.91	29.3
Temperature degree-day coefficient	K_snow	mm $℃^{-1}$ d^{-1}	0–10	4.87	6.87
Rainfall degree-day coefficient	K_rain	$℃^{-1}$ d^{-1}	0–0.05	0.03	0.05
Surface runoff coefficient	K_run	–	0–5	5	5
Rainfall scaling factor	P_max	mm	0–500	497.79	500

After convergence of the NSGA-II multi-objective optimization algorithm, 30 Pareto front optimal solutions are obtained with Nash efficiency coefficients F1 between -0.06 ~ 0.1 and log-transformed Nash coefficients F2 between -2.2 ~ 0.34 during the calibration period. For F1, the simulation obtained during the validation period is worse, but for F2 for the low-flow simulation, it is usually better. The latter can be explained by the fact that the flow is usually lower during the validation period except for the flood period (June-August) so that the flow deviation is smaller and the low-flow process can be better simulated.

To visualize the optimization results, the optimal solution of the Pareto front for the calibration cycle is shown in Figure 4. It can be observed that the NSGA-II diffusion between F1 and F2 in the multi-objective optimization scheme is quite consistent. This illustrates that multi-objective calibration allows for better exploration of the optimal region and more optimal solutions, providing flexibility in decision-making for stakeholders. Overall, the use of NSGA-II is effective for the multi-objective parameter optimization of the WetSpa model.

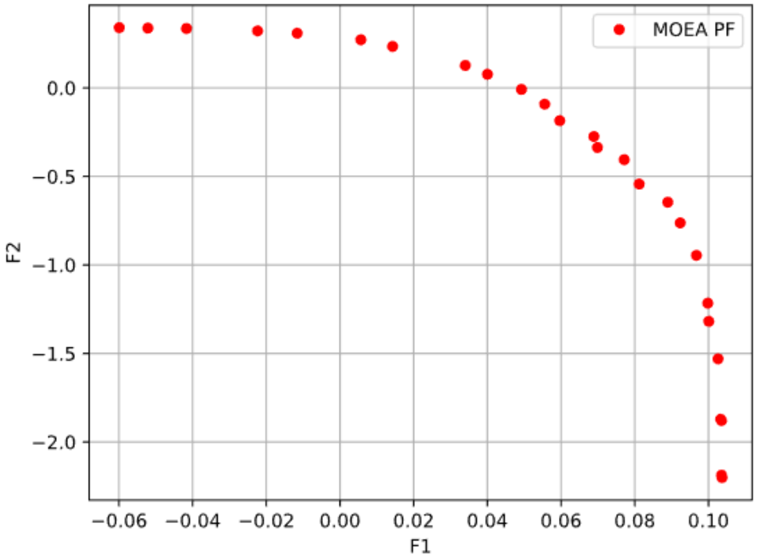


Figure 4. The Pareto frontier optimal solution

Figure 5 shows the comparison process between the simulated and observed flows in the Jianshan River basin in the validation period of 2020. It is the model result of 30 Pareto optimal solutions obtained with the NSGA-II optimal parameter set, and the simulation effect is shown in the red area. It can be seen from the figure that the hydrological-hydraulic model of this basin based on the WetSpa model has high accuracy and can describe the rainfall-runoff process of the basin more reasonably.

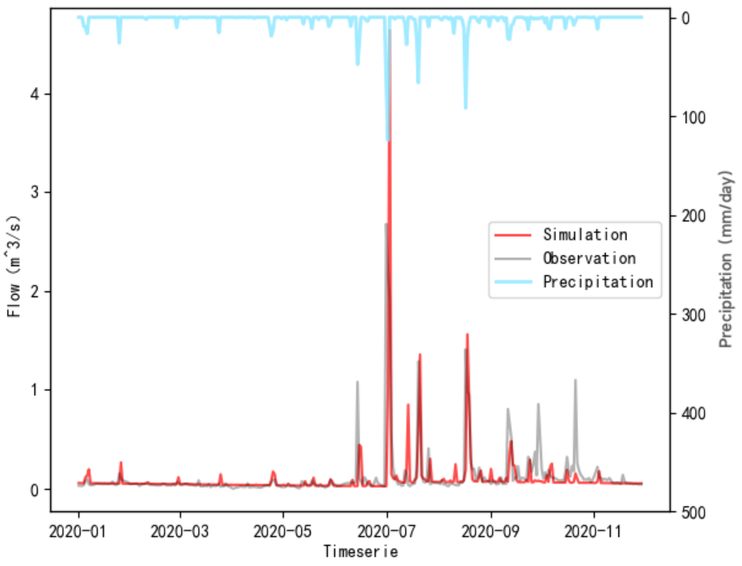


Figure 5. Comparison between simulated and measured values of the runoff process in 2020

5 CONCLUSION

In this study, based on the data of temperature, rainfall, evaporation, topography, land use, and soil in the Jianshan River basin, a distributed hydrological model, Wetspa, was used to establish a hydrological and hydrodynamic simulation system for the basin, and the daily observed flows were used to calibrate and validate the hydrological and hydrodynamic model for two years from 2019 to 2020. The parameters of the model were determined, among which the highest simulated Nash efficiency factor F1 reached 0.1 and the highest simulated low-flow factor F2 was 0.34 in 2020, and the total runoff error reached 9.7% and the peak flow error was 13.8%, with a large deviation of the simulation effect. From the comparison graph of runoff simulation and actual measurement, the reason for the large simulation error may be that the observed data of flow in September-October 2020 are large, while the corresponding observed data of rainfall are small, and the observed data of rainfall coincide with the simulated flow of the model, so it is considered that there is a certain error in the observed flow data of this period, which leads to a certain negative impact on the optimization effect of model parameters.

By combining multiple sources of data to optimize the hydrological model with multi-objective parameters, the Wetspa distributed hydrological model has good adaptability in the Jianshan River basin, where there is a lack of data. The results show that the model simulation has a certain accuracy and can reasonably describe the rainfall-production-sink flow process in the basin, which can provide a reference for integrated water environment management and water quality simulation studies.

REFERENCES

[1] Salvadore, E., J. Bronders, and O.J.J.o.H. Batelaan, Hydrological modelling of urbanized catchments: A review and future directions. 2015. 529: p. 62-81.

[2] Sood, A. and V. Smakhtin, Global hydrological models: a review. Hydrological Sciences Journal-Journal Des Sciences Hydrologiques, 2015. 60(4): p. 549-565.

[3] Allawi, M.F., et al., Review on applications of artificial intelligence methods for dam and reservoir-hydro-environment models. Environmental Science and Pollution Research, 2018. 25(14): p. 13446-13469.

[4] Sanzana, P., et al., A GIS-based urban and peri-urban landscape representation toolbox for hydrological distributed modeling. Environmental Modelling & Software, 2017. 91: p. 168-185.

[5] Bhatt, G., M. Kumar, and C.J. Duffy, A tightly coupled GIS and distributed hydrologic modeling framework. Environmental Modelling & Software, 2014. 62: p. 70-84.

[6] Helmi, N.R., et al., WetSpa-Urban: An Adapted Version of WetSpa-Python, A Suitable Tool for Detailed Runoff Calculation in Urban Areas. Water, 2019. 11(12).

[7] Jiang, D.J. and K. Wang, The Role of Satellite-Based Remote Sensing in Improving Simulated Streamflow: A Review. Water, 2019. 11(8).

[8] Wang, Z.M., O. Batelaan, and F. DeSmedt. A distributed model for water and energy transfer between soil, plants and atmosphere (WetSpa). in 21st European-Geophysical-Society General Assembly. 1996. The Hague, Netherlands.

[9] Liu, Y.B. and F. De Smedt, Flood modeling for complex terrain using GIS and remote sensed information. Water Resources Management, 2005. 19(5): p. 605-624.

[10] Mu, Q., et al., Development of a global evapotranspiration algorithm based on MODIS and global meteorology data. Remote Sensing of Environment, 2007. 111(4): p. 519-536.

[11] Srinivas, N. and K.J.E.C. Deb, Multiobjective Function Optimization Using Nondominated Sorting Genetic Algorithms. 1994. 2(3): p. 1301-1308.

[12] Deb, K., et al., A fast and elitist multiobjective genetic algorithm: NSGA-II. 2002. 6(2): p. 182-197.

[13] Nash, J.E. and J.V.J.J.o.H. Sutcliffe, River flow forecasting through conceptual models part I — A discussion of principles - ScienceDirect. 1970. 10(3): p. 282-290.

[14] Liu, Y.B. and F.D.J.U.M. Smedt, WetSpa Extension , A GIS-based Hydrologic Model for Flood Prediction and Watershed Management Documentation and User Manual. 2004.

[15] Iman, R.L., W.J.J.C.i.S.-T. Conover, and Methods, Small sample sensitivity analysis techniques for computer models.with an application to risk assessment. 1980. 9(17): p. 1749-1842.

Sub-game Perfect Nash Equilibrium under Loser Reporting

Wei Huang[1*], Chao Huang[1], Xiaoyun Pang[1] and Liang Yuan[1]
[1]School of Computer Science and Information Security, Guilin University of Electronics Technology, No. 1, Jinji Road, Guilin, China 541004
*Wei Huang: huangwei@guet.edu.cn

ABSTRACT

Parallel allocation is a decentralized mechanism for allocating indivisible objects to agents, in which agents are allowed to report their favorite objects among the remaining goods parallelly. How to maximize the interests of self-interested agents while taking into account fairness has always been a hot research topic. In this paper, we study the loser-reporting policy which ensures that the difference between the number of items taken by the agents does not exceed one. In order to make the expected benefit obtained by the agent equal to the expected benefit obtained by the agent in the sub-game perfect Nash equilibrium (SPNE) in parallel allocation, we come up with a concept of shelving disputes, and the SPNE strategy can be computed in polynomial time with respect to the number of objects.

Keywords: Parallel allocation, loser-reporting, shelving dispute, fairness

1. INTRODUCTION

Resource allocation has always been a hot topic in computer science, economics and artificial intelligence[1-2]. There are plenty of examples in our life, such as the allocation of fixed classrooms and training opportunities for the unemployed. Utility, fairness and computational complexity must be considered when distributing goods. Early related work mainly used sequential allocation[3-4], which is a simple but extremely broad mechanism. The agents sequentially declare their favorite item among the remainder based on a pre-defined sequence.

There are three dimensions of realistic resource allocation problems, divisible [6-7] or indivisible [4,8,9], centralized [10] or decentralized [11], paid [12] or unpaid allocation [8]. Here, we study the problem of allocating indivisible and unpaid items under loser-reporting policy [17,18] which is a decentralized mechanism.

Kohler and Chandrasekaran [11] proved that the SPNE strategy could be obtained by taking the last item in each other's partial order to get a sequence of items and then inverting the sequence under sequence allocation. Kalinowski and Xia [13] made a significant development on this basis, and the above method also works with two agents and any order(no longer strictly alternate). Kalinowski et al. [14] further studied the impact of strategic behavior on the complete-information extensive-form game of such sequential allocation procedures. Yin et al. [15] proposed a new called Group Dominant Resource Fairness which determines the allocations by solving a small number of linear programs. Baklanov et al.[16] proved that a PROPm allocation is guaranteed to exist for all instances, independent of the number of agents or goods. Huang et al. [17,19] first propose a parallel elicitation-free allocation protocol for allocating indivisible items to agents and prove that the SPNE strategy can be obtained by taking the last item in each other's partial order to get a sequence of items and then inverting the sequence under parallel allocation.

In this paper, we study the loser-reporting policy in parallel allocation. The allocation process [18] can be divided into several rounds. At the beginning of each round, each agent can declare an item at the same time. If more than two agents declare the same item, we consider it a dispute. We find that the introduction of loser-reporting policy in parallel allocation may reduce the expected benefit of the previous SPNE sequence. But the expected benefit under the loser-reporting policy is equal to the expected benefit after using shelving dispute. Moreover, the strategy of shelving the dispute can be given in polynomial time.

The remainder of this paper is organized as follows. We introduce background and notations in Section 2, and the loser-reporting policy is discussed in Section 3. Then we study shelving disputes in Section 4. Finally, we propose some possible directions and questions for future research in Section 5.

International Conference on Computer Graphics, Artificial Intelligence, and Data Processing (ICCAID 2022)
edited by Ruishi Liang, Jing Wang, Proc. of SPIE Vol. 12604, 126041K
© 2023 SPIE · 0277-786X · doi: 10.1117/12.2674574

2. BACKGROUND AND NOTATIONS

There are $m \geq 2$ indivisible and different items will be allocated to two agents, which is denoted by $N = \{A, B\}$, The set of items is represented by $\mathcal{O} = \{o_1, \ldots, o_m\}$. Let $\succ_A$ and $\succ_B$ be the strict priority order of agent A and agent B over $\mathcal{O}$. The utility function is additive for each agent $i(i \in N)$, $\mathcal{R}_i(o)$ represents the rank of the item o in $\succ_i$ ($\succ_i$ is the strict preference order of agent i), $u_i(o)$ represents the utility value of the item o to agent i, we define $o_1 \succ_i o_2$ if and only if $u_i(o_1) > u_i(o_2)$. In this paper, we assume that $u_i(\emptyset) = 0$ and the utility is additive, $u_i(\chi) = \sum_{o \in \chi} u_i(o)$ for any bundle $\chi \in \mathcal{O}$.

There are many fetching sequences for item set $\mathcal{O}$, Such as $\delta = (\delta_A, \delta_B) = \big((x_1, x_2, \cdots, x_n), (y_1, y_2, \cdots, y_n)\big)$ $(1 \leq n \leq |\delta|)$. $|\delta|$ represents the length of the fetching sequence, δ_i^k represents the item reported by agent $i(i \in N)$ in round k. We say strategy δ is well–defined if and only if in any round $1 \leq k \leq |\delta|$, item δ_i^k is still available, and there is no item available after round $|\delta|$. In the rest of this paper, we only consider well–defined strategies.

a) p_k represents the probability of obtaining the declared item in round k

b) $\mathcal{X}_i(\delta)$ represents the collection of items reported by agent i in parallel allocation

c) $\mathcal{O}_k$ represents the collection of items reported by all agents in round k in parallel allocation

Definition 1:(Utility function) We use the Borda counting method to represent the benefit value of the item o, $u_i(o) = |\mathcal{O}| - \mathcal{R}_i(o) + 1$

Definition 2:(Expected utility) In parallel allocation, we cannot get the concrete benefit value in most cases. But we can calculate the expected utility according to the reporting strategy of each agent. The expected revenue of any agent $i(i \in N)$ is $u_i\big(\mathcal{X}_i(\delta)\big) = \sum_{k=1}^{|\delta|} u_i(\delta_i(k)) p_k$

Theorem 1: Let $\delta = \big((x_1, x_2, \cdots, x_n), (y_1, y_2, \cdots, y_n)\big)$ represents the sequence $rev\left(allocate\big(rev(\succ_A), rev(\succ_B)\big)\right)$.

For any $1 \leq i < j \leq n$ has the following properties.

a) Either $x_i = y_i$ or $x_i \succ_A y_i \; and \; y_i \succ_B x_i$

b) $x_i \succ_B x_j \; and \; y_i \succ_B x_j$

c) $y_i \succ_A y_j \; and \; x_i \succ_A y_j$

Proof Agent A and agent B have strict preference orders for items. If $x_i \neq y_i$, then $x_i \succ_A y_i \; and \; y_i \succ_B x_i$. According to the execution process of $rev\left(allocate\big(rev(\succ_A), rev(\succ_B)\big)\right)$. We can get $x_i \succ_B x_j$ and since $y_i \succ_B x_i$, so $y_i \succ_B x_j$. In the same way, the third clause can be proved to be true.

3. LOSER REPORTING

Although the SPNE has been achieved in parallel allocation, the consideration of fairness is insufficient. If agent A is lucky enough to win every round in the drawing, Agent A will get all the items, but agent B will not get any items, which is extremely unfair. Therefore, we introduce the loser-reporting policy, which can ensure that the difference between the number of items obtained by the agent does not exceed one.

The loser-reporting policy can be described as follows, assuming that there are m indivisible items that will be allocated to A and B. We consider parallel allocation throughout the allocation process. If an item is declared by more than one agent, They draw lots to determine who obtains the item, the winner obtains the item and immediately goes silent, the loser reports and obtains an item from the remaining set and unlocks the winner's silent status, then opens the next round of item declaration until the remaining item is empty.

In this paper, we suppose agent A and agent B adopt the $rev\left(allocate\big(rev(\succ_A), rev(\succ_B)\big)\right)$ reporting strategy. Then a possible allocation process can be described as a finite non-empty triple sequence $(\delta_A^1, \delta_B^1, \zeta_1), (\delta_A^2, \delta_B^2, \zeta_2), \cdots, (s, s, \emptyset)$. $\mathcal{C}_k$ represents the remaining items in round k, $\mathcal{C}_1 = \mathcal{O}$, and $\mathcal{C}_{k+1} = \mathcal{C}_k \backslash (\{\delta_A^k\} \cup \{\delta_B^k\})$. If agent A (or agent B) is silent (i.e., not allowed to declare an item) in round k, then δ_A^k (or δ_B^k) is s, otherwise δ_A^k (or δ_B^k) is the item declared by agent A (or agent B).

Example 1: If we consider there are $m = 7$ items (i.e., $\mathcal{O} = \{o_1, o_2, o_3, o_4, o_5, o_6, o_7\}$). Suppose that agent A has preference $\succ_A$ such that $o_5 \succ o_1 \succ o_7 \succ o_4 \succ o_2 \succ o_3 \succ o_6$ and $\succ_B$ is $o_3 \succ o_7 \succ o_5 \succ o_2 \succ o_1 \succ o_6 \succ o_4$. If each agent adopts the $rev\left(allocate\left(rev(\succ_A), rev(\succ_B)\right)\right)$ reporting strategy, we can get $\delta_A = \{o_5, o_2, o_1, o_4\}, \delta_B = \{o_7, o_2, o_3, o_6\}$. The possible allocation process can be described as a directed acyclic graph shown in Figure 1. Agent A and B declare o_5 and o_7 respectively, in the first round. Then the allocation process starts the next round. Obviously, agent A and agent B have a conflict when declaring item o_2. Two branches will appear here (the left branch represents A wins, and the right branch represents B wins). We calculate that the expected benefit value of agent A is $u_A(o_1) + \frac{1}{2}u_A(o_2) + \frac{3}{4}u_A(o_4) + u_A(o_5) + \frac{1}{4}u_A(o_6) = 17.75$, and the expected benefit value of agent B is $\frac{1}{2}u_B(o_2) + u_B(o_3) + \frac{1}{4}u_B(o_4) + \frac{3}{4}u_B(o_6) + u_B(o_7)$=16.75. But the expected benefit value of the SPNE of parallel allocation of agent A is $u_A(o_1) + \frac{1}{2}u_A(o_2) + u_A(o_4) + u_A(o_5) = 18.5$. Similarly, the expected benefit value of the SPNE of parallel allocation of agent B is $\frac{1}{2}u_B(o_2) + u_B(o_3) + u_B(o_6) + u_B(o_7)$ =17. We can find that the expected benefit values of both Agent A and Agent B are less than the expected benefit in parallel allocation

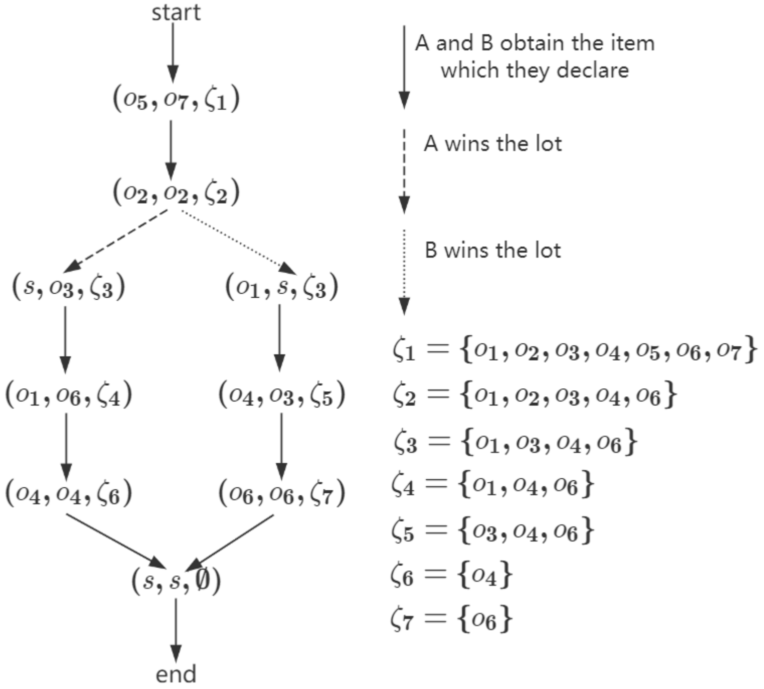


Figure 1. The allocation process when each agent adopts $rev\left(allocate\left(rev(\succ_A), rev(\succ_B)\right)\right)$.

Theorem 2 Under the loser-reporting policy, the expected benefit value of the sequence taken according to the SPNE strategy of parallel allocation is less than or equal to the expected benefit value of the SPNE of parallel allocation.

Proof Let's say the first contested position is i, and the second contested position is j. If agent A wins the lottery at point i, agent B will choose y_{i+1} according to the perfect Nash equilibrium strategy of the parallel allocation. Then A chooses x_{i+1} and B chooses y_{i+2}, until B chooses y_j. Get an item set $\{x_i, x_{i+1}, \cdots, x_{j-1}\}$.If A loses in the lottery at point i. Get an item collection $\{x_{i+1}, x_{i+2}, \cdots, x_j\}$. The expected item collection is $\left\{\frac{1}{2}x_i, x_{i+1}, \cdots, x_{j-1}, \frac{1}{2}x_j\right\}$, so the expected benefit of A and B is the same as the expected benefit value of the SPNE in the parallel allocation when there is an even number of dispute items. Suppose there are an odd number of dispute items, and let i be the last dispute point. If A wins the lottery at point i, B will choose y_{i+1} according to the SPNE strategy of the parallel allocation. Then A chooses x_{i+1}, and B chooses y_{i+2}. Until B chooses y_n, only item x_n is left. Get an item set $\{x_i, x_{i+1}, \cdots, x_{n-1}, \frac{1}{2}x_n\}$. If A loses in the lottery, A will choose x_{i+1} according to the SPNE strategy of the parallel allocation. Then B chooses y_{i+1}, Until A chooses x_n. Only item y_n, get an item set $\{x_{i+1}, x_{i+2}, \cdots, x_n, \frac{1}{2}y_n\}$. So the expected set of items is $\left\{\frac{1}{2}x_i, x_{i+1}, \cdots, x_{n-1}, \frac{3}{4}x_n, \frac{1}{4}y_n\right\}$. In the same way, the expected item set of B is $\left\{\frac{1}{2}y_i, y_{i+1}, \cdots, y_{n-1}, \frac{3}{4}y_n, \frac{1}{4}x_n\right\}$. When $i = n$. Obviously, the expected benefit is the same as the expected benefit in parallel allocation. When $i < n$ from the view of A x_n is better than y_n, but B prefers y_n to x_n. So the expected benefit is less than the expected benefit of SPNE in the parallel allocation.

4. SHELVE DISPUTE

In order for the expected benefit obtained in the loser-reporting to be equal to the expected benefit obtained in the SPNE. We can divide items into disputed items and non-disputed and then introduce the following concept.

a) $\mathcal{X} = \{x_i \neq y_i \mid 1 \leq i \leq |\delta|\}$: the set of non-dispute items that agent A picks up

b) $\mathcal{Y} = \{x_i \neq y_i \mid 1 \leq i \leq |\delta|\}$: the set of non-dispute items that agent B picks up

c) $\mathcal{C} = \{x_i = y_i \mid 1 \leq i \leq |\delta|\}$: the set of disputed items. $\mathcal{C}_k$ represent the set of the first k round of the disputed item

From the parallel allocation strategy of SPNE, the disputed items are successively found and put into the disputed area, and the remaining items are the set of non-disputed items. In the disputed area $x_i = y_i$, combining $x_i \succ_B x_j$ *and* $y_i \succ_A y_j$, so we can drive that $x_i \succ_{A\cap B} x_j$ (For both agent A and agent B,x_i is better than x_j) Agent A and B firstly declare the set of items that are not in dispute and finally declare the item in the disputed area. We can find that under parallel allocation, the expected benefit of A is $U_A = u_A(\mathcal{X}) + \frac{1}{2}u_A(\mathcal{C})$, and the expected benefit allocation of B is $U_B = u_B(\mathcal{Y}) + \frac{1}{2}u_B(\mathcal{C})$.

Theorem 3 The expected benefit of agent A and agent B in the disputed items is $\frac{1}{2}U_A(\mathcal{C})$ *and* $\frac{1}{2}U_B(\mathcal{C})$ respectively.

Proof Make the inductive assumption that the number of disputed items m is greater than or equal to two.

a) It is easy to see that the above theorem is true when $m = 1$ *or* $m = 2$.

b) Assuming that $2 \leq m < k$, the above theorem is true.

c) Now let's consider the case where $m = k$.

When k is odd, agent A and agent B both have a one-half chance of winning the draw. In round k, the expected benefit of A is $\frac{1}{2}u_A(o_k)$, and $\frac{1}{2}U_A(\mathcal{C}_{k-1}) + \frac{1}{2}u_A(o_k) = \frac{1}{2}U_A(\mathcal{C})$. Similarly, the expected benefit value of B is $\frac{1}{2}U_B(\mathcal{C})$, so the theorem is true. When k is even, suppose that agent A wins in round $k-1$, meaning that A cannot declare item in round k. The expected benefit is $\frac{1}{2}U_A(\mathcal{C}_{k-2}) + u_A(o_{k-1})$. If agent A fails in round $k-1$, meaning that the item in round k belongs to agent A. The expected benefit of A is $\frac{1}{2}U_A(\mathcal{C}_{k-2}) + u_A(o_k)$. Taking the above two cases into consideration, the expected benefit of A is $\frac{1}{2}U_A(\mathcal{C}_{k-2}) + \frac{1}{2}u_A(o_{k-1}) + \frac{1}{2}u_A(o_k) = \frac{1}{2}U_A(\mathcal{C})$. Similarly, the expected benefit value of B is $\frac{1}{2}U_B(\mathcal{C})$, So the theorem is true.

Theorem 4 Shelving disputes in the loser-reporting can achieve the result of SPNE under parallel allocation.

Proof Using the shelving dispute method, we can know that the set of x_i that is not in dispute is $\mathcal{X}$, and the set of y_i that is not in dispute is $\mathcal{Y}$. In combination with theorem 1, we can know that the expected benefit of agent A and agent B in the disputed items is $\frac{1}{2}U_A(\mathcal{C})$ *and* $\frac{1}{2}U_B(\mathcal{C})$ respectively, So we get $U_A = u_A(\mathcal{X}) + \frac{1}{2}u_A(\mathcal{C})$ and $U_B = u_B(\mathcal{Y}) + \frac{1}{2}u_B(\mathcal{C})$.

Now we offer an algorithm to find an SPNE strategy for agent A and agent B. S_k represents the kth item in S, and R_k represents the kth item in set R.

```
$Algorithm$ 1: Finding an SPNE strategy
$input$: Preference order of agents $A\ and\ B$ $(\succ_A, \succ_B)$
$output$: An SPNE strategy for agent $A$ and agent $B$ $(\delta_A, \delta_B)$
1. $\mathcal{O} \longleftarrow \{o_1, o_2, \cdots, o_{|\mathcal{O}|}\}$
2. $\succ_A \longleftarrow o_1 \succ_A o_2 \succ_A \cdots \succ_A o_{|\mathcal{O}|}$
3. $\succ_B \longleftarrow o'_1 \succ_B o'_2 \succ_B \cdots \succ_B o'_{|\mathcal{O}|}$
4. $S \longleftarrow \{o_1, o_2, \cdots, o_{|\mathcal{O}|}\}$
5. $R \longleftarrow \{o'_1, o'_2, \cdots, o'_{|\mathcal{O}|}\}$
6. $k \longleftarrow |\mathcal{O}|$
7. $\mathcal{X} \longleftarrow \emptyset$
8. $\mathcal{Y} \longleftarrow \emptyset$
9. $\mathcal{C} \longleftarrow \emptyset$
10. $for\ i \longleftarrow 1\ to\ |\mathcal{O}|$
11.     $if\ S_k = R_k$
12.         $\mathcal{C} \longleftarrow \{S_k\} \cup \mathcal{C}$
13.     else
14.         $\mathcal{X} \longleftarrow \{R_k\} \cup \mathcal{X}$
15.         $\mathcal{Y} \longleftarrow \{S_k\} \cup \mathcal{Y}$
16.     $R \longleftarrow R \setminus \{R_k \cup S_k\}$
17.     $S \longleftarrow S \setminus \{R_k \cup S_k\}$
18.     $k \longleftarrow |R|$
19. $\delta_A \longleftarrow rev(\mathcal{X}) + rev(\mathcal{C})$
20. $\delta_B \longleftarrow rev(\mathcal{Y}) + rev(\mathcal{C})$
21. return $(\delta_A, \delta_B)$
```

Example 2: See Example 1. Apply Algorithm 1. We can find an SPNE strategy for agent A and agent B (δ_A, δ_B) according to steps 2-20.

- $\mathcal{X} = \{o_5\}, \mathcal{Y} = \{o_7\}, \mathcal{C} = \{\emptyset\}$
- $\mathcal{X} = \{o_5\}, \mathcal{Y} = \{o_7\}, \mathcal{C} = \{o_2\}$
- $\mathcal{X} = \{o_1, o_5\}, \mathcal{Y} = \{o_3, o_7\}, \mathcal{C} = \{o_2\}$
- $\mathcal{X} = \{o_4, o_1, o_5\}, \mathcal{Y} = \{o_6, o_3, o_7\}, \mathcal{C} = \{o_2\}$
- $\delta_A = \{o_5, o_1, o_4, o_2\}$,$\delta_B = \{o_7, o_3, o_6, o_2\}$

After adopting the shelving dispute algorithm, the expected benefit value of agent A is $u_A(\mathcal{X}) + \frac{1}{2}u_A(\mathcal{C}) = 18.5$, and the expected benefit value of agent B is $u_A(\mathcal{Y}) + \frac{1}{2}u_A(\mathcal{C}) = 17$.

5. CONCLUSIONS

In order to achieve the SPNE of the parallel allocation fairer, we propose the loser-reporting, which guarantees the difference between the number of items obtained by the agent does not exceed one. We study the loser-reporting under parallel allocation, propose several theorems and proofs around the perfect information game problem of the allocation of indivisible goods, and introduce the concept of shelving dispute to the expected benefit under loser-reporting is equal to the expected benefit of the SPNE under parallel allocation.

However, when the number of agents is more than two and the number of items sharply increases, By constructing the game tree, we can see that the scale of the problem becomes larger. Whether the three agents can reach SPNE is our next major work.

ACKNOWLEDGMENT

This work is supported by the National Natural Science Foundation of China, the Guangxi Natural Science Foundation, and the Guangxi Project of technology base and special talent under grants No.61105039, No.61662011, 2020GXNSFAA297186, AD19110022, and Innovation Project of GUET Graduate Education (Issue Number 2021YCXS064, 2021YCXS055).

REFERENCES

[1] G. Bombini, N. D. Mauro, S. Ferilli, and F, Esposito, “Classifying agent behaviour through relational sequential patterns,” in Agent and Multi-Agent Systems: Technologies and Applications, 4th KES International Symposium, KES-AMSTA 2010, Gdynia, Poland, June 23-25, 2010, Proceedings. Part I, P. Jedrzejowicz, N. T. Nguyen, R. J. Howlett, et al., Eds., Lecture Notes in Computer Science 6070, 273–282, Springer (2010).

[2] E. Budish and E. Cantillon, “The multi-unit assignment problem: Theory and evidence from course allocation at harvard,” American Economic Review 102, 2237–71 (2012).

[3] T. Sandholm, “Algorithm for optimal winner determination in combinatorial auctions,” Artificial Intelligence. 135(1-2), 1–54 (2002).

[4] S. Bouveret and J. Lang, “A general elicitation-free protocol for allocating indivisible goods,” in IJCAI 2011, Proceedings of the 22nd International Joint Conference on Artificial Intelligence, Barcelona, Catalonia, Spain, July 16-22, 2011, T. Walsh, Ed., 73–78, IJCAI/AAAI (2011).

[5] S. Bouveret and J. Lang, “Manipulating picking sequences,” in ECAI 2014 - 21st European Conference on Artificial Intelligence, 18-22 August 2014, Prague, Czech Republic – Including Prestigious Applications of Intelligent Systems (PAIS 2014), T. Schaub, G. Friedrich, and B. O’Sullivan, Eds., Frontiers in Artificial Intelligence and Applications 263, 141–146, IOS Press (2014).

[6] Y. Chen, J. K. Lai, D. C. Parkes, and A. D. Procaccia, “Truth, justice, and cake cutting,” in Proceedings of the Twenty-Fourth AAAI Conference on Artificial Intelligence, AAAI 2010, Atlanta, Georgia, USA, July 11-15, 2010, M. Fox and D. Poole, Eds., AAAI Press (2010).

[7] Y. J. Cohler, J. K. Lai, D. C. Parkes, and A. D. Procaccia, “Optimal envy-free cake cutting,” in Proceedings of the Twenty-Fifth AAAI Conference on Artificial Intelligence, AAAI 2011, San Francisco, California, USA, August 7-11, 2011, W. Burgard and D. Roth, Eds., AAAI Press (2011).

[8] W. Huang, L. Zhang, Y. Huang, and J. Lou, “Allocating indivisible objects with a parallel method insensitive to identities,” IEEE Access 5, 22880–22891 (2017).

[9] W. Huang, W. Huang, and D. Cai, “Finding EFL and EQL Allocations of Indivisible Goods,” 2020 International Conference on Computer Vision, Image and Deep Learning (CVIDL) (2020).

[10] S. Bouveret, U. Endriss, and J. Lang, “Fair division under ordinal preferences: Computing envy-free allocations of indivisible goods,” in ECAI 2010 - 19th European Conference on Artificial Intelligence, Lisbon, Portugal, August 16-20, 2010, Proceedings, H. Coelho, R. Studer, and M. J. Wooldridge, Eds., Frontiers in Artificial Intelligence and Applications 215, 387–392, IOS Press (2010).

[11] D. A. Kohler and R. Chandrasekaran, “A class of sequential games,” Operations Research. 19(2), 270–277 (1971).

[12] W. Huang, H. Liu, G. Dai, and A. Abraham “A tractable multiple agents protocol and algorithm for resource allocation under price rigidities,” Applied Intelligence: The International Journal of Artificial Intelligence, Neural Networks, and Complex Problem-Solving Technologies, 43(3):564-577 (2015).

[13] T. Kalinowski, N. Narodytska, and T. Walsh, “A social welfare optimal sequential allocation procedure,” in IJCAI 2013, Proceedings of the 23rd International Joint Conference on Artificial Intelligence, Beijing, China, August 3-9, 2013, F. Rossi, Ed., 227–233, IJCAI/AAAI (2013).

[14] T. Kalinowski, N. Narodytska, T. Walsh and L. Xia, “Strategic behavior in a decentralized protocol for allocating indivisible goods,” in Proceedings of the 4th international workshop on computational social choice 12, 251-262 (2012)

[15] S. Yin, S. Wang, L. Zhang, C. Kroner “Dominant Resource Fairness with Meta-Types” in IJCAI 2021 Proceedings of the Thirtieth International Joint Conference on Artificial Intelligence, Montreal, Canada, August 19-27,2021, Z. Zhou., 486-492, IJCAI(2021)

[16] A. Baklanov, P. Garimidi,V. Gkatzelis and D. Schoepflin, “PROPm allocations of indivisible goods to multiple agents,” in IJCAI 2021 Proceedings of the Thirtieth International Joint Conference on Artificial Intelligence, Montreal, Canada, August 19-27,2021, Z. Zhou., 38-44, IJCAI(2021)

[17] W. Huang, J. Lou, and Z. Wen, “A parallel elicitation-free protocol for allocating indivisible goods,” In Seventh Multidisciplinary Workshop on Advances in Preference Handling (MPREF-13), (2013).

[18] W. Huang, C. Huang, Zhi.Xu et al., “Optimistic Manipulation under Allocation Policy of Loser Reporting for Multi-Agent Systems,” Proc. SPIE 12168, International Conference on Computer Graphics, Artificial Intelligence, and Data Processing (ICCAID 2021)

[19] W. Huang, D. Cai, Y. Lu, et al., “Computing the Subgame Perfect Nash Equilibriums in Parallel Allocation Indivisible Items,” Journal of Physics Conference Series, 1621:012073 (2020).

High-risk Areas Identification of Transmission Lines Based on Historical Warning Information

Fei Wang*, Lingqi Kong
Zhiyang Innovation Technology Co., Ltd., Zibo, 255000, China
* Corresponding author: wangfeichn@163.com

ABSTRACT

With the upgrade of transmission line maintenance technology, the visual remote inspection of transmission line channel is widely used. However, the application of these marked data is currently in the stage of statistical analysis and report form, a deeper data mining work has not been carried out, such as the identification of areas with high incidence of alarm. This paper presents a method of analysis of high risk area of transmission line channel based on historical early warning data, which can be applied to the field of transmission line maintenance. By preprocessing the transmission line visual alarm data, using the improved *k-means* algorithm, analyze the clustering results, and finding out the series of modeling and data analysis. In addition, when determining the initial points, we propose the maximum and minimum longitude and latitude coordinates of the alarm data set, divide a certain number of longitude and latitude grid, obtain the candidate data within each grid, and the method of obtaining the initial point set after screening. Based on the initial point set, the alarm area distribution is reasonable, and the regional stability does not drift. This paper can identify the visual alarm data of transmission lines with high alarm incidence areas, and provide effective data support for maintenance personnel to guide the deployment of human resources and ensure the safe operation of transmission lines.

Keywords: areas with high hidden risk; historical warning data; data processing; transmission line channel; data analysis; intelligent transportation and inspection

1. INTRODUCTION

With the upgrade of transmission line maintenance technology, the visual remote inspection of transmission line channel is widely used. At present, the automatic identification of visual information has been realized and the alarm objects appear in the image, such as tower crane, mountain fire, excavator, etc. However, the application of these marked data is currently in the stage of statistical analysis and report form. For example, the statistical analysis of the alarm information of a single device or multiple devices in a designated area completes the collection from data to information, but a deeper data mining work has not been carried out, such as the identification of areas with high incidence of alarm.

In addition, the identification of the areas with high alarm incidence based on the above alarm identification results is helpful to improve the work efficiency of the transmission line maintenance personnel, because the alarm data is equipped with the longitude and latitude coordinate information and is independent of each other. Therefore, the cluster analysis is an effective means to achieve the above technical objectives. There are many existing cluster analysis algorithms, which typically include the following algorithms: dividing clustering, given a set of *N* objects, dividing the method to build *K* partitions of data, where each partition represents a cluster (class); hierarchical clustering, divided into condensed and split methods, for example, starting with each object as a cluster, merging close objects step by step from bottom up, until the iterative stop condition is met. The split method process; density-based clustering, as long as the density (the number of objects or data points in the "neighborhood") exceeds a certain threshold. Because the alarm data has not only the longitude and latitude coordinate information, but also the alarm times, the above clustering method of division and clustering is the best method to realize the identification of areas with high frequency of alarm. For the cluster order *K* to be generated, *K* objects will be randomly selected as the initial points, and the clustering results are sensitive to the initial points. As for the application scenarios of the identification of the high incidence of alarm areas, the division and clustering results of the above different initial points are somewhat different, resulting in the density and drift in the high incidence of alarm areas[1].

This paper provides an efficient and reliable identification method of alarm high incidence areas, and can provide a reasonable and reliable initial point collection, provide technical support for the identification of alarm high incidence areas, make the calculation results stable, reasonable distribution, and the obtained alarm high incidence areas stable and

International Conference on Computer Graphics, Artificial Intelligence, and Data Processing (ICCAID 2022)
edited by Ruishi Liang, Jing Wang, Proc. of SPIE Vol. 12604, 126041L
© 2023 SPIE · 0277-786X · doi: 10.1117/12.2674562

not drift. The results of this paper can provide the information of the high incidence area for the transmission line maintenance personnel.

2. THE HISTORICAL EARLY WARNING DATA IS PROCESSED TO OBTAIN THE INITIAL POINTS OF REGIONAL DIVISION

2.1 The latitude and longitude grid for the data

Based on the latitude and longitude coordinates of the visual alarm data of the transmission line channel and the required number of regions *K*, the longitude and latitude and longitude grid is divided, where *K* is specified, the division cluster is represented by *K-means* algorithm as a typical representative, and the number of regions *K* corresponds to the cluster order *K* in the division cluster algorithm[2].

Based on the visual alarm data of the transmission line channel, the maximum longitude lng_{max}, minimum longitude lng_{min}, maximum latitude lat_{max} and minimum latitude lat_{min} in the alarm data coordinate set are calculated respectively. The longitude difference as shown in the formula (1), and the latitude difference as shown in the formula (2). Where the number of alarm areas *K*, when calculating the number of latitude and longitude grid ranks *N*, if the square root of *K* is an integer, namely *SQRT (K)* is an integer, as shown in the formula (3); if the square root of *K* is not an integer, then set *INT (SQRT (K))*, set as shown in the formula (4). The latitude and longitude grid division step size was calculated, longitude step as shown in the formula (5) and latitude step as shown in the formula (6).

$$lng_{diff} = lng_{max} - lng_{min} \tag{1}$$

$$lat_{diff} = lat_{max} - lat_{min} \tag{2}$$

$$N = SQRT\ (K) \tag{3}$$

$$N = INT\ (SQRT\ (K)) + 1 \tag{4}$$

$$lng_{step} = lng_{diff} / N \tag{5}$$

$$lat_{step} = lat_{diff} / N \tag{6}$$

The latitude and longitude lines required for calculating the latitude and longitude grid, *N-1* meridian are calculated from the longitude direction, Combining the meridian corresponding to the initial maximum minimum longitude, *N + 1* line, Its warp value is successively $\{lng_{min}, lng_{min}+lng_{step}, lng_{min}+2*lng_{step}, lng_{min}+3*lng_{step},..., lng_{min}+(N-1)* lng_{step}, lng_{max}\}$, The same latitudinal direction yields *N + 1* latitude lines, Its latitude value are in order $\{lat_{min}, lat_{min}+lat_{step}, lat_{min}+2* lat_{step}, lat_{min}+3* lat_{step},..., lat_{min}+(N-1)* lat_{step}, lat_{max}\}$. Based on the above operation, the latitude and longitude grid containing *N * N* areas is divided into the longitude and latitude lines. The step length mentioned in the paper refers to the calculation of the latitude and longitude as the plane coordinates; the calculation error of the plane coordinates and the calculation of the latitude as the plane coordinates; *SQRT (K)* refers to the *K* square root calculation; *INT (SQRT (K))* refers to the *SQRT (K)* and discard the decimal part[3].

2.2 Area division of initial points based on grid candidate data

Each grid was traversed in turn to obtain the candidate data in each grid, the candidate data was ranked by the number of alarms, and the first *K* bar was taken as the initial point. Candidate data refers to the selected alarm data with longitude and latitude coordinates and the number of alarms. This paper starts with the grid with the least longitude and latitude, and traverses each grid in turn. You can also traverse from the other four vertices, or go clockwise or counterclockwise from the center point, but none of them has much impact on the results. No matter how to traverse, as long as there is no omission, the candidate data will finally be sorted according to the number of alarms, take the front K bar, the results are the same, and the time complexity is *o(n)* [4].

Based on the longitude and longitude grid obtained in section 2.1, the traversal search is conducted in turn, that is, from the grid with the least longitude and latitude, that is, the grid containing (lng_{min}, lat_{min}) coordinates, successively in the longitude direction. After the search, the latitude value is increased, and then in the longitude direction again until the traversal is completed. Based on the set traversal sequence, look for the candidate data, record the initial number of candidate data $K_{init} = 0$: if a grid has 1 or more alarm data, that is, the longitude and latitude coordinates of one or more alarm data falls in the grid, record the coordinates of the alarm data in the grid and the alarm number of candidate data is 1; if a grid has no alarm data, skip the current grid and continue to search in the next grid in order. Determine the initial point set includes: based on the candidate data search method, after the grid traversal, if $K_{init} \geq K$ is met, stop the search and return the recorded K_{init} bar data; if $K_{init} < K$ after traversing all the grids, add N by 1, repeat the traversal process until the grid time meets $K_{init} \geq K$. The resulting $\mathrm{K_{init}}$ bar candidates are sorted according to the number of alarms, and the top K bar candidate data with the largest number of alarms is taken as the initial point set[5].

For alarm high incidence of cluster identification determines the initial point set, and calculate the alarm high incidence of stable does not drift, determine the initial point, uniform distribution, reasonable value, for the cluster division, to provide reasonable initial points, the final alarm area distribution is reasonable, and the latitude and longitude coordinates as plane coordinates for calculation, improve the calculation efficiency, and the error is within the acceptable range.

3. DETERMINE THE NUMBER OF AREAS WITH HIGH RISK

This chapter provides a visual alarm area division method of determining the number of transmission line channels, with a large number of visual alarm data of transmission line channel to train the model offline advice and online decision acquisition, automatically determine the number of areas, through the off-time training and online decision separation, solve the problem of real-time calculation, based on the determined value of transmission line channel visual alarm area identification, find out of the alarm area coverage radius in accordance with the patrol radius of transmission line maintenance personnel, bring convenience for the transmission line maintenance personnel[6].

Based on the visual alarm data of transmission line channel, the recommended value is obtained by offline training, and determine the value of the visual alarm area division of transmission line channel by online decision. The specific analysis process is: count the alarm data n, the proposed initial value is set to $n / 2$; use the initial value to identify the high warning area; make the comparison of the ratio between the radius of the high alarm area and the patrol radius of the maintenance personnel to test the maximum confidence interval value, and the process is repeated until the recommended value is obtained. The recommended value is the initial value to identify the area of high alarm incidence. Comparing the ratio of the alarm high incidence area radius and the patrol radius of maintenance personnel. If the maximum value is greater than the confidence interval, if it is lower than the minimum value of the confidence interval, and the process is repeated until the determined value is obtained. Offline training means when the model is initialized or when the alarm data increases significantly, not before every online decision; $n / 2$ means $n / 2$ when n is even and $(n-1) / 2$ when n is odd. Upper and dip refer to updip as shown in the formula (7) ; dip as shown in the formula (8) .

$$k_i = INT\ ((k_{i-1} + k_{i-2}) / 2) \tag{7}$$

$$k_i = INT((k_{i-1} + k_{i-3}) / 2) \tag{8}$$

Can automatically determine the number of visual alarm areas for the transmission line channel, the identification of the coverage radius of the radius of the transmission line maintenance personnel brings convenience; the constructed model adopts off-line training and online decision-making mode to obtain the determined value, which greatly reduces the real-time calculation amount; adopts the dichotomy algorithm, which greatly improves the calculation speed of the recommended value for offline training. Identify the high alarm areas of transmission line visual alarm data, and provide effective data support for maintenance personnel to guide human deployment and ensure the safe operation of transmission lines. Based on the clustering machine learning algorithm, unsupervised learning without data marking can accurately identify the areas with high alarm; with improved *k-means* clustering algorithm, the data of millions of magnitude can be calculated within a few minutes, providing technical support for the real-time identification of high-incidence warning areas of transmission lines.

4. WEIGHT MARK AND DETERMINE THE AREA OF HIDDEN DANGER

This chapter provides a visual method to identify the areas with high alarm incidence of transmission line channels. By weighting the warning data and using the unsupervised clustering machine learning algorithm, it solves the problems of poor division accuracy and slow calculation speed of a large number of alarm data area division, which brings convenience to the transmission line maintenance personnel. Conduct preprocessing of transmission line visual alarm data and weight mark the processed transmission line visual alarm data; cluster the improved *k-means* algorithm based on the weight marked data; analyze the clustering results and find the cluster with the largest weight, namely the identification result[7].

Preprocessing refers to the separation of a certain type of alarm data from the data source, the data only retain the longitude and latitude attributes, and all the data in the same order. Weight mark refers to combine the data with the same longitude attribute value and the same latitude attribute value into a piece of data, increase the weight attribute, and the number of data bars is taken as the weight value of the data. The improved *k-means* algorithm means that the data involved in the clustering has weights, and the weight values of the weight of the member data are involved in each iteration. The cluster with the largest weight refers to the sum of the member data weight values within each cluster, and the cluster with the largest value is the cluster with the largest weight[8].

This paper proposes a method to identify the areas of high incidence of transmission hidden dangers by using the data of historical hidden dangers of transmission line channels, so as to provide a decision basis for the intelligent transportation and inspection of transmission lines. Can realize the higher level use of data, complete from information to knowledge, better play to the value of data, the transmission channel intelligent inspection application scenarios belong to the pioneering work, for maintenance personnel to provide effective data support to assist decision-making, such as strengthening the inspection of some lines, ensure the safe operation of transmission lines.

5. THE EXAMPLE ANALYSIS

5.1 Example of determining the number of area divisions

In the visual image alarm data of a local transmission line, all 3,265 pieces of alarm data with weight of the crane, among which 2,786 pieces of data in the last three months need to be identified in the area of high alarm. The maintenance personnel of the transmission line require that the radius of high alarm area is 5000m, and the upper and lower error is 5%. This example will automatically determine the number of alarm zones divided based on the above data. The conditions can be obtained, 1) the number of offline data is 3265; 2) the online data is 2786; 3) the maintenance radius R=5000m; 4) the confidence interval is [0.95,1.05]. In particular, for the number of alarm data, the alarm data is processed by weight, that is, the coordinate data of the same longitude and latitude is merged into one, and the number of occurrence of the data is represented by the weight value[9].

The first step is for the offline training, The specific steps are as follows: take the proposed initial value $k_2= (3265-1) / 2=1632$, Identify the areas with high alarm incidence; The radius of the high warning area is 1956m, $1956/5000=0.3912<0.95$, A dip is required, $K_3 = INT(1 + 1632) / 2=816$; Set the recommended value to 816 to identify the alarm high incidence area; The radius of the high alarm area is 17658m, $17658/5000=3.5316>0.95$, Up-exploration needs to be performed, $K_4 = INT(1632 + 816) / 2=1224$; Repeat with the above steps, At $k_{11}=1118$, The radius of the high warning area is 5041m, $17658/5000=1.0082 \in [0.95, 1.05]$, The value of 1118 is the recommended value.

The second step is the online decision-making, The specific steps are as follows: identify the alarm area based on the recommended value 1118, The radius of the high warning area is 4658m, $4658/5000=0.9316<0.95$, Need to drop down; Take k value 1117 to identify high alarm area, The radius of the high alarm area is 4732m, $4732/5000=0.9464<0.95$, Need to drop down; Take the k value 1116 to identify the high alarm incidence area, The radius of the high warning area is 4798m, $4798/5000=0.9596 \in [0.95, 1.05]$, The value of 1116 is the determined value.

This example automatically determines the number of alarm areas, so that the identified high alarm areas meet the maintenance radius of transmission line maintenance personnel.

5.2 Identify an example of areas with a high incidence of hidden dangers

There are 204,134 pieces of visual alarm data of a transmission line, etc. The alarm types in the image include 12 different types, such as crane, bulldoz, suspended matter and mountain fire. The alarm data includes 25 fields, including the alarm self-increase ID, alarm occurrence time, alarm content, longitude information, latitude information, subordinate line, and image storage ID. This example, based on the above data, identifies the data of the crane appearing in the visual alarm image of the transmission line, and the number of area divisions is set to 10. This example filters 204134 pieces of data and retains only the longitude and latitude information of the crane data appearing in the alarm content. A total of 37481 pieces of data are obtained and sorted in ascending order by the longitude attribute values. The data examples as shown in the Table 1[10].

Table 1 Example of the preprocessed alarm data

Sequence number	longitude	latitude	Sequence number	longitude	latitude
1	115.33386	36.56522	6	115.46275	36.57120
2	115.33386	36.56522	7	115.46275	36.57120
3	115.33386	36.56522	…	…	…
4	115.33386	36.56522	37486	121.66307	37.19030
5	115.33386	36.56522	37487	121.66307	37.19030

The weight marks the data. The marking method is to keep the data with different numbers of the same content, and adds a new field to record the repeated times of the data. For example, 1-5 pieces of data in the sample data will generate a new data "115.33386 36.56522 5". Process the data to obtain 3265 pieces of data with weight, and part of the data after the weight mark is as shown in the Table 2.

Table 2 Example of data with weight markers

Sequence number	longitude	latitude	weight	Sequence number	longitude	latitude	weight
1	115.33386	36.56522	5	6	115.90220	35.84560	48
2	115.46275	36.57120	12	7	115.90410	35.85006	21
3	115.75532	35.36612	4	…	…	…	…
4	115.80576	34.95025	9	3264	121.64983	37.20164	18
5	115.88972	36.57209	3	3265	121.66307	37.19030	6

From the 3265 data, randomly selected "to divide area number" bar data, namely 10 data as the initial cluster centroid, all data for the first iteration, iteration process not heavy, after the completion of 10 clusters, a new centroid for each cluster, calculate the weight of the member data need to participate in the calculation. If the first cluster has 423 members, take the longitude attribute as an example, the calculation formula is as shown in the formula (9):

$$R = \sum_{i=1}^{423} x_i w_i / \sum_{i=1}^{423} w_i \qquad (9)$$

This step is repeated until the cluster to which all data members belong no longer changes, where the data converges and all centroids do not change. The obtained clustering results calculate the sum of the weights of each cluster, and the cluster with the largest weight is the required result. In this case, cluster 7 has the largest weight and the weight value of 1264. Therefore, the area covered by the latitude and longitude coordinates of the data members under cluster 7 is the area with high incidence of crane alarm.

In the example, first to the data to merge and weight mark, the data involved in cluster analysis from 37481 to 3265, and the weight value is introduced into the centroid calculation process, the calculation speed, greatly improve the comparison, with the traditional *k-means* clustering algorithm for multiple tests, from an average of 125 seconds to 9 seconds, can quickly provide accurate and real-time alarm area data information for the transmission line maintenance personnel.

6. CONCLUSION

As most areas have realized the transmission line channel visual remote patrol, and can automatically identify the hidden dangers in the transmission channel information, such as mechanical, foreign body, fireworks class, resulting in a large number of marked data, and predictable with the introduction of more technology and fusion, accumulation of channel hidden danger data will soon multiply or exponentially increased. But the use of this kind of data, most of the application stay in the degree of statistical analysis, report, such as a single device or specified multiple equipment hidden dangers of the information statistical analysis, only completed from the data to the information of the collection, but a deeper level of large data analysis means to further enhance the knowledge of information work. This is because the value density of channel hidden danger data is very low, and single point analysis or short-term data analysis cannot produce actual value. However, channel hidden danger data naturally has marked attributes, which provides the possibility for the improvement of value density through big data analysis.

Based on the historical hidden danger data of the transmission channel and combined with the big data analysis means, this paper realizes the identification of the areas with a high incidence of the transmission hidden danger, which can be used for intelligent transportation and inspection assistance decision-making. This paper has been used in multiple field application, there are some shortcomings, such as data dimensions used is less, especially for the data related to natural disasters applied less, subsequent continuous iterative optimization of the model, such as combining hidden danger distribution map, such as wind, ice, dirt, minefield, dancing, bird heat map, according to different hazard characteristics of the corresponding weight customization, on the premise of guarantee model generalization ability to achieve the effect of different regional differentiation.

REFERENCES

[1] Ren, R., Zhang, L., Liu, L., & Yuan, Y.. (2021).Two auvs guidance method for self-reconfiguration mission based on monocular vision.IEEE Sensors Journal, 21(8), 10082-10090.

[2] Foti, G., Guerriero, M., Faccioli, N., Fighera, A., & Carbognin, G..(2021).Identification of bone marrow edema around the ankle joint in non-traumatic patients: diagnostic accuracy of dual-energy computed tomography. Clinical Imaging, 69(7), 341-348.

[3] Zhiyang Innovation Technology Co., Ltd., Wang Fei, Zhang Wanzheng, etc. A method for determining the initial point of visual alarm area division of transmission line channel: ZL201910745222.9 [P].2020-06-26.

[4] D Dan, Ge, L. , & Yan, X. . (2019). Identification of moving loads based on the information fusion of weigh-in-motion system and multiple camera machine vision. Measurement, 144, 155-166.

[5] Amini M H, Mallahzadeh A (2020). Analyzing Radiated Susceptibility of Superconducting Microstrip Transmission Line Under Plane Wave Excitation. IEEE Transactions on Electromagnetic Compatibility, PP(99):1-8.

[6] Schindler L , Roux P L , Fourie C J(2020) .Impedance Matching of Passive Transmission Line Receivers to Improve Reflections Between RSFQ Logic Cells. IEEE Transactions on Applied Superconductivity, PP(99):1-1.

[7] Zhiyang Innovation Technology Co., Ltd., Wang Fei, Zhan Xingang, etc. An identification method for the high incidence area of visual alarm of transmission line channels: ZL201910700272.5 [P].2020-04-07.

[8] Wu Y , H Xu, Gong X , et al(2020).A Ladder Transmission Line Model for the Extraction of Ultralow Specific Contact Resistivity--Part I: Theoretical Design and Simulation Study. IEEE Transactions on Electron Devices, PP(99):1-8.

[9] Aleksandrov V V, Branitskii A V, Grabovsky E V, et al(2021).Simulations of the Evolution of the Heterophase Electrode of a Vacuum Transmission Line During the Passage of a High Current Pulse. Plasma Physics Reports, 47(4):355-361.

[10] X Tao, Zhang D, Wang Z, et al (2020). Detection of Power Line Insulator Defects Using Aerial Images Analyzed With Convolutional Neural Networks. IEEE Transactions on Systems, Man, and Cybernetics: Systems, 50(4):1486-1498.

Research on Student Psychological Crisis Monitoring Based on Big Data

Zi Wang*[1,2], Dahuai Yu[3]
[1]School of Maxism, Hohai University, China
[2]Businiss School,Changshu Institute of Technology, China
[3]School of Maxism, Hohai University, China
* Corresponding author: wangzi@cslg.edu.cn

ABSTRACT

The study investigated the mental health status, learning and living behaviors of college students, and found that there were significant differences in their mental health levels among college students with different learning and living behaviors. Taking sleep, late return, communication with classmates, study, absenteeism and sudden illness as information collection indicators, a psychological crisis monitoring and early warning system based on big data technology is established to conduct two-level monitoring and early warning, realize dynamic tracking management and accurate monitoring and early warning of college students' mental health status, and improve the level of mental health education in colleges and universities.

Keywords:big data, Psychological crisis,Monitoring and early warning

1. INTRODUCTION

Mental health In recent years, college students' psychological crisis events have occurred from time to time, which has a great impact on their physical and mental health, especially due to the COVID-19, college students' normal learning, communication and life rhythm are often disrupted. At the National Promotion Meeting for College Students' Mental Health Education, it was emphasized that "scientific identification, real-time early warning, professional consultation and proper response are the key to doing a good job. Process management should be strengthened, psychological evaluation should be comprehensively covered and accurately applied, and early warning management should be all-weather and one-stop" [1]. Precise monitoring and early warning of college students' psychological crisis is an important task of college mental health education. The development of big data technology provides new ideas and methods for colleges and universities to solve the psychological crisis of college students. Liang Cheng and others [2] proposed to integrate big data thinking into college psychological crisis early warning work by improving data awareness, improving relevant systems, and establishing a data platform. Wang Wei et al. [3] proposed the construction path of the psychological early warning system based on the structural analysis of the system and the characteristics of big data. Wang Fang et al. [4] believed that the college students' psychological early warning system based on big data technology was divided into three parts, namely, the mental health survey system, the dynamic evaluation system of psychological conditions and the real-time monitoring system of psychological changes, and they tested it. Li Meng et al. [5] built a college students' psychological crisis warning model based on social media big data. Yu Feng et al. [6] used big data technology to establish an intelligent evaluation system for mental health and achieve early warning of psychological crisis. In general, the existing research focuses on platform construction, such as data collection indicators, analysis and processing platform, monitoring and early warning system, etc. The collection indicators are complete and lack of focus. The data analysis is complex, which is difficult to make simple judgments, and it is difficult for ordinary mental health workers to complete. Therefore, further optimizing data collection indicators and establishing a more convenient, accurate and timely psychological crisis monitoring and early warning platform will help solve the work difficulties faced by colleges and universities.

International Conference on Computer Graphics, Artificial Intelligence, and Data Processing (ICCAID 2022)
edited by Ruishi Liang, Jing Wang, Proc. of SPIE Vol. 12604, 126041M
© 2023 SPIE · 0277-786X · doi: 10.1117/12.2674610

2. RESEARCH METHODS

2.1 Research object

A questionnaire was randomly distributed to students in a university, and 5871 valid questionnaires were received. Among them, there are 3142 male students and 2729 female students according to gender; According to grade, there are 942 senior students, 672 junior students, 2468 sophomores and 1789 freshmen.

2.2 Research tools

2.2.1 Symptom Checklist

There are 90 items in the scale. Each item adopts a 5-level scoring system, and 10 factors are used to reflect 10 aspects of psychological symptoms, including somatization, obsessive compulsive symptoms, interpersonal sensitivity, depression, anxiety, hostility, terror, paranoia, psychosis and others.

2.2.2 Self compiled questionnaire

A questionnaire was prepared to investigate the psychological health of college students under the normalization of the epidemic situation, to understand the students' learning and living behavior, including sleep, late return, communication with classmates, school work, absenteeism and sudden illness.

2.2.3 Data processing

Excel is used for data processing, and SPSS and Python are used for data analysis.

3. RESEARCH RESULTS

3.1 Mental Health Status of Different College Students

3.1.1 The analysis of the self-evaluation scale of college students with different sleep conditions

The analysis of variance of the factors in the self-evaluation scale of college students with different sleep conditions is shown in Table 1. It can be seen that college students with different sleep conditions have extremely significant differences in each factor of the self-assessment scale. Further pairwise comparison showed that there were significant differences in various factors among college students with different sleep conditions ($p<0.001$).

Table1. Analysis of the differences of self rating scales in different sleep situations

Factor	Better	General	Poor	F
Somatization	1.16	1.26	1.44	269.83
Obsession	1.46	1.65	1.87	229.13
Relationship	1.34	1.51	1.73	228.92
Depressed	1.23	1.38	1.64	328.09
Anxious	1.25	1.38	1.59	257.81
Hostile	1.19	1.31	1.49	216.09
Terror	1.17	1.28	1.42	259.92
Paranoia	1.21	1.32	1.51	215.71
Psychopathic	1.19	1.34	1.54	234.09
Other	1.21	1.37	1.67	283.72
Score	1.24	1.39	1.59	243.17

3.1.2 Analysis of the self-evaluation scale of college students who did not return at night

The results of ANOVA of the factors in the self-evaluation scale of college students who did not return at night are shown in Table 2. It can be seen that college students who did not return at different nights had extremely significant differences in each factor of the self-evaluation scale. Further pairwise comparison found that there was no significant difference between "no" and "less" students in terror factors ($p=0.07$), but there were significant differences in other factors ($p<0.05$).

Table2. Analysis on Self rating Scale of College Students Who did not Return at Night

Factor	None	Less	More	F
Somatization	1.26	1.33	1.47	39.15
Obsession	1.63	1.69	1.88	17.66
Relationship	1.50	1.56	1.78	26.22
Depressed	1.38	1.45	1.67	36.68
Anxious	1.37	1.46	1.65	39.33
Hostile	1.30	1.39	1.55	52.68
Terror	1.27	1.31	1.49	25.35
Paranoia	1.32	1.38	1.58	29.67
Psychopathic	1.31	1.39	1.61	43.45
Other	1.36	1.44	1.67	36.87
Score	1.38	1.45	1.64	43.26

3.1.3 Analysis of college students' self-assessment scale

The ANOVA of each factor in the college students' self-evaluation scale is shown in Table 3. It can be seen that college students with different communication situations have extremely significant differences in each factor of the self-assessment scale. Further comparison between two groups showed that there were significant differences in each factor among college students ($p<0.01$).

Table3. Difference Analysis of Self rating Scale for Communication with Students

Factor	Normal	Less	Not good	F
Somatization	1.24	1.45	1.63	171.12
Obsession	1.61	1.62	1.86	184.65
Relationship	1.46	1.75	1.95	164.27
Depressed	1.35	1.86	1.97	168.63
Anxious	1.36	1.63	1.82	188.74
Hostile	1.28	1.56	1.62	192.65
Terror	1.22	1.50	1.75	175.29
Paranoia	1.34	1.60	1.74	172.71
Psychopathic	1.29	1.64	1.81	194.76
Other	1.34	1.67	1.91	187.86
Score	1.35	1.68	1.71	189.88

3.1.4 Analysis of Self assessment Scale of College Students

The results of variance analysis on each factor of the self-evaluation scale of college students with different academic conditions are shown in Table 4. It can be seen that their differences in each factor of the self-assessment scale are extremely significant. Further comparison between two groups found that there were significant differences in various factors among college students with different academic conditions ($p<0.05$).

Table4. Analysis of Self rating Scale for Different Academic Conditions

Factor	None	Less	More	F
Somatization	1.28	1.31	1.41	19.12
Obsession	1.63	1.70	1.91	16.43
Relationship	1.52	1.57	1.72	18.21
Depressed	1.48	1.55	1.71	21.63
Anxious	1.47	1.48	1.64	19.33
Hostile	1.39	1.37	1.56	17.68
Terror	1.26	1.34	1.48	15.34
Paranoia	1.34	1.35	1.59	19.66
Psychopathic	1.33	1.43	1.65	9.46
Other	1.35	1.46	1.56	16.84
Score	1.38	1.45	1.54	13.29

3.1.5 Analysis of the self-evaluation scale of college students with different absenteeism

The analysis of college students' self-evaluation scale with different absence rates is shown in Table 5. It can be seen that their differences in each factor of the self-assessment scale are extremely significant. It was further found that there was no significant difference between "no" and "less" students in the factors of compulsion and terror ($p=0.243$, $p=0.053$), and there were significant differences in other factors among students with different absenteeism ($p<0.05$).

Table5. Analysis of Self assessment Scale for Different Absenteeism

Factor	None	Less	More	F
Somatization	1.27	1.41	1.46	39.62
Obsession	1.64	1.73	1.98	16.46
Relationship	1.52	1.62	1.74	28.25
Depressed	1.37	1.58	1.76	22.64
Anxious	1.31	1.46	1.84	29.54
Hostile	1.28	1.43	1.62	17.43
Terror	1.33	1.45	1.58	16.87
Paranoia	1.32	1.37	1.56	18.43
Psychopathic	1.37	1.46	1.63	16.41
Other	1.38	1.43	1.59	17.83
Score	1.32	1.42	1.58	14.25

3.2 The relationship between college students' mental health and their learning and living behavior

We use python to analyze data and establish a random forest model, as shown in Table 6. It can be seen that factors such as sleep and classmate communication have different effects on each factor of the self-assessment scale. The higher the score, the greater the influence of this factor on this factor.

Table6. The relationship between college students' mental health and their learning and living behavior

Factor	Influence factor					
	Sleep	Sudden disease	Classmate communication	Truancy	Not returning at night	Academic Studies
Somatization	0.032	0.019	0.052	0.016	0.012	0.005
Obsession	0.035	0.014	0.047	0.018	0.011	0.005
Relationship	0.037	0.017	0.036	0.014	0.007	0.003
Depressed	0.033	0.015	0.059	0.017	0.008	0.007
Anxious	0.031	0.018	0.045	0.013	0.013	0.008
Hostile	0.042	0.012	0.052	0.008	0.012	0.011
Terror	0.041	0.021	0.041	0.009	0.006	0.004
Paranoia	0.029	0.026	0.038	0.011	0.005	0.006
Psychopathic	0.027	0.019	0.039	0.013	0.004	0.003
Other	0.026	0.014	0.041	0.016	0.012	0.009
Score	0.028	0.023	0.044	0.008	0.014	0.006

3.3 Discussion and analysis

In today's rapid development, social competition is becoming increasingly fierce, the pace of life is also accelerating, and college students are facing increasing pressures in learning, life, employment and emotion. The resulting mental health problems are increasingly prominent. Relevant research shows that the mental health level of college students is affected by many factors, including gender, major and place of origin [7]. At the same time, family economic conditions, family structure, parental relations, family atmosphere, childhood life experience and other family factors also have an important impact on the mental health of college students [8].

From the analysis, college students with different learning and living behaviors have different levels of mental health. Sleep, late return, communication with classmates, study, absenteeism and sudden illness affect the mental health level of college students. The worse the sleep, the more late return, the less communication with classmates, the more failed students, the more absenteeism and the more serious the disease, the lower their mental health level, the more need for college counselors, psychological centers, parents and peer groups to pay attention. At the same time, the research also shows that, for different aspects of college students' mental health, the impact of sleep, late return, communication with classmates, school work, truancy and sudden illness is different. Relatively speaking, sleep, communication with classmates and other situations have a greater impact, sudden illness has a greater impact on somatization, and academic work has a relatively small impact on mental health. Good interpersonal relationship is an important aspect of people's life. For college students, they are at an important stage of learning knowledge, understanding the society and adapting to the society. Good classmate relationship has a great impact on their learning and growth, meeting the needs of safety, communication and respect, and better achieving self-development and adapting to the society. Compared with stable factors such as gender, major and family, college students' learning and living behavior status changes in real time, which can more accurately reflect their mental health changes.

In the work, according to the actual needs, colleges and universities have generally established a number of student information platforms, such as the educational administration system, student work system and accommodation system. Relying on existing information platforms, college students' learning and life behavior information is relatively easy to observe and obtain. Through the integration and analysis of these information through large data technology, dynamic tracking and real-time monitoring and early warning of students' psychological state can be achieved, In order to find the students who encounter psychological crisis in time, colleges and universities can provide psychological assistance in a timely manner, carry out crisis intervention, and help students solve the problems they encounter. Monitoring and early warning of college students' psychological crisis using big data technology can remedy the problems of poor timeliness, low efficiency and insufficient personnel in traditional college psychological crisis monitoring and early warning, and ensure the physical and mental health of college students and the stability of the college.

4. THE CONSTRUCTION OF COLLEGE STUDENTS' PSYCHOLOGICAL CRISIS MONITORING AND EARLY WARNING SYSTEM

Take advantage of big data, give play to peer groups, counselors, psychological centers and other forces, constantly improve the traditional college students' psychological crisis monitoring and early warning methods, and further build a college students' psychological crisis monitoring and early warning system based on big data technology.

4.1 Optimize information collection indicators

Through analysis, it can be found that in college students' learning and living behaviors, sleeping, being late, communicating with classmates, studying, being absent from class, and sudden illness can help us better identify the changing state of students' mental health, and can be identified as an important indicator for psychological crisis monitoring and early warning, so as to avoid large and comprehensive indicators, and thus improve work efficiency and accurate early warning. At the same time, the existing information is also collected, such as psychological evaluation data, psychological counseling, family economic situation and other family factors.

4.2 Build psychological data platform

We rely on the existing information systems such as psychological evaluation, educational administration, student work, accommodation, and psychological counseling to build a psychological data platform, and use big data technology for integration and analysis to achieve data sharing and effective use of various information systems. We have improved and perfected the existing information system to further enhance the data. In case of absenteeism, the teacher needs to upload the absenteeism list in time, or the student can scan the code to sign in using the educational administration system to show the absenteeism in real time. Some information is collected by counselors, head teachers, psychological centers and peer groups, such as sleep, relationship with classmates, illness and other emergencies. A hierarchical and reasonable information collection and viewing system is established to make up for the shortcomings of other systems.

4.3 Establish a two-level monitoring and early warning mechanism

According to the analysis and different degrees, the indicators are quantified and divided into two levels for early warning. The system automatically alerts according to the collected information. For different levels of monitoring and early warning, establish corresponding secondary intervention mechanisms. The first level early warning is mainly intervened by the college to provide assistance and support, and the school psychological center provides professional guidance; The secondary early warning is jointly intervened by the college and the school's psychological center. The college mainly provides support and tracking management. The school's psychological center provides professional intervention. When necessary, it invites experts outside the school to conduct joint consultation and do a good job of referral.

5. CONCLUSION

To sum up, college students with different learning and living behaviors have different levels of mental health. Sleep, late return, communication with classmates, study, truancy and sudden illness reflect the mental health of college students, especially sleep, communication with peers, which has a great impact on the mental health of college students. These factors are taken as information collection indicators for data integration and analysis, The two-level monitoring and early warning can realize the dynamic tracking management and accurate monitoring and early warning of college students' mental health, timely discover and identify the potential psychological crisis of college students, improve the level of mental health education in colleges and universities, and maintain students' physical and mental health, family happiness, and campus security and stability.

REFERENCE

[1] Qian Yuting. Construction of mental health education system for students in higher vocational colleges [J]. Journal of Heilongjiang Teachers' Development College, 2022,41 (11): 74-76

[2] Shen Honghao, Dai Binrong. Family and school work together to prevent adolescent psychological crisis [J]. Journal of Yancheng Normal University (Humanities and Social Sciences Edition), 2022,42 (06): 63-69

[3] Liu Min. Research on the Psychological Intervention Mechanism of International Students in Colleges and Universities in the Post epidemic Era [J]. Research on International Student Education Management, 2022 (01): 82-91

[4] Ye Qin. Practical Exploration on the Legal Treatment of Students with Serious Mental Disorders in Colleges and Universities [J]. Scientific Consultation (Educational Research), 2022 (10): 33-35

[5] Lei Rong, Guo Ping. On the intervention strategy of college students' psychological crisis [J]. Western Academic Journal, 2022 (19): 142-145

[6] Tao Shasha. Methods and Approaches of Family School Co education in College Students' Psychological Crisis [J]. Data, 2022 (10): 177-179

[7] Huang Xiaohui. Establishing and Improving the Psychological Crisis Intervention System in Colleges and Universities [J]. Education (Higher Education Forum), 2022 (27): 60-63

[8] Ma Shilong, Guo Lanchunlei. Research on Psychological Crisis Intervention of College Students [J]. Shanxi Youth, 2022 (17): 181-183

Proc. of SPIE Vol. 12604 126040M-2

Research on an improved RGB-D camera SLAM odometer algorithm

Hongxia Cui *[a], Hu Xue [a], Hanqing Hu [a]
[a]College of Information Science and Technology, Bohai University, Jinzhou Liaoning, China
[*] Corresponding author: cuihongxia@bhu.edu.cn

ABSTRACT

With the rapid development of sensors and computer vision, simultaneous localization and mapping plays an important role in the field of intelligent robots. However, traditional odometer from ORB-SLAM2 methods based on a RGB-D camera still face the problem of low accuracy and tracking failure in complex environments with obvious irregular motion, which lead to lower stability and reliability. Aiming at solving the above problems, an odometer method is proposed by introducing three-view information to improve the localization precision of current frame. The method establishes a framework for RGB-D odometer. The co-view relationships were developed between three frames based on feature matching method. Thus, the BA optimization model is constructed based on the tracked relationship among three frames. In this way, the problems of accumulated errors can be decreased by constraints of three-view information. Experiments on the public TUM data set show that the trajectories from the proposed method is more accurate and robust than those of tradition ORB-SLAM2.

Keywords: Simultaneous localization and Mapping, RGB-D camera, Visual Odometer, Feature Extraction

1. INTRODUCTION

Many visual SLAM systems have been proposed during the recent years. Visual Odometer (VO) is the front-end of SLAM systems[1], which uses the image streams acquired by single or multiple cameras to roughly estimate camera motion trajectory. Currently, this technology is widely used in the fields of autonomous driving, mobile robotics and 3D reconstruction, etc. The vision sensors of visual odometer mainly include monocular cameras, binocular cameras and RGB-D cameras[1,2,3]. The RGB-D cameras are able to obtain depth information, which have been widely used by visual Odometer. Currently, many RGB-D cameras have been developed such as Microsoft's Kinect camera and Intel's RealSense series of cameras. The algorithms of visual odometry can be divided into two categories: the direct method based on photometric-based and feature-based method from minimize reprojection error. The direct method is of lower accuracy and robustness in complex lighting environments under assumption of constant gray scale. However, feature-based method depends on the tracked number and length of feature points, which has the advantage of insensitivity of illumination. However, the tracking accuracy and robustness of the visual odometer from feature based method would be influenced by motion blur, matching accuracy between corresponding feature points and motion changes, etc. To obtain high-precision and more robust odometer, Rual et al[2] proposed the famous ORB-SLAM2 framework to track motion trajectory in real time and simultaneously reconstruct 3D virtual surrounding environment by adopting monocular, binocular or RGB-D cameras. Although true-scale 3D information can be reconstructed by using binocular and RGB-D sensors, the ORB-SLAM2 algorithm still suffers from feature tracking failure or insufficient accuracy caused by irregular motion and feature matching errors. Liu et al[4] proposed an improved method by increasing depth-deficient feature points and inter-frame matching points in pose and position estimation process. Dong et al[5] designed an optimization algorithm that updated the depth information of local map points by increasing the number and accuracy of local map points. Liu et al[6] proposed a visual odometer for a multi-camera system that used constraints between multiple views to improve the accuracy of pose estimation and optimization. Jin et al[7] proposed a SLAM approach to improve robustness by increasing face features to improve motion constraint. The stereo camera PL-SLAM system is also used by fusing point features and line features to improve the robustness of point feature SLAM[8-11]. The algorithms of visual Odometer based on line features would increase the dimensionality of nonlinear optimization, which lead to an increase in computational time consumption and also suffer from the problems of lacking of line features and irregular motion. It is still a key problem of vision SLAM[12-21] to improve accuracy and robustness caused by the problems of motion mutation and feature tracking. In this paper, the traditional ORB-SLAM2 odometer method is improved by associating the constraint information among the three views to enhance the correlation among frames, features and map points. The rest of this paper is organized as follows: Section II presents an overview for RGBD visual odometer principle. In Section III, we propose the improved method in detail. Section IV introduce the experiments and give analysis and Section V concludes the paper.

International Conference on Computer Graphics, Artificial Intelligence, and Data Processing (ICCAID 2022)
edited by Ruishi Liang, Jing Wang, Proc. of SPIE Vol. 12604, 126041N
© 2023 SPIE · 0277-786X · doi: 10.1117/12.2674619

2. RGB-D VISUAL ODOMETER PRINCIPLE

The RGB-D image acquisition device often includes two cameras, a binocular structured-light sensor and an infrared projector projecting infrared scatter spots. Two infrared cameras acquire the infrared images in the same time. The depth map is developed by parallax map processed by detecting the infrared scatter spots in the infrared image. The RGB-D has been widely used in mobile robots and other fields. The feature-based vision odometry from RGB-D sensor mainly includes feature point extraction, feature matching and pose and position estimation, as shown in Figure 1.

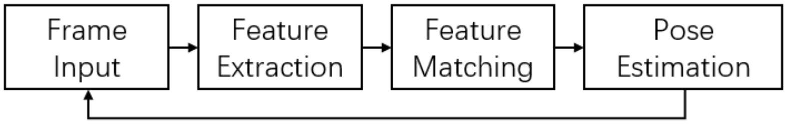


Figure 1. Basic process of visual odometer

2.1 Feature point extraction and matching methods

ORB can be considered as the combination of the FAST key point detector and BRIEF descriptor. the FAST key points can be detected in four steps. Taking the image point p as the central point, the pixel points are detected in the circle around point p. The difference in grey values between p and the pixel points in the field are computed. If there are enough pixels in the field around the candidate image point p, p would be considered as a Fast feature point which owns the enough grey difference. Then, multi-scale FAST feature points are acquired by adopting the image pyramid strategy in order to overcome the problems of lacking scale invariance for the traditional Fast feature points. Further, the gray centroid in the local region of the feature point p can computed by Equation (1). The main direction of the feature point can be derived by Equation (2).

$$u = \frac{m_{10}}{m_{00}}, v = \frac{m_{01}}{m_{00}} \tag{1}$$

$$\theta = arctan(m_{01}/m_{10}) \tag{2}$$

where, $m_{00} = \sum_{x,y} I(x,y)$, $m_{01} = \sum_{x,y} xI(x,y)$, $m_{01} = \sum_{x,y} yI(x,y)$, $m_{01} = \sum_{x,y} xyI(x,y)$; x, y are the pixel coordinates in the x- and y-axis of a pixel point, respectively . $x, y \in [-r, r]$, where r is the radius of the local area around the point p.

Further, N pairs of points around a feature point are randomly selected to form a binary feature descriptor BRIEF by the comparison results of N point pairs[22]. Feature matching is then conducted based on BRIEF descriptor. There are many famous feature matching methods include violent methods, Fast Approximate Nearest Neighbor (FLANN), Bag of Words (BOW) matching and other algorithms. Since the Steer BRIEF descriptor[16] used in this paper is a binary operator, this paper takes the Hamming distance between features (Hamming distance) as the similarity evaluation, as shown in Equation (3), which is a violent matching method.

$$D\left(F_{i,k}, F_{j,l}\right) = \sum_m b_m \oplus c_m \tag{3}$$

Where, $F_{i,k} = b_1, b_2 \ldots. b_{255}$; $F_{j,l} = c_1, c_2 \ldots. c_{255}$; m=1,2..255. $F_{i,k}$ and $F_{j,l}$ are the descriptors of the kth feature point in the ith image and the from the corresponding point feature point in the jth, respectively, $\oplus$ denotes the XOR operation. The larger $D\left(F_{i,k}, F_{j,l}\right)$ means the lower degree of similarity and the two points are considered to be a pair of corresponding points if the similarity reaches 50%.

2.2 Odometer estimation algorithm

Usually, visual odometerer adopts a frame-to-frame estimation method, which can be divided into estimation and optimization. The PnP solution from 3D-2D[14] and ICP solution from 3D-3D are often used to estimate the rotation matrix R , and translation matrix t based on geometric transformation relationship. In this paper, the RGB-D cameras are considered, which can directly obtain the image and depth information of the spatial object points. Thus, the 3D spatial coordinates (world coordinate system) of the feature points can be obtained by simple triangulation. Therefore, the spatial coordinates of feature points generated from the previous frame are used to estimate the pose of the current frame by the PnP solution method, as shown in Equation (4).

$$\begin{bmatrix} u \\ v \\ 1 \end{bmatrix} = \tau K \begin{bmatrix} R & t \\ 0_3^T & 1 \end{bmatrix} \begin{bmatrix} X_w \\ Y_w \\ Z_w \\ 1 \end{bmatrix} \tag{4}$$

Where, (u, v) is the coordinates of the feature point without distortion, K is the internal orientation element and $[X_w\ ,\ Y_w\text{,}\ \ Z_w]^T$is the vector of object point coordinates. Theoretically, the orientation parameters of the image can be determined with three image points and the corresponding object points. Indeed, four possible solutions would be calculated and another pair of points is needed for verification to determine the final positive solution.

Further, ORB-SLAM2[7] odometer uses the bundle adjustment method to optimize the camera pose and position as well as the spatial location of the feature points. To optimize the 3D coordinates of the object point observed in consecutive frames and the pose and position of the second frame, the cost function can be derived in Equation (5). By solving this cost function, the estimated parameters can be optimized aided by the g2o[18,19] model algorithm.

$$f = \frac{1}{2}\sum_{i=1}^{m}\sum_{j=1}^{n}\left\|z_i - h\left(\varepsilon_i, p_j\right)\right\|^2 \tag{5}$$

Where z_{ij} is the pixel coordinate corresponding to the observation of the feature point p_j at the image with the pose $\varepsilon_{i,}\varepsilon_i \in SE(3)$ expressed by the Lee algebraic representation ; h represents the observation equation.

3. THE IMPROVED SLAM ODOMETER FROM THREE FRAMES

3.1 Algorithmic framework and process

The traditional ORB-SLAM2 framework supports monocular, binocular and RGB-D cameras. In this paper, the co-view relationship among three views are used to improve the ORB-SLAM2 odometer algorithm, where the three views refer to continuous three frames including the first frame, the second frame and the current frame. The framework of the improved algorithm is shown in Fig. 2 and the specific steps are as follows.

Step 1: Taking the current frame as input image, feature points and descriptors are extracted for the current frame. The feature points in the current frame are selected for all map points observed in the first and the second frames by matching method based on Equation 4. Thus, the feature points in the current frame are classified as co-view map points and non-co-view map points. The co-view relationships can be developed between three frames.

Step 2: Judge the number of matches and the three-view optimized process is withdrawn if the matches are insufficient. Otherwise, the co-view map points are projected to the current frame at the pose estimated by the motion model, the feature point would be detected in the neighbour area, which is noted as the matching point.

Step 3: The tracked relationship is constructed by the above process, and the BA optimization model is constructed based on this relationship among image points and object points. A go2 graph is derived by taking the poses and positions of the three frames and map points as optimization vertices, the observation relationship between map points and poses as observation edges. The optimal solutions of vertices would be solved under the three-view constraint. Besides, the tracked relationship is subsequently updated according to the optimization results in order to optimize the robustness and accuracy of the whole visual odometer.

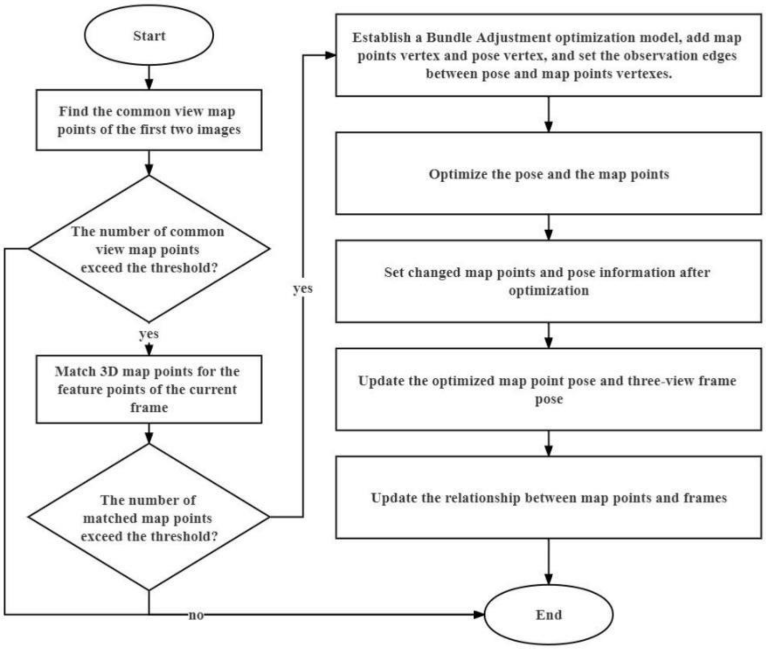

Start
Find the common view map points of the first two images
The number of common view map points exceed the threshold?
yes
Match 3D map points for the feature points of the current frame
The number of matched map points exceed the threshold?
yes
no
Establish a Bundle Adjustment optimization model, add map points vertex and pose vertex, and set the observation edges between pose and map points vertexes.
Optimize the pose and the map points
Set changed map points and pose information after optimization
Update the optimized map point pose and three-view frame pose
Update the relationship between map points and frames
End

Figure 2. Algorithmic framework in this paper

3.2 Pose Estimation and Optimization

Assume that the consecutive three frames are F_{k-2}, F_{k-1}, F_k and the current frame is F_k. The set of map points that can be observed in each frame are M_{k-2}, M_{k-1}, M_k. Let the set of co-visual map points be $M_{common} = M_{k-2} \cap M_{k-1}$,the map point P in M_{common} be $= \left[X_{m,w}, Y_{m,w}, Z_{m,w}\right]^T$, the rough rotation matrix of the current frame estimated from the constant velocity motion model be R_{cw} , and the camera position vector be t_{cw}. This co-viewing map point is projected to the initial image point $m_c(u_c, v_c)$ of the current frame from Formula (6).

$$\begin{bmatrix} u_c \\ u_c \\ 1 \end{bmatrix} = \tau' K \begin{bmatrix} R_{cw} & t_{cw} \\ 0_3^T & 1 \end{bmatrix} \begin{bmatrix} X_{m,w} \\ Y_{m,w} \\ Z_{m,w} \\ 1 \end{bmatrix} \tag{6}$$

where, τ' is the scale factor.

Then, the size of the search feature range is determined with the feature point as the candidate point. Further, the feature points in the neighbour area are selected and Hamming distances are computed between each point and the homonymous image points from the same map point on the first and second frames, respectively. The feature point with the smallest Hamming distance is determined as the corresponding point of the map point P. Besides, the rotational consistency is also used to eliminate the mis-match. Therefore, the tracking accuracy of points of three consecutive frames is improved. Further, optimization of pose of the current frame is also conduced based on bundle adjustment and the cost function is constructed by minimize the projection error of the corresponding image points of the co-view map points on three consecutive frames, as shown in Equation 7 derived from Equation 5.

$$f' = arg \min_{T'_w} \frac{1}{2} \sum_{i=1..n} \left\| \frac{u_{i,} - K T'_w m_{iw}}{\sigma_{p_i}^2} \right\|^2 \tag{7}$$

Where K is the internal camera matrix. T'_w is the pose matrix with $T'_w = \begin{bmatrix} R'_w & t'_w \\ 0_3^T & 1 \end{bmatrix}$,the original estimation of T'_w is $T_w = \begin{bmatrix} R_w & t_w \\ 0_3^T & 1 \end{bmatrix}$. u_i denotes the coordinates of the feature point in the pixel coordinate system. m_{iw} indicates the 3D coordinates of the ith map point in the world coordinate system. $\sigma_{p_i}^2$ denotes the variance of the feature points. Thus, the visual odometerer algorithm of the traditional ORB-SLAM2 from two-view tracking is improved by the optimization of pose of the current frame from three-view tracking information.

4. EXPERIMENTAL ANALYSIS

In this paper, the absolute trajectory error ATE (as shown in equation (6))[21] is used to evaluate the accuracy of the odometer.

$$ATE \quad = \sqrt{\frac{1}{N}\sum_{i=1}^{n}\left\| log(\, T_{gt,i}^{-1} T_{esti,i})^{\vee}\right\|_2^2} \tag{8}$$

Where, $T_{esti,i}$ represents the pose at the ith time epoch estimated by the odometer, and $T_{\mathrm{gt},i}$ represents the pose of the ground truth on the same epoch, i=1.. N.

The experiment was conducted on a laptop PC which was installed the Ubuntu 18.04.6 operating system. In terms of hardware, the CPU is Intel i7 6700 and the memory capacity was 8G. Two visual odometers were realized based on ORB-SLAM2 and our propose algorithm on the public dataset TUM public RGBD of freiburg1_room and freiburg1_desk datasets [20], respectively. Ten repetitive tracking experiments were conducted for each dataset. The results of ATE of the two algorithms can be acquired based on Equation (9). It can be seen in Fig. 3 and Fig. 4 that the two algorithms differ significantly, the visual odometer of our proposed method owns smaller absolute trajectory error of tracking trajectories than those of traditional ORB-SLAM2 algorithm. The trajectories of 10 odometers are visualized in Figure 5. As shown in Fig. 5(a), the algorithm in this paper is very stable, and the trajectories of 10 times almost coincide with the real trajectories. As shown in Fig. 5(b), there is a large differences between the trajectories from the ORB-SLAM2 and those of the ground truth. Especially, there are tracking loss in the trajectories from ORB-SLAM2 odometers.

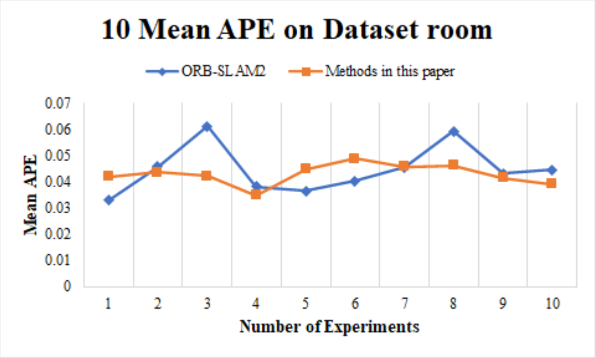


Figure 3. freiburg1_Odometer estimation accuracy of room data set

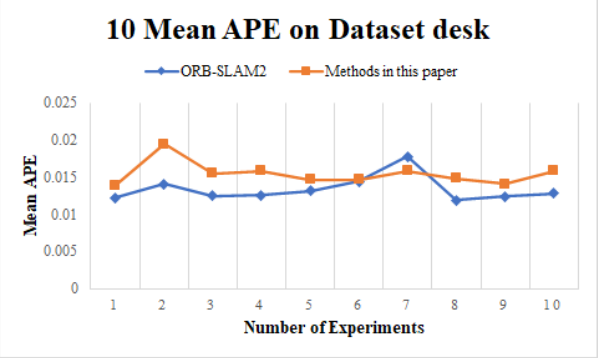
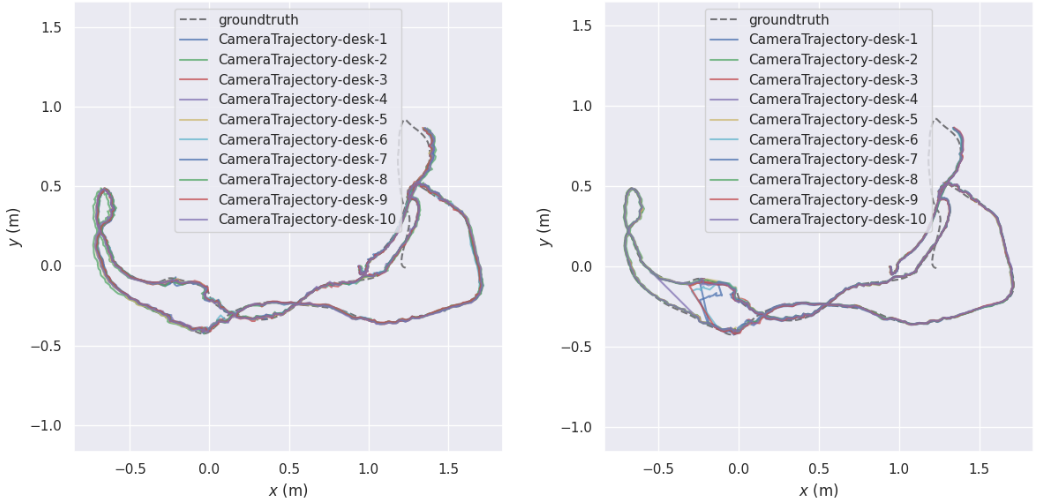


Figure 4. freiburg1_Odometer estimation accuracy of desk data set

Proc. of SPIE Vol. 12604 1260401-7



(a) (b)

Figure 5. freiburg1 desk tracking track, (a) the algorithm in this paper, (b) ORB-SLAM2

5. CONCLUSION

This paper presents an improved odometerer method for traditional RGBD ORB-SLAM2 odometerer based on three frames of information. Experiments on the TUM public data sets show that the absolute errors between the proposed algorithm and the real trajectory are smaller and the tracking trajectories are smoother compared to those of ORB-SLAM2. Especially, the experiments demonstrate that the odometerer accuracy can be improved and the tracking failure problem can be solved by using the three-view information. The next step will be considered to further optimize the constraints between feature information of multi-view geometry to improve the accuracy and robustness of odometerer estimation.

ACKNOWLEDGMENTS

This paper is supported by Soft Science Project of Liaoning Provincial Department of Science and Technology, China (No.2022JH4/10100052)

REFERENCES

[1] Mur-Artal, R., Montiel, J. M. M. and Tardos, J. D., "ORB-SLAM: a versatile and accurate monocular SLAM system," IEEE Transactions on Robotics. Papers 31(5), 1147-1163 (2015).

[2] Mur-Artal, R. and Tardós, J., "ORB-SLAM2: An Open-Source SLAM System for Monocular, Stereo, and RGB-D Cameras," IEEE Transactions on Robotics. Papers 33(5), 1255-1262 (2017).

[3] Rublee, E., Rabaud, V., Konolige, K. and Bradski, G., "ORB: An efficient alternative to SIFT or SURF Monocular, Stereo, and RGB-D Cameras," 2011 International conference on computer vision, 2564-2571 (2011).

[4] Liu, H. and Cheng, Q., “Improved RGB-D visual odometer based on ORB feature,” Manufacturing Automation. Papers 44(07), 56-59+106 (2022).

[5] Dong, J., Jiang, Y. and Han, Z., “RGB-DSLAM Visual Odometry Optimization Algorithm,” Journal of Harbin University of Science and Technology. Papers 25(06), 157-164 (2020).

[6] Liu, P., Geppert, M., Heng, L., Sattler, T. and Geiger, A., "Towards robust visual odometry with a multi-camera system," proc. of the 2018 IEEE/RSJ Int. Conf. on Intelligent Robots and Systems, 1154-1161 (2018).

[7] Jin, B., Bei, S., Yin, M. and Li, L., "A low drift visual odometry based on point and plane," J. Chongqing Inst. Technol. Papers 36(06), 88-95 (2022).

[8] Gomez-Ojeda, R., Moreno, F. A., Zunige-Noel, D., Scaramuzza, D. and Gonzalez-Jimenez, J., "PL-SLAM: A stereo SLAM system through the combination of points and line segments," IEEE Transactions on Robotics. Papers 35(3), 734-746 (2019).

[9] Gao, C., Huang, Y., Zhao, B. and Hu, X., "Stereo visual odometer using point and line features," Optical Instruments. Papers 43(04), 19-27 (2021).

[10] Huang, P., Cao, Z. and Huang, J., "A RGB-D visual odometry method based on line features," Journal of Chinese Inertial Technology. Papers 29(03), 340-349 (2021).

[11] Cheng, M., Ding, L. and Zhang, Y., "Fast PL-SLAM Algorithm Based on Improved Keyframe Extraction Strategy," Acta Electron. Sin. Papers 50(03), 608-618 (2022).

[12] Peng, H., Dong, X., Li, T. and Fan, Y., "Semi-direct SLAM algorithm based on vision sensor," Transducer and Microsystem Technologies. Papers 41(06), 114-114+21 (2022).

[13] Meng, X., Gao, W. and Hu, Z., "Dense RGB-D SLAM with multiple cameras," Sensors. Papers 18(7), (2018).

[14] Zhu, J. and Cheng, Q., "Improvement of kinect performance in RGB-D visual odometer," CAAI Transactions on Intelligent Systems. Papers 15(05), 943-948 (2020).

[15] Zhang, F., Li, Q., Wang, T. and Ma, T., "A robust visual odometry based on RGB-D camera in dynamic indoor environments," Measurement Science and Technology. Papers 32(4), (2021).

[16] Kim, C., Kim, P., Lee, S. and Kim, H. J., "Edge-based robust rgb-d visual odometry using 2-d edge divergence minimization," IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS). 1-9 (2018).

[17] Li, Y., Brasch, N., Wang, Y., Navab, N. and Tombari, F., "Structure-slam: Low-drift monocular slam in indoor environments," IEEE Robotics and Automation Letters. Papers 5(4), 6583-6590 (2020).

[18] Kummerle, R., Grisetti, G., Strasdat, H., Konolige, K. and Burgard, W., "G2O: A general framework for graph optimization," proc. of the 2011 IEEE Int. Conf. on Robotics and Automation. Papers 2011, 3607-3613 (2011).

[19] Fang, B. and Zhan, Z., "A visual SLAM method based on point-line fusion in weak-matching scene ," International Journal of Advanced Robotic Systems. Papers 17(2), (2020).

[20] Sturm, J., Engelhard, N., Endres, F., Burgard, W. and Cremers, D., "A benchmark for the evaluation of RGB-D SLAM systems," proc. of the 2012 IEEE/RSJ Int. Conf. on intelligent robots and systems. Papers 2012, 573-580 (2012).

[21] Li, D., Liu, S., Xiang, W., Tan, Q. and Hu, Y., "A SLAM System Based on RGBD Image and Point-Line Feature," IEEE Access. Papers 9, 9012-9025 (2021).

[22] Calonder, M., Lepetit, V., Strecha, C. and Fua, P., "Brief: Binary robust independent elementary features," European conference on computer vision, Berlin, Heidelberg. Papers 2010, 778-792 (2010).

Market Power Monitoring and Mitigation Mechanism of Spot Market under New Power System

Yifeng Liu[1], Yang Tang[1], Meiting Liu[1], Fangmei Bie[2], Yuliang He[3*], Yuxin Zhang[3]
[1]State Grid Hubei Electric Power CO., LTD, Wuhan, Hubei, 430077, China
[2]State Grid Hubei Electric Power Company Limited Economic Research Institute, Wuhan, Hubei, 430077, China
[3]School of Electrical and Electronic Engineering, Huazhong University of Science and Technology, Wuhan, Hubei, 430074, China
[*]Corresponding author's e-mail: venn.ellen@foxmail.com

ABSTRACT

In recent years, with the continuous development of China's electric power spot market construction, the market power problem brought by the structure of China's power generation-side has gradually become severe, and has evolved as the focus issue in the process of power market reform. The injection of renewable sources and the participation of distributed generation resources have brought profound influence on the power system and the power market, and it is urgent to establish and complete the market power monitoring and mitigation mechanism of China's electricity spot market. This paper primarily discusses the market power monitoring methods and mitigation mechanisms, then introduces the mature experience of typical international power markets in market power monitoring and mitigation mechanisms. Secondly, this paper summarizes the existing power monitoring and mitigation mechanisms in China, and analyzes the development and challenges of market power monitoring and mitigation mechanism construction under the new power system. We aim to provide some reference for the evaluation, monitoring and mitigation of market power in a series of practical aspects.

Keywords: electricity market, new power system; market power; market surveillance; market power monitoring and mitigation mechanisms.

1 INTRODUCTION

Since the "Certain Opinions of the CPC Central Committee and The State Council on Further Deepening the Reform of the Electric Power System" and its supporting documents were issued, a new round of further reform of the power system has been comprehensively launched [1]. The reform emphasized the decisive role of the market in the allocation of resources, and has achieved remarkable results in promoting the construction of the power market in recent years. In August 2017, General Office of National Development & Reform Commission and General Department of National Energy Administration jointly issued the "Notice on Carrying out the Pilot Work of Power Spot Market Construction", deploying the construction of eight pilot electricity spot markets [2]. Currently, Southern region (starting from Guangdong), Mengxi, Zhejiang, Shanxi, Shandong, Fujian, Sichuan and Gansu have all started the spot market settlement trial operation. In 2022, National Development & Reform Commission and National Energy Administration jointly issued the "Guiding Opinions on Accelerating the Construction of a National Unified Electricity Market System", clearly pointing out the necessity to speed up the formation of a power market system with features of "unified and open, competitive and orderly, secure and efficient, and improved governance" [3]. The file also put forward two stage construction goals, which are "preliminary buildup a national unified power market system by 2025", and "basically buildup a national unified power market system to meet the requirements of the new power system by 2030". The introduction of a series of reform measures will continue to consolidate the critical position of the market in China's power resource allocation.

The operational efficiency of the power market will determine the efficiency of power industry to a great extent, thus having a great impact on the sustainable development of the national economy. However, due to multiple factors such as the high concentration of power generation enterprises and the lack of flexible resources, there are still certain defects in China's power market, among which is the abuse of market power. The concept of market power, which originated in economics, originally refers to the ability of enterprises to manipulate market prices away from the fully competitive level and obtain excess income from their behaviors [4]. At the present stage, since the five major power generation groups of China have occupied most of the market share, leading to the excessive concentration of the generation side, the abuse of market power emerges broadly in the spot market. Abuse of market power from generators can bring about serious

International Conference on Computer Graphics, Artificial Intelligence, and Data Processing (ICCAID 2022)
edited by Ruishi Liang, Jing Wang, Proc. of SPIE Vol. 12604, 126041O
© 2023 SPIE · 0277-786X · doi: 10.1117/12.2674578

consequences, not only worsening the market competitiveness, distorting the price signal, thus causing inappropriate long-term investments, but also leading to the transfer of social welfare from consumers to producers, remarkably lowering the market operational efficiency, ultimately affecting the overall allocation efficiency of social resources [5]. Common methods to exercise market power used by generators include:

① Physical withholding: generators intentionally reduce the quoting capacity, or the "nominal available capacity" by modifying certain technical parameters of units or falsely reporting faults and maintenance, so as to cut down the aggregate supply in power market, and obtain excess income by raising the market clearing price. Power producers often confuse physical withholding with authentic unit failures to evade regulation [6].

② Economic withholding: generators deliberately quote part of its capacity at a price significantly higher than its marginal generation cost, so that this part of capacity will miss the clearing margin. Therefore, it will reduce the aggregate supply and raise the market clearing price, so that the clearing capacity of the generator can gain a higher income.

③ Collusion: improper collusion between groups or within the group of generators, to modify its original quotation, in order to maximize the profits of the collusion alliance[5].

Currently, there is an international consensus on the importance of introducing market power monitoring and mitigation mechanism to the power spot market. Lakić et al. presented detailed introduction of the concept, the exercising approaches and the detection methods of market power [4]; GRAF et al. elaborated the market regulation structure and market power monitoring and mitigation system of typical American electricity market [7]; Guo et al. proposed a mitigation method based on bidding capacity control for the quoting behavior of generators [8]; Salarkheili and Foroud proposed a market power evaluation framework of power generators based on the supply equilibrium function model [9]; Moiseeva, Hesamzadeh and Biggar analyzed the market power exercising of flexible sources with a high ramp rate [10]; Shafie-khah, Moghaddam and Sheikh-El-Eslami put forward a new type of market power control model and research framework based on the profit analysis of implicit and explicit collusion [11]; Bigerna et al. analyzed the correlation between the intensity of the social control and the strength of the market power and the change of the electricity price, on the basis of the variation of the supply & demand situation of the Italian power market during the Covid-19 pandemic [12].

Among Chinese scholars, multiple literature summarize the organization structure and market power monitoring and mitigation methods in typical power markets [5,13,14]; Xue et al. proposed the concept of generalized congestion and generalized market power, and emphasized the great importance of introducing regulation institution to guarantee the competitiveness and operational efficiency of the electricity market [15]. Jiang et al. summarized the structural index test system, and comprehensively investigated the principle of setting mitigation reference price in American electricity markets [16]; Bao et al. investigated and concluded the supply & demand situation, market transaction mechanism and market power handling approaches of the Nordic power market [17], and summarized the enlightenment and reference to the construction of Chinese power market; Xie et al. proposed the risk prevention methods, and constructed the general framework of expert system based on logical deduction [18]; Dong et al. applied LASSO regression algorithm to screen out the indicators based on the structural index test system, and adopted a specific algorithm to identify the market power exercising behavior of generators [19].

Most of the current studies concentrated on the market power of fossil-fired units. However, in recent years, as wind power, photovoltaic and other renewable sources continue to penetrate, the structure of generation side confronts a significant transformation, and scholars have begun to pay attention to the market power issues under high penetration of renewables, carrying out certain empirical researches. Bahn, Samano, and Sarkis focused on the transfer and expansion of renewables capacity, conducting a comprehensive analysis of the "merit-order effect"(MoE) and the market power exercising of oligopoly generators [20]. Results showed that the structure of renewables owners will significantly affect the changing direction of market clearing price; Li, Tan, and Zhang analyzed the exercising of market power during the implementation of renewable portfolio standard(RPS) [21]; Zhu et al. established the improved Lerner Index which could reflect the overall market power of thermal units in the power market and reverse market, and applied the combined optimization model to simulate the market clearing, measuring the overall market power of thermal units [22].

This paper comprehensively discusses the market power monitoring methods and the mitigation mechanism, then summarizes the mature experience in typical power markets on the basis of in-depth investigation. Secondly, we conclude the existing market power monitoring and mitigation mechanism in China, and ultimately analyze the features of market power under the new power system, providing certain reference for market power evaluation, monitoring and mitigation in China's power market.

2 MARKET POWER MONITORING AND MITIGATION MECHANISM

Evaluation and monitoring the market power is the direct approach to screen out the power generators which have the potential to exercise market power or have exercised market power in the process of quotation. Market power monitoring methods are divided into two major systems. One is the structural index test system. By constructing a series of indicators, the system can quantitatively analyze the overall structure of the market, the market power potential of each power generator, and the impact of quoting strategy on the income of generators from three dimensions of "ex-ante", "in-process" and "ex-post". The other is the conduct and impact test system. By observing the specific quoting behavior of producers and analyzing their influence on the market clearing results, the system can judge whether a power producer has exercised its market power in the process of quotation.

2.1 Structural index test system

Structural testing aims to assess the supply & demand situation in each region and the market power potential of each generator, thus providing a factual basis for market power mitigation. In practice, regulators tend to use the ex-ante indicators for structural testing solely.

2.1.1 Ex-ante indicators

Before the market quotation begins, regulators apply a variety of indicators to assess the overall structure of the market and the resource endowment of power generators, then analyzing the market power potential for each generator. Commonly used ex-ante indicators are Herfindahl-Hirschman Index, Head Market Share, Residual Supply Index and Three Pivotal Suppliers Index.

① Herfindahl-Hirschman Index

Herfindahl-Hirschman index (HHI) is often used to measure the market concentration of a specific industry. It refers to the sum of the squares of market share held by each competitor in this industry, and is often amplified by 10,000 times in practical application:

$$HHI = \sum_{j=1}^{N} \left(100 * S_j\right)^2 \tag{1}$$

In the power industry, N is for the total number of power generators; S_j is for the market share of the power generator j, which can also be replaced by the installed capacity or the market clearing quantity of the power generator. The higher the HHI value, the higher the market concentration and the greater the monopoly degree of this industry. As a static indicator to measure the overall market structure, HHI will not be affected by the distribution of enterprises' number and size, thus having decent sensitivity characteristics.

② Head Market Share

Head Market Share refers to the sum of the market share of several largest companies in a certain industry, which can also be used to calculate the installed capacity or the market clearing quantity of power generators in the power industry. The number of the companies used to calculate Head Market Share is usually taken as 4 or 8. It is generally believed that when the TOP-4 value is greater than 65%, the industry has an oligopoly structure.

③ Residual Supply Index

Residual Supply Index (RSI) measures how well the combined installed capacity of a generator's competitors can meet the power demand in a certain region, that is, the abundance of the remaining installed capacity after subtracting the tested generator relative to the predicted load, thus indirectly measuring the "critical degree" of the tested generator. The RSI value is calculated as follows:

$$I_{RSIi} = \frac{\sum_{j=1}^{N} Q_j - Q_i}{D} \tag{2}$$

In the above formula, Q_j is the total installed capacity of the generator j, and D is the forecast value of the power load in the region. If a generator's RSI value is greater than 1 (usually 1.05 in practice), the power supplier's critical degree is relatively weak and has less impact on the market clearing result; If the RSI is less than 1, the supplier is regarded as an indispensable supplier in this region.

④ Three Pivotal Suppliers Index

Three Pivotal Suppliers Index can evaluate the indispensability of a generator and judge whether it is dominant in a more accurate way. The test is conducted for all units in a transmission area except the top two generators with the largest installed capacity. By calculating the Three Pivotal Suppliers Index(TPSI), regulators can analyze whether a unit can significantly alleviate a transmission congestion, thus judging whether it has the potential to exercise market power. The application of TPSI in typical electricity markets will be detailed in Chapter 3.

2.1.2 In-process indicators

During the market transaction, the regulators monitor and evaluate the quoting behavior of the power generators through a series of indicators, and then analyze whether the power generators have exercised the market power. Commonly used in-process indicators include: high-bid capacity ratio, bid-cost markup, capacity withholding ratio and so on.

2.1.3 Ex-post indicators

At the end of the market transaction, the regulator analyzes the influence of the quoting behavior of the power producers on the market clearing results by using a series of ex-post indicators, and measures the income of the power generators. Commonly used ex-post indicators are: Lerner Index, the clearing rate of high-bid capacity, the success degree of bid strategy and so on.

It is worth noting that during the quoting process and the ex-post stage, regulators prefer to use the conduct and impact testing system to evaluate the behavior of power generators, which will be detailed in the next session.

2.2 Conduct and Impact test system

Compared with the structural index test system, Conduct and Impact Test method is relatively more intuitive. It consists of two parts: the Conduct Test and the Impact Test. Its core is to set a reasonable “Conduct Test threshold” and “Impact Test threshold”. The purpose of the Conduct Test is to judge whether the quotation of a generator unit constitutes the exercise of market power. The testing method is to verify whether the quotation of the generator unit exceeds the preset “Conduct Test threshold” in the power market. If so, the unit fails the conduct test, and a further impact test shall be taken to determine whether the unit has had a significant impact on the market clearing results by verifying whether the quoting behavior puts the market clearing price beyond the preset “Impact Test threshold”. If so, the quotation of the tested unit fails both the Conduct Test and the Impact Test system, and must be mitigated in the subsequent trading process.

Since the threshold value of the conduct test and the impact test is artificially stipulated, Isolated System Operators(ISO) can formulate Conduct and Impact Test methods with local characteristics according to a series of information such as the market structure, supply-demand ratio, transmission congestion situations and so on.

2.3 Market power mitigation mechanism

After utilizing the two testing systems above to evaluate and monitor the potential market power of the generators and the behavior of the market power exercising during the quoting process, the operator can screen out a number of generators which need to accept the mitigation. The mainstream method of market power mitigation is the “reference price correction method”: the operator replaces the original price that needs to accept the mitigation with the preset reference price to ensure the fairness and competitiveness of the market clearing results.

The level of the reference price directly demonstrates the accuracy and rationality of the market power mitigation mechanism. An appropriate reference price can not only improve the market competitiveness, but also avoid excessive mitigating, so that the dynamic distribution of social welfare can maintain a virtuous cycle. At present, the major setting principles of reference price are as follows [7]:

① Reference price based on unit cost: *Reference price = (marginal cost + variable O&M cost) * 110%*, where the marginal cost includes marginal fuel costs, additional costs about grid management and greenhouse gas, and opportunity costs for certain resources. The 10% premium is intended to cover the long-term investment cost of the unit.

② Reference price based on historical unit quotations: Generally, the weighted average bid price of the unit in the past 90 days (without mitigation) is taken, and adjusted according to the fuel price fluctuations during that period.

③ Reference price based on historical market clearing prices: Select the market clearing prices at the lowest 25% (or 50%) in the past 90 days, take the average value as the reference price, and make an adjustment according to the fuel price fluctuations during that period.

④ Reference price based on historical marginal nodal prices: Select the marginal nodal prices at the lowest 25% (or 50%) in the past 90 days, take the average value as the reference price, and make an adjustment according to the fuel price fluctuations during that period.

3 MARKET POWER MONITORING AND MITIGATION MECHANISM OF TYPICAL ELECTRICITY MARKETS

Currently, with the mature regulatory organization structure, the American regional power markets possess relatively well-developed system and abundant practical experience in market power handling. Besides, there are unique measures of control in the UK and Nordic electricity markets. This chapter introduces the application of different market power monitoring and mitigation methods in typical power spot markets, in order to provide some reference for the market power monitoring and mitigation mechanism of China's regional power market.

3.1 Structural index test system

The structural index test system focuses on the assessment of the transmission congestion status and the generators' endowment in advance and provides the evidence for the potential mitigation. The typical power markets currently using the structural index test system are CAISO and PJM in the United States.

3.1.1 CAISO

CAISO has formed a market power mitigation mechanism since 2008, and an internal market regulatory department is responsible for the market supervision. CAISO uses a Three-Pivotal Supplier Test to assess whether a specific transmission constraint is structurally competitive or noncompetitive [23]. Firstly, operators will carry out an "all-constraints run" in the day-ahead and real-time markets to identify transmission congestion during scheduling. Secondly, the operators screen out the non-competitive transmission constraints through the following methods:

Assuming a constrained path j, the unit in set K can provide counter-flow to alleviate that constraint. Further assuming that the subset P of K contains all the units belonging to the three critical generators, and the set $F = K \backslash P$ refers to the set of generators involved in K but not in P, i.e., the "non-critical" counter-flow provider.

The Residual Supply Index of the transmission line is:

$$RSI_{i,j} = \frac{\overline{CF_j}(F)}{CF_j(K)} \tag{3}$$

Among them, $CF_j(K)$ is the sum of the supply available of counter-flow for all generating sets, and $\overline{CF_j}(F)$ is the sum of that of the "non-critical" generating sets. If the RSI is less than 1, the transmission congestion fails the critical supplier test. The combination of the tested pivotal suppliers varies by different markets: In the day-ahead market, the pivotal supplier is a combination of three generators that controls the top three amount of counter-flow available for certain transmission congestion, while in the real-time market, the combination refers to the generators controlling the top three amount of counter-flow that can be held for certain transmission congestion.

Finally, regulators are going to mitigate the bids of the pivotal power generators who have fail the Three-Pivotal Supplier Test. The reference price of mitigation is the higher value of the Default Energy Bid (DEB) and the competitive Locational Marginal Price (LMP). The latter one refers to the LMP minus its non-competitive congestion component, which is the summation of the products of the shadow price of all non-competitive transmission congestion in the region and its power distribution transfer factor. If the market electricity price increases due to insufficient power supply, the competitive LMP of a competitive region will be higher than the DEB, thus ensuring a normal price signal of the congestion. For the reference price of the unit, the regulator may apply a cost-based method, a market price-based method, or a negotiatory method [7]. In particular, a premium compensation is given for generating units which have been in mitigation of more than 80% over the past 12 months.

3.1.2 PJM

The monitoring and mitigation of market power in PJM is solely handled by external market regulators, while the independent system operator (ISO) does not take the initiative in this regard [24]. Regulators first use the residual supply index, price-markup index, net income index, market share index, etc., as structural indicators to analyze the overall competitiveness of the market and to evaluate the rationality of the market structure. In addition, similar to CAISO, the Three-Pivotal Supplier Test was applied to the day-ahead market and real-time markets in PJM to assess whether certain transmission congestion requires joint alleviation by several critical suppliers. Before conducting a Three-Pivotal Supplier Test, the regulator sorts each supplier by counter flow supply from large to small. In the first iteration of the test, combine the third largest supplier with the two largest ones, then subtract the summed supply of above three suppliers from the total counter-flow supply and divide it by the total amount of counter flow *D* required for the constraint. The calculation formula is as follows:

$$RSI_3 = \frac{\sum_{i=1}^{N} S_i - \sum_{i=1}^{2} S_i - S_j}{D} \tag{4}$$

Let *j* = 3 in the first iteration, indicating that the third largest supplier is tested together with the two largest ones. If the index is greater than 1, the three largest suppliers pass the test, and all the remaining lower-ranked suppliers (*j* = 4,5,6,…, N) also pass it. Otherwise, the three suppliers tested are all critical suppliers that are able to alleviate the transmission constraints. If so, undergo the next iteration. Each new iteration tests the combination of the next supplier (*j* = 4,5,6,..., N) and the two largest suppliers. If the index is less than 1, it means that the combination of the tested supplier and the two largest suppliers fail the test. The iteration continues until the combination of the two largest suppliers and a specific supplier *j* passes the test. The remaining lower-ranked suppliers are also identified as non-critical suppliers.

For all critical power generators, the bid of their units offering counter-flow mitigation will be mitigated by a "cost-based" reference price, namely 110% of the marginal production costs, including fuel costs, operations costs, carbon emissions costs, etc. [7].

3.2 Conduct and Impact test system

The conduct and impact test system focuses on the evaluation of the power generators' quotation behaviors and their impact on the market clearing results in the process of market transactions, providing a factual basis for the potential mitigation. The typical electricity market currently using the conduct and impact test system is NYISO and MISO in the United States.

3.2.1 NYISO

NYISO has set up a dual regulatory mechanism with an internal regulator and an external one. The latter refers to Potomac Economics, a market regulatory services provider for independent operators of several regional electricity markets in the United States. NYISO applies a conduct and impact test system [25]. The conduct and impact test thresholds fluctuate depending on the transmission congestion in the area of the tested unit. If the transmission shadow price in an area is greater than $0.04 / MWh, the area is defined as a "constraint area". Generators located in the "constrained area" will accept stricter test thresholds than those located in the "unconstrained area".

NYISO sets a corresponding test threshold for the economic withholding behaviors of power generators. In areas with no transmission constraints, the test threshold is relatively loose: if the unit minimum energy bid exceeds $100 / MWh or three times the reference price, the quotation cannot pass the conduct test. The impact test verifies whether the quotation makes the market clearing price increase by USD $100 / MWh or more than 200%.

In a constrained region, the following test conditions should be met in addition to the above test thresholds. In the conduct test, if:

$$P_i > Ref\ Price_i * \left(1 + \frac{2\% * Average\ Price\ 8760}{Constrained\ Hours}\right) \tag{5}$$

Then the quotation P_i cannot pass the conduct test. "$Ref\ Price_i$ "is the reference price of unit i, and "$Average\ Price$ 8760" refers to the average clearing price of the regional day-ahead market (or real-time market) over the past 12 months, and "$Constrained\ Hours$" is the total number of hours when the region is in the transmission congestion status over the past 12 months.

If a quotation fails the conduct test, a further impact test is carried out. If the market clearing price:

$$P_{ei} > P_e * \left(1 + \frac{2\% * Average\ Price\ 8760}{Constrained\ Hours}\right) \tag{6}$$

Then the quotation P_i cannot pass the impact test.

For the physical withholding behaviors of the generators, if the capacity withheld by a specific unit exceeds: ① 10% of the capability of the generator; ② 100MW of the capability of the generator; ③ 5% of the total output capacity of a market party and its affiliates; ④ 200MW of the total output capacity of a market party and its affiliates, then the behavior of the unit is determined as physical withholding. For units located in the constrained region, the test threshold is more stringent. Capacity withheld exceeds: ① 10% of the capability of the generator; ② 50MW of the capability of the generator; ③ 5% of the total output capacity of a market party and its affiliates; ④ 100MW of the total output capacity of a market party and its affiliates, will be judged as physical withholding.

Finally, the regulator will mitigate the market power of the chosen generators by replacing the original bid with reference prices, including cost-based, market price-based, and other principles.

3.2.2 MISO

Like NYISO, Potomac Economics assumes regulatory responsibilities for the MISO, which also applies conduct and impact testing methods to execute supervision [26]. MISO first divides its whole jurisdiction into narrow-constrained area (or dynamic narrow-constrained area) and broad-constrained area based on transmission constraints' status. Similarly, for different regions, regulators set different thresholds for the physical withholding and economic withholding behavior of power generators, and the test thresholds in the narrow-constrained area will be much more stringent [7]. Finally, regulator replaces the original bid which needs mitigation with a reference price based on the historical quotation to participate in market clearing, to enhance the market competitiveness.

Overall, MISO and NYISO share a comparative similar market power monitoring and mitigation mechanism. Hence, this paper will not elaborate excessively.

3.3 Other market power monitoring and mitigation mechanisms

There are also certain electricity markets whose market power monitoring and mitigation mechanism cannot be simply classified as the above two systems, but take into account both, or focus on the ex-post measures. The following is a brief introduction.

3.3.1 ERCOT

Potomac Economics, the external regulator of ERCOT, uses a unique two-step market mitigation mechanism that can be considered as a combination of the structural index testing system and the conduct and impact testing system [27]. The two-step method first evaluates the transmission congestion status within the region, classifies all the constraints as "competitive" ones and "noncompetitive" ones, and analyzes the static market structure using structural index such as HHI. At the in-process stage, ERCOT also adopts the conduct and impact testing method. First, the reference marginal price of the region (i.e., the test threshold) is calculated through the clearing procedure, and then all the quotations exceeding this threshold will be mitigated, while the price signal of the congestion caused by the shortage of power supply is still guaranteed.

3.3.2 The UK Electricity Market

Unlike the regional electricity market in US, the UK electricity market tends to support free market competition rather than intervene the trend of prices, and generators' market power is restrained by the standard framework of UK antitrust law. The only regulator in the UK electricity industry is the Office of Gas and Electricity Markets (Ofgem), which regulates the gas and electricity industry uniformly [6]. Ofgem chooses not to interfere overly before or during the market transactions. Instead, it analyzes the clearing price afterward, traces back to the abnormal price fluctuations, and then takes compulsory measures such as investigations and fines on the generator sets which seek improper sources of income. In serious cases, the license of the generator will be revoked.

3.3.3 The Nordic Electricity Market

The Nordic power market is currently operated by the Nordic Power Exchange and conducts internal market supervision. It also publishes quarterly market supervision reports to disclose the overall situation of the market along with the quotation and manipulation of market members during the period. The investigation results and disposal opinions of violations will also be released. Due to the large number of entities in the Nordic power market and the high dispersion of the power generation side, the competitiveness between different power generators remains a high degree. Meanwhile, the operator has sufficient experience in calculating the marginal cost of the generator sets, so as to have a clear judgment of the unit quotation behavior. Moreover, the financial transactions mechanism in the Nordic electricity market is considerably perfect, and most of the transactions can be predetermined through financial contracts in advance, which greatly suppresses the impact of price fluctuations in the spot market. Therefore, the Nordic Power Exchange tends to undertake control measures in the post-event stage, to implement extremely severe punishment for the violators, so as to effectively restrain the market power exercise motivation of power generators [20].

4 OVERVIEW OF MARKET POWER IN CHINESE POWER MARKET AND CORRESPONDING HANDLING MECHANISM

Since the announcement of "Notice on Carrying out the Pilot Work of Power Spot Market Construction" [2], the eight domestic spot pilot markets have all started the trial operation of the spot market settlement. This chapter summarizes the overall situation of the market power in Chinese power market and sorts out the corresponding market power handling mechanism.

4.1 Overview of market power in Chinese power market

From the perspective of the generation side, the competitiveness of the provincial power markets is still far from insufficient. Some provinces and cities still have more than 80% of the generation enterprises that are directly or jointly controlled by the five major power generation groups. The highly monopolized generation side provides generators with enormous bargaining power, bringing about huge market power potential. Taking the southern region as an example, the structure of power source in southern provinces remains relatively homogeneous, leading to the lack of complementary resources, and the optimization potential of the generation side is significantly limited. In the short term, the growth rate of the installed capacity will still exceed that of power load, resulting in the continuous decline of the unit utilization hours, along with the consensus on the generation side. to "abandon the quantity and maintain the price" [28]. As for installed capacity structure, at the end of 2021, China's installed capacity was 2.38 billion kW, of which full-caliber thermal power units were 1.3 billion kW, accounting for more than 50%. The installed capacity of coal-fired units was 1.11 billion kW, accounting for 46.6%, and the quantity of electricity output from them accounted for 60.0% of the total power output [29]. With coal-fired units acting as the major power supply source, the generation-side pattern will last for the coming time. Meanwhile, the flexibility resources among the system are relatively scarce. The proportion of the installed capacity of gas power units remains at a low extent, and the spatial mismatch between supply and load is widely present in certain provinces. The above factors offer preconditions for the short-term supply & demand imbalance and the serious transmission blockage, which may result in significant market power potential.

In general, market power exists widely in Chinese spot markets, and the phenomenon of market power exercising by thermal units is pretty common. Moreover, there is a lack of effective monitoring and mitigation measures at the practical level, so it is urgent for the coordinated development of top-level design and actual operation.

4.2 Market power handling mechanism in Chinese spot markets

At the present stage, the Chinese pilot power spot markets are still in the process of construction. The market operation and trading mechanism need to be improved, and the framework of market power assessment, monitoring, and mitigation need to be promoted in the top-level design dimension in order to establish an effective and feasible implementation plan. Currently, there are only three pilot markets, Southern region (starting from Guangdong), Shandong, and Gansu, that have explicit monitoring and mitigation handling schemes for the abuse of market power in the rules. The following will elaborate on the handling schemes for the exercise of market power in these markets.

4.2.1 Southern region electric power market (starting from Guangdong)

According to the *Basic Rules of Operation of Guangdong Electric Power Market* [30], Guangdong power market adopts the conduct and impact test system. Among them, the conduct test is to compare the energy quotation of the generator set with the reference price of the market power test. If the unit quotation does not exceed the test reference price, it will be deemed to pass the test; Otherwise, it will be deemed to fail the conduct test, and a further impact test should be executed. The impact test is to replace the energy quotation with the reference price of the generator set, re-calculate the market clearing results, and measure the changing value of the generation income of the unit. The unit fails the impact test if the changing value of its income exceeds the preset threshold value. The quotation of the unit that has passed the conduct and impact test is regarded as a valid quotation and can directly participate in the market clearing process. However, the generator set that has failed the test can participate in the market clearing only when replacing the unqualified bid with the corresponding reference price.

The reference prices and income changing thresholds in the above test standards shall be recommended by the Market Management Committee and approved by the competent government authorities and energy regulatory authorities.

Guangdong has clarified the market power monitoring and mitigation mechanism, and formulated specific implementation methods in the rules. However, the rules do not put forward the formulation plan and principles for the reference price of market power detection of each unit, which will cause a certain degree of trouble at the specific operational level.

4.2.2 Shandong electric power market

According to the *Trading Rules of Shandong Province Electric Power Spot Market* [28], the market power supervision measures adopted consist of two parts: the market power mitigation mechanism based on capacity control before the event (hereinafter referred to as "Ex-ante regulation") and the market power correction mechanism based on price impact test after the event (hereinafter referred to as "Ex-post regulation").

① Ex-ante regulation

The purpose of ex-ante regulation is to identify the generators with market power potential in the day-ahead market stage through the structural index test, and then regulate their bidding capacity on this basis. Specifically, before the day-ahead market clearing, the regulator should first set the residual supply index threshold ρ_0^{RSI} (the current value is 1.05); Meanwhile, calculate the residual supply index of each generator. If the residual supply index exceeds the threshold, the generator will be deemed to possess market power potential.

$$\rho_j^{RSI} = \frac{S_0 - S_j}{D_0} \tag{7}$$

ρ_j^{RSI} is the residual supply index of the generator j, S_0 is the total generation capacity of the generation side, S_j is the summation of the installed capacity of all the units available of the generator j, and D_0 is the forecast load of the trading period.

Subsequently, the regulator divides the generation capacity of generators with market power potential into critical bidding capacity S_j^{CBC} and regulated bidding capacity S_j^{RBC}, namely:

$$S_j^{CBC} + S_j^{RBC} = S_j \tag{8}$$

The following formula is for calculating the regulated bidding capacity S_j^{RBC} :

$$S_j^{RBC} = \left[\frac{S_j}{D_0} - \left(\frac{S_0}{D_0} - \rho_0^{RSI}\right)\right] * D_0 \tag{9}$$

Furthermore, calculate the must-be-cleared capacity S_j^{MBC}:

$$S_j^{MBC} = \frac{D_0}{S_0} * S_j^{RBC} \tag{10}$$

The must-be-cleared capacity of each unit is ranked from low to high according to the quotation. All units that provide the must-be-cleared capacity become must-run units, and the must-be-cleared capacity will no longer participate in the market bidding.

Finally, set the bidding capacity constraint of the generators, and integrate this constraint into the market clearing procedure:

$$S_j^{minCBC} + S_j^{MBC} \leq \sum_{i \in \Omega_{j,m}} G_{i,m}^G \leq S_j^{CBC} + S_j^{MBC} \tag{11}$$

The *m* represents the index of the trading period, S_j^{minCBC} is the minimum output of the critical bidding capacity S_j^{CBC}, and $G_{i,m}^G$ is the clearing capacity of the unit *i* during the period *m*.

② Ex-post regulation

After the real-time market clearing, the regulatory authorities will judge whether the conditions for carrying out the ex-post regulation are met according to the average clearing price. If the corresponding conditions are met, the excess income from the generation side will be recovered and returned to the consumers. Ex-post regulation is divided into two parts: current-day regulation and historical regulation. The former targets the clearing results of the real-time market of the current day, while the latter targets the clearing results of the past 7 days (including the current day).

First, regulators need to set a benchmark electricity price P_i^{BEN} for each unit:

$$P_i^{BEN} = C_i * (1 + \pi) \tag{12}$$

C_i is the approved marginal production cost of each unit and π is the excess return rate (generally 10%) to cover the long-term cost of investment.

Secondly, regulators need to set up a trigger price P^{REF} for the ex-post regulation. After the day-ahead market clearing, regulators will assess the trigger conditions of the regulatory procedure of the current day: if the average price over the current day is higher than the trigger price P^{REF} times a multiple λ^{TD}(currently set as 1.15), then the regulatory procedure of the current day will be initiated; Otherwise, the trigger conditions of the historical regulatory procedure will be further assessed: if the average price over the past 7 days (including the current day) is higher than the trigger price P^{REF} times a multiple λ^{TW} (currently set as 1), then the historical regulatory procedure will be initiated.

If the above test conditions are triggered, the regulator needs to conduct the following "price impact test" for all power generators (take the current-day regulation as an example): firstly replace the price of all generator sets with the benchmark electricity price P_i^{BEN}, and re-calculate the market to get the average price $\bar{P}_{j,t}^{AVE}$; secondly, compare it with the actual average price P_t^{AVE} on that day, and calculate the contribution factors ω_j of the generator *j*:

$$\omega_j = max\left(\frac{\sum_{t=1}^{T}\left(P_t^{AVE} - \bar{P}_{j,t}^{AVE}\right)}{\sum_{t=1}^{T}\sum_{j=1}^{N}\left(P_t^{AVE} - \bar{P}_{j,t}^{AVE}\right)}, 0\right) \tag{13}$$

N represents the total number of generators, *t* indicates the trading period, and *T* represents the total number of trading periods. Subsequently, the entire excess incomes of the generation side ΔR are calculated as:

$$\Delta R = \sum_{t=1}^{T} G_t * (P_t^{AVE} - \bar{P}_t^{AVE}) \tag{14}$$

G_t is the clearing electricity in period *t*. Based on the contribution factors of each generator, the apportionment amount of each generator is calculated as:

$$\Delta R_j = \mu * \Delta R * \omega_j \tag{15}$$

In the formula above, the apportionment coefficient μ is currently set to 1. As the market develops, the social welfare transfer caused by the exercise of market power can be further corrected by enlarging this coefficient, which is conducive to inhibiting the abuse of market power.

Finally, the excess income will be returned to the consumers:

$$\Delta R_g = \mu * \Delta R * \frac{Q_g}{Q} \tag{16}$$

ΔR_g indicates the return fee to the market consumer *g*, Q_g indicates the daily traded electricity of the consumer in the spot market on that day, and Q is the summation of daily traded electricity of all the market users Q_g.

In summary, Shandong electric power market adopts a "hybrid" mechanism similar to the ERCOT: firstly use structural index test to screen out the generators with potential of market power, and constrain their bidding capacity; secondly, apply impact test and return the excess income to the consumer-side, which not only realizes the market power mitigation but also completes the compensation transfer of social welfare.

4.2.3 Gansu electric power market

The Implementation Rules of Gansu Electric Power Spot Market Transaction clarify the market power monitoring and mitigation mechanism, which is similar to Shandong [32]. It consists of ex-ante price detection mechanism and ex-post price mitigation mechanism.

① Ex-ante price detection mechanism

Consistent with Shandong, if the generator fails to pass the residual supply index test, the regulator will calculate the total bidding quantity of the startup unit during each period on the operation day. On this basis, the current quotation of the startup unit under the generator is calculated and accumulated. If the cumulative value exceeds 50% of the total bidding quantity, the generator is regarded as an abuser of market power in the day-ahead quotation stage. The ex-ante mitigation measures for the generators are: all quotations higher than the predetermined reference price will be replaced, then participate in the day-ahead and real-time market clearing, and serve as the basis for market settlement. The Gansu market has clearly given the calculation method of testing the reference price in advance: the arithmetic average value of the clearing price of the day-ahead spot market during the peak-load period (18:00-22:00) in the past 10 days (excluding the current day).

② Ex-post price mitigation mechanism

Regulators first set up an ex-post threshold P^{Post}. If the market clearing price is higher than 1.5 times the threshold, then the market clearing price fails the test and need to be corrected: the unit whose quotation is higher than the ex-post threshold will be modified to the threshold itself, then re-calculate the market clearing result, and the new result is taken as the final.

5 NEW CHALLENGE OF CONSTRUCTING MARKET POWER MONITORING AND MITIGATION MECHANISM UNDER CHINA'S NEW POWER SYSTEM

With a large amount of renewable sources widely penetrating, the structure of the new power system will be transformed significantly. The proportion of electricity output from renewables such as wind and solar will greatly increase, and non-fossil energy in primary energy consumption will occupy the dominant position. Meanwhile, the thermal units will be gradually transformed from the main power supply to the regulatory power and the backup sources, with the system's demand for peak and frequency regulation resources markedly increasing. Chen et al. first proposed an optimal path to achieve a carbon-neutral power system with negative CO2 reduction costs in China by the middle of this century based on a high-definition simulation model, and the transformation trends will profoundly affect the market power issues of the spot market [33].

Firstly, the low marginal production cost characteristic of renewables makes it always be in a priority position in the dispatching sequence, which makes the aggregate supply curve moves to the right, leading to a decline in the clearing price. This phenomenon is also known as the "merit-order effect" of renewable units [20]. Considering the regional policies to support the priority consumption of renewables output, along with the subsidies which leads to negative market prices more frequently, the market clearing price will undergo a significant downward trend in the future. Furthermore, as the power output of coal-fired units is gradually compressed, their current bargaining power in the spot market will be significantly weakened. The literature shows that under 80% of renewables penetration, Chinese thermal units' annual online hours will be observably reduced to around 1400 hours, with only 16% of the unit utilization, which means thermal units can only recover most of its cost in a short time through the scarce price mechanism, or through its revenue in the capacity market, auxiliary service market, causing the lack of long-term investment incentives[20]. Various results show that the market power endowment distribution on the generation side will be rearranged.

Secondly, the most distinguished characteristics of the new power system are: high proportion of renewables, high proportion of power electronic equipment, low rotational inertia, and considerable randomness [34]. The major reason for the randomness is the volatility, intermittence and unpredictability of the output of renewable units. Therefore, the new power system will be equipped with a higher proportion of flexible sources, such as gas-fired units and energy storage system. When the output fluctuation of renewable units leads to a short-term imbalance between the power supply and demand, the flexible sources will be adjusted in time to maintain the stability of the system. Such resources are expected to possess considerable market power potential in the future power market, which has not been fully evaluated at present. In addition, when the output vacancy period of renewables and load peak period overlap (for instance, photovoltaic generation equipment will stop working after the sunset, while the residential power load will rise dramatically), the flexible sources will own a larger share in market bidding capacity, which motivates them to exercise market power in order to gain huge revenue in a short period of time [7]. Such rapid-adjustment resources are in a key position for certain periods, yet the regulatory measures for their market power exercise still need to be investigated and improved.

Thirdly, the impact of renewables penetration on market price is not only determined by the merit-order effect, but also related to the ownership of renewables capacity. The study shows that under perfect competitive conditions, the introduction of 5000MW wind turbines into the Ontario electricity market will reduce the clearing price by about 30%, while the price falls by only 7% when the same capacity is solely introduced into the largest generator[20]. This indicates that the abuse of market power by oligopolistic generators will greatly offset the merit-order effect brought by renewables, thus hindering the downward trend of market price, and further increasing the social welfare transferred to the generation side. This phenomenon further proves the need for regulators to evaluate and control the market structure of the generation side.

Finally, with the transformation of the new power system continues to advance, the market volume will be greatly expanded, and the market entities will continue to increase. However, limited by their resources and capability, it is difficult for the market operators to carry out a series of responsibilities such as analyzing the overall market structure, implementing the market power monitoring and mitigation methods, and evaluating the market transaction results alone. Therefore, we believe that the introduction of a third-party regulator would be considerable when conditions are ripe. Mature experience of typical power market shows that introducing an independent, professional, impartial, and open external regulator will provide the maximum support to market operators in market power monitoring and mitigation, market risk assessment, market information disclosure supervision, and market transaction result evaluation [35].

6 CONCLUSION

This paper summarizes the theory and the framework of market power monitoring and mitigation, introduces the mature experience of typical power markets in the current situation of market power in the assessment and prevention, and finally presents the development and challenge of constructing market power monitoring and mitigation mechanism for the Chinese new power system.

ACKNOWLEDGMENTS

This work is supported by State Grid Hubei Electric Power CO. Science and Technology Project (Design, research and application of key technologies of provincial electricity market trading system under the unified national electricity market, 521538210002).

REFERENCES

[1] The Central Committee of the Communist Party of China, the State Council. (2015). "Several Opinions of the CPC Central Committee and The State Council on Further Deepening the Reform of the electric Power System" (Zhongfa [2015] No. 9). [online]. https://news.ncepu.edu.cn/xxyd/llxx/52826.htm (7 October 2022).

[2] General Office of National Development and Reform Commission, General Department of National Energy Administration. (2017). "Notice on Carrying out the Pilot Work of the Construction of Electricity Spot Market" (Development and Reform Office Energy [2017] No. 1453). [online]. http://www.nea.gov.cn/2017-09/05/c_136585412.htm. (7 October 2022).

[3] National Development and Reform Commission National Energy Administration. (2022). "Guiding Opinions on Accelerating the Construction of a National Unified Electricity Market System" (National Development and Reform Commission [2022] No. 118). [online] http://www.gov.cn/zhengce/zhengceku/2022-01/30/content_5671296.htm. (7 October 2022).

[4] Lakić, E. et al. (2017). The review of market power detection tools in organised electricity markets. In 2017 14th International Conference on the European Energy Market (EEM) (pp. 1-6). IEEE.

[5] Chen, D., Jing, Z., and Shi, J. (2019). Control of market power in the electricity market. In: Proceedings of the 2019 Academic Annual Meeting of the Electricity Market Professional Committee of the Chinese Society for Electrical Engineering and the National Electricity Trading Institutions Alliance Forum, China SiChuan Chengdu. pp.56-69(in Chinese).

[6] Chen, Q. et al. (2018). Review on Market Power Monitoring and Mitigation Mechanisms in Foreign Electricity Markets. SOUTHERN POWER SYSTEM TECHNOLOGY, 12(12): 9-15(in Chinese).

[7] Graf, C. et al. (2021). Market Power Mitigation Mechanisms for Wholesale Electricity Markets: Status Quo and Challenges. Work. Pap. Stanf. Univ.

[8] Guo, H. et al. (2019). Market power mitigation clearing mechanism based on constrained bidding capacities. IEEE Transactions on Power Systems, 34(6), 4817-4827.

[9] Salarkheili, S. and Foroud, A. A. (2013). Market power assessment in electricity markets: supply function equilibrium‐based model. International Transactions on Electrical Energy Systems, 23(4), 553-569.

[10] Moiseeva, E., Hesamzadeh, M. R., and Biggar, D. R. (2014). Exercise of market power on ramp rate in wind-integrated power systems. IEEE Transactions on Power Systems, 30(3), 1614-1623.

[11] Shafie-khah, M., Moghaddam, M. P., and Sheikh-El-Eslami, M. K. (2016). Ex‐ante evaluation and optimal mitigation of market power in electricity markets including renewable energy resources. IET Generation, Transmission & Distribution, 10(8), 1842-1852.

[12] Bigerna, S. et al. (2022). COVID-19 lockdown and market power in the Italian electricity market. Energy Policy, 161, 112700.

[13] Zhang, W. et al. (2021). Overview of Market Power Mitigation Measures in North American Electricity Market. GUANDONG ELECTRIC POWER, 34(04): 24-33(in Chinese).

[14] Zhong, J. et al. (2018). Overview of Market Power Regulation and Mitigation Measures in Electricity Markets. In: Chinese Society of Electrical Engineering Proceedings of the 2018 Academic Annual Meeting of the Electricity Market Professional Committee and the National Electricity Trading Institutions Alliance Forum, China Shanghai. pp.167-175(in Chinese).

[15] Xue, Y. et al. (2010). A Research Framework for Genaralized Congestions and Market Power. Automation of Electric Power Systems, 34(21): 1-10(in Chinese).

[16] Jiang, X. (2020). Reference Price Setting for Market Power Mitigation in U.S. Electricity Markets. ZHEJIANG ELECTRIC POWER, 39(12): 85-89(in Chinese).

[17] Bao, M. et al. (2017). Review of Nordic Electricity Market and Its Suggestions for China. Proceedings of the CSEE, 37(17): 4881-4892(in Chinese).

[18] Xie, J. et al. (2020). Research on the risk prevention method of market power in electricity market. Price Theory and Practice, (12), 49-53(in Chinese).

[19] Dong, L. et al. (2021). Identification of Market Power Abuse in Spot Market of Chinese Electric Market. Proceedings of the CSEE, 23(41): 8397-8408(in Chinese).

[20] Bahn, O., Samano, M., and Sarkis, P. (2021). Market power and renewables: The effects of ownership transfers. The Energy Journal, 42(4).

[21] Li, L., Tan, Z., and Zhang, E. (2014). Research on market power in the implementation of renewable portfolio standard. Power System Protection and Control, 42(12): 106-112(in Chinese).

[22] Zhu, W. et al. (2019). Market Power Assessment of Thermal Power Units Considering Renewable Energy Consumption. Power System and Clean Energy, 35(12): 74-82(in Chinese).

[23] CAISO. (2019). California Independent System Operator Corporation Fifth Replacement Electronic Tariff. [online]. http://www.caiso.com/Documents/Section39-MarketPowerMitigationProcedures-asof-Sept28-2019.pdf. (7 October 2022).

[24] Monitoring Analytics L. (2019). 2019 State of the Market Report for PJM. [online]. https://www.monitoringanalytics.com/reports/PJM_State_of_the_Market/2019/2019-som-pjm-volume2.pdf. (7 October 2022).

[25] New York Independent System Operator I. (2019). NYISO Tariffs. [online]. https://nyisoviewer.etariff.biz/ViewerDocLibrary/MasterTariffs/9FullTariff.pdf. (7 October 2022).

[26] Midcontinent Independent System Operator I. (2013). Market monitoring and mitigation measures of this module D. [online]. https://www.misoenergy.org/ (7 October 2022).

[27] Potomac Economics L. (2018). State of the market report for the Ercot wholesale electricity markets. Virginia: Potomac Economics, LTD. [online]. https://www.potomaceconomics.com/documents/2018-state-of-the-market-report/ (7 October 2022).

[28] Liang, Z. et al. (2017). Discussion on Pattern and Path of Electricity Spot Market Design in Southern Region of China. Automation of Electric Power Systems, 41(24): 16-21(in Chinese).

[29] China Electricity Council. (2022). 2021-2022 National Electricity Supply and Demand Situation Analysis and Forecast Report. [online]. https://gxj.sxxz.gov.cn/gyjj/202201/t20220130_3729704.html. (7 October 2022).

[30] Southern Energy Regulatory Bureau. (2018). Basic Rules for Guangdong Electricity Market Operation. [online]. http://nfj.nea.gov.cn/adminContent/initViewContent.do?pk=402881e55992395f0159b57e91880036. (7 October 2022).

[31] Shandong Supervision Office of National Energy Administration. (2020). Basic Rules for Shandong Electricity Market Operation. [online]. http://sdb.nea.gov.cn/tzgg/content_1451. (7 October 2022).

[32] Gansu Electric Power Trading Center. (2021). "Implementation Rules of Gansu Electric Power Spot Market Trading" (Interim V2.2 for Settlement Trial Operation). [online]. https://news.bjx.com.cn/html/20210430/1150527.shtml. (7 October 2022).

[33] Chen, X. et al. (2021). Pathway toward carbon-neutral electrical systems in China by mid-century with negative CO2 abatement costs informed by high-resolution modeling. Joule, 5(10):2715–2741.

[34] Zhang, Y. (2021). Digital transformation and system mechanism construction are keys: Interview with Sun Zhengyun, Chief information Officer of State Grid Corporation of China. State Grid Corporation of China, (05),24-27(in Chinese).

[35] Potomac Economics L. (2022). RTO Market Monitoring. [online]. https://www.potomaceconomics.com/practice-areas/rto-market-monitoring/. (7 October 2022).

[36] Liao, K. et al. (2020). Analysis on Regulatory Principles of Manipulative Behavior for European and American Electricity Markets and Enlightenment to China. Automation of Electric Power Systems, 44(14): 1-8(in Chinese).

Data Driven Z-FFR Physical Modeling

Wenbin Xiong[1], Zhangchun Tang[2], Pan Liu [3,*], Qiang Gao[1], Yan Shi[2], Fanyu Qu[4], Chencheng Liu[5], Cheng Liu[3,6]

[1]Nuclear and Radiation Safety Center, MEE. Beijing, 102445, China.
[2]University of Electronic Science and Technology of China, Chengdu, 611731, China.
[3]Tianfu Innovation Energy Establishment, Chengdu, 610299, China.
[4]China National Nuclear Power Co., Ltd. Beijing, 100045, China.
[5]Central South University of Forestry and Technology, Changsha, 410000, China.
[6]Electrical Engineering, Tsinghua University, Beijing, 10084, China.
[*] Corresponding author: liupanccsu2020@163.com

ABSTRACT

The Z-FFR (Z-Pinch Driven Fusion Fission Hybrid Reactor) is an important innovative design concept. The high uncertainty of the operating process of the pulsed power unit and the physical process of fusion and the absence of some theoretical and experimental conditions make it difficult to establish a high-precision mechanistic model, and it is difficult to obtain an accurate mathematical model of a complex, dynamic system. A data-driven physical modelling approach is urgently needed to replace the mechanistic models obtained with the aid of extensive simulations and experiments. The approach includes the creation of functional modules, the packaging of sub-modules, the configuration of module interfaces and the configuration of analytical models. Based on the actual needs of Z-FFR design and operation monitoring, the online analysis can be autonomously configured to accommodate different experimental data through machine learning, enabling anomaly detection, trend prediction, model design evaluation and operation assessment during the experimental process.

Keywords: Z-FFR; Data driven; Physical modeling; Machine learning

1. INTRODUCTION

The key core of Z-FFR is the physical process of fusion, which is very complex [1-4], and the introduction of new tools such as artificial intelligence and big data [5-8] allows for the physical modelling of Z-FFR using machine learning and a data-driven approach.

Due to the high uncertainty of the operating process of pulsed power systems and fusion processes and the absence of some theoretical and experimental conditions, it is difficult to establish a high precision mechanistic model for complex, dynamic systems with accurate mathematical models, and a set of data-driven models is urgently needed to replace the mechanistic models obtained from a large number of simulations and experiments. In addition, the traditional data analysis work of each model relies on the designer to manually import the data package into commercial software (Origin, Matlab, Excel, etc.) and then perform a series of tedious operations to generate charts and conclusions after point selection or programming, and the algorithm model and visualization means are relatively single and solidified, and the huge repetitive workload makes the overall automation of data analysis very low; from data import to The whole set of data analysis process from data import to conclusion chart generation cannot be intuitively displayed and recorded, resulting in poor traceability of the data mining work as a whole; at present, there is no complete set of deep learning data analysis mining tools from data pre-processing, feature extraction to analysis model design and training for the characteristics and needs of experimental data.

2. Z-FFR CONFIGURATION AND CHARACTERISTICS

The overall design of the Z-FFR [9] is shown in Figure 1 and consists mainly of the Z-Pinch drive, the fusion ignition system and the sub-critical reactor. The sub-critical energy reactor Keff<<1, no critical safety accidents and easy to achieve non-energetic residual heat safety, no radioactive leakage accidents, good inherent safety nature and no need for off-site emergency systems. the Z-FFR is capable of burning U-238 and Th-232 and can start from natural uranium, thus can be a millennium energy source with less resource constraints. It can be used on a large scale as soon as the fusion

International Conference on Computer Graphics, Artificial Intelligence, and Data Processing (ICCAID 2022)
edited by Ruishi Liang, Jing Wang, Proc. of SPIE Vol. 12604, 126041P
© 2023 SPIE · 0277-786X · doi: 10.1117/12.2674673

Proc. of SPIE Vol. 12604 126041P-464

neutron source technology is up to scratch. Subcritical energy reactors with an energy amplification factor of M > 10, and even up to 20 or more, can thus significantly reduce the requirements for fusion technology and can facilitate the early application of fusion energy technology. Z-FFRs are economical due to their simple fuel fabrication and cycle, long refuelling cycles and low and easy disposal of nuclear waste (most fission products have a short half-life), and are uniquely proliferation-resistant as they do not require a uranium-plutonium cycle and do not rely on uranium isotope separation technology. The minor actinides produced are largely transmuted, making it a very clean and environmentally friendly energy system.

Create functional modules, including parameter picking module, data pre-processing module, feature extraction module, dimensionality reduction analysis module, correlation analysis module and machine learning module; parameter picking module loads all models in the database with different models. The correlation analysis module includes double clustering analysis, association rule algorithm, typical correlation analysis module; machine learning module includes recurrent neural network, convolutional neural network, long and short time neural network, multilayer perceptron neural network module; users choose different algorithm modules to achieve data analysis according to actual needs. Double clustering analysis means that by extracting the time domain and frequency domain feature parameters of different time series data, forming a time series data-feature parameter matrix with time series data as the horizontal axis and feature parameters as the vertical axis, and clustering the rows and columns of the matrix respectively at the same time, it can analyse the classification of the same feature parameters under different time series data, the classification of different feature parameters under the same time series data, from high uncertainty.

The data table of each system in the issuance, the parameters to be analyzed are selected; the data pre-processing module pre-processes the data to meet the subsequent analysis requirements before using machine learning for feature learning; the feature extraction module uses data time domain analysis, frequency domain analysis and time-frequency domain analysis methods to extract the time domain features and frequency domain features of the Z-FFR time series signal.

The modular modelling approach based on data-driven Z-FFR operation and design evaluation forms a set of easy-to-operate, autonomously configurable and adaptable real-time analysis processes for Z-FFR design and operation data through functional module development, algorithm module packaging, parameter interface configuration, analysis model configuration and visual interface display, which can achieve Z-FFR model monitoring and trend prediction. In summary, the Z-FFR should be a perfect energy system in terms of performance.

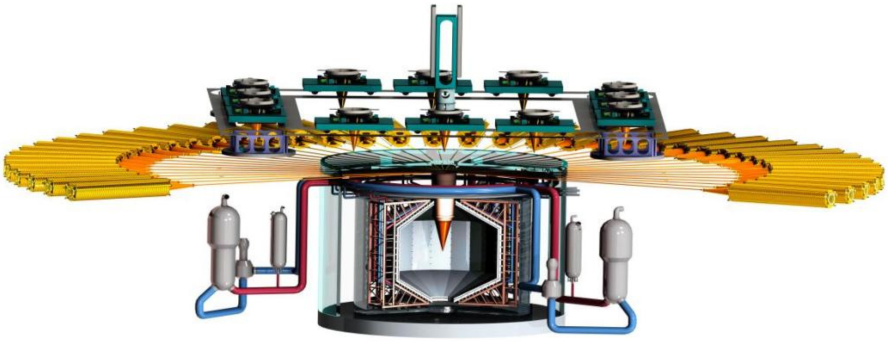


Figure 1. Z-FFR general structure diagram.

3. DESIGN PRINCIPLE AND PROCESSES OF MODELING

Data-driven physical modelling has been studied in depth in the aerospace field [10]. The dimensionality reduction analysis module reduces the dimensionality of the features through the feature extracted. The sub-module package is a modular package of the developed functional modules; each functional module contains multiple algorithm modules, the data pre-processing module contains cleaning, noise reduction, missing data completion, automatic removal of bad data,

data smoothing processing module. The feature extraction module contains mean, variance, rms, peak, skewness, cliffness, peak indicator, pulse indicator, fundamental frequency, multiplicity, spectral cliffness module; the dimensionality reduction.

The procedure for extracting valuable local feature information that can characterise the state of the launch vehicle from the time-series data of the same measurement point across rounds or between multiple measurement points of the same round is as follows.

1.For time-series data of the same round, the columns of the double-clustering matrix represent different measurement points, i.e. measurement point 1, measurement point 2, measurement point 3 ......This constitutes the correlation analysis matrix between data from different measurement points at the same time.

2.Cluster analysis of both rows and columns, with rows being clustered to analyse the correlation of the same indicator across all measurement points, and columns being clustered to analyse the correlation between all time and frequency domain indicators corresponding to a given measurement point.

3.After clustering the rows and columns separately, a number of classes are formed in the rows and columns, with parameters in the same class being correlated and parameters in different classes being dissimilar.

4.Uses mean square residuals to measure the consistent correlation between classes in rows and classes in columns. Assuming that class A in a row contains the element Xi , i = 1 ,2 ,3,...,n, and class B in a column contains the element Yj , j = 1 ,2 ,3,...,m, with n and m being positive integers, denoting the mean of Xi and denoting the mean of Yj , then the mean of class A and class The mean square residual between class A and class B is defined as [10].

$$\delta_{A,B} = \sum_{i=1}^{n}\sum_{j=1}^{m}\left[\left(X_i - \overline{X}\right) - \left(Y_j - \overline{Y}\right)\right]^2 \tag{1}$$

The mean square residual between the two classes is thus obtained, and the consistent relevance of the classes in the rows and columns is measured by the mean square residual; the smaller the mean square residual, the higher the relevance of the classes.

5.Based on the double clustering results in step 4, similar measurement point time series data and similar metrics are mined from the high uncertainty Z-FFR's same-shot multi-test and same-shot multi-test time series data for design and operational evaluation.

This modeling approach defines a framework for Z-FFR data analysis, which enables the preservation of the data in the database. The historical and real-time data is subjected to data pre-processing, dimensionality reduction analysis, correlation analysis and machine learning to achieve anomaly detection and health assessment to assist commanders in making launch decisions. The steps are as follows.

S1: Parameter selection: scan and load the data tables of all models and systems in the database, and select the parameters to be analyzed, this method is applicable to both the horizontal analysis of multiple parameters of the same type of experiment with different hair times and the vertical analysis of the same parameters of different experiment types.

S2: Pre-processing: Z-FFR experimental run history data and real-time data are time-series data, there may be anomalies, while the amount of data is large, before using machine learning for feature learning, a series of pre-processing of data is required, data cleaning, data noise reduction, due to telemetry data transmission and electromagnetic interference, further missing data patching, bad data since dynamic clearance, data smoothing.

S3: Feature extraction: using time domain analysis, frequency domain analysis, time-frequency domain analysis techniques to extract Z-FFR time series signal parameter features, such as pressure, temperature slow variables in the time domain for analysis, vibration signals using frequency domain feature analysis.

S4: Dimensionality reduction analysis: As high-dimensional feature parameters have information redundancy that is not conducive to further analysis, dimensionality reduction is required. Thousands of parameter measurement points are collected on the Z-FFR, in order to discover the implied relationships between measurement points faster and reduce the model training and testing time to meet the need for real-time analysis, principal component analysis and equal metric mapping can be performed on the selected Z-FFR parameters. Streamlining of high-dimensional feature parameters and reduction of feature dimensionality to enable faster fitting of the model with fewer parameters.

S5: Correlation analysis: Parameters within each subsystem of Z-FFR are often highly correlated, and even parameters between systems are implicitly correlated because of conduction. In order to reveal the correlation between multi-dimensional feature parameters and make feature parameter selection more targeted, selecting feature parameters with strong correlation for mutual prediction can achieve higher prediction accuracy, further Z-FFR design and operational data are The correlation analysis is carried out to realize the correlation analysis of the same measurement point across rounds and multiple measurement points of the same round.

S6: Machine learning: predictive analysis of specific features, establishment of mapping relationships between multiple measurement points of the same round or between different rounds of the same measurement point, trend prediction, anomaly detection, design and evaluation. Algorithm modules include: recurrent neural network, convolutional neural network, long and short term memory network, and multi-layer perceptron neural network. The output of the network is one of the temperature measurement points. The trained network can predict the future temperature of the temperature measurement point by inputting past or current time parameters, and then compare it with the actual temperature to determine the health of the subsystem in which the temperature measurement point is located, assisting designers and experimenters in the design and operation of the control system. Overall, the flow chart for this method is shown in Figure 2.

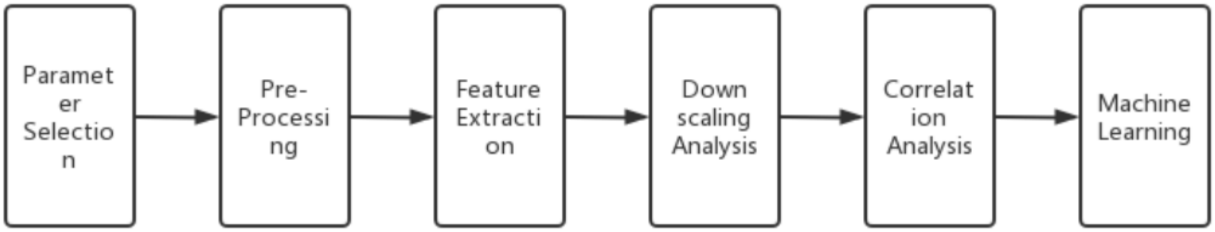


Figure 2. Data-driven Z-FFR physical modelling flow chart.

4. MATHEMATICAL EQUATIONS

In terms of drives, the data-driven approach will be superior to the traditional domain. A comparison of drive parameters is shown in Table 1.

Table 1. Comparison of drive parameters.

Parameter Name	Traditional[1]	New approach in this paper
Number of switches	100,000	1,000,000
Convergence current	25MA	60MA
Module Voltage	50kV	100kV
Time-delay	100ns	150ns

From here it can be seen that the data drive can be designed with more switches up to the order of 1,000,000, so it is possible to prepare a drive with more than 50MA and a delay time of 150ns. Also, Table 2 gives a comparison of the physical process parameters.

Table 2. Comparison of physical process parameters.

Parameter Name	Traditional[1]	New approach in this paper
Radiation energy	3MJ	15MJ
Radiation temperature	250eV	300eV
Neutron yield	10^{12}-10^{13}	10^{19}

Using data-driven physical modeling, it is clear that the radiative energy is increased by a factor of 5. At the same time, the neutron yield will increase by a factor of 1-10 million, reaching the threshold for fusion ignition.

5. CONCLUSION

With the data-driven modeling approach of this paper, a solid foundation was laid for achieving fusion ignition. In terms of switching control, the number has increased tenfold to over 1,000,000. The convergence current can reach 50 MA with a delay time of 150 ns. In the fusion process, the radiation energy is increased by a factor of 5. In particular, the neutron yield is increased by a factor of 1 million to 10 million, with a number of 10^{19}, reaching the fusion ignition threshold.

The physical modelling method used in this paper adopts a double clustering analysis that can automatically obtain valuable local information and discard noise representing randomness, which is suitable for online analysis of missing and inaccurate data; the data-driven data analysis method for Z-FFR physical modelling can complete the correlation analysis and prediction of Z-FFR parameters without complex a priori knowledge of the mechanism model; a complete set of data analysis This method can be used to analyse the Z-FFR parameters of the same experiment in a horizontal comparison, or to analyse the same parameters of different experiments in a vertical comparison, so as to evaluate and predict the operation status of the Z-FFR equipment and models from multiple dimensions. This method allows for the evaluation and prediction of the operational status of Z-FFR equipment and models from multiple dimensions.

ACKNOWLEDGMENTS

The authors would like to thank the reviewers and editors for their hard work and the corresponding author **Liu Pan** for proofreading the English manuscript. This work was supported in part by the National Natural Science Foundation of China under Grant 51875087.

REFERENCES

[1] D. B. Sinars, S. A. Slutz, M. C. Herrmann, et al. Measurements of magneto-Rayleigh-Taylor instability growth during the implosion of initially solid metal liners, Phys. Plasmas 18, 056301, 2011.

[2] A. B. Zylstra, O. A. Hurricane, D. A. Callahan, et.al, Burning plasma achieved in inertial fusion, Nature, 601, 542-548, 2022.

[3] X. J. Peng, C. A. Liu, X. M. Shi, Nuclear energy future and Z-Pinch drive fusion fission hybrid reactor. chapter 1, National Defense Industry Press. Beijing.

[4] R. D. McBride, M. R. Martin, R. W. Lemke, et al., Beryllium liner implosion experiments on the Z accelerator in preparation for magnetized liner inertial fusion, Phys. Plasmas 20, 056309 (2013).

[5] J. Degrave, F. Felici, J. Buchli, et.al, Magnetic control of takamak plasmas through deep reinforcement learning, Nature, 602, 414-419, 2022.

[6] Pan Liu, Yinzhao Zhou, WenBin Xiong, et.al, Big data-aided study of the physical process of volume ignition. You have 2022 6th International Conference on Mechanics, Mathematics and Applied Physics (ICMMAP 2022) (ICMMAP 2022).

[7] Pan Liu, Gongjie Liu, Jie Liu, et.al, Big data collaborative artificial intelligence and high-performance computing to drive physical design of fusion. 2022 2nd International Conference on Big Data and Intelligence Algorithms (BDIA 2022).

[8] Gaoyang Liu, Dongfang Peng, Wenbion Xiong, et.al, Physical design of fusion target with edge computing. 2022 6th International Conference on Electrical, Mechanical and Computer Engineering.

[9] X. J. Peng, C. A. Liu, X. M. Shi, Nuclear energy future and Z-Pinch drive fusion fission hybrid reactor. chapter 6, National Defense Industry Press. Beijing.

[10] G. Wang, H. Geng, H. Xu, et.al, A modular modeling approach based on a data-driven launch vehicle health assessment model. Chinese invention patent, CN111160393.

An improved Approach to Privacy Data Protection in the Body Area Network

Wei-xing WANG [1], Liuqin YE [2]
1.School of Artificial Intelligence, The Open University of Guangdong, Guangzhou 510091, Guangdong, China; 2. Beihai Vocational College, Guangxi Beihai, 536000, Guangxi, China
[1]wxwang007@126.com;

Abstract

Wireless Body Area Network (WBAN) is radio frequency-based wireless network technology. It is currently widely used in sports, medical health and other fields. WBAN often transmits the user's body monitoring data, such as heart rate, pulse, blood pressure and other personal privacy data. These sensitive data are bound to face a serious risk of data leakage or malicious tampering. In order to solve the privacy security of user data in WBAN environment, this paper proposes an improved privacy protection strategy. This strategy selects a portion of the region so that the phantom nodes are evenly distributed around the source nodes, and the adjacent packet space has a certain angle, in addition to increasing the diversity of the path from the source node to the base station by selecting the peer node. The simulation results show that the strategy can provide better privacy data protection without adding too many child nodes, and extend the security time.

Keywords- Wireless body area network, Privacy data protection, Data security

1. Introduction

Wireless Body Area Network[1] (referred to as WBAN)is a communication network centered on the human body, with the purpose of collecting various physiological parameters of the human body, and, consisting mainly of sensors with perceptual functions and a body master station (or WBAN coordinator). The WBAN coordinator acts as a network cluster, acts as a medium, and is also a gateway between the WBAN and the external network (3G, 4G, WiMAX, Wi-Fi), which stores data and ensures its secure transmission and exchange. The sensor nodes in WBAN can not only perceptually collect the relevant physiological parameters of the human body, but also obtain the environmental information around the human body. Then the collected data information is sent to mobile terminals and other related equipment, and finally through the Internet to the collected data information sent to the telemedicine service center and other related equipment, the medical center according to different needs of the data processing and analysis, the specific WBAN architecture[2] as shown in Figure 1. WBAN is currently widely used in sports, medical health and other fields, and is increasingly becoming a research hotspot[3].

Since most of the data collected by the wireless body area network is user privacy information, it is vulnerable to intrusion by malicious attackers during transmission. Through attack methods such as eavesdropping and traffic analysis, intruders pose a great threat to the security of user data privacy in the wireless area network. When an attacker attacks the user's privacy data, the attacker can not only obtain the relevant parameter configuration in the network environment, but also obtain the user's personal sensitive privacy data, then the attacker can potentially obtain various service information, and then grasp the user's various behaviors. Therefore, it is necessary to protect the privacy and security of user data during the transmission of wireless body area network data. In order to solve the privacy security of user data in the body area network, this paper proposes an improved data security transmission mechanism for the security and privacy protection of data transmission in the wireless body area network, the main idea is to divide the communication node network, divide it into several communication ring networks, and apply the K anonymity scheme to the perception layer and transmission layer of the hierarchical model, through multiple K anonymization processing, to ensure that the user data at each layer is relatively safe in probability. This method can minimize the probability of privacy data being stolen, thereby achieving the safe transmission of user privacy data in the wireless body area network.

The security of data information can be ensured through content encryption and anonymity, while the background information such as the communication mode and geographical location of nodes will still be exposed to attackers. Because the exposure of location privacy of source nodes in sensor networks inevitably threatens the security of monitored targets, the protection of location privacy of data source nodes is an urgent problem to be solved. These

International Conference on Computer Graphics, Artificial Intelligence, and Data Processing (ICCAID 2022)
edited by Ruishi Liang, Jing Wang, Proc. of SPIE Vol. 12604, 126041Q
© 2023 SPIE · 0277-786X · doi: 10.1117/12.2674623

security and privacy issues have become the key issues restricting the deployment and application of WSNs. According to the different capabilities of attackers, the source location privacy protection protocols can be divided into two categories: the source location privacy protection protocols against global traffic attackers and the source location privacy protection protocols against local traffic attackers. For the former, the attacker can only monitor the situation of the whole small area network, which is not applicable to the large-scale sensor network. The existing research works mainly forward data packets from phantom nodes to base stations through the shortest path, which is relatively single and easy to cause overlap on the path. Therefore, this paper proposes Phantom routing based on area and multi-node selection (PRAMS) for the second attacker. This protocol can make the selected PHANTOM nodes keep a certain Angle and distance, distribute around the source node evenly enough, and diversify the routing path through multi-node selection, which greatly reduces the possibility of overlapping paths. Considering the attacker with stronger visual ability, this protocol reduces unnecessary flooding by shielding the node's choice of route in the visible area, and recovers the node state after the detection target leaves. Simulation results show that the proposed protocol can provide better security performance and less energy consumption.

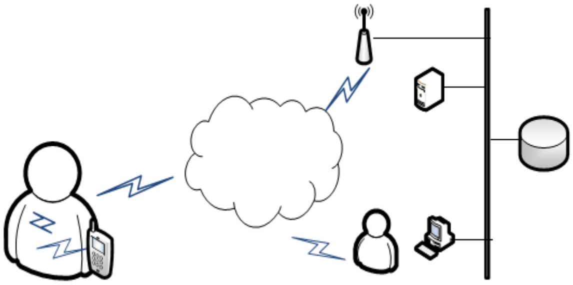


Figure 1. Wireless body area network architecture

2. Related work

Nowadays, in view of the data transmission privacy security issues of wireless body area networks, data encryption technology is often used to protect users' private data. However, due to the limited energy, resource and data storage capabilities of the wireless body area network, the existing password mechanism has been difficult to meet the needs of the wireless body area network. In order to ensure the security of the user's private data in the body area network, it is necessary to introduce other new technologies under the coordination of cryptography to achieve the safe transmission of user privacy data. Summarizing the existing related technologies for the security and privacy protection of wireless body area network data transmission, it can be roughly divided into the following types: password and key management technology, secure transmission privacy protection protocol design, authentication technology, secure routing technology, intrusion detection technology and biometric technology, anonymity technology, and fuzzy technology.

At present, for cryptography, a document[4] proposes a key management scheme based on time and location, in order to achieve the safe transmission of user privacy data in the wireless body area network, the protocol combines symmetric cryptography algorithms with active secret sharing mechanisms. The literature[5] proposes a key protocol (PSKA) based on physiological signal information, which provides a guarantee for the secure communication behavior between nodes in the body area network. The literature [6] proposes a node secure communication protocol for anonymous authentication and key negotiation. For the secure routing technology, literature[7] proposes a secure routing protocol based on feedback information, literature[8] proposes a secure routing protocol based on geographic information, literature[9] proposes a secure routing protocol based on cryptographic algorithms, literature [10] proposes a secure routing protocol based on multipath transmission, literature [11] proposes a secure routing protocol based on hierarchy, and literature[12] proposes a secure routing protocol based on specific attacks. For the application of biometric technology and intrusion detection technology, the literature[13] proposes role-based intrusion detection technology and integrates biometric technology, which can effectively ensure the safe transmission of data between WBAN sensor nodes and cluster nodes. Aiming at anonymity technology, the literature [14] studies the mechanism of traffic analysis attack under global listening, and proposes a statistical anonymity protection method Profit. For the fuzzy technical scheme, the literature proposes a modified fuzzy library scheme.

3. Network model

The network model is divided into three layers. The top layer is the application layer, which is generally played by a medical data center, and its role is mainly to store, analyze and analyze all the user data collected, and use the results for actual medical diagnosis. The middle layer is the transport layer (called the cluster head node layer), which is mainly responsible for collecting the user data of the lower layer and sending the results to the application layer. The lowest layer is the perception layer (collection node layer), which is mainly responsible for sensing the collection of relevant information with users and transmitting useful information to the upper layer network.

This article assumes that the WBAN system is a hierarchical system consisting of N sensor nodes, which is a collection of all sensor nodes $S = \{s_1, s_2, \cdots s_N\}$, $N \gg M$.

According to the traditional sensor node clustering method, it is assumed that the entire WBAN network is composed of M independent sensor node clusters, i.e. $CH = \bigcup_{i=1}^{M} CH_i$ .Each of these clusters is a non-empty true subset of the WBAN node set S, i.e. A $CH_i \subseteq S$. And all clusters are mutually exclusively divided, i.e. $CH_i \cap CH_j = \varnothing, (i \neq j, j = 1, 2, \cdots M)$. We use $|CH_i|$ to indicate the number of valid sensor nodes in the cluster.

For the second layer, suppose that each node cluster CH_i has a cluster head node h_i , which is responsible for storing, forwarding, and processing user physiological signs data transmitted from nodes inside the cluster and other cluster head nodes, and finally sending the results it collects to the medical data processing center. The cluster header node collection can be represented as $H = \{h_i, h_2, \cdots h_M\}$.

4. System model

4.1. network model

This paper presents an improved network model. A large-scale homogeneous wireless sensor network is deployed in a large monitoring area, and the sensor nodes are randomly and evenly distributed in the monitoring target area. Once a monitoring object is found, the node closest to the object will periodically send the monitoring results to the base station in the form of data packets until the attacker finds the target or the monitoring target leaves the monitoring area. The following assumptions are made for the whole network:

(1) The network is connected, that is, any two nodes in the network can communicate through multi hop transmission;

(2) There is only one base station in the network, and there is only one monitored target at the same time. The target has mobility and will leave the monitoring area every certain time. The nearest node automatically becomes the data source node after monitoring the target, and sends the information to the base station in the form of data packets;

(3) Each node in the network has a unique identification ID and knows its own location information, which is recorded as (Ni. X, Ni. Y).

4.2. Attacker model

The target of the attacker is to trace the routing path backward until he finds the data source node. This paper assumes that the attacker has the following characteristics:

(1) The attacker has enough storage space and powerful computing power. The attacker can quickly detect the sending node and move to the node location;

(2) The attacker monitors data packets near the base station and determines the location of the next hop node through the monitored data packets; The attacker only has the ability of local traffic analysis, and its listening range is equivalent to the communication range of the sensor;

(3) The data content is secure, and the attacker cannot obtain the encrypted information in the data package.

5. Solution

The improvement strategies proposed in the study are mainly divided into three stages.

5.1. Network initialization

In the network initialization stage, each node needs to obtain the location information of the neighbor node and the minimum hop count information from itself to the base station. After the end of this phase, each node i stores the finite flood hop value hw of the active node, the radius r of the visible area, the minimum hop value from itself and the neighbor node to the base station, and the location coordinates of the neighbor node. After that, the base station broadcasts the flooding information Sink_Msg to the whole network, including the message type, node ID, node location coordinates, and hop count information hb from the base station, with an initial value of 0. After receiving this message, the node will add HB by 1. If the node receives the message for the first time or the hop count information is smaller than the information stored by itself, the hop count information hi, b = hb and the information of the neighboring node will be updated, and then the message will be forwarded.

5.2. Limited flooding

When the target is detected in the nearby area, the data source node broadcasts the message Source_Msg to the nodes within its hw hop range, which is similar to the process of network flooding. It includes the message type, node ID, the value of Angle α_i, and hs indicates the hop count of the message. The initial value of all the above parameters is 0. When the message reaches each forwarding node, the count is increased by 1. If the count reaches *hw*, the node no longer broadcasts the message. When node i receives a broadcast message, it records the value of αi in the message and calculates the value of its value of αi according to formula (1), and then forwards it to the neighboring node. If *hs* is less than or equal to r, the node information is found in the set i.arent based on the ID number in the message. If the node is found in *i.arent,* the node is marked. All nodes marked in i.arent cannot be selected as the next hop forwarding node in the multi-node selection forwarding phase, so nodes in the visible zone are masked by nodes outside the zone. In particular, if the monitored object leaves the monitoring range of the source node, the source node sends a broadcast message with hop number r+1. After receiving the message, the node unmarks the sending node in i.parent.The Limited Flooding Algorithm is represented by Table 1.

α_i will be calculated by formula (1).

$$\alpha_i = arccos\frac{H^2+h_{i,s}^2-h_{i,b}^2}{2\times H\times h_{i,s}^2} \quad (1)$$

TABLE 1 Limited Flooding Algorithm

```
Algorithm1 limited flooding algorithm

case limited flooding:
  if (hs<=hw)
    if (node I receives the message for the first time) then
        record information about the neighboring node;
       hs = hs +1;
         computing Angle $\alpha_i$ ;
        broadcast the modified message;
    else logs the information about the neighbor node and discards the message;
    if (hs<=r) then
        search for neighbor node information in the set i.arent;
        if (find the neighbor node) then
             mark a neighbor node in the routing list;
    else stop the flood
randomly walk every node in the region in order to get a set i.child;
```

5.3. Directed routing

At the end of finite flooding, the area around the source node is divided into n parts, n is an even number, the Angle of each is, and these areas are defined as A1, A2, A3...... An, A1 is adjacent to An. The source node selects a region Ai when sending data packets, and decides the selection range of the next sending region according to the currently selected region. The number of selected regions is shown in formula (2).

$$m=\begin{cases}\frac{n}{2},\ \frac{n}{2}\ \text{is an old number}\\ \frac{n}{2}+1,\ \frac{n}{2}\ \text{is an even number}\end{cases} \tag{2}$$

The number of intervals between the current area and the selected area is $k=\frac{n-m+1}{2}$, the minimum interval Angle is $\Delta\beta=(k+1)\times\theta$, and the selection range of the sending area is $k=\frac{n-m+1}{2}$. In particular, if the source node sends the packet for the first time, it randomly selects one of the N regions as the sending region. The packet sent by the source node contains forwarding hop number hx and Angle range [(i-1)θ,iθ], and hx follows random distribution [hmin, hmax]. If node I receives a packet, it randomly selects a node from i.s_child whose Angle is in the Angle range of the packet, and then forwards the packet to the node. This process is repeated until the packet is forwarded H times. The area around the source node is divided into 8 parts, the minimum interval Angle is $\pi/2$, A2 is the current area, then the selection range of the next packet sending area is {A4,A5,…,A8}. From Equation (3), we can obtain the minimum spacing angle $\Delta\theta_{\min}=\frac{\pi}{2}$ as n goes to infinity.

$$\lim_{n\to\infty}\left(\frac{\frac{n}{2}+a}{2}\right)\bullet\frac{2\pi}{n}=\frac{\pi}{2},\ a\in\{0,1\} \tag{3}$$

Selecting the sending area of the next hop packets from the selection area ensures that the sending Angle between adjacent packets is at a minimum. In this way, the phantom nodes generated by adjacent packets are separated by a certain distance, which increases the difficulty of backward tracing and prolonging the security time.The Directed Routing Algorithm is represented by Table 2.

TABLE 2 Directed Routing Algorithm

```
Algorithm 2 directed routing algorithm

case directed routing:
    select the send region from the selectable field;
    random number $h_x \subseteq [hop_{\min},\cdots,hop_{\max}]$;
    if (hi,s<hx) then
      if ($i.child \neq \varnothing$) then
        select a neighbor node from i.hild in the Angle range as the next node and send the message;
      else  sent message to the sink through the brother selection routing algorithm;
```

5.4. Multiple nodes select forwarding

After the forwarding of the directed route is complete, the multi-node forwarding route selection stage is entered. The node I that receives the packet randomly selects one of the nodes of the two sets i.Prest and i.Baxter as the next hop forwarding node. The forwarded data packet contains the equal-distance node marker field equal and hop limit value HL. Every time the next hop node is selected from the set I. Bundle, equal is set to 1 and the value of HL is reduced by 1. When HL is 0, only nodes are selected from I. bundle to make the packet reach the base station quickly. Too many packets are forwarded through equidistant nodes, which increases the network delay and energy overhead. In order to prevent this protocol from expanding the range of node forwarding paths by selecting nodes from i.Rother set, the possibility of overlapping paths is reduced, and the security time is prolonged.

6. Experimental verification

The proposed strategy was validated using a comparative approach. Assume that there are 10,000 nodes evenly distributed in an area of 6000m x 6000M. The communication radius of each node is 100m. The average number of neighbors per node is 8.64. The number of neighbors of a small number of nodes is 3. The listening radius of the attacker is equivalent to the communication radius of the node. The radius of the viewing area is 600m.

The results show that the security time increases with the increase of the number of directed walk hops, because the distance from the phantom node to the base station becomes longer and longer, the transmission path becomes more complex, and the attacker needs more time to find the location of the source node. At the same time, a larger number of phantom nodes are generated, which can reduce the possibility of overlapping paths. Compared with the other strategies, this policy improves the security time by 58.6% and 36.8% on average. Therefore, this policy provides better security.

7. Conclusion

Data in all types of devices in WBAN is sensitive to users. Attackers can steal user privacy data passed by WBAN, from which information such as the user's identity, location, and health can be inferred. Therefore, WBAN has hidden dangers of user privacy leakage. TThis article proposes an improved privacy data protection policy in WBAN. Through directed routing and multi-node selective forwarding, the strategy increases the difficulty of the attacker to reverse chase the source node, prolongs the security time of the source node, and effectively protects the location privacy of the source node. In order to ensure the privacy of data during data transmission, the strategy also effectively designs a privacy protection mechanism for user data. This strategy keeps energy consumption as low as possible while maintaining high performance. In the future, how to establish an intrusion detection model and prevent attack behavior is a valuable research direction.

References.

[1] Mainanwal V, Gupta M, Upadhayay S K. A survey on wireless body area network: Security technology and its design methodology issue[C]// International Conference on Innovations in Information, Embedded and Communication Systems. Coimbatore: IEEE Press, 2015:1-5.

[2] GONG Ji-bing, WANG Rui, CUI Li. Research Ddvances and Challenges of body Sensor network[J]. Journal of Computer Research and Development, 2010,47(5): 737-753. (in Chinese)

[3] LIU Lu, XUE Xiu-qin, LUO Xian-lu. Architecture and Challenges of Wireless Body Area Network[J]. Computer Knowledge and Technology, 2012, 8(29): 6918-6920. (in Chinese)

[4] ZHENG G, FANG G, SHANKARAN R, et al. Multiple ECG fiducial points based random binary sequence generation for securing wireless body area networks[J]. IEEE Journal of Biomedical and Health Informatics, 2017, 21(3): 655-663.

[5] Venkatasubramanian K K, Banerjee A, Gupta S K S. PSKA: usable and secure key agreement scheme for body area networks[J]. Information Technology in Biomedicine, 2010, 14(1): 60-68.

[6] SHEN J, GUI Z, JI S, et al. Cloud-aided lightweight certificateless authentication protocol with anonymity for wireless body area networks[J]. Journal of Network and Computer Applications, 2018, 106: 117-123.

[7] Eu Z A, Tan H P, Seah W K G. Design and performance analysis of mac schemes for wireless sensor networks powered by ambient energy harvesting[J]. Ad Hoc Networks, 2011, 9(3): 300-323.

[8] Chen X Q, Makki K, Yen K. Sensor network security: A survey[J]. IEEE Communications Surveys & Tutorials, 2009, 11(2): 52-73.

[9] Ozturk C, Zhang Y, Trappe W. Source-location privacy in energy-constrained sensor network routing[C]// in SASN'04: Proceedings of 2004 ACM Workshop on Security of Ad Hoc and Sensor Networks, Washington, DC, USA, October 2004.

[10] Kamat P, Zhang Y, Trappe W, et al. Enhancing source-location privacy in sensor network routing[C]// in ICDCS'05: Proceedings of the 25th International Conference on Distributed Computing Systems, Ohio, USA, June 2005.

[11] Wang W P, Liang C, Wang J X. A source location privacy protocol in WSN based on locational angle[C]// IEEE International Conference on Communications. Washington: IEEE Computer Society, 2008: 1630-1634.

[12] Chen J, Fang B X, Yin L H, et al. A Source-Location Privacy Preservation Protocol in Wireless Sensor Networks Using Source-Based Restricted Flooding[J]. Chinese Journal of Computers, 2010, 33: 1736-1747.

[13] TAO Z L, Liu Y B, Li C X. Strategy of source-location privacy preservation in WSNs based on phantom single-path routing[J]. Journal of Chongqing University of Posts and Telecommunications, 2013, 25(2): 178-183.

[14] HE D, ZEADALLY S, KUMAR N, et al. Anonymous authentication for wireless body area networks with provable security[J]. IEEE Systems Journal, 2017, 11(4):2590-2601.

Application Research of geological Data Acquisition and Sharing System based on mobile terminal for cross-sea bridge

Xiaopeng Shi*[a], Guohe Guo[b], Tiancheng Liu[a], Fang Pan[b], Dongchao Luo[a]
[a] CCCC Highway Bridges National Engineering Research Centre Co., Ltd, Beijing, 100120, China;
[b] Guangdong Provincial Highway Construction Corporation, Guangzhou, 511447, China.
* Corresponding author: shixiaopengql@163.com

ABSTRACT

Cross-sea bridges have the characteristics of long strips, and the geological conditions along the route are complex. The geological data information collected by the survey is related to the quality of the entire bridge project. The geological information in the current survey stage is mostly recorded in paper logs. The design unit will design the structure after obtaining the paper geological information, and make design adjustments based on the actual geological information during the actual construction process. This puts forward higher requirements for the convenient, efficient, accurate collection and sharing of geological information, and the software based on mobile terminals can just meet these requirements. In view of the current problems in the collection and sharing of bridge engineering geological information, this paper designs an information solution for the collection and sharing of cross-sea bridge geological data, designs the system architecture and application functions in detail, and develops a cross-sea bridge geological information collection and sharing solution. The Android and IOS terminal apps build a bridge engineering geological system with B/S architecture through the front-end and back-end separation mode, and establish a standard geological database. Through the application of system functions in practical engineering projects, the digital collection and transmission management of geological data is realized, which greatly improves the transmission and sharing efficiency of geological data in the process of survey, design and construction, and provides a reference for the realization of digital management of bridge geological information.

Keywords: monile terminal, bridge, geology, sharing, system

1. INTRODUCTION

As a kind of long linear banded engineering, the cross-sea bridge engineering needs more geological data than the general bridge engineering, the scope of investigation needs to be wider, and the geological situation is more complex. Geological data information collected by investigation is related to the quality of the whole bridge project. At present, the geological data collection of bridge is mainly recorded in paper logs, and the design unit can only start the structural design after obtaining the paper geological information. In addition, in the actual construction process, the stratum data of each pile foundation is not exactly the same as that of the survey borehole, which requires the actual geological data information to be shared and transmitted in time, so as to facilitate the design unit to make rapid design adjustment. The traditional methods of geological information collection and transmission cannot meet these requirements in timeliness and other aspects, so it is necessary to combine with new computer information science and technology to develop digital geological data information collection and management technology [1].

With the rapid development of IOS, Android, HarmonyOS and other operating systems, driven by the continuous upgrading of mobile communication technology, smart phones, tablets and other smart mobile terminals have entered a stage of blowout development. Software applications based on mobile terminals have increased significantly, and the public has become more and more aware of mobile terminals. Among them, mobile phone has become the most concerned tool for the public, and also the most frequently used tool every day. With the increasing convenience and powerful functions of smart phones, APP has become an important starting point for enterprises to realize business applications and improve work efficiency. The main advantages of APP include comprehensive and timely information, strong continuity, anytime service, interactive number and high flexibility. Geological data collection and sharing based on intelligent mobile terminal is currently a hot research and development direction [2].

Qiang Lili, Zhang Hua [3]~[5] et al. developed a field geological data acquisition and management system based on Android and Web based on the characteristics of geological data acquisition. Zhou Chang 'an [6] analyzed the current situation of engineering investigation quality information management, and built an engineering investigation quality

International Conference on Computer Graphics, Artificial Intelligence, and Data Processing (ICCAID 2022)
edited by Ruishi Liang, Jing Wang, Proc. of SPIE Vol. 12604, 126041R
© 2023 SPIE · 0277-786X · doi: 10.1117/12.2674603

information management system on this basis. Sun Yan et al. [7] designed and developed the spatial database management system of coal geological exploration GIS from the perspective of coal field exploration, development and scientific research data application. Shuai Xunbo et al. [8] developed corresponding information sharing software based on the characteristics of natural gas geological information from the perspective of data integration application and sharing. Huang Dazhong et al. [9] analyzed the deficiencies in data acquisition and management of railway geological drilling at the present stage, and proposed a data acquisition solution based on Android mobile terminal. To sum up, at the current stage, the information-based collection and management of geological data is mainly concentrated in coal, water conservancy, civil and industrial construction industries, and there are few researches in the field of Bridges, and there is a lack of in-depth research on data sharing among all participating units such as survey, design and construction.

In view of the problems existing in geological data collection and sharing of survey, design and construction units, and based on the detailed investigation of the platform construction scheme of the existing geological information system, from the point of view of data collection, management and sharing, this paper carries out the research on the construction of the geological data collection and sharing platform of the cross-sea bridge based on mobile terminal. This work is of great scientific significance and engineering value for realizing timely and efficient sharing of geological data among different units.

2. OVERALL SYSTEM DESIGN

The main purpose of the geological data acquisition and sharing system for cross-sea Bridges is to build and develop a safe and usable software system based on the hardware system according to the characteristics of bridge survey, design and construction process and the user's operational requirements. The system should have good reliability, security, maintainability, expansibility, portability and manageability. There should also be integration and interoperation.

2.1 Overall architecture design

Through technical research and field practice, the background of the mobile terminal and B/S architecture developed by the system in this paper is determined. The overall architecture of the cross-sea bridge geological data acquisition and sharing system is shown in Figure 1 below, which mainly includes the infrastructure layer, data resource layer, service support layer, business application layer and user layer.

The infrastructure layer mainly includes servers, networks, security, storage, mobile terminals and other external environments and facilities that support system operation. The data resource layer mainly includes basic data, geological collection data, document data and system business data. Service support layer mainly includes data storage, data query, message middleware and other services. The business application layer mainly includes project management and other functions, providing various information function applications for bridge geological data collection and sharing. At the user level, all kinds of personnel, including survey personnel, design personnel, construction personnel and supervision personnel, need geological data information to cooperate with the operation.

Proc. of SPIE Vol. 12604 126041R-78

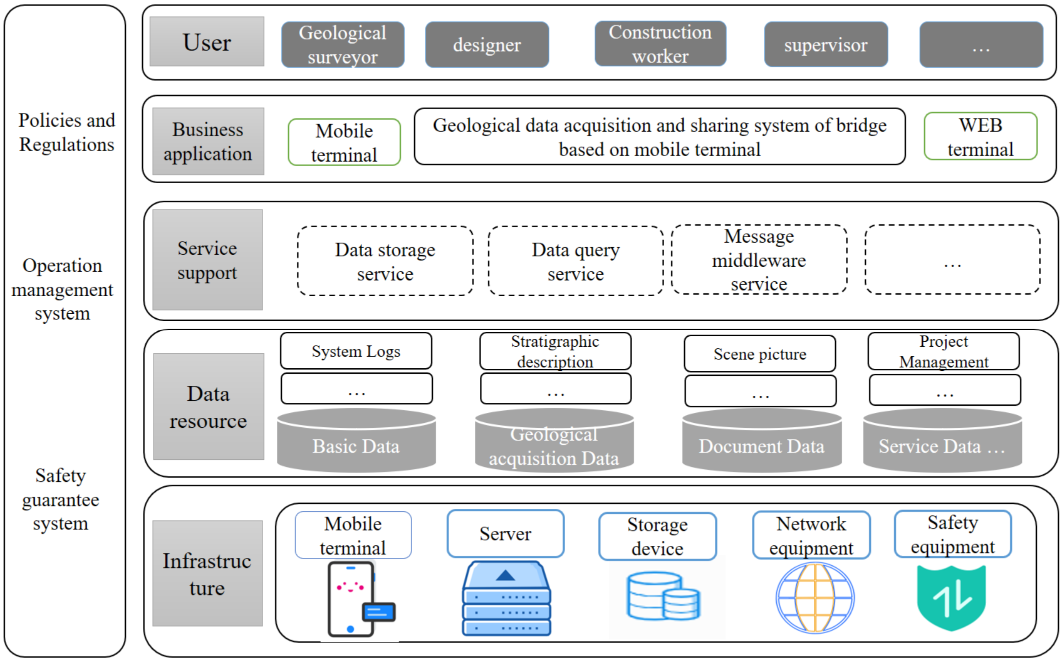


Figure 1. Figure of system architecture

2.2 Functional architecture design

The functions of the cross-sea bridge geological data collection and sharing system based on mobile terminal mainly include project management, user management, drilling management, layered management, rock and soil physical and mechanical property management, collaborative message management, file management, statistical analysis, system management and other functional modules. Normal and safe use of mobile apps requires a corresponding background system to provide services. The two complement each other and cannot be achieved without one. Mobile App functions mainly include drilling data collection and query, geological stratified data collection and query, rock and soil physical and mechanical properties collection and query, collaborative messages, statistical analysis, address book, archive data, login and logout, etc. The WEB back-end management system not only provides services to the mobile terminal, but also supports batch import and export functions such as drilling and geological stratification. The system functional architecture is shown in Figure 2.

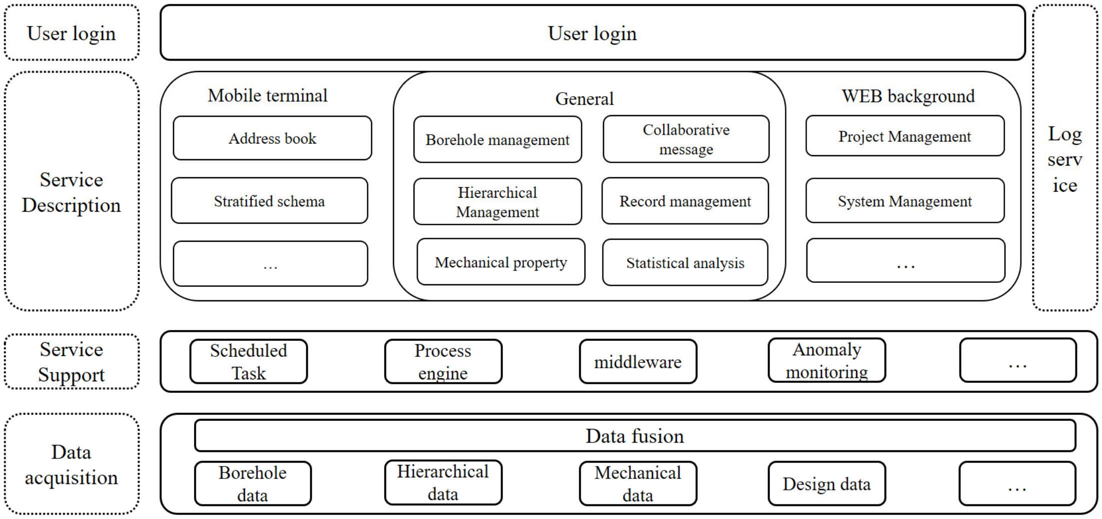


Figure 2. Figure of functional architecture design

3. SYSTEM FUNCTION REALIZATION AND APPLICATION

3.1 System implementation

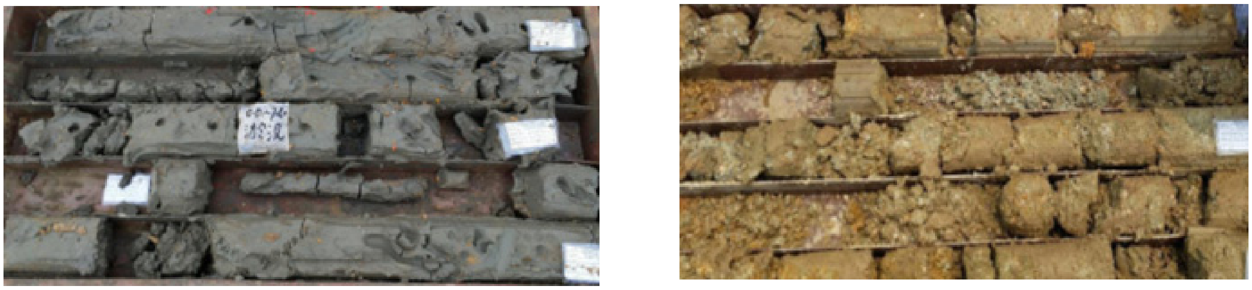
The hybrid development technology of React Native and Vue is adopted on the mobile terminal, which effectively guarantees the user experience. After the development and testing, the installation package of Android and iOS platforms can be released at the same time to realize the simultaneous operation of a set of code at both ends. Spring Boot framework of open source lightweight Java platform is selected for the development of the background system on the WEB side. Its out-of-the-box and convention are superior to configuration and strategy, which can greatly improve the development efficiency of the platform. The technical stack on the end mainly uses Bootstrap and Thymeleaf template engines to realize the development of visual layers. Spring Framework and Apache Shiro technology are integrated in the back-end to ensure the system has good scalability and stability. The geological data involved in the system is managed by Mysql database, and the persistence layer is developed by Apache MyBatis, Hibernate Validation and Alibaba Druid, and the continuous integration and delivery of the system platform is realized by Docker and Jenkins.

System security management is a very important part in the process of system operation and use, which can be subdivided into user authentication, permission management, data storage security management, etc. User identity authentication is carried out by the combination of account password, graphic verification code and SMS verification code. Role-Based Access Control (RBAC) is adopted for permission management. Different users can control data operation permissions accurately to each button according to their roles. Data storage security management is mainly based on encryption algorithm to protect the data as the core assets of the system, to prevent sensitive data from being leaked, lost, tampered or maliciously used.

The timeliness of geological data sharing requires that when geological data information changes, it can be pushed to relevant personnel in time. Thanks to the development of mobile Internet, the instant message push and convenient browsing mode of mobile terminal have been recognized by users. In this paper, the system implements accurate App push according to the user's APP account management system, and at the same time, With the help of the open development interface between wechat and Dingding, the push sharing of geological information on wechat public account, enterprise wechat and Dingding is realized. spring-boot-starter-mail is integrated in the background of the system to realize the mail sharing push of geological information, and the short message sharing push of geological information is realized based on the signature short message template of Ali Cloud.

3.2 System Application

Huangmaohai cross-sea Bridge is located in Huangmaohai sea area, across the east channel of Yamen Waterway. In the early data table name, there is no adverse geological action affecting the project safety, such as karst, debris flow, collapse, underground caves, ground collapse and ground cracks.

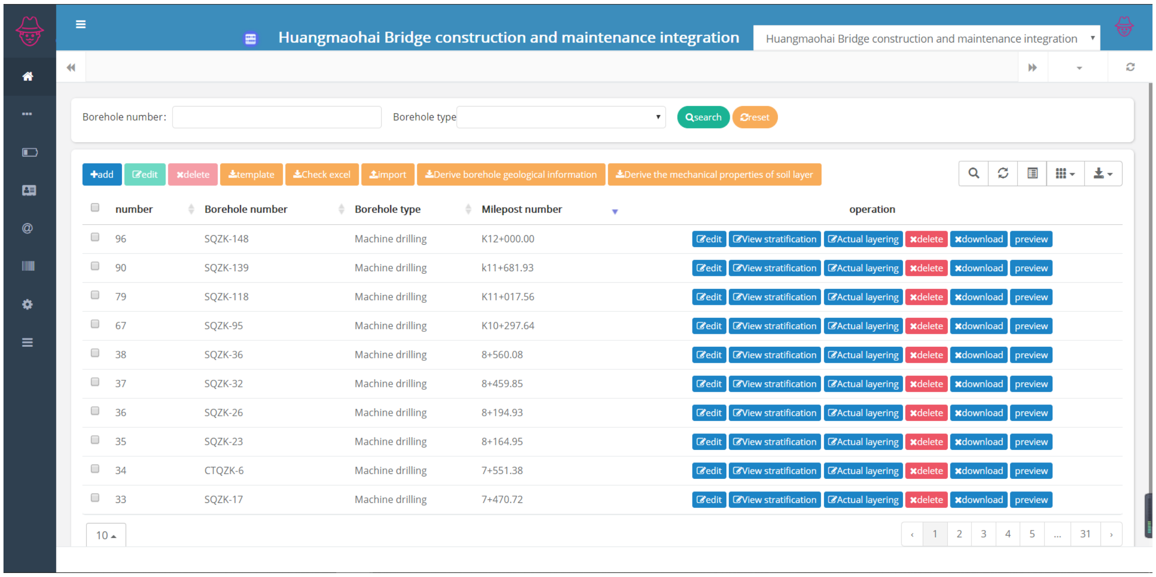


(a) Silt and silty soft soils

(b) Weathered rock

Figure 3. Representative picture of typical rock and soil of Huangmaohai Bridge

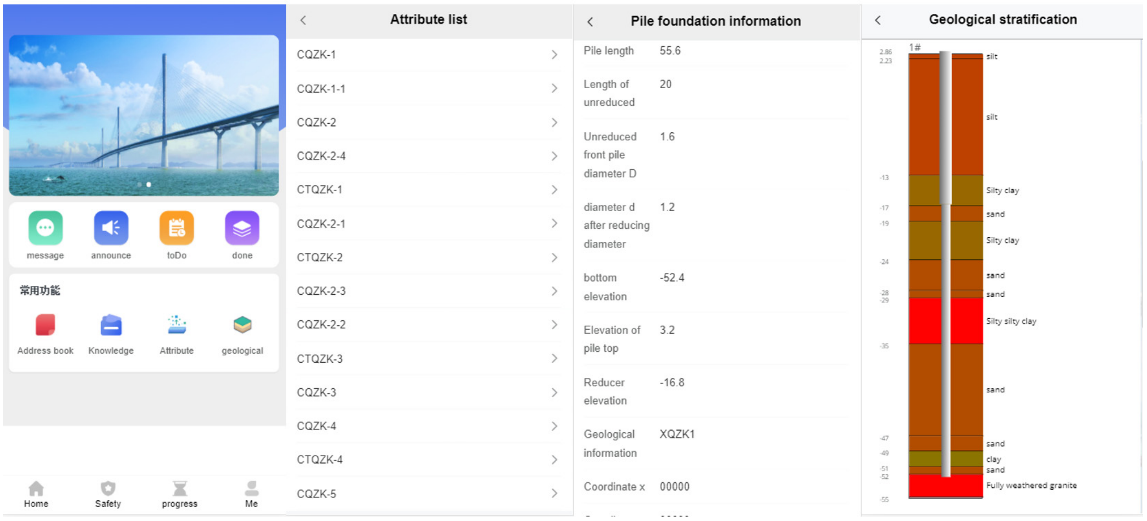
WEB side background maintenance project basic information, user information, system logs, collaborative message management, and at the same time, it can also realize the batch import and export of drilling and geological stratification, so as to form a standard geological database for the management of geological data information format. Figure 4 shows the background application effect on the WEB side.


≡
Huangmaohai Bridge construction and maintenance integration
Huangmaohai Bridge construction and maintenance integration ▾
Borehole number :
Borehole type ▾
search
reset
+add
edit
×delete
template
Check excel
import
Derive borehole geological information
Derive the mechanical properties of soil layer

□	number	Borehole number	Borehole type	Milepost number	operation
□	96	SQZK-148	Machine drilling	K12+000.00	edit View stratification Actual layering ×delete ×download preview
□	90	SQZK-139	Machine drilling	k11+681.93	edit View stratification Actual layering ×delete ×download preview
□	79	SQZK-118	Machine drilling	K11+017.56	edit View stratification Actual layering ×delete ×download preview
□	67	SQZK-95	Machine drilling	K10+297.64	edit View stratification Actual layering ×delete ×download preview
□	38	SQZK-36	Machine drilling	8+560.08	edit View stratification Actual layering ×delete ×download preview
□	37	SQZK-32	Machine drilling	8+459.85	edit View stratification Actual layering ×delete ×download preview
□	36	SQZK-26	Machine drilling	8+194.93	edit View stratification Actual layering ×delete ×download preview
□	35	SQZK-23	Machine drilling	8+164.95	edit View stratification Actual layering ×delete ×download preview
□	34	CTQZK-6	Machine drilling	7+551.38	edit View stratification Actual layering ×delete ×download preview
□	33	SQZK-17	Machine drilling	7+470.72	edit View stratification Actual layering ×delete ×download preview

Figure 4. WEB side background application effect

Field data collection and sharing are mainly based on mobile terminals. In the survey stage, users input and query geological information relying on borehole data; in the construction stage, users input and query actual geological information relying on pile foundation data. Borehole information includes borehole number, borehole type, borehole depth, commencement date, completion date, etc. The survey stratification information and the actual geological stratification information of pile foundation mainly include stratum number, stratum number, soil code, soil name, etc. The main information of rock and soil physical and mechanical properties includes age origin, standard value of lateral friction resistance, characteristic value of foundation bearing capacity, standard value of rock saturated compressive strength, standard value of rock natural compressive strength, recommended value of compression modulus, recommended value of base friction coefficient, etc. The mobile terminal can automatically generate a stratification diagram according to the geological information in the investigation and construction stage, and visually display the geological stratification state of borehole and pile foundation. After collecting or modifying geological information, users can automatically push the information to designated personnel through APP push, wechat public account, enterprise wechat, Dingdou, email, short message, etc., according to the template and personnel configuration in collaborative management, so as to improve the efficiency of geological information sharing. The application effect of mobile terminal is shown in Figure 5 below.



Attribute list

- CQZK-1
- CQZK-1-1
- CQZK-2
- CQZK-2-4
- CTQZK-1
- CQZK-2-1
- CTQZK-2
- CQZK-2-3
- CQZK-2-2
- CTQZK-3
- CQZK-3
- CQZK-4
- CTQZK-4
- CQZK-5

Pile foundation information

Pile length	55.6
Length of unreduced	20
Unreduced front pile diameter D	1.6
diameter d after reducing diameter	1.2
bottom elevation	-52.4
Elevation of pile top	3.2
Reducer elevation	-16.8
Geological information	XQZK1
Coordinate x	00000

Geological stratification

Figure 5. Mobile application effect

4. CONCLUSION

This paper analyzes the existing problems and difficulties in the collection and sharing of bridge engineering geological information, designs a geological data collection and sharing system based on mobile terminal from the perspective of data collection and sharing, describes the overall system architecture and functional architecture of the system, and develops a mobile terminal using React Native and Vue hybrid development technology. Based on the Spring Boot framework, the corresponding background management system is built, and the practical application is carried out with the Huangmaohai cross-sea Bridge as the engineering background, forming the standard geological acquisition and sharing information database of cross-sea bridge. The application results show that the system effectively improves the transmission and sharing efficiency of geological data in the process of survey, design and construction, provides a unified technical exchange platform for survey, design, construction, supervision and other majors, and provides a strong reference and guidance for the collection and sharing of geological information and intelligent management of the bridge. It is worthy of further popularization and application in bridge geological data collection and sharing projects.

REFERENCES

[1] Chen Yulong, Zhang Yuning. Research on informatization and automatic mapping technology of engineering geology [J]. Water Resources and Hydropower Technology, 2018, 49(4):9.

[2] Jia Liqiong, Long Na, Shang Yuntao. A new model of geological data service based on mobile intelligent terminal [J]. China Mining, 2018, 27(4):5.

[3] Qiang Lili. Research and Implementation of Key Technologies of Intelligent Terminal Communication Survey and Design System Based on Android [D]. Beijing University of Posts and Telecommunications,2017.

[4] Zhang H. Design and implementation of field geological data sharing and visualization system based on Android and Web [D]. Zhejiang University,2019.

[5] Tang Lianghui. Design and implementation of Database Module of Intelligent Terminal Communication Survey and Design System Based on Android [D]. Beijing University of Posts and Telecommunications,2016.

[6] Zhou Chang 'an. Construction and Empirical Study of Engineering Investigation Quality Information Management System [D]. Chongqing University,2020.

[7] Sun Yan, Li Bin, Luo Qun. Design of spatial database management system of coal geological exploration GIS [J]. Science Technology and Engineering, 2021.

[8] Shuai Xunbo, Shi Wenchang, Feng Mei, et al. Construction of a natural gas geological information sharing platform based on dynamic storage [J]. Natural Gas Industry, 2021, 41(9):9.

[9] Huang Dazhong, Wen Xiaopeng, Li Honggang. Research on data acquisition and Management System of railway geological Drilling based on mobile terminal [J/OL]. Railway standard design :1-5[2022-11-23].

Electric Power Enterprise Digital Integrated Office Platform Based on J2EE Architecture

Miaomiao Tian[a], Xianhuang Hu*[a], Xiaoyang Hu[a], Wenhui Su[a], Lei Xiao[a], Xiong He[a]
[a]Wuling Power Corporation Ltd., Changsha, Hunan, 410004, China
* Corresponding author: hu_hxh@wldl.com.cn

ABSTRACT

The improvement and upgrade of the electric power enterprise integrated management system are faster and faster. In the selection of the system architecture, the selected architecture is often analyzed according to the actual needs, which is to establish the advantages of the system architecture and reduce the cost of subsequent development and upgrade. With the increasing workload of operation and maintenance of more and more terminal equipment and network management day by day, the defects of an insufficient number of operation and maintenance personnel and backward management means will become more prominent. This paper applies the B/S interactive structure and the MVC three-tier architecture. The development platform is based on J2EE. SQL Server is applied as the database server, and UML modeling tools are adopted in the modeling process. The system before design was investigated by a detailed survey of the user's needs. The implementation of the system effectively improves the efficiency of its infrastructure project management, making the company's infrastructure management more scientific and reducing the business operating costs.

Keywords: Electric power enterprise, Management system, System architecture, Analytic hierarchy process, J2EE

1. INTRODUCTION

The actual needs of the digitally integrated office platform of the electric power enterprises and the mastery of the operation habits of front-line operation and maintenance operators, as well as the mastery of the actual situation of the continuous growth of equipment in electric power companies, that is, it is necessary to design an integrated management system that can meet both the needs of personnel and the needs of company development. It helps realize the process of the internal operation of the power company, and the intensive management of human and financial resources. It also gives full play to the advantages of information technology to bring convenience to the personnel engaged in related business [1].

Comparing the most widely used B/S and C/S architectures, namely browser-server and client-server, and the widely used client-server model before, the products of browser-server architecture transfer data processing to the rear server. It does not need users to install client software. The system can be accessed and operated only through the browser, which obviously reflects the more convenient features [2]. In addition, because of the centralized processing by the server, the client (browser) of B/S architecture does not need to upgrade the system. Only the server needs to update, which reduces the tedious work of the upgrade process and greatly reduces the maintenance cost in the later period [3].

Literature[4] mainly studies the development of power enterprise management, the design principles, objectives, and information generation of computer management information systems for large-scale water conservancy projects, as well as the development and application of project management software in this field through the study of power grid companies. According to the above discussion, we can find that China has not yet formed a complete system in the field of electric power enterprise management information system. The research field of the WEB-based electric power enterprise management information system is relatively blank, but this research field is bound to be a mainstream trend in the process of electric power information construction. Literature [5] designs a set of digital office platforms for an electric power company in a certain area. Based on the actual business needs of the unit, the system uses B/S architecture and ASP. NET technology, and mainly realizes the functions of procurement process management, contract management and project schedule management. The application of the system has played a great role in improving the management level of the unit's power supply project.

The main advantage of B/S architecture is that the client it serves is a browser, which only needs to manage the server, so that the digitally integrated office platform of power enterprises can run stably and safely. Moreover, the front end of B/S architecture is a server, which is in line with the user's computer operation habits and easy to maintain; the technical

International Conference on Computer Graphics, Artificial Intelligence, and Data Processing (ICCAID 2022)
edited by Ruishi Liang, Jing Wang, Proc. of SPIE Vol. 12604, 126041S
© 2023 SPIE · 0277-786X · doi: 10.1117/12.2674632

expansion of the system is also relatively easy, which can leave ample space for the future development of the power grid [6].

Through the J2EE development platform, SQL 2017 database and B/S three-tier structure, this paper can provide users with a digital integrated office platform for power enterprises, which integrates all kinds of resources. Users only need to log in through the front-end browser to view the power grid operation data in the background of the system, and can also receive the alarm from the system and carry out related operations. Moreover, the cost required for the development of the system is low, and the system is convenient for the subsequent system upgrade and data volume increase caused by the expansion of the power system's scale in the future.

2. ANALYSIS OF THE ADMINISTRATIVE LEVELS OF MANAGEMENT STRUCTURE IN ELECTRIC POWER ENTERPRISES

2.1 The analytic hierarchy process

The process of using the fuzzy analytic hierarchy process is displayed in Figure 1:

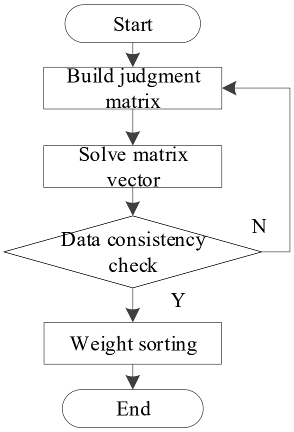


Figure 1 Fuzzy analytic hierarchy process

(1) Establish a perfect and scientific index model aiming at the whole index system and construct a fuzzy judgment matrix;

(2) Calculate and analyze the fuzzy judgment matrix to solve a ranking vector of the fuzzy judgment matrix;

(3) Check the overall consistency;

(4) Sort the weight of each element [8].

2.2 Management structure indicator weights

The basic steps of calculating index weight by the fuzzy analytic hierarchy process are as follows:

The construction of the fuzzy consistent judgment matrix can directly show the matrix of fuzzy relationship, which is a general term for matrices with similar characteristics [8]. Assuming that the calculation method of the binary comparison matrix ϕ is, if $\phi_i + \phi_j = 1$ in the formula, then it can be considered that there is a complementary relationship between ϕ and the fuzzy matrix. Where, ϕ_{ij} means the membership degree of the index α_j compared to the index R [9].

If $\phi_{ij} = 0.5$, it means that α_i and α_j are equally important;

If $\phi_{ij} < 0.5$, then it means that α_j is more important than α_i. If ϕ_{ij} is smaller, then the relative importance of α_i is higher;

If $\phi_{ij} > 0.5$, then it means that α_i is more important than α_j . If ϕ_{ij} is larger, then the relative importance of α_i is higher.

Then , we should solve the priority vector θ of the fuzzy judgment matrix. Let $\phi = \left(\phi_{ij}\right)_{\max}$ be the fuzzy complementary judgment matrix, and take the following ones:

$$\eta = \left(\theta_{ij}\right)\sum_{i=1}^{n}\phi_{ij} - 0.5 \tag{1}$$

$$\varpi_{ij} = \left(\phi_{ij} - \eta\right)^{\max} \tag{2}$$

Where, η is the fuzzy consistent matrix corresponding to matrix ϕ , and the deviation matrix of matrix ϕ is ϖ .

The consistency check is carried out for the fuzzy complementary judgment matrix ϕ to determine $\varpi \in 0$: if $\eta < 0$, the consistency of the matrix can reach a good level. If the consistency level fails to meet the standard, it needs to be corrected. In reality, the value of ϖ should be selected based on the actual situation. The smaller the value selected, the higher the ability of bid evaluation experts is required [10].

The sorting formula is as follows:

$$\rho_{ij} = \frac{1}{\alpha^2 + \beta^2 + n}\sum_{i=1}^{n}\phi_{ij} \tag{3}$$

Where, $\rho = \left(i_1, i_2, \cdots, i_n\right)$ is the sorting vector of ρ .

3. TEST ANALYSIS

3.1 Test platform

The digital integrated office platform of electric power enterprises adopts the simulation environment, and the settings are presented in Table 1.

Table 1 Simulation platform configuration

Settings	Server	User port
Processor	SPARC，64 kernel	Intel i5，3.5GHz
Memory	1TB	16G
Hardware	8TSAS	1TB SSD
System	OracleSolari 10	64 bits windows

The realization of the system function module of the electric power enterprise digital integrated office platform can be quickly repaired when the equipment has problems. The implementation process of this system needs to adopt the mature information and science and technology in the world to build its information management structure. It is to lay a solid foundation for successfully replacing the manual processing mode in the power dispatching and monitoring industry and improving the management efficiency after it is put into use [11]. Its main functions also include data communication with other departments in the power system, such as the transmission work area, substation work area, marketing department, etc., which can realize stable and reliable business data flow and instruction transmission. It can also strengthen the initiative of fault handling, and record the processing process and results in the system to promote the follow-up process. A large number of power company system data stored in its server, especially the data of power dispatching monitoring and management, will play a huge role in promoting business development and industry development. When the amount of data reaches a certain level, the data mining technology can be adopted to quickly search the internal data of the system. The analysis of common data can be of great help in the study of system problems [12].

3.2 Analysis of test results

In the actual application of the electric power enterprises' digital integrated office platform, at least 300 users should be called. However, in the actual test, the feasibility of this test method is relatively low. In view of this situation, Load runner software is applied to complete the performance test. This software is a simulation test. After the software test is completed, the test results are displayed in Table 2.

Table 2 Performance test

The number of users	100	250	500
Response time/ms	22	35	51
CPU utilization rate /%	4.5	7.9	11.2

The application performance test based on the J2EE technology is displayed in Figure 2. 500 users are simulated, and the performance of the system is presented in the case of simultaneous use.

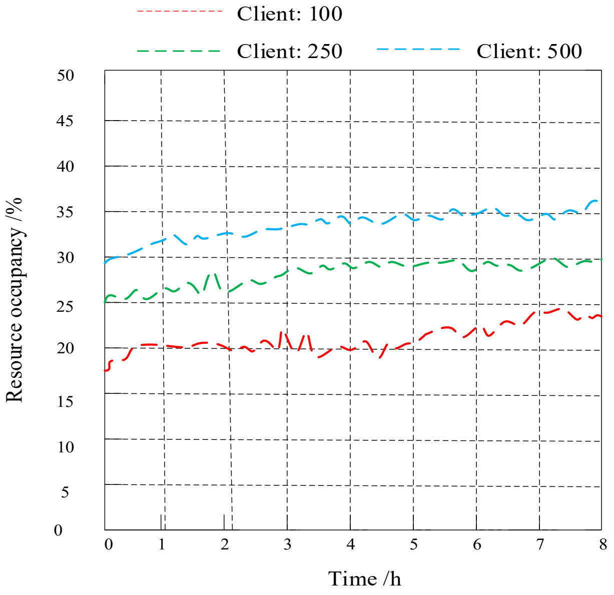


Figure 2 Use performance test analysis

The test time is 8 hours. The test must be the frequency of reading and writing operations of the server. After the test, the resource occupancy rate of the system is maintained at 17-37%, not more than 40%. The results indicate that the system can still run stably under the concurrent use of 500 people.

4. CONCLUSION

In this paper, a digitally integrated office platform is established for electric power enterprises. The design and development of the system apply the J2EE architecture and the MVC design pattern. Firstly, the fuzzy hierarchy analysis is applied to analyze the management architecture of power enterprises, and the weight calculation is carried out according to the demand indicators. Finally, the platform design is completed. Through the experimental analysis, the system simulates the long-term test of different user orders in the simulation environment, which proves that the system can meet the common office of more than 500 users. The system resource occupation is kept at a low level, which can meet the actual needs and play a key role in the comprehensive management information construction of the electric power enterprises.

The use of the database also needs to be further optimized. The system designed in this subject adopts the current more advanced database, and uses data mining technology, which can realize the comprehensive management of huge amounts of data, and also enable the power dispatching and monitoring personnel to obtain the desired data steadily. However, the system must be able to automatically clean the data because of the continuity of the work of the power dispatching system and the wide range of business. This kind of cleaning is similar to anti-virus software that cleans the registry and can achieve the same cleaning of computer programs. It intends to ensure that the digitally integrated office platform of power enterprises can serve power grid management for a long time.

REFERENCES

[1] Melamed, Dan, Skokan, Bryan, Zenkowich, Mathew, Kocher, Dan. Estimating Uncertainty for aLife-Cycle Cost Management Program[J]. Cost Engineering. 2018 (11):51-62

[2] Srinivas Medida. IEEE Trans on Power System. Krishna V Prased. SCADA-EMS on theInternet.. 2018.37[46] Qiu Bran. Web-based SCADA Display Systems for Access via Internet. IEEE Transactions on Power Systems. 2018.27-28

[3] Aichun Zhang, Weikang Lu. Research on Computer Innovation Development Strategy in Internet+ Age[A]. Proceedings of 2018 4th International Conference on Mechatronic Systems and Materials Application (ICMSMA 2018) [C]. 2018,28

[4] Hu Yu. Application Status and Prospect of Power Dispatching Automation System[J]. Science and Technology Innovation, 2019 (5): 152-153

[5] Liu Ying Design and Implementation of Integrated Monitoring System for Power Communication Network [J] Building Materials and Decoration, 2018 (21): 296

[6] Guangzejing Research on optimization and improvement of power communication network integrated resource management system [J] Software, 2020, 41 (01): 189-191. 2018 (05)

[7] Li Min Design and Implementation of Power Communication Network Integrated Management System [J] Electronic Design Engineering, 2018, 26 (15): 103-107

[8] Yang Linhui Design and application of data acquisition framework for power communication network management [J] Automation and Instrumentation, 2019 (01): 167-170

[9] Chen Yuan Design and Implementation of Integrated Monitoring System for Power Communication Network [J] Electronic Technology and Software Engineering, 2018 (04): 29

[10] Huangfu Dashuang Research and Application of Sub query in SQL Server Database [J] Computer knowledge and technology 2020(28):196-197

[11] Han Jiying Design and Implementation of Data Visualization Based on JAVA [J] Scientific and technological innovation 2020(32):75-76

[12] Yang Zhihao Solution to the problem of page jump garbled code in Java Web programming [J] Electronic production 2020(20):321-323

Analysis and Research on students' classroom behavior data based on GCN

Minchao Ban[a], Mingwei Tang[a,*], Jie Zhou[a], MingFeng Zhao [b,*], Zhiming Xiao[a]
[a]School of Computer and Software Engineering, Xihua University, Chengdu China.
[b]China Mobile Group Design Institute Co., Ltd. Sichuan Branch, Chengdu 610045, China.
bmc@stu.xhu.edu.cn, tang4415@126.com, jiezhou_xhu@163.com, xzm@stu.xhu.edu.cn
*corresponding author

Abstract

The analysis and study of students' classroom behavior can help develop students' abilities and improve teachers' teaching, and has been one of the key issues closely followed by the education community. In recent years, graph convolutional neural networks (GCN) have been widely used in various fields with outstanding success. Therefore, this paper proposes a GCN-based approach for modeling and analyzing student classroom behavior data. The experimental results show that the method helps to develop students' ability and improve teachers' teaching level.

Keywords-component; Graph convolutional neural networks; Students' classroom behavior; Deep learning model

1. Introduction

Currently, many deep learning models can directly extract features from datasets, and these models are widely used by everyone. Model classification is facilitated by the presence of most of these features in the input data. There is also a significant role for deep learning models in natural language processing [1-3], scene text recognition [4], action recognition [5], image classification [6], audio classification [7], etc. The computer vision community now heavily relies on Convolutional Neural Networks (CNNs) [8]. The convolutional neural network is used to recognize student behavior based on preprocessed data. A classification algorithm is used to classify student behavior based on features extracted from the network. In order to recognize student behavior, it is imperative to preprocess the data. The importance of recognizing facial information should be higher than general student behavior. Due to the fact that these preprocessing steps cannot be automated in practical applications, preprocessing is an integral part of the research process for student behavior recognition technology.

Student classroom behavior reflects the state and learning efficiency of students and is closely related to the quality of classroom instruction. Teaching quality in turn reflects the level of educational quality of a school. Improving the quality of education is not only a critical issue within funny, but also an important issue of common concern to the government, the public and other stakeholders. Therefore, student classroom behavior is one of the important factors affecting teaching quality and is one of the hot topics receiving common attention from all walks of life.

For a long time, the evaluation of the quality of higher education has mainly focused on the input of resources, relatively ignoring the "added value" of students after entering university, i.e. the academic performance of college students. There are three mainstream evaluation views in the practice of higher education quality evaluation: reputation view, resource view and output view. The traditional evaluation view is that the quality of higher education is determined by external factors such as reputation and resources of universities. The evaluation criteria of higher education mainly focus on the input and supply of educational resources, such as the strength of teachers, the investment of educational funds, the number of courses offered, the area of school buildings, library collections, etc. However, while the development of students' cognition, skills and attitudes is one of the core criteria to measure the level of quality of education [9]. This is because the feedback data collected on student topics through big data analysis can serve to improve the quality of teaching and learning. Students' classroom behavior is also one of the most important factors affecting the quality of teaching and learning; therefore, this paper analyzes and proposes countermeasures for students' classroom behavior.

Classroom behavior management training facilitates teachers' classroom management practices and promotes improved social and academic achievement of students. Improvements in academic achievement can be explained to some extent by increases in the amount of time students spend on individual behavioral tasks in the classroom. Improving teachers' classroom management is expected to significantly improve student achievement. Therefore, the data we collected on

International Conference on Computer Graphics, Artificial Intelligence, and Data Processing (ICCAID 2022)
edited by Ruishi Liang, Jing Wang, Proc. of SPIE Vol. 12604, 126041T
© 2023 SPIE · 0277-786X · doi: 10.1117/12.2674596

student classroom behaviors were further modeled using GCN. The analysis of the results can provide directions for developing students' abilities and improving teachers' teaching to better serve the field of pedagogy.

In recent years, machine learning has been rapidly developed in various fields [21-25]. Deep learning, as one of the core technologies supporting the development, has made great achievements in both academic and industrial fields. Many scholars have used deep learning techniques to model and analyze various types of data and textual information, and have achieved better results. However, the complexity of students' classroom behaviors makes it a bottleneck in analyzing academic classroom behavioral data information with the help of deep learning techniques.

There are many methods to analyze student classroom behavior data. Among them, traditional methods based on machine learning mainly use handcrafted bags of words, dictionaries, and other features to train classifiers (e.g., support vector machines [10]). Although these methods achieve better performance, they require a lot of manual engineering processes, which can lead to loss of accuracy and greatly affect the performance of data analysis results. Compared to traditional methods, neural networks have gained widespread attention as they can automatically learn data features.

This year, neural network-based models have shown good performance in processing individual student behavioral data and are widely used in classification tasks based on individual student behavioral data. For example, recurrent neural networks (RNN) [11], long short-term memory (LSTM) [12], and gated recurrent units (GRU) [13] to learn word embedding vectors of student classroom behavioral data sequences and then model them by attention mechanism to generate corresponding weighted scores and obtain hidden features of the corresponding sequences. However, these methods are time-consuming and parallel operations are not convenient to implement. Worse still, they are prone to gradient disappearance or explosion during model training. A gated convolutional neural network (CNN) with a special gating mechanism is then designed to solve the parallel operation problem [14]. Although CNN-based models enhance students' behavioral learning in the classroom, they do not handle long-distance relational and graph data information well in most cases, and the relationships between individuals are not effectively learned. Graph convolutional neural networks (GCNs) [15-19] are used in various fields because they can better process graph result data and learn feature information between data sequences. Therefore, this paper proposes a GCN-based model for analyzing students' classroom behavior.

2. RELATED WORK AND THEORY

2.1 Embedding layer

The word embedding layer maps each piece of data w_i in a given sentence student subject behavior dataset into a low-dimensional real-valued vector space, and we use the pre-trained language model $BERT_{(base)}$ [20] to obtain a fixed word embedding for each piece of data (Word embedding), Bert has been widely used in NLP in recent years, and the algorithm has good feature representation capability in word embedding due to pre-training on a large corpus. For each input data, we splice it into a sequence form like $\left[[CLS], S, [SEP]\right]$, where$[CLS]$ is a special token located in front of two segments and $[SEP]$ is a special symbol used to split two different sentences, and then transform it into a vector X as input. After processing by BERT, we can obtain a set of word embedding vectors $h_x \in \mathbb{R}^{d_w}$, where $d_w$is the dimension of the word embedding.

$$h_x = BERT(X)$$

2.2 GCN layer

Suppose the undirected graph $\mathcal{G}(S) = (V, E)$, where N $v_i \in V$ nodes, edges $\left(v_i, v_j\right) \in E$, adjacency matrix $A \in R^{(N\times N)}$ and a degree matrix $D_{ii} = \sum_j A_{ij}$, each node has its own features, and the features of these nodes form a matrix $X \in R^{(N\times D)}$. In this paper, we first convert the student classroom behavior data information into graph structure data information, and then apply GCN to extract features. In which, the GCN is propagated between layers as follows:

$$H^{l+1} = \sigma\left(\widetilde{D}^{-\frac{1}{2}}\tilde{A}\widetilde{D}^{-\frac{1}{2}}H^l W^l\right)$$

Here, $\tilde{A} = A + I_N$ is the adjacency matrix of the undirected graph G with self-connections added. I_N is the unit matrix$\widetilde{D}_{ii} = \sum_j A_{ij}$ and W^l is a layer-specific trainable weight matrix. $\sigma(.)$ denotes the activation function. $H^{(l)} \in R^{N\times D}$ denotes the output features of the lth layer, $H^{(0)} = X$.

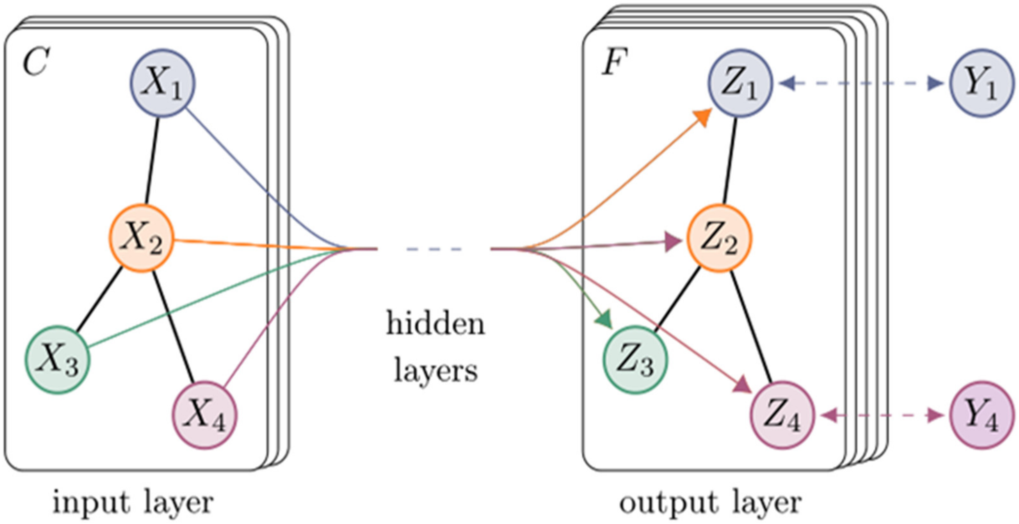


Figure 1. Graph convolutional neural networks Framework

The graphical convolutional neural network is shown in Figure 1. Through several layers of the GCN, the features of the nodes in the input layer aggregate information from neighboring nodes and the state changes from X to Z. Finally, the hidden features H required for the model are obtained.

2.3 Output layer

The hidden representation H of student subject behavior data features is obtained from the GCN layer as the final output for classification, and then, input to a dense layer, and a softmax function is used to predict the student subject behavior analysis, as follows.

$$\bar{y} = softmax(WH + b)$$

Here $\bar{y}$ is the final student subject behavior analysis result, $W \in \mathbb{R}^{|C| \times d_h}$ and $b \in \mathbb{R}^{|C|}$are the weight matrix and bias obtained in training, and $|C|$denotes the number of classifications.

2.4 Model training

In the model training process, we use the loss function is used to calculate the degree of deviation between the true and predicted values of the target value Y in the test set, and train the optimization parameters using the optimizer to minimize the loss function. The loss function consists of cross-entropy loss and L2 regularization and is defined as shown below.

$$loss = -\sum_{i \in D} y_i \log(\bar{y}_i) + \lambda ||\Theta||^2$$

Here D denotes the training dataset, y_i and $\bar{y}_i$ denote the true sentiment polarity and predicted sentiment polarity of aspect words, respectively. λ denotes the coefficient of L2 regularization, and Θ denotes all the parameters used. Besides, we use the dropout strategy to avoid overfitting during the model training process.

3. EXPERIMENTAL SETTINGS

In our experiments, we use the pre-trained model BERT-base English version to initialize the word embeddings separately. The initial values of all weight matrices are sampled from the uniform distribution $U(-0.01,0.01)$, and all offset values are set to 0. The dimensionality of the hidden representation is set to 300, and the GCN layer is set to 2. The batch size is set to 64, and the maximum length of the input data is 85. We set the dropout rate to 0.1 to prevent overfitting.

4. EXPERIMENTAL RESULTS AND DATA ANALYSIS

The experimental results are shown in Figure 2. In this experiment, the best performance of the experimental results was obtained when the number of layers of GCN was set to 2. The information from student classroom behavior data was modeled and used for prediction. The results show that the analysis and study of students' classroom behavior can determine the attitude of learning in class, which can help to improve students' performance and guide teachers' teaching style. This suggests that improving student classroom behavior can help improve student achievement and teacher teaching and research.

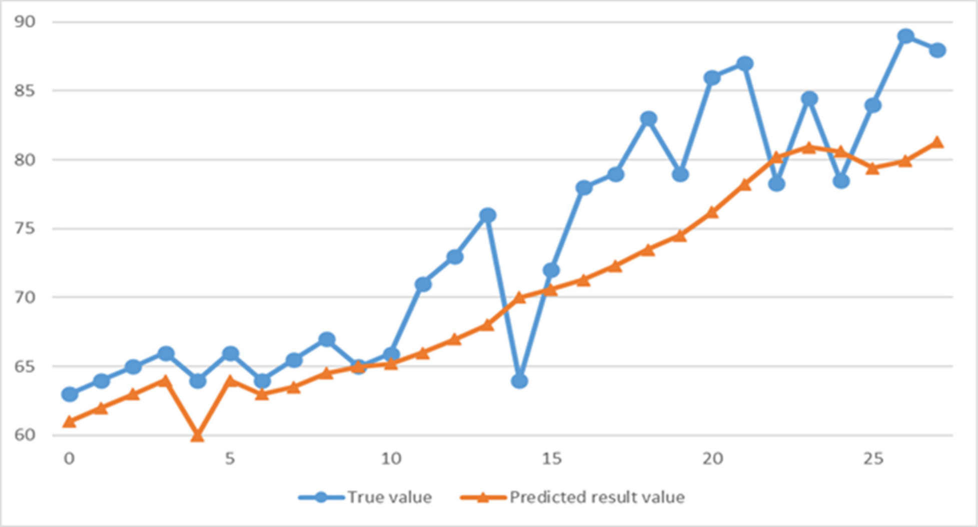


Figure 2. GCN Experiment’s results

5. CONCLUSION

The paper proposes an analysis and research on students' classroom behavior data based on GCN. It has been demonstrated that the proposed model performs extremely well in experiments.

The analysis and study of information on student behavior in the classroom is very important for the educational community. This contributes to the development of student competencies and the improvement of teacher instruction. Of course, research on individual student behavior data outside of the classroom may also have important implications for classroom rationale and experimentation. This will be the next step of our research.

ACKNOWLEDGMENT

This work is supported by the Scientic Research Funds project of Science and Technology Department of Sichuan Province (No. 2019YFG0508,2019GFW131, 2023**), the National Natural Science Foundation of China (No. 61902324), Funds Project of Chengdu Science and Technology Bureau (No.2017-RK00-00026-ZF,) and Science, Technology Planning Project of Guizhou Province (No. QianKeHeJiChu-ZK[2021]YiBan319), the Xihua University Education and teaching reform project (No: xjjg2021049, xjjg2021115) and the Sichuan Province's Higher Education Talent Training and Teaching Reform Project for 2021-2023 (No: JG2021-929).

REFERENCES

[1] Zhao Ziguo, Tang Mingwei, Tang Wei; Wang Chunhao, Chen Xiaoliang. Graph convolutional network with multiple weight mechanisms for aspect-based sentiment analysis. NEUROCOMPUTING, 2022, 06:1-16.

[2] Wang, Xiaodi, Tang, Mingwei, Yang, Tian, Wang Zhen. A novel network with multiple attention mechanisms for aspect-level sentiment analysis. KNOWLEDGE-BASED SYSTEMS, 2021, 08:1-16.

[3] LITMAN R, ANSCHEL O, TSIPER S, et al. Scatter: selective context attentional scene text recognizer[C]. proceedings of the IEEE/CVF conference on computer vision and pattern recognition, Seattle, WA, USA, 2020: 11962-11972.

[4] WANG Y, XIE H, ZHA Z J, et al. Contournet: Taking a further step toward accurate arbitrary-shaped scene text detection[C]. proceedings of the IEEE/CVF conference on computer vision and pattern recognition, Seattle, WA, USA, 2020: 11753-11762.

[5] LI Y, JI B, SHI X, et al. Tea: Temporal excitation and aggregation for action recognition[C]. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Seattle, WA, USA, 2020: 909-918.

[6] MUNRO J, DAMEN D. Multi-modal domain adaptation for fine-grained action recognition[C]. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Seattle, WA, USA, 2020: 122-132.

[7] HERSHEY S, CHAUDHURI S, ELLIS D, et al. CNN architectures for large-scale audio classification[C]. International Conference on Acoustics, Speech and Signal Processing (ICASSP), New Orleans, LA, USA, 2017:131-135.

[8] LECUN Y, BOSER B, DENKER J S, et al. Backpropagation applied to handwritten zip code recognition[J]. Neural computation, 1989, 1(4): 541-551.

[9] Liu WB. The Study of the Causes and Countermeasures of Classroom Problem Behavior Producing in Colleges or Universities. 2016 4TH INTERNATIO NAL CONFERENCE ON SOCIAL SCIENCES RESEARCH (SSR 2016).

[10] Herman KC, Reinke WM, Dong N, Bradshaw CP. Can Effective Classroom Behavior Management Increase Student Achievement in Middle School? OURNAL OF EDUCATIONAL PSYCHOLOGY, 2022, 114(1): 144-160.

[11] Thorsten Joachims. Transductive inference for text classification using support vector machines. In ICML '99 Proceedings of the Sixteenth International Conference on Machine Learning, pages 200{209, 1999.

[12] G. Arevian, Recurrent neural networks for robust real-world text classification, in: IEEE/WIC/ACM International Conference on Web Intelligence (WI'07), IEEE, 2007, pp. 326–329.

[13] S. Hochreiter, J. Schmidhuber, Long short-term memory, Neural computation 9 (1997) 1735–1780.

[14] J. Chung, C. Gulcehre, K. Cho, Y. Bengio, Empirical evaluation of gated recurrent neural networks on sequence modeling, in: NIPS 2014 Workshop on Deep Learning, December 2014, 2014.

[15] W. Xue, T. Li, Aspect based sentiment analysis with gated convolutional networks, in: Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers), 2018, pp. 2514–2523.

[16] T. N. Kipf, M. Welling, Semi-supervised classification with graph convolutional networks.

[17] C. Zhang, Q. Li, D. Song, Aspect-based sentiment classification with aspect-specific graph convolutional networks, in: EMNLP/IJCNLP (1), 2019 .

[18] B. Huang, K. M. Carley, Syntax-aware aspect level sentiment classification with graph attention networks, in: Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP), 2019, pp. 5469–5477.

[19] K. Wang, W. Shen, Y. Yang, X. Quan, R. Wang, Relational graph attention network for aspect-based sentiment analysis, in: Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics, 2020, pp. 3229–3238.7.

[20] Ke W, Gao J, Shen H, et al. Incorporating explicit syntactic dependency for aspect level sentiment classification[J]. Neurocomputing, 2021, 456: 394-406.

[21] Kenton J D M W C, Toutanova L K. BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding[C]//Proceedings of NAACL-HLT. 2019: 4171-4186.

[22] LeCun Y, Bengio Y, Hinton G. Deep learning[J]. nature, 2015, 521(7553): 436-444.

[23] Hu M, Liu B. Mining and summarizing customer reviews[C]//Proceedings of the tenth ACM SIGKDD international conference on Knowledge discovery and data mining. 2004: 168-177.

[24] Vaswani A, Shazeer N, Parmar N, et al. Attention is all you need[J]. Advances in neural information processing systems, 2017, 30.

[25] Guo D, Ren S, Lu S, et al. GraphCodeBERT: Pre-training Code Representations with Data Flow[C]//ICLR. 2021.

Prediction Method of Electronic Contract Performance Risk Based on Decision Tree Model

Hui Xiang[1]*, Yan Lu[1], Jing Zhang[1], Hongyu Wang[1], Junying Wu[2]
[1]State Grid Hebei Procurement Company, Shijiazhuang 050000, China
[2]Information and Communication Branch of State Grid Hebei Electric Power Co., Ltd, Shijiazhuang 050000, China
*hsiangd@163.com

Abstract

In the Internet environment, an electronic contract is a necessary transaction condition for business cooperation between enterprises. To avoid the risk factors in enterprise electronic contracts and promote the smooth progress of business cooperation, this paper researches the prediction method of enterprise electronic contract performance risk based on the decision tree model. On the basis of the decision tree model, the key nodes are pruned, and then the validity parameters are combined to complete the collection and analysis of enterprise electronic contract risk data. According to the distribution characteristics of risk factors, the performance reliability conditions are improved. The prediction of electronic contract performance risk is realized by solving the value range of prediction parameters. The experimental results indicate that under the decision tree model, the maximum accuracy of risk vector extraction in the process of electronic contract performance reaches 97. 5%, which has outstanding application value in avoiding risk factors and promoting business cooperation between enterprises.

Keywords: Decision tree model, Electronic contract, Performance risk, Validity parameter, Risk factor, Business cooperation

1 INTRODUCTION

A Decision tree is a tree-processing structure, in which each node represents an independent information sample. The arrangement of leaf nodes will not change during the implementation of the test.

Decision tree representation is one of the most widely used logic methods, which infers classification rules in the form of decision tree representation from a set of disorderly and irregular cases. The decision tree classification method adopts a top-down recursion mode. The attribute values are compared at the internal nodes of the decision tree. The downward branches from the nodes are judged according to different attribute values, and the conclusion is obtained at the leaf nodes of the decision tree. A path from the root of a decision tree to a leaf node corresponds to a conjunction rule, and the entire decision tree corresponds to a set of disjunctive expression rules. Decision tree is a machine learning method based on information theory. It is a graphical representation of various alternatives, and each alternative or event may lead to two or more events, leading to different results. Because the lines connecting the decision points of the analysis of various alternatives are like a fallen tree, it is called decision tree analysis.In classification analysis, the decision tree model is the most popular model. The main reason is that it is very convenient to show the results of mining in a graphical way (tree structure), and it is suitable for enterprise management departments to make decisions.

There are three reasons why decision tree technology is the main technology of classification and prediction in the field of data mining. First, the classifier constructed by decision tree is easy to understand; second, the speed of decision tree classification is faster than other classification methods; third, the classification accuracy of decision tree classification method is better than other methods.

Electronic contract refers to the agreement between the two parties reached by means of data messages. The main carriers of electronic contract are electronic data interchange (EDI) and electronic mail (EMAIL). According to the definition of the United Nations Model Law on Electronic Commerce, the former refers to the information structure standard based on the agreement to realize the electronic transmission of information from computer to computer. The latter refers to the transmission of information between computer terminals through Internet service providers. It can be seen that both of them transmit information through the network, but the compilation of EDI should be based on the standardized message form established through prior consultation, while e-mail has no standardized form. Electronic

International Conference on Computer Graphics, Artificial Intelligence, and Data Processing (ICCAID 2022)
edited by Ruishi Liang, Jing Wang, Proc. of SPIE Vol. 12604, 126041U
© 2023 SPIE · 0277-786X · doi: 10.1117/12.2674621

contract is the E-mail form of paper contract, which uses electronic pulse to transmit the transaction information needed by the associated cooperative enterprises, so that the parties can provide corresponding services according to the content of the contract text. Because the electronic contract of enterprises has a complete closed-loop chain of evidence, the electronic contract has the same legal effect as the paper contract in the process of completing business cooperation between enterprises.

Because the two parties of the electronic contract are in different places, one party can not automatically get the other party's acceptance and confirmation while using the electronic contract to make the contract behavior, so compared with the ordinary contract form, it is easy to form greater commercial risks. As an inevitable product of the Internet era, there are many risk factors in the performance of electronic contracts, such as high-risk transactions. How to accurately extract the risk vector while promoting the smooth progress of business cooperation between enterprises has become an urgent problem to be solved. To better cope with the above situation, this paper puts forward a method for predicting the performance risk of electronic contracts based on the decision tree model.

2 RISK DATA COLLECTION AND ANALYSIS OF ENTERPRISE ELECTRONIC CONTRACT

2.1 Decision tree model

The decision tree model is responsible for screening the risk data in the electronic contract of the enterprise, and can select the results according to the decision characteristics, and arrange the layout of tree nodes and branch nodes, so as to eliminate the information samples with obvious interference while maintaining the stable business cooperation relationship between enterprises [5]. A complete decision tree model must include three executive links: decision feature selection, tree node selection and branch node selection. The so-called decision feature refers to the processing principle followed by removing the risk data in the enterprise electronic contract. The tree node is the initial input position of the data information, and the branch node is the main place for information filtering. The specific execution process is presented in Figure 1.

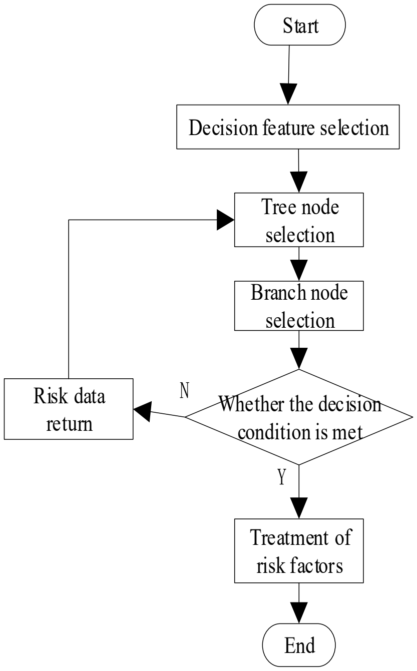

Start
Decision feature selection
Tree node selection
Branch node selection
Whether the decision condition is met
N
Risk data return
Y
Treatment of risk factors
End

Figure 1 Decision tree execution flow

When the selected contract data does not meet the basic decision conditions, these information samples will be returned to the tree node position again. In the current situation, the branch node filters and screens these data again until the information output by the decision tree model can promote the smooth progress of business cooperation between enterprises [7].

2.2 Node pruning

Decision tree technology is to obtain the data classification and prediction rules under different values of input variables and output variables through learning, and to use them to classify and predict the test set. It mainly consists of two stages. The first stage is the construction stage. The training data set is selected for learning, and the decision tree is derived. The second stage is the pruning stage. The test data set is used to test the decision tree. If the established decision tree can not answer the question correctly, we need to prune the decision tree to solve the problem of over-adaptation to the data until a correct decision tree is established. The purpose of pruning is to reduce the fluctuation due to the noise of the training set.

Decision tree pruning is proposed to solve the problem of over-approximation of training data. Pruning methods usually use statistical methods to delete the most unreliable branches (branches) in order to improve the speed of classification and recognition in the future and the ability to classify and recognize new data. In the paper, the decision tree node pruning is to achieve the purpose of effective training of enterprise electronic contract risk data. When predicting risk performance behavior, in order to avoid the occurrence of data over-fitting, the actual number of connection nodes should be controlled under the condition of less than the value of the basic decision-making organization [8]. To put it simply, node pruning is to rearrange the acquired electronic contract risk data of enterprises according to the decision tree model. When the numerical results can meet the prediction requirements of risky performance behavior, it is determined that the current pruning conditions can promote the smooth progress of business cooperation among enterprises [6]. The node pruning processing expression satisfying the decision tree model is:

$$D(s)=\sum_{a=1}^{+\infty} g_s k_s \left|\Delta A\right|^2 \tag{1}$$

Where, s indicates the decision tree node marking coefficient, a represents the initial value of the node fitting parameter, ΔA represents the unit accumulation of the enterprise electronic contract risk data, g_s represents the definition basis vector of the decision tree model, and k_s represents the risk data collection parameter. The larger the accumulation of enterprise electronic contract risk data is, the more nodes need to be pruned in the decision tree model.

2.3 Validity parameters

The validity parameter, also known as the effect index, determines the carrying capacity of the decision tree model for the enterprise electronic contract risk data. When the number of pruning nodes remains unchanged, the larger the value of the validity parameter, the stronger the carrying capacity of the decision tree model mechanism for the risk data [9]. To realize the accurate prediction of risky performance behavior, the value of validity parameter index should belong to the value interval of $(0,1]$. Let δ be the minimum value result of risk performance parameters, f be the risk identification vector in the process of enterprise electronic contract performance, and χ be the standard prediction coefficient. With the support of the above physical quantities, the simultaneous formula (1) can express the validity parameter solution result based on the decision tree model as follows:

$$H=-\sum_{\delta}\frac{D(s)}{f}\cdot\left(\chi^2-1\right) \tag{2}$$

Stipulate when that $\chi=1$, the prediction result of the performance risk of the enterprise electronic contract obtained by solving has no application effect; When $\chi<1$, the result obtained by solving indicates that the performance risk of the enterprise electronic contract is relatively high; when $\chi>1$, the result obtained by solving indicates that the performance risk is relatively low.

3 PREDICTION OF PERFORMANCE RISK OF ELECTRONIC CONTRACT

Because the two parties of the electronic contract are in different places, one party can not automatically get the other party's acceptance and confirmation while using the electronic contract to make the contract behavior, so compared with the ordinary contract form, it is easy to form greater commercial risks.When the decision tree model is used to predict the performance risk of electronic contract, it is necessary to determine the performance reliability conditions according to the distribution characteristics of risk factors, so as to solve the actual value range of prediction parameters.

3.1 Distribution characteristics of risk factors

The distribution characteristics of risk factors refer to the arrangement of data parameters that may cause risky behaviors in the process of electronic contract performance. Because the decision tree model has limited carrying capacity for data samples, when predicting risky performance behaviors, the performance characteristics of relevant influencing factors must match the pruning situation of decision tree nodes [4]. If the decision tree model is regarded as a limited but relatively large data storage space, the solution of the distribution characteristics of risk factors can be understood as the calculation of the residual value of relative risk prediction. Suppose that $c_{\min}$ represents the minimum value of the risk influence parameter in the process of enterprise electronic contract performance, $c_{\max}$ represents the maximum value of the risk influence parameter, the decision tree model stipulates that the value conditions of $c_{\min} \in (1, e)$ and $c_{\max} \in (1, e)$ are established at the same time. The simultaneous formula (2) can express the distribution characteristics of risk factors as follows:

$$X = -\sum_{c_{\min}}^{c_{\max}} \log_H \left(\frac{|\tilde{j}|}{|\overline{B}|} \right)^2 \tag{3}$$

Where, $\tilde{j}$ refers to the standard performance vector of the enterprise electronic contract, and $\overline{B}$ refers to the average value of the risk expression coefficient. To avoid the occurrence of high-risk performance behavior, the number of pruned decision tree nodes must be less than the initial number of decision tree nodes.

3.2 Performance reliability conditions

The performance reliability condition affects the ability of the decision tree model to process the enterprise electronic contract risk data. Without considering other interference conditions, the larger the assignment range of the performance reliability condition, the stronger the ability of the decision tree model to process the enterprise electronic contract risk data [3]. The so-called credibility can be understood as the degree of credibility. For the relevant cooperative enterprises, the performance ability of the electronic contract determines whether the business cooperation can proceed smoothly. Thus, in the process of predicting risky performance behavior, the value of the credibility index should not be less than zero. Compared with other performance risk prediction factors, the reliability condition can better adapt to the solution results of the distribution characteristics of risk factors, promote the smooth progress of business cooperation between enterprises, and solve the problem of unreasonable performance of electronic contracts caused by unnecessary conditions. Therefore, the precise definition of performance reliability condition is the basis for predicting risky performance behavior [1].

3.3 Predicted parameters

For the decision tree model organization, the prediction parameters determine the execution ability of the enterprise electronic contract performance risk prediction method. If the distribution characteristics of risk factors are regarded as fixed value parameters, it can be considered that in a fixed value range, the larger the value of the prediction parameters, the greater the possibility of risky behavior in the process of contract performance [2]. The calculation expression of the prediction parameter index is as follows:

$$M = {\beta y} \Big/ {\left[X \left(w' - 1 \right)^2 p \right]} \tag{4}$$

Where, β represents the electronic contract transaction parameters, y represents the execution intensity of business cooperation between enterprises, w' represents the partial derivative solution result of the performance risk characteristics, and p represents the standard performance vector of the enterprise electronic contract. So far, the calculation and processing of relevant parameters and indicators have been completed. With the support of decision tree model conditions, the accurate prediction of enterprise electronic contract performance risk has been realized.

4 CASE STUDY

4.1 Preliminary preparation

To verify the prediction ability of the selected method on the performance risk of enterprise electronic contract, the following comparative experiment is designed. Firstly, the prediction method based on decision tree model was used to monitor the transaction status of business cooperation between enterprises, and the obtained risk vector extraction accuracy value was recorded as the experimental group data. Secondly, the prediction method based on a flexible strategy is used to supervise the transaction of business cooperation between enterprises, and the obtained risk vector extraction accuracy value is recorded as the control group data. At different times, the values of the experimental group and the control group in terms of the accuracy of risk vector extraction were recorded. Values and times must be accurate. Then, the experimental results are sorted out. Finally, the experimental results are summarized.

The following table shows the specific models of the experimental equipment selected for this experiment.

Table 1 Experimental equipment

No.	Equipment components	Names
1	The host computer	I Class 9 +Core GTX1660S
2	Operating system	Windows 10
3	Processor	E5 AMD AM4
4	Signal transceiver	MT-viki HDMI

It can be concluded from Table 1 that in order to achieve the best experimental effect, the host selected in this experiment is i9-level ten-core GTX1660S; the processor is E5 AMD AM4, and the signal transceiver is MT-viki HDMI. In addition, to avoid the appearance of unfair experimental results, the experimental equipment models selected by the experimental group and the control group are exactly the same.

4.2 Experimental results

Figure 2 reflects the specific experimental values of the extraction accuracy of the risk vector of the experimental group and the control group.

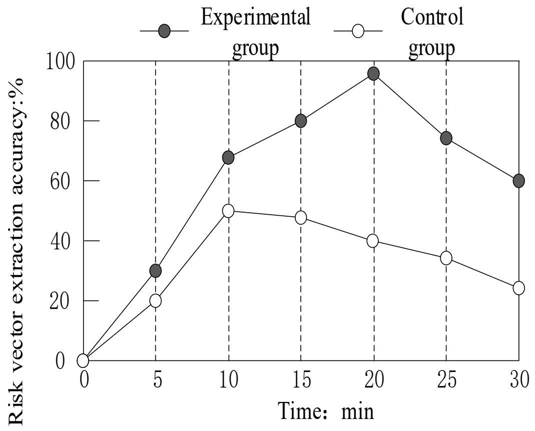


Figure 2 Extraction accuracy of risk vector

Figure 2 indicates that when the experiment time is equal to 20 min, the extraction accuracy of risk vector in the experimental group reaches the maximum value of 97.5%; when the experiment time is equal to 10 min, the extraction accuracy of risk vector in the control group reaches the maximum value of 50.1%, which is 47.4% lower than the maximum value in the experimental group. Although the difference is not obvious in the first 5 minutes, when the experimental time was 0-20 minutes, the accuracy of risk vector extraction in the experimental group showed an upward trend, while the control group had a temporary decline at the time of 10 minutes.

To sum up, under the action of the prediction method based on the decision tree model, the accuracy index of the risk vector extraction shows an obvious increasing trend. Moreover, with the change of time, the trend can be maintained for a long time, which verifies that the stability is good. This greatly reduces the risk of misprediction. Thus, it can better avoid the risk factors in the electronic contract of enterprises. Therefore, it can provide more reliable risk prediction results for decision makers, and can play a certain role in promoting the smooth progress of business cooperation between enterprises.

5 CONCLUSION

Under the action of the decision tree model, the enterprise electronic contract performance risk prediction method carries out pruning processing on relevant tree nodes according to the validity parameter value result. It determines the assignment condition of the performance reliability condition according to the calculation value of the risk factor distribution characteristics to realize the accurate calculation of the prediction parameter index. In this paper, through the experimental comparison, from the point of view of the accuracy of risk vector extraction, the effectiveness of this method is very obvious. Thus, compared with the prediction method based on a flexible strategy, the prediction method based on the decision tree model can effectively avoid the risk factors in enterprise electronic contracts, which is in line with the original intention of promoting the smooth progress of business cooperation between enterprises.

Through the above analysis of this paper, we can see that the decision tree has the characteristics of clear hierarchy, simple and clear in the process of electronic contract performance risk prediction. In particular, the decision-making problem is in a multi-stage and multi-level, it can easily express the correlation and interaction between each stage of decision-making and the overall decision-making. It provides a simple and effective decision-making method for enterprise managers in risk prediction. On the other hand, the quality of decision tree analysis mainly depends on the data and judgment. If the data and judgment are correct, the probability estimation provided will be more practical and accurate, and the reliability of the decision made through decision tree analysis will be greater. Therefore, when using decision tree model analysis, data should be collected extensively. Consult the opinions of experts and managers with rich experience, check and modify the probability distribution repeatedly, and provide a reliable basis for the final risk prediction of electronic contract performance.

With the continuous development of global integration economy, the risk of electronic contract performance faced by enterprises will be further increased. Future research will be targeted on the existing model in-depth study. The purpose is to further enhance the performance and robustness of the model, and further improve the prediction accuracy of enterprise electronic contract performance risk.

ACKNOWLEDGEMENTS

The study was supported by “Research on contract risk management technology based on contract performance data (Grant No. KJ2022-046)”, and “Research and application demonstration of key technologies and equipment of energy industry cloud network intelligent Internet of Things (Grant No. 22310302D)”.

REFERENCES

[1] Cui, J., Shi, Y. (2020). The Stage Characteristics of Risks in China's Financial Leasing Industry: An Explanation from Incomplete Contract and Evolutionary Game. Technical Economy, 39 (05), 149-155.

[2] Chen, X., Li, B., Wu, S. (2021). Supply chain decision-making considering risk-averse manufacturers under asymmetric demand information. Journal of Management Engineering, 35(06), 234-249.

[3] Gu, W., Li, H., Zhang, L. (2021). Risk Analysis and Revision Suggestions of Digital Resource License Agreement Terms in Judicial Practice. Library and Information Service, 65 (18), 15-23.

[4] Hu, B., Pan, Y., Zhou, X. (2022). The International Practice of Suspension of the Right of Early Termination of Financial Contracts and Its Enlightenment. Journal of Xiamen University (Philosophy and Social Sciences), 72 (03), 164-172.

[5] Liang, X., Jiang, A., Wang, G. (2020). Identification technology of remainder signal of sealed relay based on parameter optimization decision tree algorithm. Journal of Electronic Measurement and Instrumentation, 34 (01), 178-185.

[6] Ma, Y., Hong, H., Lin, L. (2020). Adverse Incentives for Borrowers' Performance in Online Lending: Empirical Evidence Based on "Renrendai" Data. Financial Research, 46 (05), 66-80.

[7] YAN, L., GUO, L., NING, Y. (2021). Study on the Incentive Effect of Contract Flexibility on Contractor's Performance: Taking Information Transparency as Moderator. Management Review, 33 (10), 222-236.

[8] ZHENG, X., WANG, S., JIN, L. (2021). The Impact of Relationship Conflict on Contractor's Performance: The Moderating Role of Relationship Governance. Journal of Civil Engineering and Management, 38 (02), 17-23.

[9] Zhang, Z., Wang, J., Chen, C. (2021). Research on settlement mechanism of government-authorized contract for difference in electricity spot market. Power Grid Technology, 45 (04), 1337-1346.

Rice Yield Prediction Based on LSTM and GRU

Yunqing Qiu
Department of Computer Science and Engineering, The Ohio State University, Columbus, Ohio, 43210, USA
*Corresponding author's e-mail: qiu.588@osu.edu

Abstract

Rice yield prediction is a vital problem in the national agriculture and economy. The development of deep learning overcomes the obstacles of traditional machine learning and shows superior performance in solving complicated problems. Especially for natural language processing (NLP) models such as LSTM and GRU, these models outperform the time series data, thus having great potential for complex agricultural spatiotemporal data with high dimensionality and nonlinearity. However, there is little discussion about performance of these two models in rice yield prediction. In this article, we adopted two popular NLP models to build and test 12 different model frameworks based on optimal hyperparameter configurations. And we compared model depth as well as bidirectional setting on the rice yield prediction by observing the performance of MSE losses throughout the training process. The results illustrated that both simple and complex models had outstanding fitting for small-sample training, and the depth and direction of the models did not significantly impact the performance of the experiment. But the complex model notably increases the training cost and decreases the convergence rate, implying that it's not necessarily suitable for time-series problems with small-sample data. Further, the results could provide insights into a deep learning framework construction and hyperparameter selection for subsequent studies with comparable characteristics.

Key Words: Rice Yield Prediction, NLP, LSTM, GRU, Model Comparison

1. Introduction

Agriculture, as the primary industry, occupies a fundamental position in the national economy. Grain yield is a significant indicator measuring agricultural development in agricultural field [13]. The government estimates the grain yield annually to facilitate the policy adjustment in agricultural production and financial support, as well as the integrated management in grain dispatch and resource allocation [6]. Therefore, accurate estimation of grain production appears particularly important in national financial benefits, production planning, and food security. However, under the influence of many factors, precisely estimating the grain yield is quite difficult. For example, the environmental factors of global climate and local weather have high variability and complex nonlinear effects on food production, causing difficulty in prediction accuracy [7]. Agricultural remote sensing data collected by satellites are characterized by high dimensions and unstructured data, shielding the key features to be discovered. The unknown and new environment has an uncertain influence on seed yield and crop production, making it hard to obtain the insightful recommendation with interpretability for the future exploitation. Despite various machine learning models, such as Linear Regression, SVM, and Random Forest had been proposed to tackle these problems, the limitation still exists. For instance, the regional research data is hard to determine its universal applicability and the selected features are constrained by dimension. With the wider interdisciplinary application of deep learning technology and the rapid development of high-performance computing, the deep learning models are effectively exploited and provide strong support for the yield prediction. For example, the autonomous feature extraction ability helps to obtain useful information from the involute noise. The powerful generalization ability contributes to promoting the applicability of the model, and the excellent expressiveness can deal with the high-dimensional sparse features.

2. Previous Works

Scientists have carried out lots of studies on crop yield prediction. For example, Alibabaei, *et al.* studied Bidirectional Long Short-term Memory (LSTM) and Bidirectional Gated Recurrent Unit (GRU) and used historical data including climate data, irrigation schedules, and soil moisture content to estimate end-of-season yields, ultimately validating the higher accuracy and performance of LSTM, Bidirectional LSTM, GRU as well as Bidirectional GRU [1]. Khaki, *et al.* proposed a crop yield prediction method that combines CNNs and RNNs to accurately predict maize and soybean yields across the entire corn belt in the United States, proving that the deep learning method is far superior to traditional statistical

International Conference on Computer Graphics, Artificial Intelligence, and Data Processing (ICCAID 2022)
edited by Ruishi Liang, Jing Wang, Proc. of SPIE Vol. 12604, 126041V
© 2023 SPIE · 0277-786X · doi: 10.1117/12.2674760

methods, such as Least Absolute Shrinkage and Selection Operator (LASSO), random forest, DFNN, etc. [8]. In addition, Khaki, *et al*. used the back propagation method for feature selection based on the trained CNN-RNN model, and explained the kernel variable in the deep learning model. Sun, *et al*. also proposed a multi-layer deep learning model that coupled RNN and CNN to extract spatial and temporal features. The model accurately predicted corn yield from 2013 to 2016 in the experiment, and verified the effectiveness and superiority of the proposed method compared with other methods. Shook, *et al*. constructed a model based on LSTM that used pedigree correlation measurements and weekly weather parameters to analyze genotypic responses in multiple environments. They demonstrated that LSTM partially advantages over models such as Radial Basis Function Support Vector Regression (SVR-RBF), LASSO regression, and data-driven USDA [10]. Besides, the LSTM model also uses a temporal attention mechanism to provides valuable insights of important time windows in the growing season. You, *et al*. proposed a prediction framework for real-time prediction of crop yield using remote sensing data. The framework uses a histogram-based dimensionality reduction method and a Deep Gaussian Process to eliminate spatial correlation errors. Furthermore, the framework also applies modern representation learning ideas to crop yield prediction, obtaining more effective features than commonly used manual features [12]. Gangopadhyay, *et al*. constructed a model based on a dual-attention mechanism, which can effectively learn critical information from different perspectives and help to understand the impact of weather on yield prediction based on this insight [3].

As the studies mentioned above, deep learning provides powerful predictions for crop yields that are highly dependent on meteorological data, and also provides optimized decision support for the entire agricultural planting and harvesting cycle. Through the development of data monitoring and collection technology, massive data provides support for the training of large-scale artificial neural networks, assisting in automated feature extraction, decisions making, and interpretable insights.

3. Dataset

3.1. Data Collection

The data of crop yield and environmental monitoring are from a large agricultural province in China, Guangxi Zhuang Autonomous Region. This open-source dataset is used in the 2019 Tianchi Competition held by Alibaba (https://tianchi.aliyun.com/competition/entrance/231753/introduction). Guangxi Zhuang Autonomous Region has a subtropical monsoon climate, with abundant heat, abundant precipitation, distinct dry and wet conditions, and moderate sunshine. In terms of topographic structure, it has the characteristics of relatively broken plots, complex planting structure, and prominent drought and flood. The dataset collected multi-dimensional meteorological data from ground, satellite and radar in 81 counties of Guangxi from 2015 to 2017, and the yield of early and late rice in each county in four years. Among them, the meteorological data contains 12 daily meteorological characteristics, including sunshine hours, four wind directions monitored every 6 hours, average wind speed, average precipitation, maximum temperature, minimum temperature, average temperature, relative humidity and average air pressure. These variables are illustrated as follows (Table. 1 Variables Description).

Table 1. Variables Description.

Variables	Unit	Max	Min	Mean	SD
SH	h	13	0	3.998639	3.816937
WS_{Avg}	m/s	12.6	0	1.855579	0.995302
$Prec$	mm	335.5	0	5.211333	14.09668
T_{Max}	℃	41.6	-1.5	25.78523	7.458326
T_{Min}	℃	30.7	-5.5	18.32623	6.529348
T_{Avg}	℃	34.3	-2.4	21.25261	6.704619
RH	%	100	24	79.29492	10.98642
AP_{Avg}	hPa	1035.2	893.2	988.1103	23.61646

The abbreviations stand for the following: Max: maximum, Min: minimum, SD: standard deviation, Avg: average, SH: sunshine hours, WS: wind speed, Prec: precipitation, T: temperature, RH: relative humidity, AP: air pressure. Table 1 is the data description of the dataset output by Python.

3.2. Data Preprocessing

Data quality is one of the most critical factors affecting model construction and result generation. For the statistical collection of highly challenging crop and climate-related data, appropriate data pre-processing is needed to provide a reliable basis for the implementation of subsequent algorithms. After examining the amount of missing values and considering the characteristics of high feature dimension and large sample size of the data set, we use the ' 0 ' to fill the missing values and delete the unmatching samples that do not appear in the output file. Also, for the features of wind direction, characters such as 'N' (North), ' NNW' (North-northwest) are used to represent the direction and intensity of the period, and we convert the string value into a two-dimensional vector according to its direction and intensity. Finally, the Max-min Normalization is performed for all features.

4. Model

4.1. Model Description

Compared with traditional machine learning models, deep learning models are more capable for using the hierarchical algorithmic structure of neural networks. Among the deep learning models, RNN model and its variants LSTM (Long Short-Term Memory) and GRU (Gated Recurrent Unit) models have significant effects on extracting complex rules within the time series, especially in NLP field [2]. Therefore, our research focus on LSTM and GRU models. The LSTM model is designed based on the RNN, but solving the risks of gradient disappearance and gradient explosion in the RNN model. Compared with RNN, the LSTM unit adopts cell state to preserve long-term memory, and then cooperates with three gate mechanisms (input gate, output gate and forget gate) composed of sigmoid neural network layer and dot multiplication operation to filter information, so as to achieve long-term memory control [5]. Taking the forgetting gate in the gate mechanism contained in LSTM as an example, the sigmoid function of the forgetting gate generates the value in the range of the [0,1] to determine the degree of retention of the information received in the precedent cell, and then filter the information by multiplication. GRU model proposed by Cho et al. in 2014 can be treated as a simplified version of LSTM and also prevent the gradient vanishing and gradient exploding. Compared to LSTM model, GRU is simpler in structure, like a LSTM with a forget gate [4]. On the basis of LSTM model, bidirectional LSTM is the extensions of the LSTM [9], as well as the bidirectional GRU model is the extensions of the GRU. Among them, two independent units replace the original single unit in the form of stack heap, thus the context information from both sides of the sequence sends the sequence forward and backward to the same output layer.

4.2. Framework Design

We intended to test the impacts from model depth and direction, so we designed the following (Figure 1) model framework with the optimal performance from the pre-test. The model is used to solve the many-to-one rice yield prediction problem with time-series data. The input is the pre-processed meteorological data of a county and the county ID, which is represented as a one-hot vector; the output is the early rice yield of the county in that year. The two parts of the input data are processed with different neural network layers. For the meteorological data, the LSTM or GRU layers are first implemented with the different settings of 1-3 layers and whether bidirectional or not as the experiment configuration. To prevent overfitting, a dropout function is added after each LSTM and GRU layer [1]. After that, two linear layers are connected and give the output from meteorological data. For the one-hot vector county ID, two linear layers process the input and provide the output. Consequently, the final output is integrated by the multiplication operations from two parts of the outputs.

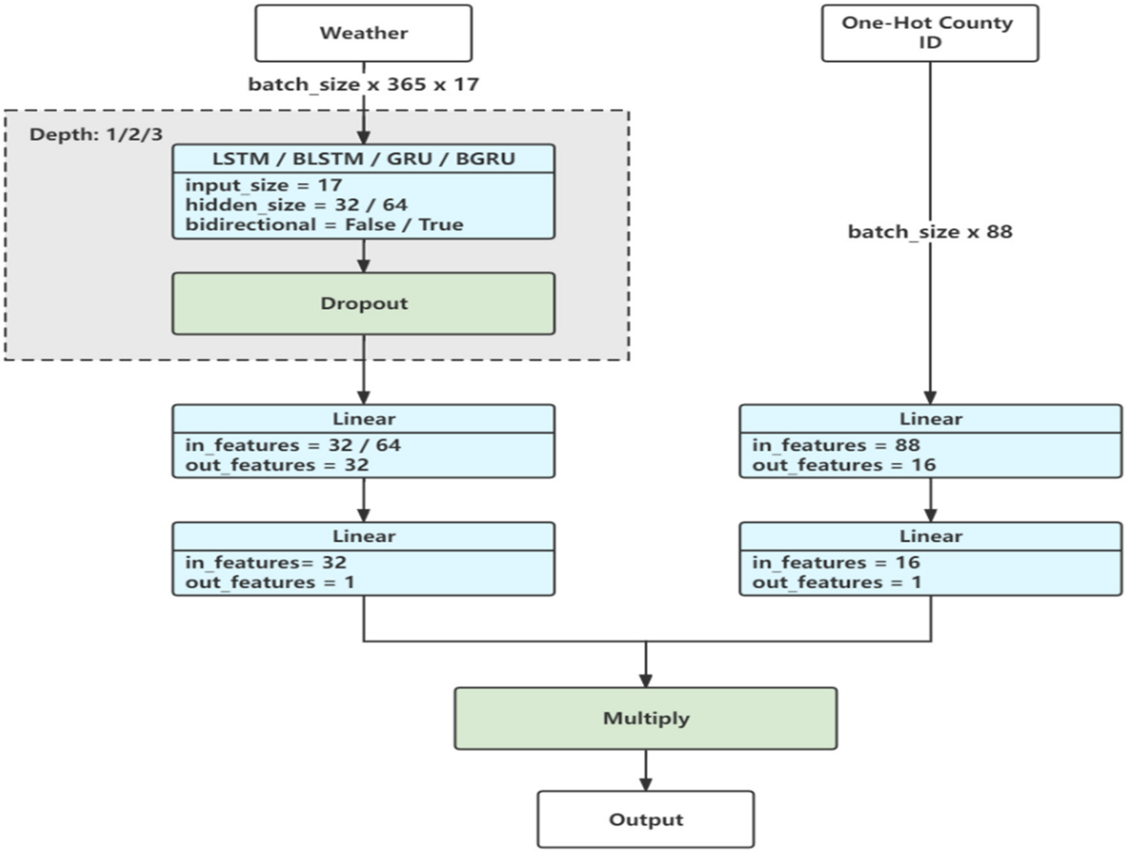


Figure 1. Model Framework

4.3. Model Training

For the division of data sets, we use the meteorological and crop yield data from 2015 to 2016 as the training set, and the meteorological and crop yield data in 2017 as the test set. After data preprocessing, the data dimension of input training is (20,365,17), of which 20 is batch size as using the batch-first mode, 365 is the number of time series and 17 is the number of features (Section 3.1 and 3.2). On the basis of the original data set, meaningless features such as time are removed, and all features of the weather is processed as standardized quantitative data.

The setting of hyperparameters is based on the characteristics of small data sample size and the optimal case of multiple parameter tuning. The network model uniformly sets the hidden layer as 2 layers and 16 nodes per layer; set the dropout ratio to 0.2. The other hyperparameters during training are defined as follows: epoch is set to 50; learning rate is set to 0.0001; momentum is set to 0.9. And the selected optimizer is the Stochastic Gradient Descent (SGD). The whole model is implemented by pytorch.

4.4. Model Evaluation

For the results of predicted yield, Mean Square Error (MSE) is adopted to evaluate the performance of the prediction power. MSE can be calculated using equation:

$$\boldsymbol{f = \frac{1}{2m}\sum_{i=1}^{m}(y_i - \hat{y}_\iota)^2}$$

where m is the total number of districts and counties, y_i is the early rice yield of the second district and county predicted by the contestant, and $\hat{y}_i$ is the actual early rice yield of the i_{th} county.

5. Results & Discussion

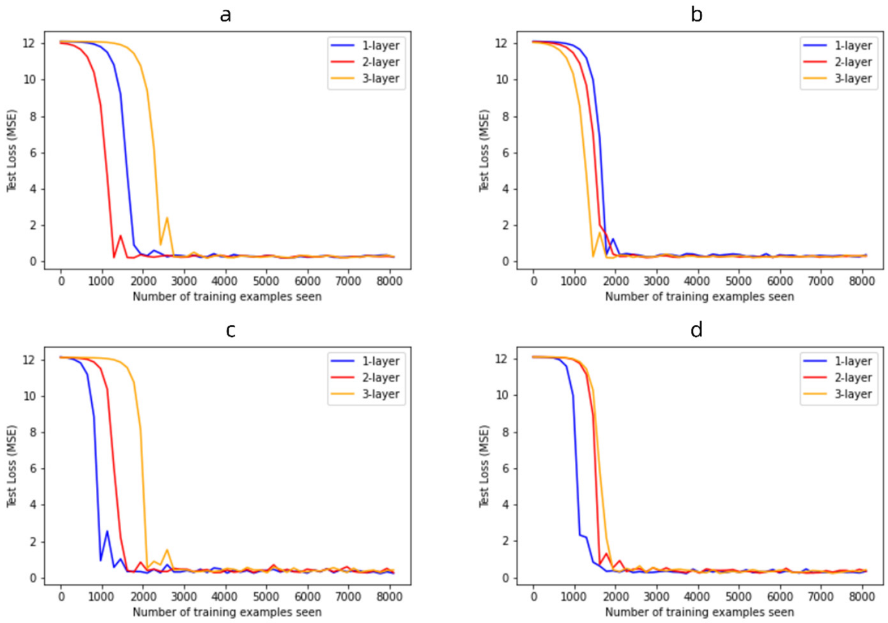


Figure 2. MSE losses on test set. (a) LSTM model; (b) bidirectional LSTM model; (c) GRU model; (d) bidirectional GRU model.

We recorded loss changes and final results on the test set throughout the training process for all cases of LSTM and GRU models. Figure 2 shows the loss value changes of the (a) LSTM, (b) bidirectional LSTM, (c) GRU, and (d) bidirectional GRU models with different depths (1-3 layers). Table 2 shows the final MSE loss results of the 12 models and the structure of the feature layers incorporated in the models. Through the observation of Figure 2, all models tend to converge after training 2000 samples. Also, according to the MSE losses in Table 2, all cases of models have achieved satisfactory training results. After the 50-epoch training, the MSE loss is reduced by 96-98% compared with the initial state. In addition, in the comparison of the four subgraphs of Figure 2, the number of neural network layers has a discernible effect on the convergence rate, especially shown from the different depths of (a) LSTM and (c) GRU. Besides, in (c) GRU and (d) bidirectional GRU models, depth directly links to the convergence rate, and models with fewer network layers converge faster.

Table 2. Test set MSE loss on models with different depths and bidirectional setting

Model	MSE	Feature size for each layer
LSTM (1 layer)	0.2358	(17,32)
LSTM (2 layer)	0.2693	(17,32), (64,32)
LSTM (3 layer)	0.2725	(17,32), (64,32), (64,32)
Bidirectional LSTM (1 layer)	0.3495	(17,32)
Bidirectional LSTM (2 layer)	0.2730	(17,32), (64,32)
Bidirectional LSTM (3 layer)	0.2931	(17,32), (64,32), (64,32)
GRU (1 layer)	0.2419	(17,32)
GRU (2 layer)	0.2797	(17,32), (64,32)
GRU (3 layer)	0.4256	(17,32), (64,32), (64,32)
Bidirectional GRU (1 layer)	0.3342	(17,32)
Bidirectional GRU (2 layer)	0.3368	(17,32), (64,32)
Bidirectional GRU (3 layer)	0.4166	(17,32), (64,32), (64,32)

Through the overall examination of the results, deep learning models are proven to show strong learning ability for high-dimensional nonlinear data with their large-scale parameters. For the complicated meteorological data, the LSTM and GRU models with different settings of depth and direction do not show considerable distinctions in the final fitting performance. Due to the limited sample size, the complexity of the problem is slightly lower, the final loss of all models leans to be similar, and all accomplish superior prediction performance.

However, the depth and direction of the models have a significant impact on the convergence rate. It demonstrates that fewer parameters will provide faster convergence when the parameter magnitude can reach the same degree of final outcomes. And over-complicated models lead to higher computational overhead and longer convergence time. Considering the computational and time costs of the models, simpler models with fewer feature layers and parameters can be assigned priority in the model selection for small-sample data problems.

6. Conclusion

In this paper, two NLP models, LSTM and GRU, are used to design the framework of the rice yield prediction model for time-series meteorological data and showed excellent prediction performance. Then, the potential impact of depth and direction in LSTM and GRU models was further experimented. According to the results, it is discovered that the change of depth and direction significantly affects the convergence rate for the small-sample data, indicating that model complexity on depth selection and bidirectional settings may bring unnecessary computational overhead.

Compared to the previous studies on the application of LSTM and GRU models in agricultural prediction, this paper improves the framework to the dataset from a different region and extends the applicability of the model. In addition, experiments were carried out from the perspective of model depth and bidirectional setting. The result showed that a variety of depth and direction combinations had similar superior prediction results, but differentiated in the effect of convergence rate. Above that, it demonstrates that for time-series problems with small-sample data, the depth and direction of the model have no significant disparities in the learning performance, for both LSTM and GRU models have good applicability. But the model complexity needs to be determined on demand, otherwise, it may bring unnecessary convergence rate reduction. Overall, this paper provides referable ideas from deep learning framework design, model selection, and hyperparameter configuration for subsequent research.

However, due to the boundaries of the dataset, the number of models compared is relatively little, so the inherent constraints of LSTM and GRU models are not presented. We consider improving the experiment from two aspects in the future. Firstly, extending the model structure. Because the feature screening ability of LSTM and GRU is comparably weak, the Attention Mechanism can be utilized to associate different positions of a single time-series input, which supports to compute a representation of the sequence [11]. Secondly, considering the complementarity between features, uncovering the correlation between diverse meteorological features, and the carries out corresponding feature fusion.

References

[1] Alibabaei, K., Gaspar, P. D., & Lima, T. M. (2021). Crop Yield Estimation Using Deep Learning Based on Climate Big Data and Irrigation Scheduling. Energies, 14(11), 3004. https://doi.org/10.3390/en14113004

[2] Cambria, E., & White, B. (2014). Jumping NLP Curves: A Review of Natural Language Processing Research [Review Article]. IEEE Computational Intelligence Magazine, 9(2), 48–57. https://doi.org/10.1109/MCI.2014.2307227

[3] Gangopadhyay, T., Shook, J., Singh, A. K., & Sarkar, S. (2020). Interpreting the Impact of Weather on Crop Yield Using Attention. 6.

[4] Gers, F. A., Schmidhuber, J., & Cummins, F. (2000). Learning to Forget: Continual Prediction with LSTM. Neural Computation, 12(10), 2451–2471. https://doi.org/10.1162/089976600300015015

[5] Hochreiter, S., & Schmidhuber, J. (1997). Long Short-Term Memory. Neural Computation, 9(8), 1735–1780. https://doi.org/10.1162/neco.1997.9.8.1735

[6] Horie, T., Yajima, M., & Nakagawa, H. (1992). Yield forecasting. Agricultural Systems, 40(1–3), 211–236. https://doi.org/10.1016/0308-521X(92)90022-G

[7] Khaki, S., & Wang, L. (2019). Crop Yield Prediction Using Deep Neural Networks. Frontiers in Plant Science, 10, 621. https://doi.org/10.3389/fpls.2019.00621

[8] Khaki, S., Wang, L., & Archontoulis, S. V. (2020). A CNN-RNN Framework for Crop Yield Prediction. Frontiers in Plant Science, 10, 1750. https://doi.org/10.3389/fpls.2019.01750

[9] Schuster, M., & Paliwal, K. K. (1997). Bidirectional recurrent neural networks. IEEE Transactions on Signal Processing, 45(11), 2673–2681. https://doi.org/10.1109/78.650093

[10] Shook, J., Gangopadhyay, T., Wu, L., Ganapathysubramanian, B., Sarkar, S., & Singh, A. K. (2021). Crop yield prediction integrating genotype and weather variables using deep learning. PLOS ONE, 16(6), e0252402. https://doi.org/10.1371/journal.pone.0252402

[11] Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, L., & Polosukhin, I. (2017). Attention Is All You Need (arXiv:1706.03762). arXiv. http://arxiv.org/abs/1706.03762

[12] You, J., Li, X., Low, M., Lobell, D., & Ermon, S. (2017). Deep Gaussian Process for Crop Yield Prediction Based on Remote Sensing Data. Proceedings of the AAAI Conference on Artificial Intelligence, 31(1). https://doi.org/10.1609/aaai.v31i1.11172

[13] Zhen, L., & Routray, J. K. (2003). Operational Indicators for Measuring Agricultural Sustainability in Developing Countries. Environmental Management, 32(1), 34–46. https://doi.org/10.1007/s00267-003-2881-1

Semantic layout aware generative adversarial network for text-to-image generation

Jieyu Huang[1a], YongHua Zhu[1b], Zhuo BI[2c] And Wenjun Zhang[2*]
[1]Shanghai Film Academy, Shanghai University Shanghai, China
[2]School of Information Technology, Shanghai Jian Qiao University Shanghai, China
[a]jieyu88@shu.edu.cn
[b]zyh@shu.edu.cn
[c]21028@gench.edu.cn
[*]wjzhang@shu.edu.cn

Abstract

Text-to-image(T2I) generation methods aim to synthesize a high-quality image which is semantically consistent with the given text descriptions. Previous (T2I) Generative Adversarial Networks generally first create a low-resolution image with rough shapes and colors, and then refine the initial image into a high-resolution image. Most stacked architecture still remains two main problems. (1) The final images generated by these methods depend heavily on the quality of the initial image. If the initial one is not initialized correctly, the resulted image seems like a simple combination of visual features from several images scales. (2) The cross-modal fusion methods about text and image that previous works widely adopted is limited in the text-image fusion process. In the paper, we propose a novel generation model, which introduce a one-stage backbone directly generate high-quality images without multi generators and a novel Semantic Layout Deep Fusion Network to sufficiently fuse text features and image features. Experiments on the challenging CUB and COCO-Stuff datasets demonstrates the ability of our model in generating images, regarding both semantic consistency with input text description and visual fidelity.

Keywords: feature fusion, one-stage generation

1. Introduction

Over the last few years, people have witnessed the remarkable evolution of Generative Adversarial Networks (GANs) [1] for diverse conditions: layout [2,3,4], scene graph [5,6,7] and text [8,9]. But natural language description is the most common and convenient medium for people to explain the world. Therefore, synthesizing photo-realistic images conditioned on linguistic description has become an active research area.

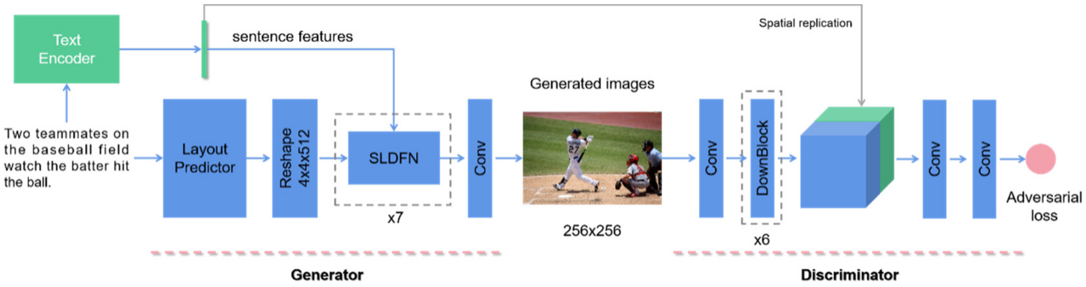


Figure 1. The framework of SLA-GAN.

Although existing multi-stage methods [10,11] achieve great process, there are two core challenges for text-to-image to solve: visual fidelity of the synthesized image and semantic consistency of the given text and the resulted image. First, the final synthesizing image highly is depended on the quality of the initial image that make the last one seems to be composed of some strange details and fuzzy shapes. Second, the current text-image fusion methods fuse text and image information less effectively. Generally, there are three common fusion methods: features vector concatenation [8], cross-modal attention mechanism[12] and Condition Batch Normalization(CBN) [13]. The direct concatenation of feature vector makes not full use of both text information and image information. Next, the text-image fusion method based on cross-modal

International Conference on Computer Graphics, Artificial Intelligence, and Data Processing (ICCAID 2022)
edited by Ruishi Liang, Jing Wang, Proc. of SPIE Vol. 12604, 126041W
© 2023 SPIE · 0277-786X · doi: 10.1117/12.2674685

attention mechanism is difficult to synthesis the higher resolution image because of the computational cost. Besides, the CBN module only apply a few times in recent works. It causes that the image features and text features are not fused efficiently in image generation process.

To solve the above problems, we propose a text-to-image generation method named as Semantic Layout Aware Generative Adversarial Network (SLA-GAN) (see Figure 2). For the first issue, we replace multi-stage method with one-stage method which can avoid the entanglements between multiple generators. Regarding the second question, our model designs a Semantic Layout Deep Fusion Network block to enhance the text-image semantic consistency.

2. Method

As shown in Figure 1, the overall network architecture of our proposed model SLA-GAN following one-stage method illustrated in [14] is composed of a generator and a discriminator. Firstly, our model has a text encoder that capture semantic feature and a pre-trained layout predictor network proposed in [15] that predict spatial layout. The text vector first fed into layout predictor to predict a scene layout. Then, we apply several SLDFN blocks to boost text-image fusion process and generate image matching text semantics. The text feature vector is replicated first and then concatenated with the image features vector. And the discriminator takes it with adversarial loss to distinguish the result image realistic and semantic consistency of text description.

2.1 Generator

To address the instability of training in stacked generator framework, we followed one-stage method that avoids the entanglements between different generators. Our method also take the adversarial loss associated with MA-GP loss [16] in training process.

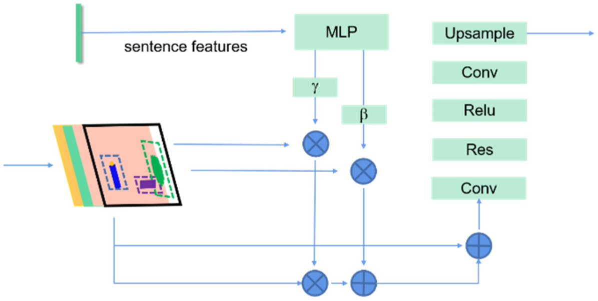


Figure 2. Structure of the SLDFN.

2.2.1 text encoder and layout predictor

Our text encoder adopts bi-directional Long Short-Term Memory (Bi-LSTM) to encode text description. In many early works, they used pre-trained model to extract semantic vector from text description during the training, which fixed the parameter can not learn better visual representation for image generation. We trained the text encoder with the image generator jointly by minimizing the Deep Attentional Multimodal Similarity (DAMS) loss [12].

We adopted the layout predictor provided by [15] that has been used to predict coarse spatial layout in given text description. Because semantic information should be added text-relevant sub-regions.

2.1.2 Semantic Layout Deep Fusion Network

We take the encoded sentence vector s and reshaped image layout map $f_0 \in \mathbb{R}^{4\times512\times512}$ from layout predictor as input of the first SLDFN block. Each SLDFN block is composed of two Convolution layers, a Residual layer, a RelU layer and a Upsample layer. It takes intermediate feature map $f_i \in \mathbb{R}^{ch_i\times\frac{h_i}{2}\times\frac{w_i}{2}}$ which fused with the sentence vector generated from i-th SLDFN as input. Since the generator in our one-stage structure needs to generate high-resolution images, it must be composed of more layers than the generator in the previous stack framework. Therefore, we have applied 7 times SLDFN to enhance the generated image visual fidelity.

2.2 Discriminator

Our method take the one-way output discriminator with MA-GP loss as our discriminator proposed in [14] to train the network.

$$
\begin{aligned}
L_D = {} & E_{x\sim P_r}\left[\max\left(0,1+D(x,s)\right)\right] \\
& +\frac{1}{2}E_{G(z)\sim P_g}\left[\max\left(0,1+D(G(e),s)\right)\right] \\
& +\frac{1}{2}E_{x\sim P_{mis}}\left[\max\left(0,1+D(x,s)\right)\right] \\
& +\lambda_{MA}E_{x\sim P_r}[(\|\nabla_x D(x,s)\|_2+\|\nabla_s D(x,s)\|_2)^p] \\
L_G = {} & E_{G(z)\sim P_g}[D(G(e),s)]+\lambda_{DA}L_{DAMSM}
\end{aligned}
$$

Where s is the sentence vector and e is immediate vector that sentence vector fused with image layout map f_i; P_{mis}, P_g, P_r denote the generated mismatching data distribution, image distribution, real data distribution; the hyper-parameter λ_{DA}, $p$and λ_{MA} are set 0.1, 2 and 6.

3. Experiment

We train our model on COCO-Stuff [17] and CUB [18] bird datasets. The COCO dataset is divided into a training set containing 80k images and a test set containing 40k images. Each image in this dataset includes five language descriptions. The CUB dataset includes 11,788 images which belong to 200 bird species, where 8,855 images are employed for training and 2,933 images employed for testing. We compare our model with the recent state-of-the-art methods in text-to-image generation and quantify the performance of the SLA-GAN in terms of Inception Score (IS) [19] and Frechet Inception Distance (FID) [20].

Table 1. Statistic score on COCO and CUB in IS and FID.

	Inception Score	Frechet Inception Distance	
	CUB	COCO	CUB
DF-GAN	4.86 ± 0.04	19.32	19.24
DM-GAN	4.75 ± 0.07	32.64	16.09
Our method	**5.13 ± 0.06**	19.83	**15.63**

3.1 Quantitative Results

Table 1 summarizes comparison results of the IS and FID. Inception Score aims to measure the quality and diversity of generated images. Higher IS denotes that the generative model synthesizes a high-quality image which belongs to a specific class more clear. FID takes the second-order information of the final layer of the inception model and calculates the similarity between the generated image and the real image. Contrary to IS, a lower FID score denotes the generated image more close to the real image. For 256×256 images, our method outperforms other methods on the CUB dataset both in IS and FID. The SLA-GAN performed a little bit worse in FID on COCO datasets, but our model still demonstrates the ability to generate complex and realistic image.

3.2 Qualitative Results

Our method compared the generated images with two the-state-of-art method, i.e. DM-GAN [9] and DF-GAN [14] as shown in Figure 3. It displays several image examples from our model trained on COCO and CUB datasets given the same text description. As shown in the left 3 columns in Figure 3, our model synthsizes images with more visual fidelity that matching the give text description. For instance, “a small bird with a grey head and black nape, with blue and grey covering the rest of its body”, our method generates a bird with more clear attribute. But, the generated image by DM-GAN reshaped its natural structure. For the COCO datasets, it can be seen that our model is able to generate realistic image with multiple objects and clear background. In the last column, the giraffe is natural and recognizable.

Table 2. Ablation study on CUB datasets.

	Inception Score	Frechet Inception Distance
w/o SLDFN	4.16 ± 0.02	23.92
Full model	**5.13 ± 0.06**	15.63

Proc. of SPIE Vol. 12604 1260401-509

3.3 Ablation Study

Table 2 shows the importance of the key components by comparing IS and FID of the ablation study. We replace the SLDFN block with Upsample block to make the ablation study on the CUB datasets. Without SLDFN block, it reduces the overall performance of the model. It is obvious that our complete model gets better scores in these indicators.

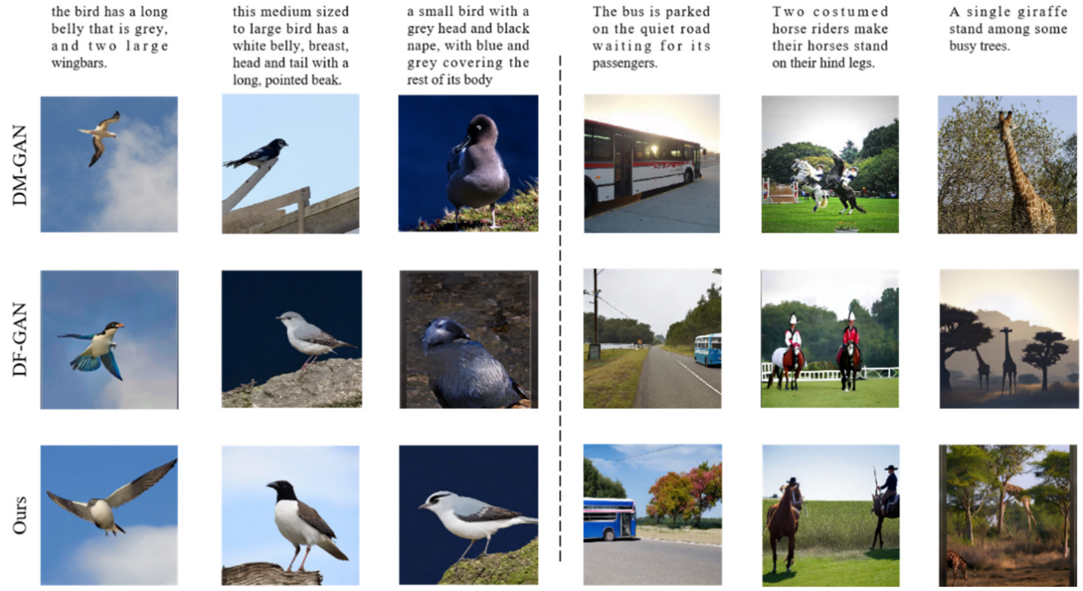

the bird has a long belly that is grey, and two large wingbars.
this medium sized to large bird has a white belly, breast, head and tail with a long, pointed beak.
a small bird with a grey head and black nape, with blue and grey covering the rest of its body
The bus is parked on the quiet road waiting for its passengers.
Two costumed horse riders make their horses stand on their hind legs.
A single giraffe stand among some busy trees.
DM-GAN
DF-GAN
Ours

Figure 3. Examples of 256×256 generated images from our proposed model and baselines on COCO datasets and CUB datasets.

4. Conclusion

In this paper, we design a one-stage model for synthesizing high-quality images from natural language text. Compared to the previous GAN models, our model allows us to generate realistic recognizable images. The qualitative result shows that compared with the baseline model, our model improves the generation quality. Besides, our proposed text-image fusion method, the Semantic Layout Deep Fusion Network block, which fuses text and image information more sufficiently.

Reference

[1] Goodfellow I, Pouget-Abadie *J*, Mirza M, et al 2020 *J*. Generative Adversarial Networks. Communications of the ACM. **63(11)** 139-144.

[2] Zhao B, Yin W, Meng L, et al 2020 *J*. Layout2image: Image generation from layout. International Journal of Computer Vision. **128(10)** 2418-243

[3] Sun, W., & Wu, T. 2019. Image synthesis from reconfigurable layout and style. *In Proc. of the IEEE/CVF Int. Conf. on Computer Vision* pp 10531-10540.

[4] Sylvain T, Zhang P, Bengio Y, et al. 2021. Object-centric image generation from layouts. *Proc. of the AAAI Conf. on Artificial Intelligence*. vol 3 pp 2647-2655.

[5] Johnson J, Gupta A, Fei-Fei L. 2018. Image generation from scene graphs *Proc. of the IEEE Conf. on computer vision and pattern recognition* pp 1219-1228.

[6] Li Y, Ma T, Bai Y, et al. 2019 *J*. Pastegan: A semi-parametric method to generate image from scene graph. Advances in Neural Information Processing Systems, **32**.

[7] Ashual, O., & Wolf, L. 2019. Specifying object attributes and relations in interactive scene generation. *In Proc. of the IEEE/CVF Int. Conf. on computer vision* pp 4561-4569.

[8] Zhu M, Pan P, Chen W, et al. 2019. Dm-gan: Dynamic memory generative adversarial networks for text-to-image synthesis *Proc. of the IEEE/CVF Conf. on computer vision and pattern recognition* pp 5802-5810.

[9] Reed S E, Akata Z, Mohan S, et al. 2016 *J*. Learning What And Where To Draw. Advances in neural information processing systems **29**.

[10] Zhang H, Xu T, Li H, et al. 2017. Stackgan: Text to photo-realistic image synthesis with stacked generative adversarial networks *Proc. of the IEEE Int. Conf. on computer visio*n pp 5907-5915.

[11] Zhang H, Xu T, Li H, et al. 2018 *J.* Stackgan++: Realistic image synthesis with stacked generative adversarial networks. IEEE transactions on pattern analysis and machine intelligence **41(8)** 1947-1962.

[12] Xu T, Zhang P, Huang Q, et al. 2018. Attngan: Fine-grained text to image generation with attentional generative adversarial networks *Proc. of the IEEE Conf. on computer vision and pattern recognition* pp 1316-1324.

[13] Ma J, Zhang L, Zhang J. 2019 *J.* SD-GAN: Saliency-discriminated GAN for remote sensing image superresolution. IEEE Geoscience and Remote Sensing Letters **17(11)** 1973-1977.

[14] Tao M, Tang H, Wu F, et al. 2022. DF-GAN: A Simple and Effective Baseline for Text-to-Image Synthesis *Proc. of the IEEE/CVF Conf. on Computer Vision and Pattern Recognition* pp 16515-16525.

[15] Zhu Y, Huang J, Ge N, et al. 2021. Text Pared into Scene Graph for Diverse Image Generation *The 5th Int. Conf. on Computer Science and Application Engineering pp 1-5.*

[16] Qiao, T, Zhang, J, Xu, D, et al. 2019. Mirrorgan: Learning text-to-image generation by redescription *In Proc. of the IEEE/CVF Conf. on Computer Vision and Pattern Recognition* pp 1505-1514.

[17] Lin, T. Y, Maire, M, Belongie, S, et al. L. 2014. Microsoft coco: Common objects in context *In European Conf. on computer vision* pp.740-755.

[18] Wah C, Branson S, Welinder P, et al. 2011 *J.* The caltech-ucsd birds-200-2011 dataset.

[19] Tim S, Goodfellow I and Zaremba W et al. 2016 *J.* Improved techniques for training gans. In Advances in neural information processing systems 2234–2242

[20] Heusel M and Ramsauer H, et al. 2017 *J.* Gans trained by a two time-scale update rule converge to a local nash equilibrium. In Advances in neural information processing systems 6626–6637

Ship Detection in SAR Image Based on Improved YOLOv5 Network

Cheng-ge FANG [1a], Ying BI [2b], Zhen-yu WU [3c], Hui WANG [4d], Zi-wei CHEN [5e]
[1]China Academy of Launch Vehicle Technology
[2]Beijing Institute of Aerospace Systems Engineering
[3]Beijing Institute of Aerospace Systems Engineering
[4]Beijing Institute of Aerospace Systems Engineering
[5]Beijing Jiaotong University
[a]fangchg126@126.com, [b] centres@foxmail.com
[c]wuzhenyumail@163.com, [d] wh5768373@126.com
[e]zwchen@bjtu.edu.cn

Abstract

In order to improve the accuracy of YOLO series algorithm in detecting small ship targets in SAR images, a target detection algorithm based on improved yolov5 is proposed in this paper. In this paper, The Multi-Scale Channel Attention Module (MS_CAM) is added to the network structure to aggregate local and global feature information in the way of channel attention, which can alleviate the problem of large semantic gap between different scales to a certain extent. In addition, the PANet fusion structure in YOLOv5 was replaced by BiFPN structure to make the network better weight of learning features. The experiment on the open RSDD-SAR data set shows that compared with the traditional method, the AP value and recall rate of the whole data set are improved.

Keywords: YOLOv5; SAR; target detection; ship detection;

1. INTRODUCTION

Synthetic Aperture Radar (SAR) [1-2] is a high-resolution imaging radar, which can obtain high-resolution radar images similar to optical photography under weather conditions with extremely low visibility. With the development of Synthetic Aperture Radar (SAR, Synthetic Aperture Radar) imaging technology [3], the resolution of SAR images has been greatly improved, and ship target detection using SAR images has become one of the important applications in the field of marine remote sensing. .

Due to the low accuracy and slow speed of the traditional SAR image recognition technology [4-5] when detecting small ships in the image, the combination of deep learning [6] and target detection tasks in recent years has made the target detection field get a rapid development. In the development of deep learning, the detection model in deep learning is usually divided into two-stage target detection model and single-stage target detection model, two-stage target detection Fast-RCNN [7], Faster-RCNN [8], etc., the R-CNN model can obtain better than CNN Faster recognition speed, and better recognition accuracy. The R-CNN model generally has higher accuracy, but the model is usually more complex, requires longer training, and the detection speed is difficult to meet the real-time requirements. The corresponding single-stage target detection models such as: SSD[9], YOLO[10], etc., only need a single network to be able to complete both positioning and classification at the same time. Such a network must be able to be trained end-to-end, and the speed of reasoning must be faster. However, the accuracy of this method will be reduced compared with the two-stage target detection.

In the remainder of this paper, Section 2 provides and related work. In Section 3, the overall structure of the benchmark model YOLOv5 are introduced in detail. The Section 4 presents the implementation of our proposed improved structure. The Section 5 gives the experimental results and Section 6 summarizes the whole paper.

2. RELATED WORK

In order to further improve the accuracy of target detection, after comparative experiments, this paper proposes a target detection algorithm based on improved YOLOV5 [11-13]. In YOLOv5, the PANet feature fusion structure [14] is used to complete the integration of feature information. However, this method does not highlight the feature representation of the foreground target, and cannot achieve good results in the detection of marine ships. Therefore, this paper introduces the

International Conference on Computer Graphics, Artificial Intelligence, and Data Processing (ICCAID 2022)
edited by Ruishi Liang, Jing Wang, Proc. of SPIE Vol. 12604, 126041X
© 2023 SPIE · 0277-786X · doi: 10.1117/12.2674533

Multi-Scale Channel Attention Module (The Multi-Scale Channel Attention Module, MS_CAM) [15] to aggregate local and global feature information in the way of channel attention, which can highlight the feature representation of foreground targets to a certain extent.

Secondly, because the feature fusion part has a greater impact on the ability to express feature information at different scales. In order for the network structure to learn the importance distribution weight of each feature layer, the original YOLOv5 PANet structure was replaced with a BiFPN structure in the feature fusion stage [16].

This paper uses the public Rotated Ship Detection Dataset in SAR Images, RSDD-SAR dataset [17], in order to compare the verification results under various polarization methods [18], it is divided into four subsets VV according to the polarization method before the experiment, VH, HH, HV. Secondly. In order to improve the generalization ability of the model, some data enhancement methods are adopted, including: mosaic data enhancement [19], random inversion, random image perspective change, color space enhancement and other operations. Finally, the recall rate (recall) and average precision (AP) of the algorithm on the test set are respectively verified [20].

Experimental results prove that the improved network structure can significantly improve the recall rate of ship target detection on various polarization modes and overall data.

3. BENCHMARK MODEL

At present, YOLO series network has been developing rapidly in the field of target detection, and YOLOv5 is the most representative of YOLO series target detection network. YOLOv5 can adjust the number of convolution kernels and the number of stack implementation of different structures of five different sizes of testing model, namely YOLOv5n, YOLOv5s, YOLOv5m, YOLOv5l, YOLOv5x; The smaller the number of model parameters, the lower the detection accuracy, but the faster the running speed. To meet he requirements of real-time detection, the YOLOv5s basic model is adopted.

Compared to other models in the YOLO series, the backbone of YOLOv5 has an SPPF structure added to the end. The structure uses different pooling kernels to complete the maximum pooling operation, and then stitches the feature maps of multiple pooled kernels to obtain the final output result. At the same time, the YOLOv5 model integrates convolution and batch normalization operations in the reasoning stage, which further improves the running speed of the network. In addition, YOLOv5 is optimized for positive and negative sample matching:

- Cross-detection layer matching: Assuming that a ground truth box can match anchors on 2 or even 3 prediction branches, then both 2 or 3 prediction branches can predict the ground truth box, that is, a ground truth box can be predicted by multiple prediction branches.

- Cross-grid matching: Suppose a ground truth box falls in a grid of a certain prediction branch, the grid has 4 neighborhood grids left, up, right and bottom, according to the center position of the ground truth box, the nearest 2 neighborhood grids are also used as prediction grids, that is, a ground truth box can be predicted by 3 meshes.

4. THE IMPROVED YOLOv5

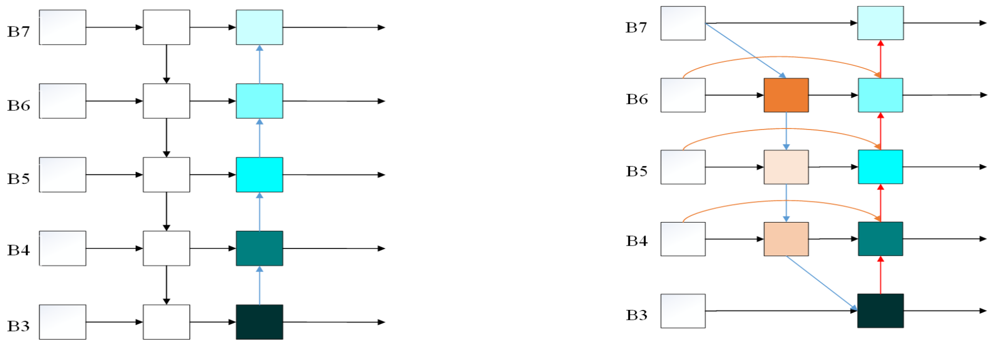
4.1 Multi-Scale Channel Attention Module

In YOLOv5, the PANet feature fusion structure is used to complete the integration of feature information. However, it does not highlight the feature representation of the foreground target, and there is still a large semantic gap between different scales. The Multi-Scale Channel Attention Module aggregates local and global feature information in the way of channel attention, which can alleviate the problem of large semantic separation between different scales to a certain extent.

4.2 Applying the BiFPN to the Model

Feature fusion has a great impact on improving the ability to express features at different scales, so scholars have conducted in-depth research on it. The feature fusion method used in YOLOv5 is PANet. On the basis of FPN, this structure shortens the information flow path and adds different branches to increase the information flow path, thereby improving the performance of the network. Specifically, it is proposed to add a bottom-up path to the top-down path in FPN, as shown in Fig. 4.1(a). Through the bottom-up path enhancement, the information propagation path is shortened, and the precise positioning information of the low-level features is utilized at the same time.

For the BiFPN structure, its structure is shown in Fig. 4.1(b). Compared with the PANet structure, BiFPN deletes those nodes with only one input edge. Its starting point is that if a node has only one input edge without feature fusion, its contribution to the feature network that aims to fuse different features will be smaller; and if two nodes are at the same level, BiFPN will start from the original input Add extra edges to output nodes to fuse more features without adding too much cost Treat each bidirectional (top-down & bottom-up) path as a feature network layer and repeat the same many times layer to achieve higher level feature fusion. In addition, different input features have different resolutions, and feature layers of different scales actually contribute differently. Therefore, BiFPN performs weighted fusion when merging feature layers, so that the network can understand the importance distribution weight of each feature layer learned. Therefore, in our model, we intend to replace the PANet in YOLOv5 with the BiFPN structure.



(a) PANet

(b) BiFPN

Figure 1. Fusion structure of different features

4.3 The improved YOLOv5s network structure

After combining the feature optimization of the above feature fusion part and attention mechanism, finally, the structure of the improved YOLOv5s is shown in Figure 4.2. In order to achieve better recognition effect with the YOLOv5 basic network, the attention module here MS_CAM the location of feature fusion added to the high-level feature map and low-level feature map of YOLOv5. Specifically, since the low-level feature map contains more noise than the high-level feature map, adding the attention module to the low-level feature map effectively improves the model's attention to the foreground target. In addition, in order to apply the BiFPN feature fusion structure to YOLOv5, the neck part of YOLOv5 is improved, as shown in the red dotted line in the figure.

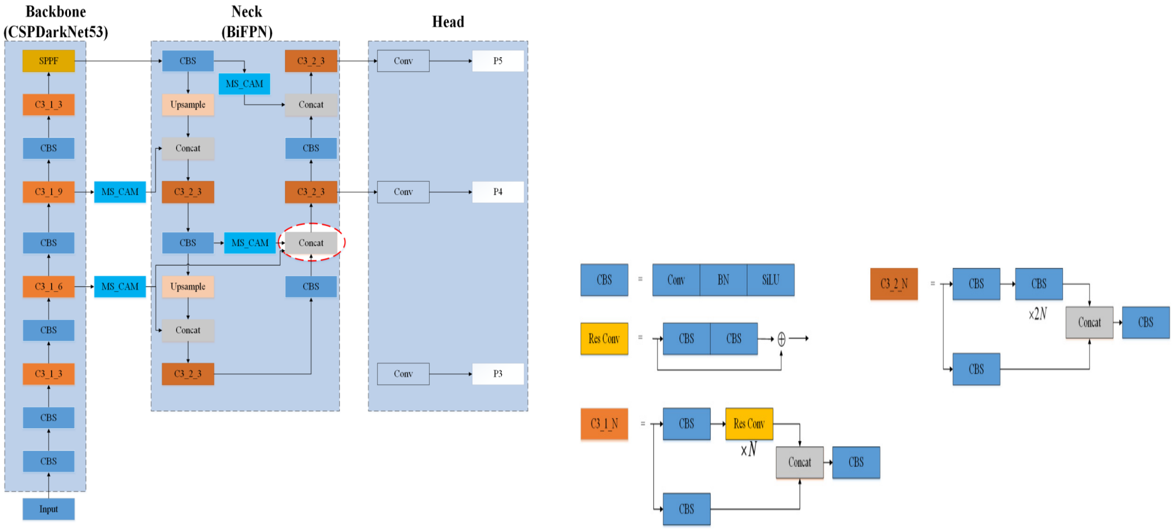


Figure 2. Improved YOLOv5s structure

5. EXPERIMENTS AND VERIFICATIONS

5.1 Experiment Settings

5.1.1 Machine Configurations

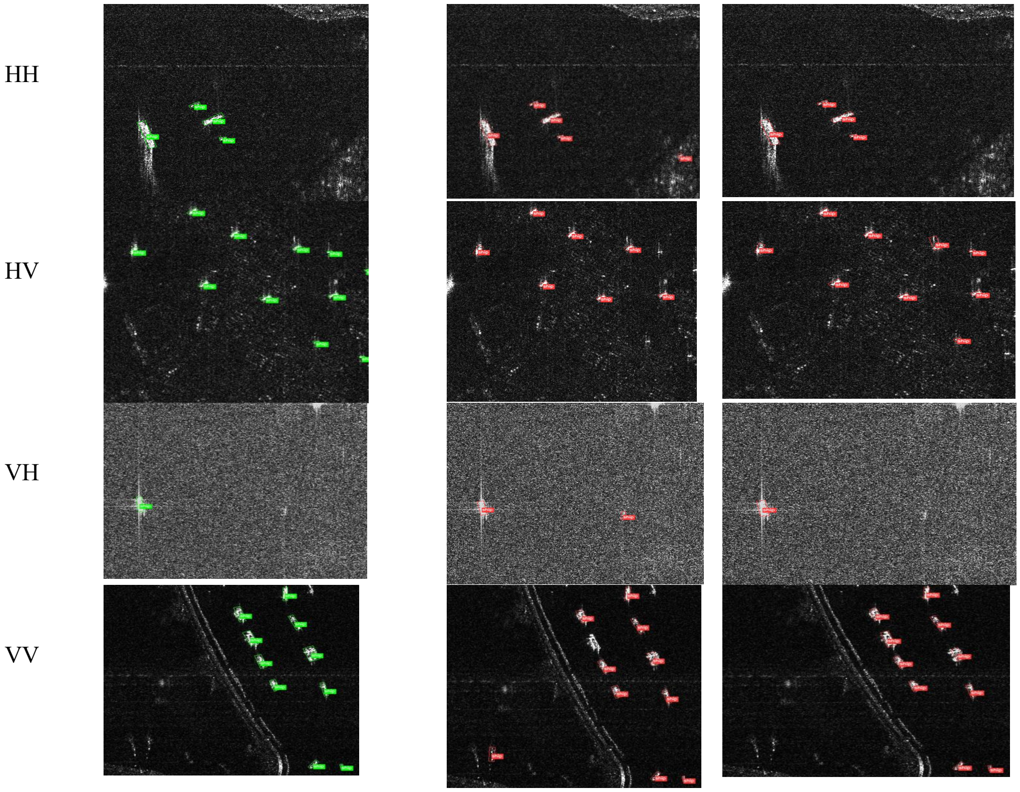
The model adopted in this paper is implemented on the pytorch 1.9.0 framework and the pycham community edition 2020.3.5 linux platform. The model training was conducted in a single GeForce RTX 3090 (24G video memory), CUDA11.1 experimental environment, and the operating system was Linux. We use the pretrained model on the dataset MSCOCO for the network weight of the backbone part of the network. The resolution of the input image in the network is 512 * 512. In the training process, the number of single training pictures (batch size) is set to 8, the SGD optimizer is used, the initial value of the learning rate is set to 0.01, the cosine annealing algorithm is used to update the learning rate, and the training batches (epochs) are set to 120.

5.1.2 Configurations of Polarization mode

First, we used the Rotated Ship Detection Dataset in SAR Images (RSDD-SAR dataset), which is rich in scenes, includes a variety of imaging modes and polarization methods, and has 10,263 ship instances. Introduce different scenes, add elements, and improve the training effect.

Before the training process, we preprocess the dataset, such as using some data enhancement methods, including: mosaic data enhancement, random inversion, random image perspective change, color space enhancement, etc. Then the data slices in the data set are uniformly cut into 512*512 resolution, and the training set and test set are divided into experiments. Finally, in order to introduce different polarization methods to better verify the effectiveness of the algorithm, we divided the training set and test set into four subsets according to four polarization methods (HH, HV, VH, VV).

5.2 Overall Performance Evaluation



(a) real label (b) label on YOLOv5 (c) Label on Improved YOLOv5

Figure 3. All polarization visualization results

Detection is performed on four polarization subsets using the YOLOv5l model and modified YOLOv5. Qualitative comparison, the visual detection results of ship targets are shown in Figure 5.1. Figure 5.1(a) shows the real labels of ships on each subset, and Figure 5.1(b) shows the target detection results of YOLOv5l on each subset. Figure 5.1(c) shows the object detection results of the improved YOLOv5 network on various subsets. We can see that the improved YOLOv5 has a lower probability of false positives than YOLOv5l. Quantitative comparison, from the comparison of recall rate and MAP, it can be seen that the improved YOLOv5 has improved in both aspects.

6. CONCLUSIONS AND FUTURE WORK

This paper presents a target detection algorithm based on improved yolov5. Our method mainly introduces attention mechanism and improves feature fusion structure to improve the accuracy of detecting small ship targets in SAR images. Through comparative analysis of the experimental results, it can be seen that the AP value and recall rate of our proposed model have been improved.

REFERENCES

[1] Brown W M, Porcello L J. An introduction to synthetic-aperture radar[J]. IEEE spectrum, 1969, 6(9): 52-62.

[2] Lan D U, Zhaocheng W, Yan W, et al. Survey of research progress on target detection and discrimination of single-channel SAR images for complex scenes[J]. Journal of Radars, 2020, 9(1): 34-54.

[3] Zhou Y, Wang W, Chen Z, et al. High-resolution and wide-swath SAR imaging mode using frequency diverse planar array[J]. IEEE Geoscience and Remote Sensing Letters, 2020, 18(2): 321-325.

[4] Hou B, Yang W, Wang S, et al. SAR image ship detection based on visual attention model[C], 2013 IEEE International Geoscience and Remote Sensing Symposium-IGARSS. IEEE, 2013: 2003-2006.

[5] Ding B, Wen G, Huang X, et al. Target recognition in SAR images by exploiting the azimuth sensitivity[J]. Remote Sensing Letters, 2017, 8(9): 821-830.

[6] Sun B, Wang X, Li H, et al. Small-target ship detection in SAR images based on densely connected deep neural network with attention in complex scenes[J]. Applied Intelligence, 2022: 1-18.

[7] Bharati P, Pramanik A. Deep learning techniques—R-CNN to mask R-CNN: a survey[J]. Computational Intelligence in Pattern Recognition, 2020: 657-668.

[8] Ren S, He K, Girshick R, et al. Faster r-cnn: Towards real-time object detection with region proposal networks[J]. Advances in neural information processing systems, 2015, 28.

[9] Liu W, Anguelov D, Erhan D, et al. Ssd: Single shot multibox detector[C], European conference on computer vision. Springer, Cham, 2016: 21-37.

[10] Redmon J, Divvala S, Girshick R, et al. You only look once: Unified, real-time object detection[C], IEEE conference on computer vision and pattern recognition. 2016: 779-788.

[11] Xu X, Zhang X, Zhang T. Lite-yolov5: A lightweight deep learning detector for on-board ship detection in large-scene sentinel-1 sar images[J]. Remote Sensing, 2022, 14(4): 1018.

[12] Yang W, Zhang Z. SAR Images Target Detection Based on YOLOv5[C], 2021 4th International Conference on Information Communication and Signal Processing (ICICSP). IEEE, 2021: 342-347.

[13] Jia W, Xu S, Liang Z, et al. Real‐time automatic helmet detection of motorcyclists in urban traffic using improved YOLOv5 detector[J]. IET Image Processing, 2021, 15(14): 3623-3637.

[14] Wang K, Liew J H, Zou Y, et al. Panet: Few-shot image semantic segmentation with prototype alignment[C], IEEE/CVF International Conference on Computer Vision. 2019: 9197-9206.

[15] Wang D, Song Y, Huang J, et al. SAR Target Classification Based on Multiscale Attention Super-Class Network[J]. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 2022, 15: 9004-9019.

[16] Tan M, Pang R, Le Q V. Efficientdet: Scalable and efficient object detection[C], IEEE/CVF conference on computer vision and pattern recognition. 2020: 10781-10790.

[17] Congan X U, Hang S U, Jianwei L I, et al. RSDD-SAR: Rotated ship detection dataset in SAR images[J]. JOURAL OF RADARS, 2022, 11(4): 581-599.

[18] Rapinel S, Betbeder J, Denize J, et al. SAR analysis of wetland ecosystems: Effects of band frequency, polarization mode and acquisition dates[J]. ISPRS Journal of Photogrammetry and Remote Sensing, 2020, 170: 103-113.

[19] Zeng G, Yu W, Wang R, et al. Research on Mosaic Image Data Enhancement for Overlapping Ship Targets[J]. arXiv preprint arXiv:2105.05090, 2021.

[20] Miao T, Zeng H C, Yang W, et al. An improved lightweight retinaNet for ship detection in SAR images[J]. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 2022, 15: 4667-4679.

Proc. of SPIE Vol. 12604 126041X-6

Machine learning-based screening of the relationship between differential genes and immune infiltration in gastric cancer

Fengjin Shang[1.2], Yaxing Liu[1.2], Luyao Chen[1.2], Haoran Zhang[1.2], Changhong Lian[1*]
[1]Department of General Surgery, Heping Hospital Affiliated to Changzhi Medical College, Changzhi, Shanxi 046000, China;
[2]School of Graduate, Changzhi Medical College, Changzhi, Shanxi 046000, China;
*Correspondence should be addressed to changhong lian; lianchanghong0029@163.com

Abstract

Gastric cancer (GC) is the fifth most prevalent malignancy in humans and one of the common causes of cancer-related deaths. In 2020, there will be 1.8 million new cases of GC and 770,000 deaths worldwide, which is the fifth and fourth highest of all malignancies, respectively. The dataset GSE103236, GSE13911, and GSE79973 was downloaded from the GEO database to comparing differentially expressed genes (DEGs) in paired carcinoma tissue and paraneoplastic tissue clinical samples by LASSO and SVM-RFE machine learning methods. Finally, we analyzed the relationship between differential genes and 22 types of immune cells by using the CIBERSORT method, and then provided new ideas for immunotherapy of gastric cancer.

Keywords: Gastric cancer; Bioinformatics; Biomarkers; Diagnosis; Treatment;

1.Introduction

Gastric cancer (GC) is the fifth most prevalent malignancy in humans and one of the common causes of cancer-related deaths. In 2020, there will be 1.8 million new cases of GC and 770,000 deaths worldwide, which is the fifth and fourth highest of all malignancies, respectively [1]. In China, gastric cancer is the second most frequent malignancy and the second cancer-related cause of death, and the incidence is trending younger [2]. The occurrence of gastric cancer follows a multi-stage, multi-factorial process, and its development is the result of a joint action of genetic susceptibility and environmental conditions. The main susceptibility factors include H. pylori infection, smoking, dietary habits, and Epstein-Barr virus (EBV) infection [3]. About 50% of gastric cancers may be triggered by environmental factors, mainly dietary habits and social behaviors [4]. Although most gastric cancers are sporadic, familial aggregation can still be observed in 10% of patients, and of these, approximately 1-3% are associated with known cancer susceptibility syndromes and/or genetic causes [5]. In recent years, the incidence and mortality of gastric cancer have decreased due to the decline in the prevalence of H. pylori and improved dietary habits, but the absolute incidence continues to rise due to the aging of the world population [6]. Although systemic chemotherapy, radiotherapy, surgery, targeted therapy and immunotherapy have all been proven effective in the treatment of gastric cancer [7], GC is a highly malignant tumor with obvious heterogeneity, high recurrence and strong metastasis, therefore the survival rate of gastric cancer is not ideal [8], in addition, the early diagnosis rate of GC is low,most GC patients are already in the local progressive stage at the time of consultation, and the surgical outcome is poor. Therefore, there is an urgent need for new biomarkers to assist us in the early diagnosis of gastric cancer and to provide us with new treatment strategies.

It has been shown that in addition to the genetic susceptibility and environmental factors mentioned above, the tumo r microenvironment (TME) is regulation and control for the invasion behavior of cancers and influences the tumor response to immuno-therapeutic [9], cells in the tumor microenvironment provide energy to other cell types in a paracrine manner and constantly stimulate the growth of tumor cells, which in turn dysregulates cell growth and creates conditions for immune escape and tumor therapy resistance [10]. TME has multiple components, including tumor cells, blood vessels, infiltrating immune cells, stromal cells, tissue fluid and cytokines [11], of which infiltrating immune cells are most relevant to the prognosis of patients with gastric cancer [12]. To this end, comparing differentially expressed genes (DEGs) in paired carcinoma tissue and paraneoplastic tissue clinical samples is important for identifying the role in the regulation of the GC immune microenvironment, in order to study the correlation between between differential genes on infiltrating immune cells for global analysis, which can identify different tumorigenic immune phenotypes and improve the predictive power of immunotherapy.

International Conference on Computer Graphics, Artificial Intelligence, and Data Processing (ICCAID 2022)
edited by Ruishi Liang, Jing Wang, Proc. of SPIE Vol. 12604, 126041Y
© 2023 SPIE · 0277-786X · doi: 10.1117/12.2674517

The role of bioinformatics-based microarray data expression analysis of gene functional expression and other aspects and applications in oncology is very broad [13], and in the area of clinical research, it can also contribute to the diagnosis and treatment of concrete diseases. We establish that there have been conducted to diagnostic genes for multiple tumor types, nevertheless there are fewer reports on machine learning to recognize the association between GC differential genes and immune cells. In this study, we performed a joint analysis in the Gene Expression Omnibus (GEO) database, screened powerful genes using Least Absolute Shrinkage and Selector Operation (LASSO) and Sup-port Vector Machine-Recursive Feature Elimination (SVM-RFE) machine learning methods, and examined differential genes by Gene Ontology (GO), Kyoto Encyclopedia of Genes and Genomes (KEGG), Disease Ontology (DO), and Gene Set Enrichment Analysis (GSEA), to explore the relationship bet ween differential ge nes and immune cells in GC patients.

2. Methods and Materials

2.1 Data Collection from the GEO Database

The microarray dataset GSE103236, GSE13911, and GSE79973 was downloaded from the GEO database. The GSE103236 dataset included 10 of GC tissue and 9 paracancer specimens. whereas the GSE13911 dataset included 38 of GC tissue and 31 cancerous side specimens. For data reliability and reproducibility, we merged the GSE103236 dataset with the GSE13911 dataset. furthermore, the combat function contained in the interior the “SVA” procedures in R was applied for the purpose of eliminate the batch effect. additionally, we also downloaded the GSE79973 dataset in the GEO database for validation of the above data, including 10 pairs of GC tissue samples and the corresponding normal tissue samples.

2.2 Differentially Expressed Genes (DEGs)

The DEG analytics was implemented with the aid ofthe Limma program [14]. aiming to evaluate the changes in gene expression, an empirical moderated t-tests were utilized. The DEGs are genes that had an absolute fold change that was higher than 1 and had an adjusted p value that was lower than 0.05.

2.3 KEGG Pathway and GO Term Enrichment Analysis

The biological properties of differentially expressed genes (DEGs), which consist of biological processes, cellular components, and molecular functions, were investigated using GO enrichment analysis. We use the “org.hs.eg.db” procedure comprised in Bioconductor to annotate the genes. The KEGG pathway enrichment analysis of DEGs was carried out by the “Cluster Profiler” Bioconductor program to determine significance pathways that are most associated with the occurrence and development of GC. The 10 most enriched functions and pathways were selected and visualized using "enrichplot" program insided Bioconductor.

2.4 DO and GSEA enrichment analysis

The enrichment properties of DEGs in disease were studied using DO enrichment analysis, For this study we used the "DOSE" program inside Bioconductor. To study the enrichment properties of DEGs in KEGG, we used the " GSEA Base" Bioconductor program.

2.5 Analysis and Screening DEGs in GC

When doing five-fold cross-alidation, a method known as LASSO and SVM-RFE were employed, separately to filter the significance genes[15.16].

2.6 Inspection of Receiver Operating Characteristics (ROC)

We used the GSE79973 dataset to validate the screened novel biomarkers, while the we used the P ROC work in the R procedures to create Receiver Operating Characteristic (ROC) curves to determine the area under the curve (AUC) for screening value genes and assessment their diagnostic meaning [17].

2.7 Immune Infiltration Research

To study the diverse levels of infiltration of immune cell types between cancerous side tissue and GC tissue. contained B cells, CD4 T cells, CD8 T cells, macrophages, neutrophages and dendritic cells, the “corrplot” program was utilize to obtain the Spearman rank correlation coefficient. To discover the connected with between immune cells and importance genes.

3.Result

3.1 Identification of DEGs and Data Preprocessing

First, we merged the GSE103236 dataset with the GSE13911 dataset and used the "limma" program for de-batching, After that, under the criteria of P-adjustment <0.05 and log2 fold-change (FC) | >1, total number of 1067 differential genes were screened, encompass 551 up-regulated genes and 556 down-regulated genes, (Figure 1.a) and (Figure 1.b) showing the volcano plot and heat map of differential genes, respectively. which show 50 up-regulated genes and 50 down-regulated genes, respectively.

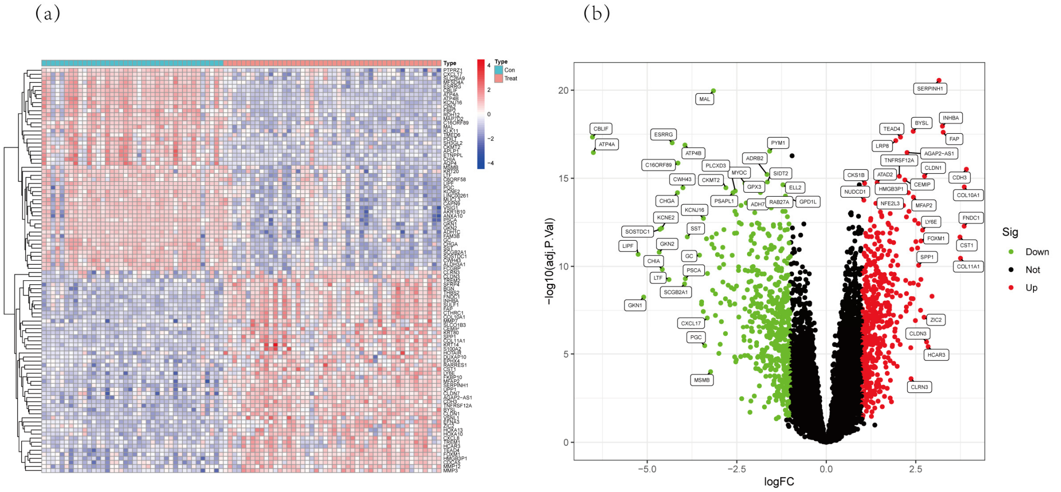


Figure 1: Identification of DEGs in GC. (a) total number of 1067 differential genes were screened, encompass 551 up-regulated genes and 556 down-regulated genes, which were shown in the heat map.(b)which were shown in the Volcanoes map.

3.2 DEGs Functional Enrichment Analysis

The Research shows that the biological process (BP) enrichment was major linked to the regulation of hormone levels，organelle fission，nuclear division。Enriched molecular function (MF) is related to the signaling receptor activator activity, receptor ligand activity, glycosaminoglycan binding。Cellular component (CC) enrichment is related to the apical part of cell, spindle and collagen−containing extracellular matrix (Figure 2.a). Protein digestion and absorption, Cytokine−cytokine receptor interaction and Cell cycles ignaling pathway were linked in KEGG analysis(Figure 2.b)。The conclusions of DO enrichment analysis showed that the differential genes were largely enriched in cell type benign neoplasm, urinary system cancer and non-small cell lung carcinoma. Notably, 16, 28 and 44 genes were significantly enriched in gastritis, stomach carcinoma and stomach cancer, respectively, which verified the reliability of our selected expression microarray data (Figure 2.c). We performed GSEA enrichment analysis of all differential genes, showing enrichment of five signaling pathways in the control and experimental groups, respectively, and the conclusions indicate that: in the control group major enriched in KEGG_DRUG_METABOLISM_CYTOCHROME_P450, KEGG_FATTY_ACID_METABOLISM, KEGG_METABOLISM_OF_XENOBIOTICS_BY_CYTOCHROME_P45, KEGG_NEUROACTIVE_LIGAND_RECEPTOR_INTERACTION and KEGG_RETINOL_METABOLISM (Figure 2.d). In the experimental group primary enriched in KEGG_CELL_CYCLE, KEGG_CYTOKINE_CYTOKINE_RECEPTOR_INTERACTION, KEGG_DNA_REPLICATION, KEGG_ECM_RECEPTOR_INTERACTION and KEGG_PROTEASOME (Figure 2.e).

3.3 Diagnostic Feature Biomarkers

We used LASSO(Figure 3.a) and SVM-RFE(Figure 2.b) to screen for genes of diagnostic value among differential genes respectively：SERPINH1, MAL, FAP, AGAP2-AS1, MYOC, BGN, ACACB (Figure 3.c). Meanwhile, we validated the expression of the screened genes in cancer and paracancer using the GSE79973 dataset to determine the reliability and accuracy of the screened genes. Among them BGN, FAP, SERPINH1 and AGAP2-AS1 were highly expressed in tumors, but the difference between AGAP2-AS1 in Paracancerous tissues and cancer tissues was not statistically significant, p > 0.05. Therefore, we excluded this gene in the next experiments. And ACACB, MYOC, MAL were lowly expressed compared to normal tissues (Figure 4).

3.4 Diagnostic Sensitivity of Valuable Genes in GC

To verify the diagnostic sensitivity of screening novel diagnostic genes for gastric cancer, we used Receiver Operating Characteristics analysis results showed that all of our screened genes showed high diagnostic value. where the AUC of ACACB is : 0.936(9 5% CI: 0.873−0.982), AUC of BGN is: 0.946(95% CI: 0.891−0.987)，AUC of FAP is : 0.953(95% CI: 0.901−0.990), AUC of MAL is: 0.975(95% CI: 0.943−0.996)，AUC of MYOC is: 0.952(95% CI: 0.906−0.989), AUC of SERPINH1is: 0.981 (95% CI: 0.958−0.996) (Figure 5). Then, to validate the authenticity of the screened genes in GC diagnosis, we re-validated the novel biomarkers using the GSE79973 dataset, and the results showed : AUC of ACACB is :0.890(95% CI: 0.720−1.000),AUC of BGN is: 0.920 (95% CI: 0.730−1.000),AUC of FAP is: 0.900(95% CI: 0.700−1.000),AUC of MAL is: 0.960(95% CI: 0.850−1.000),AUC of MYOC is: 0.830(95% CI: 0.620−0.990),AUC of SERPINH1 is: 0.890(95% CI: 0.690−1.000) (Figure 6).

3.5 Immunological infiltration

We finding the infiltration of immune cells by using the CIBERSORT method. We used the composition of 22 common immune cells in GC tissues and the relationship between immune cells. It has been shown that there are significant differences in immune cells in tumor and Paracancerous tissues, including Plasma cells, Macrophages M0, Macrophages M1, Dendritic cells resting and Mast cells resting (Figure 7.a.b.c). Immediately afterwards, we investigated the differential between the level of highly expressed genes and immune cells in gastric cancer tissues. The result is displayed BGN and Macrophages M1, Macrophages M0, Macrophages M2, T cells CD4 memory activated, Dendritic cells resting, T cells CD4 memory resting, Mast cells resting, plasma cells were statistically significant (Figure 8.a, 9.a). Gene FAP is associated with Dendritic cells resting, Macrophages M0, Macrophages M1, Macrophages M2, NK cells activated and T cells CD4 memory activated, Neutrophils, B cells naive, T cells CD4 memory resting, NK cells resting, Mast cells resting and Plasma cells were statistically significant (Figure 8.b,9.b). SERPINH1 and Macrophages M1, Macrophages M0, T cells follicular helper, Dendritic cells resting, Mast cells activated, NK cells activated, T cells CD4 memory activated, Macrophages M2, Neutrophils, Monocytes, T cells CD4 memory resting, Mast cells resting, Plasma cells were statistically significant (Figure 8.c,9.c). The consequences of the immune cell correlation analysis of the above three genes indicates that the expression of the genes was most closely related to macrophages, which showed a moderate positive correlation, followed by T cells CD4 memory activated, which showed a weak positive correlation.

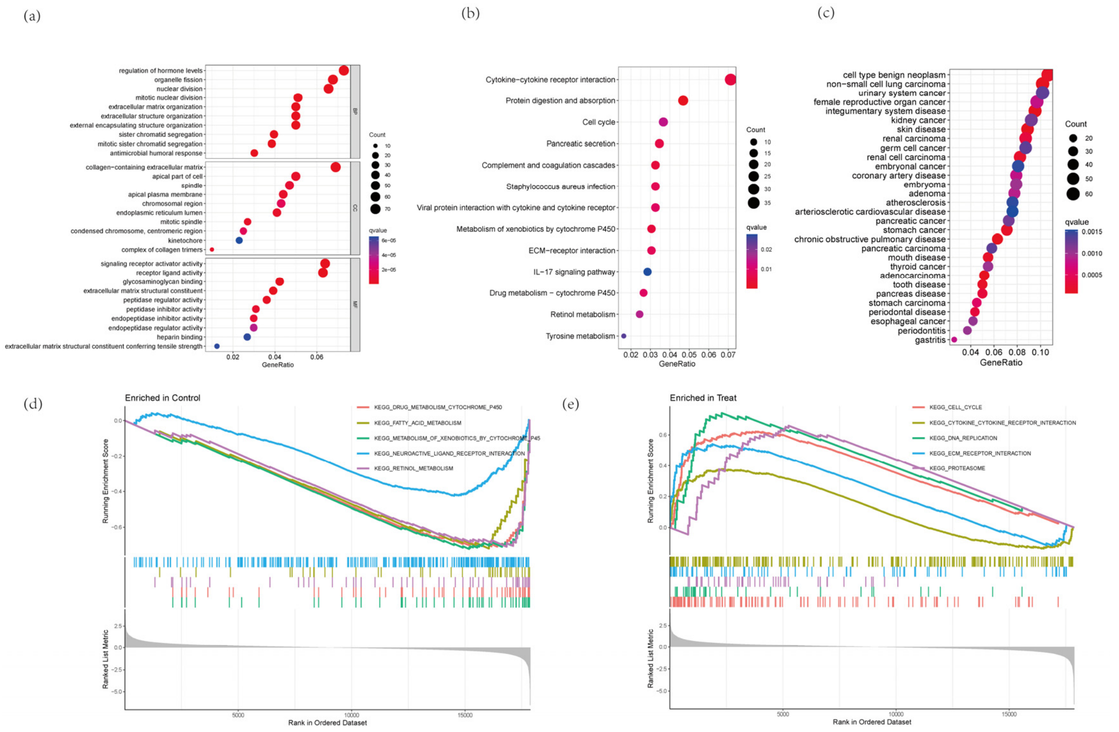


Figure 2: DEGs Functional Enrichment Analysis. (a)The Research shows that the BP, MF, CC enrichment. (b) KEGG analysis.(c) DO enrichment analysis.(d) GSEA enrichment analysis of control groups and experimental groups (e).

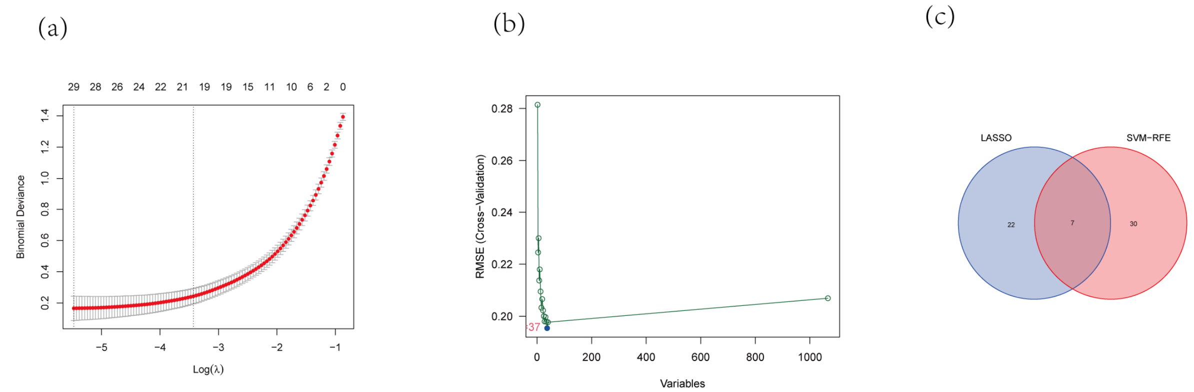


Figure 3: Diagnostic Feature Biomarkers. (a)feature selection in the LASSO. (b) selection by the SVM-RFE approach.(c) Venn diagram diagnostic markers shared via the LASSO and SVM-RFE approach.

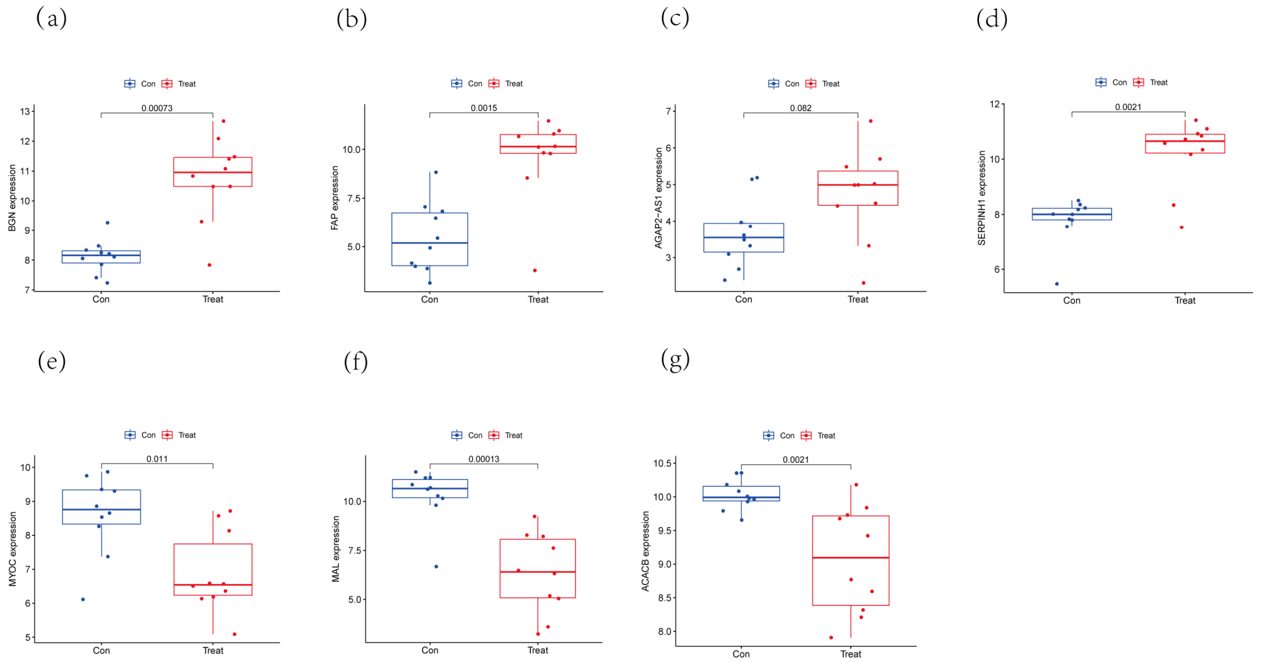


Figure 4: The Expression of the Screened Genes in Cancer and Paracancer Using the GSE79973 dataset, (a-g)

4.Discussion

Gastric cancer, as a solid tumor, is a serious threat to global health. Owing to the lack of specific diagnostic markers, , and therefore, treatment is often unsatisfactory [18]. To improve the prognosis and quality of life of gastric cancer patients, genomic, epigenomic and transcriptomic analyses of cancers using Next-generation sequence (NGS) have revealed the relationship between various malignancies and genomic information, and these studies provide new therapeutic targets for GC [19]. In addition, chemotherapy and targeted therapy have been introduced into clinical patients with gastric cancer and have significantly improved clinical outcomes, but patients' quality of life has not seen substantial improvements due to the toxicity and resistance of chemotherapy drugs [20]. The appearance of immune checkpoint inhibitors (PD-1, PD-L1 and CTLA-4) can lead to significant palliative effect on patients with advanced tumorsrs. In recent years, mismatch repair deficiencies (dMMR) of immune checkpoint inhibitors have been shown to be effective in gastrointestinal tumors [21]. However, in only a fraction of patients， immunotherapy focuses on the awakening of immunologic molecular components in order to counteracting cancerous cells in the TME. Thus, it is indispensable to In-depth research the influence of the TME on GC immunotherapy.

We screened the differential genes by database, in which ACACB, MYOC and MAL are low expressed in gastric cancer and found that ACACB can act as a novel antimetabolic biomarker in colorectal cancer [22], MYOC was confirmed in a previous algorithm study whose findings were consistent with ours [23], and MAL can inhibit gastric cancer progression through the phosphorylation pathway of STAT3 [24]. SERPINH1, BGN and FAP are highly expressed in gastric cancer, where SERPINH1 regulates Epithelial-mesenchymal transition (EMT) by the Wnt/β-catenin signaling pathway and thus promotes the development of gastric cancer [25], BGN has been shown in previous studies and our findings are consistent [26], and FAP plays a key function in many kinds of tumors involving gastric cancer [27]. In this study, we mainly want to use bioinformatics to explore the common relationship between highly expressed genes and immune cells, which in turn will provide us with new options for targeted therapy. It has been shown that carcinoma and stromal cells interact with each other and jointly promote tumor development.

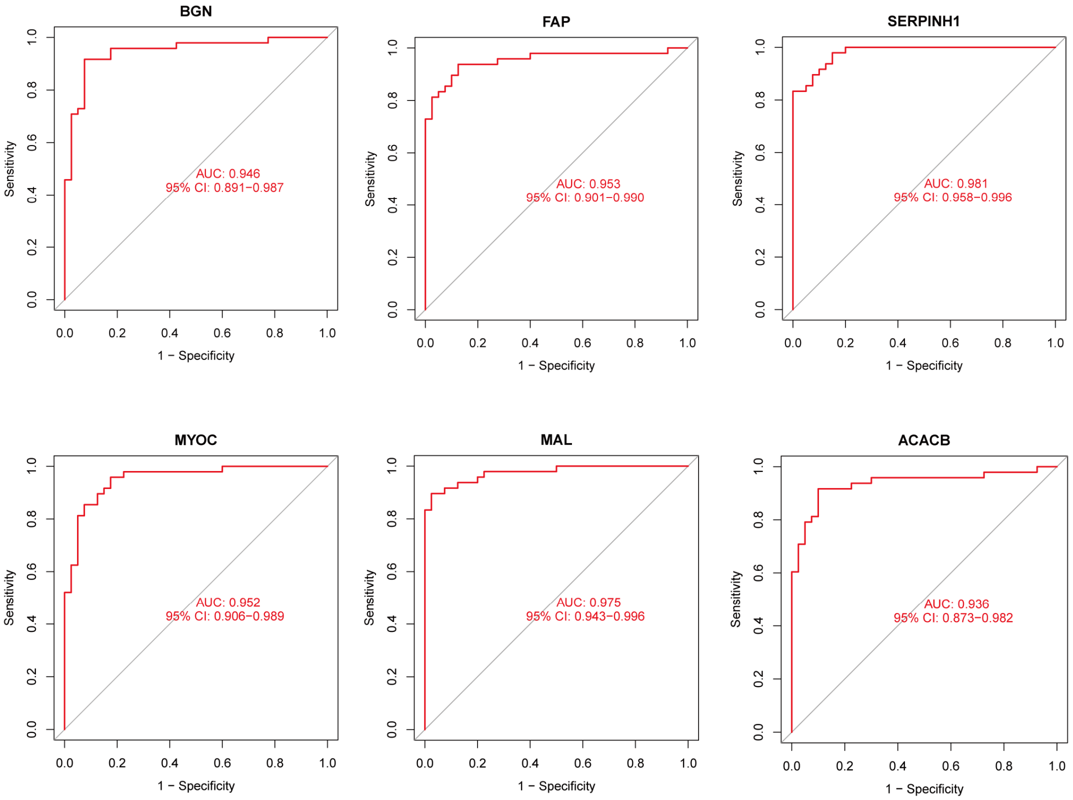


Figure 5: The diagnostic sensitivity of screening novel diagnostic genes for gastric cancer used Receiver Operating Characteristics using GSE103236 and GSE13911 datasets.

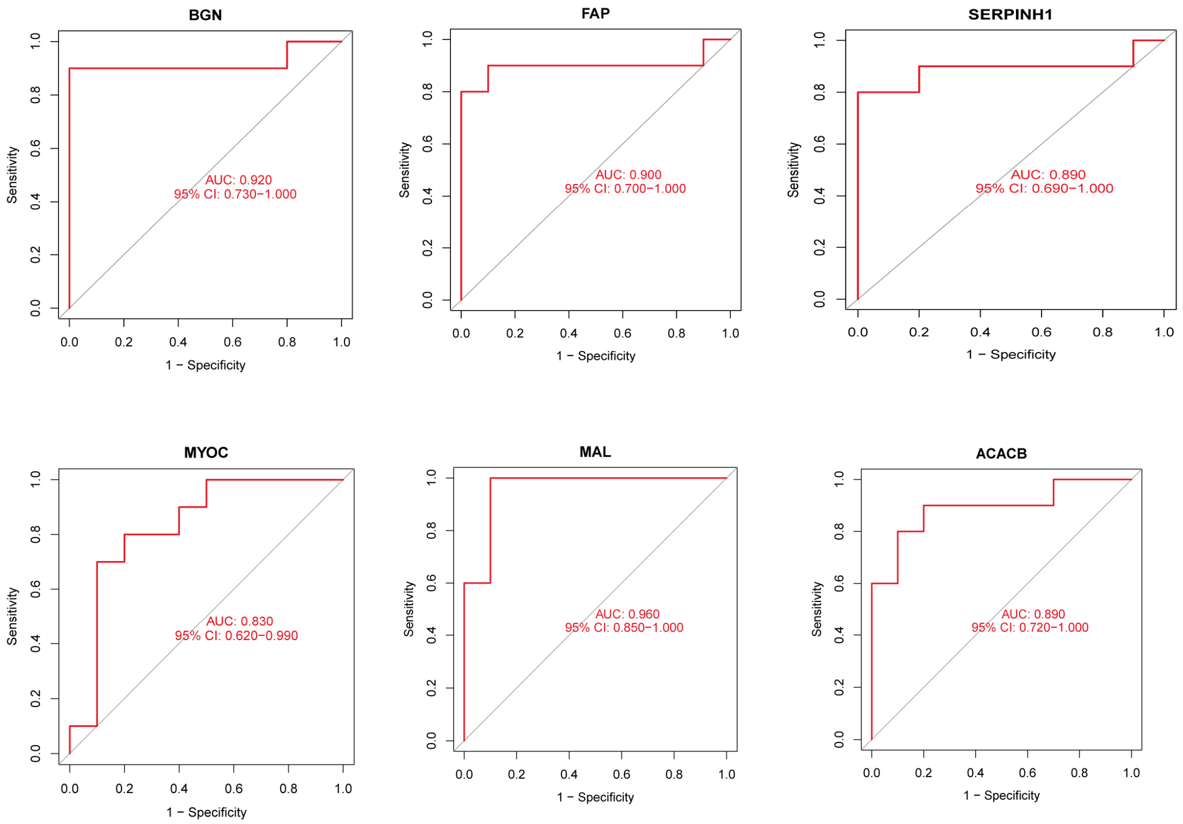


Figure 6: ROC curve of the diagnostic genes using GSE79973 datasets.

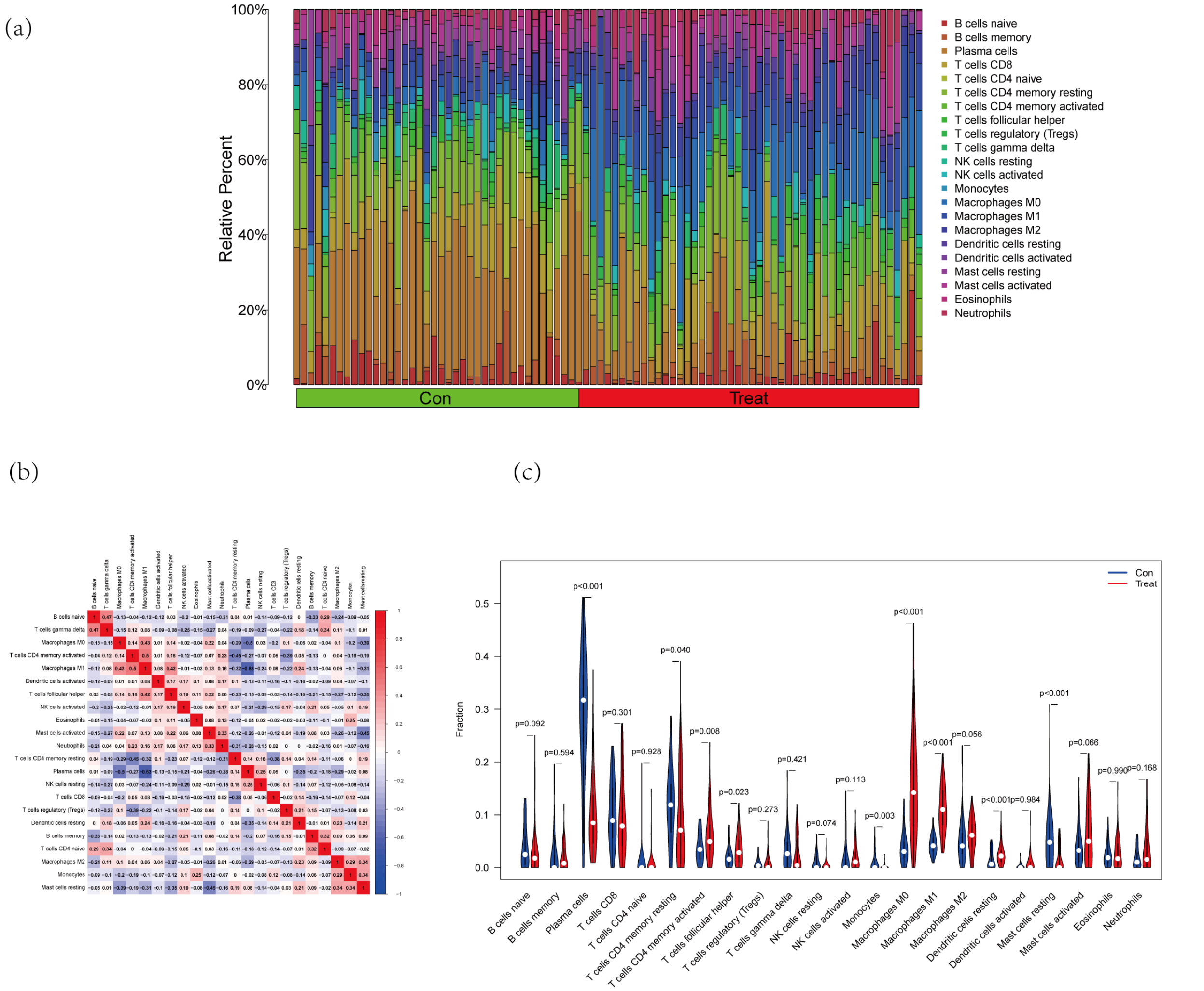


Figure 7: Immunological infiltration. (a)(b)The proportion of the 22 immune cells by the CIBERSORT algorithm.(c) Differential expression of immune cells in cancer and paraneoplastic tissues

Immune infiltration plays an significant function in promoting or inhibiting the growth of GI tumors [28]. We found that the differential gene BGN, which is highly expressed in tumors, was positively correlated with Macrophages M1, Macrophages M0, Macrophages M2, T cells CD4 memory activated, Dendritic cells resting. FAP and Dendritic cells resting, Macrophages M0, Macrophages M1, Macrophages M2, NK cells activated, and T cells CD4 memory activated were positively correlated. SERPINH1 and Mast cells activated, Neutrophils, NK cells activated, T cells CD4 memory activated, T cells follicular helper were positively correlated. All three differentially expressed genes were positively correlated with macrophage infiltration in gastric cancer, which validates the findings of the correlation between macrophages and gastric cancer prognosis as suggested by zunqiang xiao et al [29]. In addition, macrophages have been shown to stimulate angiogenesis, promote tumor cell migration and invasion, and suppress anti-tumor immunity to promote cancer development and progression in a mouse model [30]. Macrophages do not have a defined phenotype or biological activity, but rather are a collection of multiple subpopulations of cells with a wide range of functions and roles under homeostatic and pathological conditions. The multiple functions of macrophages are regulated by the following: developmental origin, tissue environment and acute microenvironment [31]. Currently, conventional macrophage-based tumor therapy has been applied clinically, including radiation therapy, chemotherapy, and immunotherapy. tumour-associated macrophages (TAM) targeted therapies, including TAM depletion, Inhibition of TAM recruitment, In addition, the macrophage role, also have a close relationship with primary and metastatic tumors and interact with other immune cells to influence tumor progression[30].

CD4 T cells play an important role in normal human immunity, rapidly activating the body's immunity in response to viral attack, forming an immune barrier and thus protecting the homeostasis of the internal environment. Studies have shown that CD4 T recognition of tumor antigens has multiple roles in immunotherapy. In mouse models, studies of CD4 T cells have shown that granzyme, MHC class II-dependent contact-dependent cytotoxicity, IFN-y and TNF-a all have certain cell-killing functions. At the same time, in addition to cognate antigen recognition via the TCR, other co-receptors may act together in order to determine the activation set point of these cells, and this particular mechanism may play a role in antitumor [31.32].

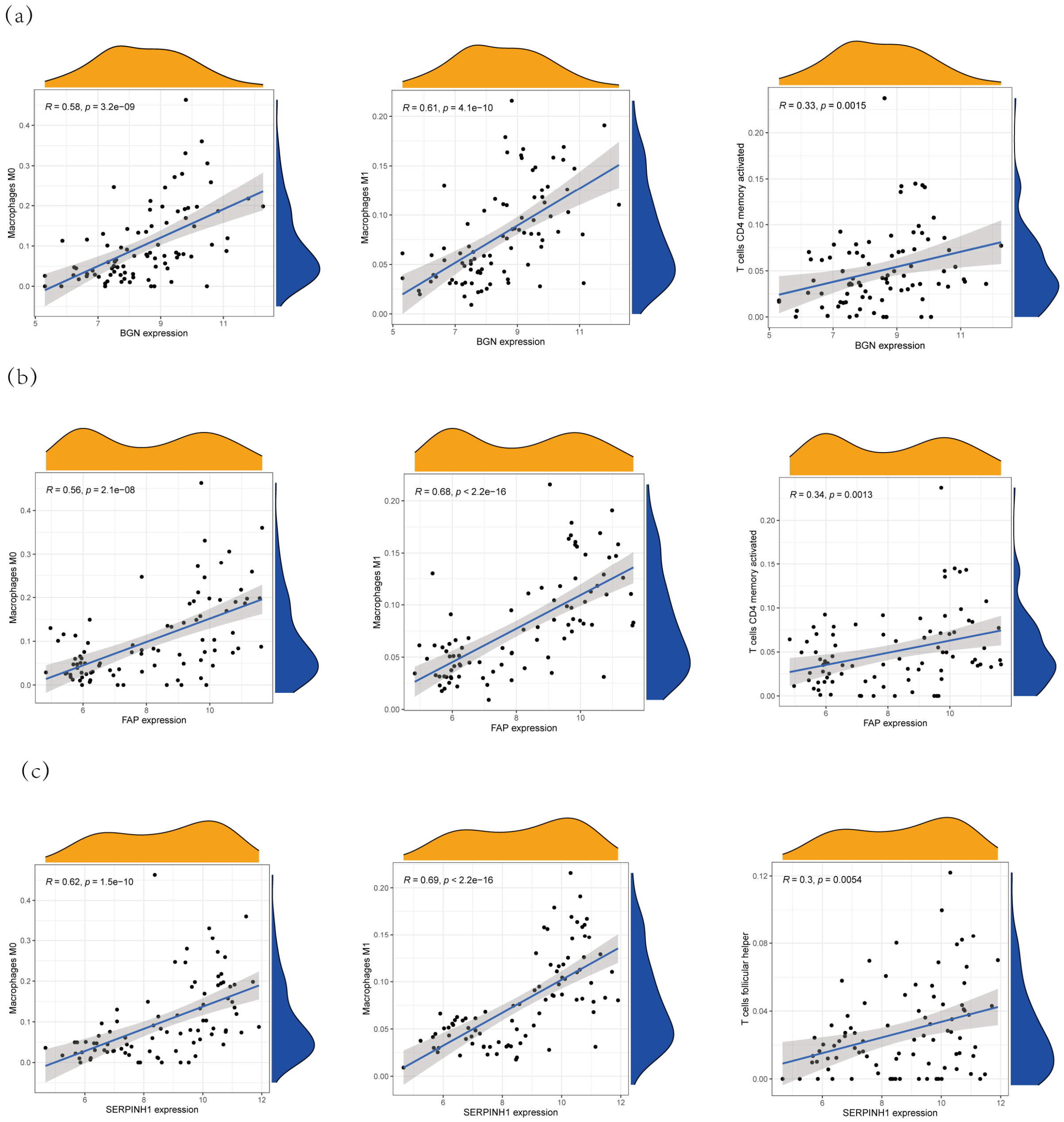


Figure 8: The correlation between (a)BGN, (b)FAP, (c)SERPINH1 and immune infiltration positively correlated with level in GC.

5. Conclusion

We found six genes with statistically significant abnormal expression in gastric cancer, whose phenotypes and functions need further study. Our clinical data and samples show that BGN, FAP and SERPINH1 genes are highly expressed in gastric cancer and are correlated with immune cells, which can be used as prognostic and diagnostic markers for patients with gastric cancer. We concluded that BGN, FAP, and SERPINH1 are involved in the regulation of the gastric cancer microenvironment, especially macrophages and CD4 T. Therefore, more basic studies are needed to further the clinical value of related genes before screening genes into the clinic.

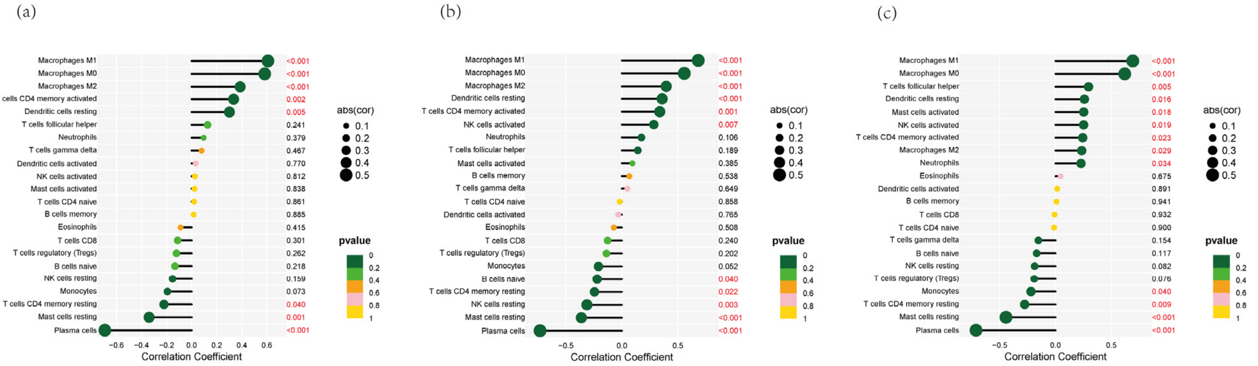


Figure 9: Relationships between (a) BGN, (b) FAP, (c) SERPINH1, and infiltrating immune cells in GC.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Authors' Contributions

Fengjin Shang and Yaxing Liu contributed equally to this work.

References

[1] Sung H, Ferlay J, Siegel RL, et al. Global Cancer Statistics 2020: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries. *CA Cancer J Clin*. 2021;71(3):209-249. doi:10.3322/caac.21660

[2] Cao W, Chen HD, Yu YW, Li N, Chen WQ. Changing profiles of cancer burden worldwide and in China: a secondary analysis of the global cancer statistics 2020. *Chin Med J (Engl)*. 2021;134(7):783-791. Published 2021 Mar 17. doi:10.1097/CM9.0000000000001474

[3] Fukayama M, Abe H, Kunita A, et al. Thirty years of Epstein-Barr virus-associated gastric carcinoma. *Virchows Arch*. 2020;476(3):353-365. doi:10.1007/s00428-019-02724-4

[4] Machlowska J, Baj J, Sitarz M, Maciejewski R, Sitarz R. Gastric Cancer: Epidemiology, Risk Factors, Classification, Genomic Characteristics and Treatment Strategies. *Int J Mol Sci*. 2020;21(11):4012. Published 2020 Jun 4. doi:10.3390/ijms21114012

[5] Spoto CPE, Gullo I, Carneiro F, Montgomery EA, Brosens LAA. Hereditary gastrointestinal carcinomas and their precursors: An algorithm for genetic testing. *Semin Diagn Pathol*. 2018;35(3):170-183. doi:10.1053/j.semdp.2018.01.004

[6] Gullo I, Grillo F, Mastracci L, et al. Precancerous lesions of the stomach, gastric cancer and hereditary gastric cancer syndromes. *Pathologica*. 2020;112(3):166-185. doi:10.32074/1591-951X-166

[7] Joshi SS, Badgwell BD. Current treatment and recent progress in gastric cancer. *CA Cancer J Clin*. 2021;71(3):264-279. doi:10.3322/caac.21657

[8] Gao JP, Xu W, Liu WT, Yan M, Zhu ZG. Tumor heterogeneity of gastric cancer: From the perspective of tumor-initiating cell. *World J Gastroenterol*. 2018;24(24):2567-2581. doi:10.3748/wjg.v24.i24.2567

[9] Bader JE, Voss K, Rathmell JC. Targeting Metabolism to Improve the Tumor Microenvironment for Cancer Immunotherapy. *Mol Cell*. 2020;78(6):1019-1033. doi:10.1016/j.molcel.2020.05.034

[10] Oya Y, Hayakawa Y, Koike K. Tumor microenvironment in gastric cancers. *Cancer Sci*. 2020;111(8):2696-2707. doi:10.1111/cas.14521

[11] Kaymak I, Williams KS, Cantor JR, Jones RG. Immunometabolic Interplay in the Tumor Microenvironment. *Cancer Cell*. 2021;39(1):28-37. doi:10.1016/j.ccell.2020.09.004

[12] Hennequin A, Derangère V, Boidot R, et al. Tumor infiltration by Tbet+ effector T cells and CD20+ B cells is associated with survival in gastric cancer patients. *Oncoimmunology*. 2015;5(2):e1054598. Published 2015 Jun 3. doi:10.1080/2162402X.2015.1054598

[13] Hephzibah Cathryn R, Udhaya Kumar S, Younes S, Zayed H, George Priya Doss C. A review of bioinformatics tools and web servers in different microarray platforms used in cancer research. *Adv Protein Chem Struct Biol*. 2022;131:85-164. doi:10.1016/bs.apcsb.2022.05.002

[14] Ritchie ME, Phipson B, Wu D, et al. limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res*. 2015;43(7):e47. doi:10.1093/nar/gkv007

[15] Cai W, van der Laan M. Nonparametric bootstrap inference for the targeted highly adaptive least absolute shrinkage and selection operator (LASSO) estimator [published online ahead of print, 2020 Aug 10]. *Int J Biostat*. 2020;/j/ijb.ahead-of-print/ijb-2017-0070/ijb-2017-0070.xml. doi:10.1515/ijb-2017-0070

[16] Sanz H, Valim C, Vegas E, Oller JM, Reverter F. SVM-RFE: selection and visualization of the most relevant features through non-linear kernels. *BMC Bioinformatics*. 2018;19(1):432. Published 2018 Nov 19. doi:10.1186/s12859-018-2451-4

[17] Robin X, Turck N, Hainard A, et al. pROC: an open-source package for R and S+ to analyze and compare ROC curves. *BMC Bioinformatics*. 2011;12:77. Published 2011 Mar 17. doi:10.1186/1471-2105-12-77

[18] Onoyama T, Ishikawa S, Isomoto H. Gastric cancer and genomics: review of literature. *J Gastroenterol*. 2022;57(8):505-516. doi:10.1007/s00535-022-01879-3

[19] Cancer Genome Atlas Research Network. Comprehensive molecular characterization of gastric adenocarcinoma. *Nature*. 2014;513(7517):202-209. doi:10.1038/nature13480

[20] Wagner AD, Syn NL, Moehler M, et al. Chemotherapy for advanced gastric cancer. *Cochrane Database Syst Rev*. 2017;8(8):CD004064. Published 2017 Aug 29. doi:10.1002/14651858. CD004064.pub4

[21] Solomon BL, Garrido-Laguna I. Upper gastrointestinal malignancies in 2017: current perspectives and future approaches. *Future Oncol*. 2018;14(10):947-962. doi:10.2217/fon-2017-0597

[22] Hong HJ, Shao Y, Zhang S, et al. ACACB is a novel metabolism-related biomarker in the prediction of response to cetuximab therapy inmetastatic colorectal cancer [published online ahead of print, 2022 Aug 25]. *Acta Biochim Biophys Sin (Shanghai)*. 2022;10.3724/abbs.2022121. doi:10.3724/abbs.2022121

[23] Nation JB, Cabot-Miller J, Segal O, Lucito R, Adaricheva K. Combining Algorithms to Find Signatures That Predict Risk in Early-Stage Stomach Cancer. *J Comput Biol*. 2021;28(10):985-1006. doi:10.1089/cmb.2020.0568

[24] Geng Z, Li J, Li S, et al. MAL protein suppresses the metastasis and invasion of GC cells by interfering with the phosphorylation of STAT3. *J Transl Med*. 2022;20(1):50. Published 2022 Jan 29. doi:10.1186/s12967-022-03254-5

[25] Tian S, Peng P, Li J, et al. SERPINH1 regulates EMT and gastric cancer metastasis via the Wnt/β-catenin signaling pathway. *Aging (Albany NY)*. 2020;12(4):3574-3593. doi:10.18632/aging.102831

[26] Shi Y, Qi L, Chen H, et al. Identification of Genes Universally Differentially Expressed in Gastric Cancer. *Biomed Res Int*. 2021; 2021: 7326853. Published 2021 Jan 21. doi:10.1155/2021/7326853

[27] Fitzgerald AA, Weiner LM. The role of fibroblast activation protein in health and malignancy. *Cancer Metastasis Rev*. 2020;39(3):783-803. doi:10.1007/s10555-020-09909-3

[28] Yang S, Liu T, Cheng Y, Bai Y, Liang G. Immune cell infiltration as a biomarker for the diagnosis and prognosis of digestive system cancer. *Cancer Sci*. 2019;110(12):3639-3649. doi:10.1111/cas.14216

[29] Xiao Z, Hu L, Yang L, et al. TGFβ2 is a prognostic-related biomarker and correlated with immune infiltrates in gastric cancer. *J Cell Mol Med*. 2020;24(13):7151-7162. doi:10.1111/jcmm.15164

[30] Cassetta L, Pollard JW. Targeting macrophages: therapeutic approaches in cancer. *Nat Rev Drug Discov*. 2018;17(12):887-904. doi:10.1038/nrd.2018.169

[31] DeNardo DG, Ruffell B. Macrophages as regulators of tumour immunity and immunotherapy. *Nat Rev Immunol*. 2019;19(6):369-382. doi:10.1038/s41577-019-0127-6

[32] Oh DY, Fong L. Cytotoxic CD4$^+$ T cells in cancer: Expanding the immune effector toolbox. *Immunity*. 2021;54(12):2701-2711. doi:10.1016/j.immuni.2021.11.015

Transient overvoltage suppression strategy for new energy units based on the electronic search algorithm

Zhihao Tian, Ning Mi, Shibin Bai, Fei Meng, Gang Liu, Tong Li, Wenchao Zhang, Xiwen Cui *
State Grid Ningxia Electric Power Co.LTD., Ningxia Daily News Building, No. 47 Zhongshan
South Street, Xingqing District, Yinchuan City, Ningxia, 750001, China
* Corresponding author. Tel.: +86 157-1116-6086.
E-mail address: 1594535384@qq.com

Abstract

This paper proposes an electronic search algorithm-based transient overvoltage suppression method for the transient overvoltage problem caused by DC faults during the outgoing transmission of new energy clusters from high-voltage DC. Firstly, the relationship between the control parameters of the rotor-side converter and the transient voltage rise of the doubly-fed wind turbine is determined by combining the transient overvoltage principle; secondly, the control parameters are optimized and searched based on the electronic search algorithm. Finally, by simulation, the electronic search algorithm is proved to be effective in terms of convergence and parameter optimization as well as the applications of transient overvoltage suppression.

Keywords: DC outfeed system; transient overvoltage; doubly-fed wind turbine; control parameter optimization; electronic search algorithm

1. Introduction

To achieve the goal of carbon peaking and carbon neutrality, since 2008, several kilowatt-level new energy bases have been accomplished in Northwest China, causing the region to have the highest renewable energy percentage in China's power system [1]. A number of large new energy bases are distributed in the "Three Norths" region of China. Located at the end of the power grid whose frame is weak, the energy bases are thousands of kilometers away from the central and eastern load centers. High voltage direct current (HVDC) transmission technology which has good technical and economic advantages in terms of capacity, distance, and flexibility can transmit new energy electricity from the "Three Norths" region [2]. When a typical AC fault or DC fault occurs in the DC transmission system, there is a risk of wind turbine chain off-grid or equipment damage due to transient overvoltage in the wind power convergence area.

When transient overvoltage occurs, the aptitude of the voltage is low at first, then increases to a high value. After the phase change failure, the DC near-zone voltage drops greatly due to the absorption of a large amount of reactive power, and the near-zone fan enters the low-wear link. In the low penetration link, reactive power compensation devices without automatic voltage regulation and the presence of a large amount of reactive power in the filters cause a reactive surplus and then lead to transient overvoltage [3]. After the AC short-circuit fault, the DC current decreases rapidly and cannot recover instantly at the same time as the fault removed, resulting in the reactive power consumption of the converter being less than the reactive power generated by the filter. Reactive power surplus and transient overvoltage will happen [4].

[1, 2] suggest that reducing the reactive power output and increasing the active power output during the low penetration process and restoring the active power as soon as possible after the low penetration process. According to [3, 4], the method of optimizing the internal loop control parameters for doubly-fed wind turbine is qualitatively analyzed. The reactive internal loop proportionality coefficient in the rotor-side converter control can be reduced appropriately to control the reactive power output of the turbine. If the proportional coefficient can be quantitatively analyzed, it will be more conducive to grasp the reactive power output of the turbine. In [5], a double-fed wind turbine rotor-side converter double closed-loop control system is selected as the research object, and the identification of control parameters is realized by using a sine cosine optimization algorithm as the identification algorithm. In [6], the PI parameters of the grid-side controller and rotor-side controller are optimized based on a frequency domain identification method using a direct superposition of pseudo-random signals as the excitation method for the doubly-fed wind turbine controller. If the wide applicability of the step change of manual setting reference value to the controller can be considered, more extensive use will be brought to the method. In [7], a stepwise parameter identification strategy based on an optimization-seeking

International Conference on Computer Graphics, Artificial Intelligence, and Data Processing (ICCAID 2022)
edited by Ruishi Liang, Jing Wang, Proc. of SPIE Vol. 12604, 126041Z
© 2023 SPIE · 0277-786X · doi: 10.1117/12.2674787

Proc. of SPIE Vol. 12604 126041Z-529

algorithm is proposed for the grid-connected inverter of a permanent magnet direct-drive wind turbine generator, which enables the identification of the current inner loop ratio, the integral parameters, and the reactive current support coefficients of the dual closed-loop control system of the grid-side converter. If the algorithm with a better global optimization effect can be adopted, hopefully the identification accuracy will be improved. In reference [8], a multilayer neural network model trained by the standard BP algorithm is used to identify low penetration parameters. If the training samples are increased, the training efficiency of the neural network will be effectively improved and the available neural network model will be obtained quickly.

The Electronic search algorithm (ESA) is a meta-heuristic intelligent optimization algorithm proposed in 2017 [9], which has already been applied in the field of power systems [10,11]. The electron search algorithm is based on the orbital motion of electrons around the nucleus of an atom and maps the process that the electrons around the nucleus will gradually change their orbits to reach the highest energy level molecular state with the process of solving the optimal solution for the objective function to achieve the effect of finding the global optimal solution.

Although some achievements have been made in existing researches, most DFIG transient analysis and fault traversal methods are designed for AC power grid faults, which are not fully applicable to the fault condition of wind power DC external transmission system commutation failure. Moreover, researches on transient overvoltage suppression mainly focus on the DC system side. There are few researches on the active suppression of transient overvoltage by using the reactive voltage regulation capability of wind turbines.

The paper will address the problem of transient overvoltage triggered by AC faults, and achieve the suppression of transient overvoltage by optimizing the wind turbine control parameters in the BPA based on electronic search algorithms and co-simulation with MATLAB and BPA, under the condition of meeting the reactive power demand of wind turbines.

2. Control parameters of double-fed wind turbine

2.1. Main Structure of Double-fed Wind Turbine

As a mainstream wind turbine type, the double-fed wind turbine has a stator connected to the grid and a rotor connected to the grid through a rotor-side converter (RSC) and a grid-side converter (GSC). The main structure of the double-fed wind turbine consists of a wind turbine, a gearbox, a double-fed induction generator, RSC and GSC [12], whose structural schematic is shown in Fig. 1.

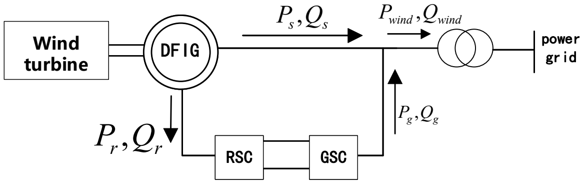


Fig. 1. Structure diagram of DFIG.

In Fig. 1, the arrow points towards the power flow direction, P_s and Q_s are the active power and reactive power emitted by the stator; P_r and Q_r are the active power and reactive power emitted by the rotor; P_g and Q_g are the active and reactive power obtained by the GSC from the grid.

The reactive power Q_{wind} output to the grid from the turbine:

$$Q_{\text{wind}} = Q_{\text{s}} - Q_{\text{c}} \tag{1}$$

As Q_c is small enough and can be ignored, regard Q_{wind} approximately equals to Q_s. Considering GSC has strong influence to Q_s, ignore the influence of GSC and analyzes the control characteristics of RSC.

2.2. Relationship between Transient Overvoltage and RSC Control Parameters

The RSC controls the rotor excitation current to realize the variable speed and constant frequency and voltage operation of the doubly-fed wind turbine, and the active power and reactive power are regulated independently by decoupling control. The typical control block diagram of RSC for a doubly-fed wind turbine is shown in Fig. 2, which contains two parts: the power outer loop and current inner loop.

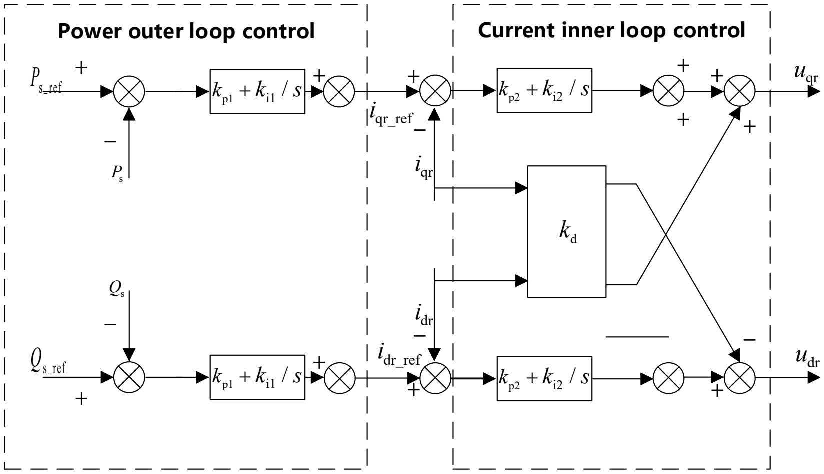


Fig. 2. Block diagram of a typical RSC control system.

P$_{\mathrm{s\_ref}}$ and Q$_{\mathrm{s\_ref}}$ are the reference values of stator active power and reactive power, respectively. P$_{\mathrm{s}}$ and Q$_{\mathrm{s}}$ are the measured active power and reactive power flowing through the stator. i$_{\mathrm{dr}}$, i$_{\mathrm{qr}}$, u$_{\mathrm{dr}}$ and u$_{\mathrm{qr}}$ are the reference values of rotor current and voltage on the d and q axis. i$_{\mathrm{dr\_ref}}$ and i$_{\mathrm{qr\_ref}}$ are the reference values of the stator currents on the d and q axis. k$_{\mathrm{p1}}$, k$_{\mathrm{i1}}$ are the power outer loop control parameters, and k$_{\mathrm{p2}}$, k$_{\mathrm{i2}}$ is the current inner loop control parameters. is the d and q axis coupling term.

To facilitate the study of the effect of RSC parameter changes on the system, the RSC circuit can be transformed into rotor impedance. To further analyze the internal and external loop control parameters, use Davinan Theorem to transform the RSC circuit to a voltage source and impedance connected in series. The RSC Davinan equivalent circuit is shown in Fig. 3.

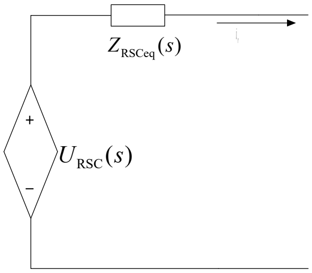


Fig. 3. The RSC circuit Davinan equivalent circuit.

The voltage source and rotor impedance are expressed by equations (2) and (3), where u$_{\mathrm{r}}$ is the rotor voltage, i$_{\mathrm{r\_ref}}$ is the reference rotor current, i$_{\mathrm{r}}$ is the rotor current, i$_{\mathrm{e}}$ is the difference rotor current, which can be defined as $i_{\mathrm{e}}=i_{\mathrm{r\_ref}}-i_{\mathrm{r}}$.

$$U_{\mathrm{RSC}}(s)=\frac{\left(sk_{\mathrm{p1}}+k_{\mathrm{i1}}\right)\left(sk_{\mathrm{p2}}+k_{\mathrm{i2}}\right)\left(P_{\mathrm{s\_ref}}+\mathrm{J}Q_{\mathrm{s\_ref}}\right)}{s^2+\left(sk_{\mathrm{p1}}+k_{\mathrm{i1}}\right)\left(sk_{\mathrm{p2}}+k_{\mathrm{i2}}\right)} \tag{2}$$

$$Z_{\mathrm{RSCeq}}(s)=\frac{s\left(sk_{\mathrm{p2}}+k_{\mathrm{i2}}\right)}{s^2+\left(sk_{\mathrm{p1}}+k_{\mathrm{i1}}\right)\left(sk_{\mathrm{p2}}+k_{\mathrm{i2}}\right)}-\left(s-\mathrm{j}\omega_{\mathrm{r}}\right)\left(L_{\mathrm{r}}-L_{\mathrm{m}}{}^2/L_{\mathrm{s}}\right) \tag{3}$$

L_s and L_r are the stator and rotor winding self-inductance, respectively. L_m is the mutual inductance of rotor and stator windings. ω_r is the speed of the rotor.

Substituting $s = J\omega$ into equation (3), the rotor impedance can be expressed as:

$$Z_{\mathrm{RSCeq}} = \frac{\mathrm{j}\omega\left(\mathrm{j}\omega k_{\mathrm{p2}} + k_{\mathrm{i2}}\right)}{(\mathrm{j}\omega)^2 + \left(\mathrm{j}\omega k_{\mathrm{p1}} + k_{\mathrm{i1}}\right)\left(\mathrm{j}\omega k_{\mathrm{p2}} + k_{\mathrm{i2}}\right)} - \mathrm{j}(\omega - \omega_{\mathrm{r}})\left(L_{\mathrm{r}} - L_{\mathrm{m}}^2 / L_{\mathrm{s}}\right) \tag{4}$$

The relationship between the RSC inner and outer loop control parameters and the rotor impedance is obtained by partial derivatives of Z_{RSCeq} concerning the control coefficients k_{p1}, k_{i1}, k_{p2} and k_{i2}, which can be expressed as：

$$\begin{cases} \dfrac{\partial Z_{\mathrm{RSCeq}}}{\partial k_{\mathrm{p1}}} > 0 \\ \dfrac{\partial Z_{\mathrm{RSCeq}}}{\partial k_{\mathrm{i1}}} < 0 \\ \dfrac{\partial Z_{\mathrm{RSCeq}}}{\partial k_{\mathrm{p2}}} > 0 \\ \dfrac{\partial Z_{\mathrm{RSCeq}}}{\partial k_{\mathrm{i2}}} < 0 \end{cases} \tag{5}$$

From equation (5) it is clear that the partial derivatives of rotor impedance Z_{RSCeq} to k_{p1} and k_{p2} are greater than 0; the partial derivatives of the other two coefficients k_{i1} and k_{i2} are smaller than 0.

The rotor impedance is negatively related to the rotor current i_r, while the stator reactive power is positively related to the rotor current d-axis component i_{dr} [13]. In this way, it can be further concluded that the stator reactive power Q_s has a negative correlation with the rotor impedance Z_{RSCeq}, as shown：

$$\frac{\partial Q_{\mathrm{s}}}{\partial Z_{\mathrm{RSCeq}}} = \frac{\partial Q_{\mathrm{s}}}{\partial i_{dt}} \frac{\partial i_{dt}}{\partial i_{\mathrm{r}}} \frac{\partial i_{\mathrm{r}}}{\partial Z_{\mathrm{RSCeq}}} < 0 \tag{6}$$

Combining with the formula $\Delta U = Q / S$, it can be seen that the transient overvoltage ΔU is correlated with k_{p1}, k_{p2}, k_{i1} and k_{i2}.

3. Optimization of wind turbine reactive power control parameters

3.1. Electronic search algorithm

The electron search algorithm is based on the orbital motion of electrons around the atomic nucleus, mapping the process by which the electrons around the nucleus will gradually change orbits to reach the highest energy level of the molecular state to the process of solving for the optimal solution of the objective function, to achieve the effect of finding the global optimal solution. The specific steps are as follows.

Stage 1: Atomic diffusion stage. The candidate solutions are randomly distributed in the search space, and each candidate solution is similar to an atom. By releasing or absorbing a specific amount of energy, electrons will realize transitions between orbitals.

Stage 2: The orbital transition stage. The electrons around each atomic nucleus will move to higher energy level orbitals to realize the transition of electrons. Each atomic nucleus can be represented as:

$$\begin{cases} e_i = N_i + (2 * rand - 1)\left(1 - 1/n^2\right) \cdot r \\ n \in \{2,3,4,5\} \\ rand \in [0,1] \end{cases} \tag{7}$$

N_i is the current position of the atomic nucleus and is a uniform random number within the range of [0,1], n represents the nearby energy level that each electron around the atomic nucleus can occupy, and r is the orbital radius corresponding to the electron at the initial energy level. Once the electrons occupy these new orbits, the electrons located at the highest energy level (i.e., the best fit value) around the atomic nucleus will be selected as the best electrons.

Therefore, one electron around each nucleus will be selected as the best electron, which can be used to repositioning the corresponding atom in the next step.

Stage 3: Repositioning of the nucleus. The position of the new nucleus is determined by the energy distribution of the emitted photons. For each atomic nucleus, the standard form of the nuclear relocation formula can be expressed in a vector representation as:

$$\vec{D}_{\mathrm{k}} = \left(\vec{e}_{\mathrm{best}} - \vec{N}_{\mathrm{best?}}\right) + Re_{\mathrm{k}} \otimes \left(1/\vec{N}^2_{\mathrm{best?}} - 1/\vec{N}^2_{\mathrm{k}}\right) \tag{8}$$

$$\vec{N}_{\mathrm{new\_k}} = \vec{N}_{\mathrm{k}} + Ac_{\mathrm{k}} * \vec{D}_{\mathrm{k}} \tag{9}$$

In the formula, the symbol $\otimes$ indicates vector-by-vector multiplication element by element. In each iteration, the leap distance D_k is calculated from the current best nuclear position N_{best} of each atomic nucleus, the best electron around the nucleus e_{best}, and the current nuclear position N_k affected by the energy constant Re_k of the Rydberg Equation. Then, the calculated D_k is used in combination with the above formula to calculate the new nuclear position N_{new}, whose convergence speed is determined by the acceleration coefficient Ac. This process is carried out on all atomic nuclei to achieve the repositioning of all atoms to the global optimum point.

Among them, the acceleration coefficient Ac and the energy constant Re of the Rydberg Equation play an indispensable role in the iteration of the atomic nucleus position. After the first set of nuclei have been located, the acceleration coefficient Ac and the energy constant Re of the Rydberg Equation will be iterated by using the OTM method, which can be expressed as follows:

$$Re_{\mathrm{k+1}} = Re_{\mathrm{k}} + \left(Re_{\mathrm{best}} + \sum_{n}^{i=1} \frac{Re_{\mathrm{i}} / f_{N_i|\ Re_i}}{1/f_{N_i|\ Re_i}}\right)/2 \tag{10}$$

$$Ac_{\mathrm{k+1}} = Ac_{\mathrm{k}} + \left(Ac_{\mathrm{best}} + \sum_{n}^{i=1} \frac{Ac_{\mathrm{i}} / f_{N_i|\ Ac_i}}{1/f_{N_i|\ Ac_i}}\right)/2 \tag{11}$$

Where n is the number of atoms, Re_i and Ac_i are the algorithm coefficients of the available kernels, $f_{N|Rei}$, $f_{N|Aci}$ are the fitness function values of these cores, Re_{best} and Ac_{best} are the algorithm coefficients corresponding to N_{best}, Re_{k+1} and Ac_{k+1} are the iterative values of the Rydberg energy constant and the accelerator coefficients, respectively. The trajectory information of all relocated atoms is used to iteratively navigate all other atoms in order to converge to the global optimum. In this algorithm, the maximum number of iterations is selected as the termination criterion.

3.2. Parameter Optimization Based on Co-Simulation of PSD-BPA and MATLAB

In this paper, PSD-BPA and MATLAB are used to optimize parameters based on an electronic search algorithm. The core idea is to carry out power flow calculation on the grid through PSD-BPA to obtain the transient overvoltage voltage rise and the reactive power output value of the wind turbine. The power flow files are read through MATLAB and imported into the electronic search algorithm to update the objective function value, to realize the optimization of parameters k_{p1}, k_{p2}, k_{i1} and k_{i2}. The RSC control parameters of the wind turbine in PSD-BPA are modified according to the optimization results, and the power flow calculation is carried out again. The iteration is repeated until the maximum number of iterations is reached or the objective function value is optimal. The specific steps are as follows:

1) Set the number of iterations and initialize the parameters. In the search space, the total number of electrons m and the problem dimension d will be determined. The initial positions of electrons in the search space are randomly initialized.

2) Calculate the value of the fitness function corresponding to the nucleus and the electron. Wherein, the fitness corresponding to the atomic nucleus is given by formula (7), and the fitness corresponding to the electron is obtained by formula (8). In this paper, the objective is to minimize the difference between the peak value of transient overvoltage at the DC transmission terminal and the reference value of transient overvoltage, and the wind power control parameters to be optimized are taken as the optimization variables to establish the transient overvoltage coordination, optimization model. The objective function is as follows:

$$J\left(k_{\mathrm{p2}},k_{\mathrm{i1}},k_{\mathrm{p2}},k_{\mathrm{i2}}\right)=\min\left(U_{\max}-U_{\mathrm{ref}}\right)$$

3)Update the nuclear position N_{new}. Taking the minimum value of the fitness function of the electron and nucleus for each nucleus, and updating the position of the nucleus in combination with equation (9).

4)The comparison results in an optimal position. If $N_{new} > N_{best}$, the N_{best} value will not be updated; on the contrary, if $N_{new} < N_{best}$, the N_{best} value is updated with N_{new}.

5)Determine whether the maximum number of iterations is reached. If not, the iteration continues and repeats steps 2) to 4); If the maximum number is reached, output the optimal solution.

4. Simulation analysis

In this paper, the wind power of an actual grid is selected as an example via the extra-high voltage outgoing grid, as shown in Fig. 4. The PSD-BPA platform is used for simulation, and to verify the suppression of transient overvoltage by the wind turbine control parameter optimization scheme based on the co-simulation of PSD-BPA and MATLAB proposed in this paper.

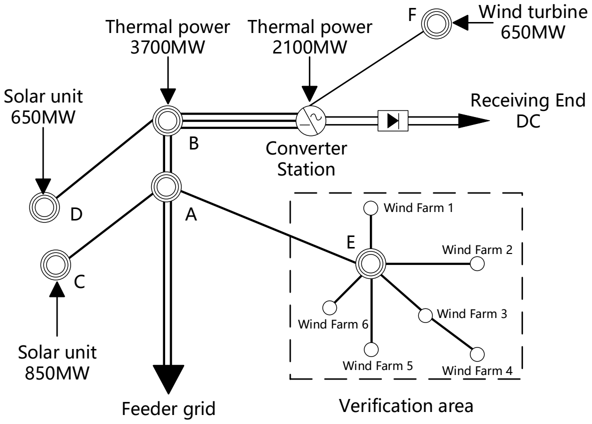


Fig. 4. Diagram of the verification grid in this paper.

As shown in Figure 4, a large number of new energy units are connected to the convergence stations C, D, E, and F in the near area of an HVDC transmission terminal; four thermal power units are directly connected to the converter station, with a total installed capacity of about 2700 MW. In practice we discover that the wind turbine near the convergence station E has the most serious transient overvoltage at the end after the DC phase change failure fault, thus the wind farms near the convergence station E were selected as the research object.

The calculation shows that the three-phase short-circuit fault is set at 1s under the mode of 2100MW of thermal power start-up, 2000MW of DC transmission power, 0.6 wind power simultaneous rate, and 0.8 photovoltaic simultaneous rate in the converter station, and the bus transient overvoltage of the converter station is 1.33p.u. The bus transient voltage growth at the end of each wind farm which connected to the low voltage side of the convergence station E is shown in Table 1.

Table 1. Transient terminal voltage of each wind field after a three-phase short-circuit fault.

Wind Farm	Transient overvoltage at machine terminal /p.u.
1	1.346
2	1.353
3	1.344
4	1.338
5	1.346
6	1.332

From the data in Table 1, it can be concluded that the transient overvoltage at the machine end of wind farm 2 is more serious, so the parameter optimization method based on the ESA algorithm proposed in this paper is used to find the best RSC control parameters for this wind farm to achieve the transient overvoltage suppression effect.

According to Fig. 5, the transient terminal voltage of the wind plant drops from 1.352 p.u. to 1.283 p.u.; the transient overvoltage drops from 0.352 p.u. to 0.283 p.u. According to the demand about the high penetration process in GB/T 19963.1-2021, “the technical regulations for wind power plant connected to the power system part 1: onshore wind”, the wind power plant should ensure that as the synchronized voltage rises to 1.25 to 1.3 times of the nominal voltage, the generators can continuously operate 500 ms without losing synchronism.

Applying the optimized RSC control parameters and the same fault to the other wind farms connected to the convergence station E, the transient overvoltage of the converter station bus is shown as 1.269 p.u., which is much lower than before optimization. The transient overvoltage at the machine end of each wind farm is shown in Table 2.

From the comparison of the data in table 2, it can be concluded that the RSC control parameters optimized by the ESA have a suppression effect on the transient overvoltage.

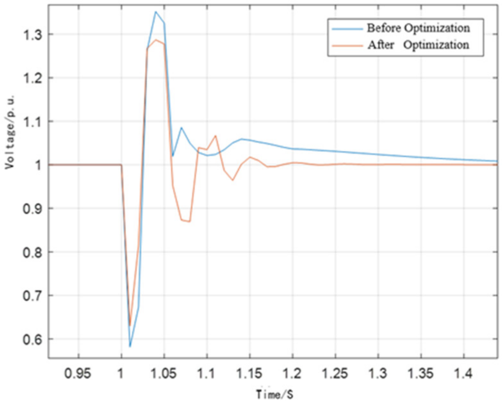


Fig. 5. Comparison curve of terminal voltage before and after optimization.

Table 2. Transient overvoltage at each wind farm before and after optimization after three-phase short-circuit fault.

Wind Farm	Transient overvoltage at machine terminal before optimization /p.u.	Transient overvoltage at machine terminal after optimization /p.u.
1	1.346	1.274
2	1.353	1.281
3	1.344	1.274
4	1.338	1.267
5	1.346	1.272
6	1.332	1.263

5. Conclusion

In this paper, a method is proposed to optimize the control parameters of the wind turbine based on the ESA by using the co-simulation of PSD-BPA and MATLAB to investigate the transient overvoltage problem caused by typical AC faults during the transmission of new energy from large clusters via HVDC. Simulation verification shows that after the optimization of the proposed strategy and algorithm parameters, the transient overvoltage suppression effect of the example system is significantly improved, and the following conclusions can be drawn.

1) The RSC control parameters of the doubly-fed wind turbine have a strong influence on its own reactive power generation Qs, while the turbine's reactive power affects the transient overvoltage voltage rise value; therefore, the RSC control parameters of the doubly-fed turbine have a strong correlation with the transient overvoltage voltage rise value.

2) The electron search algorithm simulates the step process of electrons between energy levels so that the candidate solution can approaches to the global optimal solution. Compared with the classical particle swarm algorithm and genetic algorithm, the electronic search algorithm has good convergence and computational accuracy in the process of wind turbine control parameter optimization. The optimization results can effectively suppress the transient overvoltage, which is of good engineering significance.

Acknowledgements

This work is supported by the Natural Science Foundation of Ningxia (State Grid Ningxia Electric Power Co.LTD.) (Research on the stability characteristics and control strategy of Ningxia power grid with high proportion of new energy and multi-DC transmission terminals/No. B329NX210003).

References

[1] ZHOU X X, CHEN S Y, LU Z X, et al. Technology Features of the New Generation Power System in China[J]. Proceedings of the CSEE, 2018, 38 (07): 1893-1904+2205.

[2] XIN B A, GUO M Q, WANG S W, et al. Friendly HVDC Transmission Technologies for Large-scale Renewable Energy and Their Engineering Practice[J]. Automation of Electric Power Systems, 2021, 45 (22): 1-8.

[3] CAO S S, ZHANG W C, WANG M, et al. Study on Fast Analysis Method Transient Fundamental Frequency Overvoltage of Wind Turbine Generators in Sending System when Serious Power Disturbances Occur in Large-capacity UHVDC[J]. High Voltage Engineering, 2017, 43 (10): 3300-3306.

[4] TU J Z, ZHANG J, ZENG B, et al. HVDC Transient Reactive Power Characteristics and Impact of Control System Parameters During Commutation Failure and Recovery[J]. High Voltage Engineering, 2017, 43 (07): 2131-2139.

[5] Yin C, Li F. Reactive Power Control Strategy for Inhibiting Transient Overvoltage Caused by Commutation Failure[J]. IEEE Transactions on Power Systems, 2021, 36(5): 4764-4777.

[6] Zheng Z, Ren J, Xiao X, et al. Response mechanism of DFIG to transient voltage disturbance under commutation failure of LCC-HVDC system[J]. IEEE Transactions on Power Delivery, 2020, 35(6): 2972-2979.

[7] PAN X P, WEN R C, JU P, et al. Decoupling Estimation of Parameters in Rotor Side Controller of DFIG-based Wind Turbine by Frequency Domain Method[J]. Automation of Electric Power Systems, 2015, 39 (03): 634-638

[8] ZHAO K, ZHANG Z X, ZHOU Y, et al. Low Voltage Ride-through Control Paraments Identification Method for Permanent Magnet Synchronous Generator Based on Multi-layer Neural Network[J]. Shandong Electric Power, 2022, 49 (04): 1-6.

[9] Atomic orbital search: A novel metaheuristic algorithm. Appl Math Model 2021, 93, 657-683.

[10] WANG L P, LI N N, YAN X R, et al. Cascade reservoirs' joint optimal operation of power generation based on improved electro-search algorithm[J]. Control and Decision, 2020, 35 (08): 1916-1922.

[11] HUANG T X, CHEN B, GUAN X, et al. Reactive Power Optimization of Power System Based on Electronic Search Algorithm[J]. Electronic Science and Technology, 2019, 32 (01): 58-61+71.

[12] WU X, GUAN Y J, NING W, et al. Mechanism of Interactive Effect of RSC Parameters in DFIG on SSO and its Application[J]. Power System Technology, 2018, 42 (08): 2536-2544.

[13] Zheng Z, Ren J, Xiao X, et al. Response mechanism of DFIG to transient voltage disturbance under commutation failure of LCC-HVDC system[J]. IEEE Transactions on Power Delivery, 2020, 35(6): 2972-2979.

Re-exploring the Effect of Bilingualism on Inhibitory Control: Application of Survival Analysis in Flanker Task

Geqi Qi*, Fangfang Qin, Xuedi Zhang, Ting Yun
School of Psychology
Inner Mongolia Normal University
Hohhot, Inner Mongolia, China
* Corresponding author: qigeqi@imnu.edu.cn

Abstract

The existence of bilingual advantage has been under debate because of the inconsistent findings across different studies. Two important factors contributing to this controversy are traditional measurements of bilingualism and the granularity of behavioral data in cognitive tasks. In this study, we address this issue by operationalizing bilingual experience as a multifaceted spectrum and using the Cox proportional hazard model for analyzing behavioral data. Unlike the traditional analysis, the method enabled the inclusion of both correct and incorrect responses. And the results showed that behavioral performance got better with the increased overall bilingual level in the flanker task. Additionally, we found that the active L2 usage time in the immersiona environment was a more reliable predictor of bilingual advantage than the overall bilingual level of traditional analysis.

Keywords-bilingual; inhibitory control; executive functions; flanker task

1. Introduction

Psychological studies in bilingualism have established that the continued practice of language control (e.g. switching between two languages) generalizes to non-linguistic cognitive domains[1,2]. However, numerous studies carried out in recent years indicated that they did not replicate these findings and therefore questioned the robustness of the bilingual advantage [3,4]. It is generally described as a bilingual advantage. However, many studies in recent years have reported that they failed to replicate these findings, and therefore questioned the robustness of the bilingual advantage [1].

Traditionally, studies on the impact of bilingualism have been compared to those of monolinguals. However, language experience of the bilinguals can be highly diverse, making the results of these studies vary from one to another. Thus, one of the main factors that make it difficult to get consistent results on bilingual advantage is the heterogeneity of the bilingual population. To resolve this issue, researchers are beginning to explore bilingualism as a continuum and examine the influences of various individual factors within bilingual population [1]. Currently, there are two main instruments that are used to assess the level of bilingualism: the Language Experience and Proficiency Questionnaire (LEAP-Q) and the Language History Questionnaire (LHQ). A recently developed Language and Social Background Questionnaire (LSBQ) has many features similar to those of the previous two instruments, but may also provide a composite factor score which represents the overall level of bilingualism. [5].

A bilingual advantage in cognitive tasks is usually evaluated by comparing average response times between task conditions and between groups of participants. However, the premise of this assessment is to assume that the data is normally distributed, which is not the case. To solve this problem, data, deviating from the mean by more than two standard deviations are usually deleted to give a normal distribution for each participant. Also, the traditional approach cannot take both accuracy and response time into account. Any trial for which the response is incorrect is deleted from the response time analysis. Therefore, if one group responded less accurately, there will be fewer data points in the response time analysis. Hence, the analysis cannot take this disparity into account, and the information contained in the incorrect trials is lost [6].

Analytical issues described above can be addressed by using Cox proportional hazard model (Cox PH model) to analyze behavioral responses to cognitive tasks. Suppose participant A is good for inhibitory control and takes X time to successfully complete a particular trial in the incongruent condition. We can assume that participant B, who is bad at inhibitory control, takes longer to answer correctly than participant A. It's also possible that participant B takes the same length of time (or less) compared to participant A, only with an incorrect response. It is not probable that participant B can

International Conference on Computer Graphics, Artificial Intelligence, and Data Processing (ICCAID 2022)
edited by Ruishi Liang, Jing Wang, Proc. of SPIE Vol. 12604, 1260420
© 2023 SPIE · 0277-786X · doi: 10.1117/12.2674550

answer correctly in less time than participant A in this trial. The Cox PH model allows us to capture this by including both the time needed to respond correctly and the time needed to respond incorrectly. The response time in an incorrect trial is censored and construed as the minimum time necessary to give a correct response in this trial.

The goal of this study is to investigate the relation between inhibitory control and the overall level of bilingualism using both traditional analysis and Cox PH model. LSBQ was used to measure the diverse bilingual experience of the participants, detailing not only the duration of second language (L2) exposure, but also the degree of active use of L2. A composite factor score representing the overall bilingual level was calculated for each participant. To make a distinction between the effect of bilingualism and that of other factors such as age of acquisition (AoA), sex and intelligence, all factors were incorporated into the analysis. For universal comparability, flanker task was used to assess inhibitory control of the participants.

2. Method

2.1 Participants

Sixty five healthy, right handed bilingual adults participated in the study (Table 1) [7]. All participants completed Raven's Standard Progressive Matrices task (RVMT) for the control of intelligence and nonverbal spatial reasoning ability [8].

TABLE 1. Distribution in gender, age and RMVT score

	Gender	Age	RVMT
Min.	Female (n=49)	18.00	32.00
Max		52.00	52.00
Mean		32.76	42.20
SD		7.92	4.89
Min.	Male (n=15)	23.00	33.00
Max		43.00	54.00
Average		29.13	42.93
SD		5.83	5.73

Bilingual experiences were measured using LSBQ (Table 2). All participants spoke English as their L2, while using a variety of native language (L1). The Oxford Quick Placement Test (QPT) results indicate that all participants had a good command of English. (Table 2). The distribution of the LSBQ scores used in this study is shown in Table 2. Factors like AoA in L2 (F1), L2 exposure/use in a home setting (F2) and social community setting (F3) were treated as measures of the duration of L2 language exposure. The number of years actively using the L2 (F4) and the amount of time actively using the L2 in immersion environment (F5) were considered as factors related to degree of active use of L2. The composite factor score was calculated by adding the LSBQ factor scores weighted by the variance for each factor.

2.2 Task

Participants completed flanker task, which was presented by E-Prime 2.0 Professional. They were asked to respond to the direction of a red target arrow, encircled with white symbols on a black background. The task consisted of six blocks, with 72 trials in each of them. Three of the blocks were 'mixed', and the remaining three blocks were 'congruent', 'control' and 'neutral' respectively. The numbers of congruent and incongruent trials in the mixed blocks were same. In a congruent block, the flank arrows were aligned with the target arrow. In the neutral block, double-sided arrows surrounded the arrow. In the control block, a single arrow appears in the centre of the screen. Trial sequence was randomized across all blocks. Flanker effect was defined as the contrast between the congruent condition and incongruent condition within the mixed block, with the congruent condition was set as a baseline. For the mixing cost, the congruent block was fixed as the reference level and compared to the congruent Mixed block. Lastly, for the facilitative effect, the baseline was neutral and contrasted with the congruent block.

TABLE 2. BILINGUAL EXPERIENCE SCORES

	QPT (%)	F1	F2	F3	F4	F5	Composite Factor Score
Min.	51.7	0	-7.15	10.77	0.96	0.1	2.02
Max	100	22	16.7	74.53	30.08	287.89	24.28
Average	88.39	8.31	2.55	51.66	10.24	59.28	10.41
SD	10.93	4.65	5.09	11.38	5.05	60.93	4.04

2.3 Analysis

In the traditional analysis, stepwise multiple linear regression was used to model the additive effect for all predictors, thereby identifying the specific effect of bilingual experience. In the new method of analysis, we exploited the Cox PH model to capture both response time and accuracy within a single flanker task performance, and explore the additive effect of all predictors.

3. RESULTS

3.1 Traditional analysis of behavioral response

Accuracy rates were high under all conditions (97.84%, 1.44). The response time (RT) is shown in Table 3. Incorrect trials and trials using RTs of less than 200ms were excluded from the analysis, resulting in the elimination of 2.2% of the overall trials. We performed separate analyses on the three effects of interest, namely the flanker effect, facilitation effect, and mixing cost.

To determine the explanatory value of the predictors, models were constructed using a stepwise multiple linear regression analysis for each of the three age-specific effects depending on their age, sex, RVMT scores, and composite factor scores. As shown in Table 3, the analysis produced two regression analysis models for each effect, with the second model yielding the higher percent of variance explained (flanker effect: 4%, $R^2 = 0.04$; facilitation effect: 16%, $R^2 = 0.16$; mixing cost: 5%, $R^2 = 0.05$). Model validity was determined using a Durbin–Watson D statistic., with D = 1.47 for flanker effect, D=2.16 for facilitation effect, and D=2.48 for mixing cost. The value of t was associated with an error probability below 0.05 for the model variables (age, sex, RVMT, and composite factor score). As the standardized coefficients show, it is age that has the highest explanatory value of the three effects. However, contributions of age and composite factor score were significant only for the facilitation effect. According to the tolerance indicators and VIF, there was no collinearness between the model variables.

TABLE 3. STEPWISE MULTIPLE LINEAR REGRESSION ANALYSIS

			Flanker Effect			
		Step1			Step2	
	b	SE	β	b	SE	β
Age	0.27	0.27	0.13	0.29	0.27	0.14
Sex	-1.08	4.76	-0.03	-0.45	4.80	-0.01
RVMT	0.18	0.39	0.06	0.19	0.39	0.06
Composite Factor Score				-0.49	0.50	-0.13
ΔR^2		0.02			0.04	
			Facilitation Effect			
		Step1			Step2	
	b	SE	β	b	SE	β
Age	1.21	0.58	0.26*	1.29	0.57	0.276*
Sex	-10.73	10.26	-0.13	-8.37	10.19	-0.10
RVMT	1.61	0.85	0.23	1.67	0.84	0.239*
Composite Factor Score				-1.82	1.06	-0.21
ΔR^2		0.122*			0.16	
			Mixing Cost			
		Step1			Step2	
	b	SE	β	b	SE	β
Age	0.50	0.48	0.14	0.55	0.48	0.15
Sex	-1.69	8.55	-0.03	-0.10	8.56	0.00
RVMT	0.20	0.71	0.04	0.24	0.70	0.04
Composite Factor Score				-1.23	0.89	-0.18
ΔR^2		0.02			0.05	

* $p<0.05$

In order to exploit the effects of more specific bilingual experience factors on behavioral performance, another stepwise multiple linear regression analysis was conducted with the bilingual experience factors, namely the Qpt score, AoA in L2 (F1), L2 exposure/use in a home setting (F2) and social community setting (F3) were treated as measures of the duration of L2 language exposure. The number of years actively using the L2 (F4) and the amount of time actively using the L2 in an immersion environment (F5). The analysis also produced two regression analysis models for each effect, with the second model yielding the higher percent of variance explained. (flanker effect: 17%, $R^2 = 0.17$; facilitation effect: 23%, $R^2 = 0.23$; mixing cost: 10%, $R^2 = 0.10$). Model validity was determined using a Durbin–Watson D statistic, with D = 1.80 for flanker effect, D=2.20 for facilitation effect and D=2.47 for mixing cost. The t value has been associated with an error probability of less than 0.05 for the model variables. As shown by the standardized coefficients, only F5 (i.e. The amount of time actively using the L2 in immersion environmento) had significant and the greatest explanatory value for the flanker effect. According to the tolerance indicators and VIF, there was no collinearness between the model variables.

3.2 Cox proportional hazard model

The optimal Cox PH modisfor the data are summarized in Table 4. As would be expected, performance was better in the congruent condition ($\chi^2(1) = 1274.88$, $p < 0.0001$). The best predictor of performance was the trial type. (congruent vs. incongruent), followed by RMVT score (χ^2 (1) = 326.36, $p < 0.0001$). Male participants seemed to yield better performance than female participants ($\chi^2(1) = 82.23$, $p < 0.0001$), whereas younger participants yield better performance than older participants (χ^2 (1) = 57.00, $p < 0.0001$). There was also a trend for participants with a higher composite factor score to do better than those with a lower composite factor score ($\chi^2(1) = 11.95$, $p < 0.005$), indicating behavioral performance got better with increased overall bilingual level.

TABLE 4. THE COEFFICIENTS OF THE COX PROPORTIONAL HAZARD MODEL FITTED TO THE TIME TO CORRECT RESPONSES

	Coefficient					
	Coefficient	SE	Chi_sq	Df	p	RR (95%CI)
Age	-0.009	0.001	57.001	1.000	0.000	0.991(0.989~0.993)
Sex	-0.213	0.023	82.233	1.000	0.000	0.808(0.772~0.846)
RVMT	0.032	0.002	326.356	1.000	0.000	1.033 (1.030~1.037)
Trial Type	0.700	0.020	1274.883	1.000	0.000	2.014(1.938~2.093)
Composite Factor Score	0.007	0.002	11.945	1.000	0.003	1.007(1.003~1.012)

3.3 Inadequate bilingual group vs. Adequate bilingual group

We further used a composite factor score as a criterion to divide participants into groups with different levels of bilingualism. Since all of the participants in this study were bilinguals, we first grouped them into five equal bins according to their composite factor scores (Table 5). The categorical classifications were validated by the significant chi-square analysis ($\chi^2(4) =64$, $p < 0.0001$). The minimum and maximum composite factor scores for each quantile are shown in Table 5. Individuals with a composite score below 9.18 are classified as an inadequate bilingual group, while those with a composite score above 10.75 are classified as an adequate bilingual group. Those in between can be classified as not highly differentiated or rejected.

As we can see from the Cox PH analysis showed that adequate bilingual group yields better performance than the inadequate bilingual group ($\chi^2(1) = 107.15$, $p < 0.0001$), indicating thn aat participants from adequate bilingual group respond correctly faster than those from the inadequate bilingual group (Fig.1).

TABLE 5. COMPOSITE FACTOR SCORES DISTRIBUTED IN FIVE QUANTILES

	Quentile1	Quentile2	Quentile3	Quentile4	Quentile5
Num	12	14	13	13	12
Min	2.02	7.24	9.45	***10.75***	14.06
Ave	4.97	8.53	10.04	12.17	16.52
Max	6.77	***9.18***	10.66	13.77	24.28

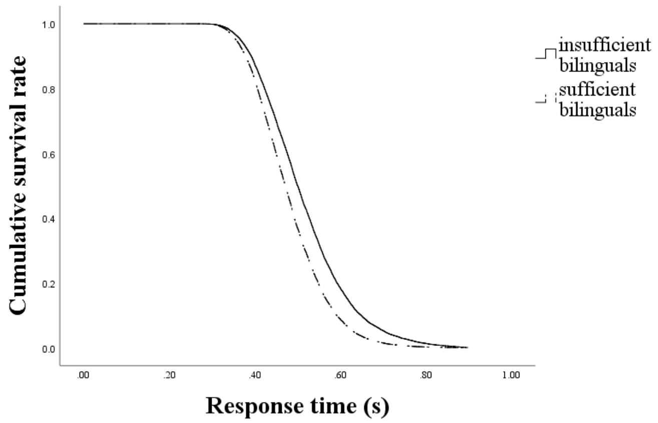


Figure1. Cumulative survival rate of response time during flanker task.

4. DISCUSSION

The present study examined the relation between inhibitory control and bilingualism using both traditional analysis and the Cox PH model. The results provide insight into the current bilingual advantage controversy by specifically pointing to the importance of capturing multiple aspects of bilingual experiences and the use of an analytical method that includes all responses without excluding incorrect trials.

As we can see from traditional analysis (Table 3), participants demonstrated the expected behavior for all measured contrasts (flanker effect facilitation effect and mixing cost). However, they were not modulated in accordance with their overall level of bilingualism (i.e. composite factor score). Previous studies indicated that test and retest reliability of flanker task, and utilization of RT difference scores was typically low. [4]. Thus, behavioral measures like comparing flanker effect can have the inadequate granularity to capture patterns of inhibitory control.

However, when we looked for the effect of specific bilingual experience factors on the task performance, the active L2 usage time in the immersion environment had a significant explanatory value of the flanker effect. This suggests that the intensity of L2 usage may be a more reliable predictor of the bilingual advantage than the overall bilingual level in the flank task. Thus, the more time bilinguals spend actively using L2 in an immersion setting, the more likely they will show significantly better performance during cognitive tasks. Indeed, previous studies have suggested that immersion in a second language environment can be seen as an explanation of variation in situations where a bilingual advantage is observed. It is believed that immersion in a second language leads to an increased inhibition of the first language. Immersion increases the need for enhanced cognitive monitoring, leading to overall improvement in executive functions. [5].

In contrast to the traditional analysis, the Cox PH analysis allowed us to capture accuracy and speed within a single analysis without having to discard inaccurate responses. It also includes all the data points without breaking model assumptions. By retaining all potentially informative data, we are able to confidently interpret model parameter estimates. The composite factor score predicted better performance overall (Table 4). This is consistent with previous studies, which also reported an overall advantage for bilingual children compared with their monolingual peers in the coaching task. It is worth noting that previous studies with twice the amount of trials in the cognitive task as other studies also observed bilingual advantage with traditional analysis. The effect of such a design and the considerable amount of deleted data will need to be examined in the future [3]. When we used LSBQ categorical classification method to partition the participants into inadequate bilingual and adequate bilingual groups (Table 5), there were no significant differences between the two groups in flanker effect, facilitation effect, and mixing cost. However, we can see from the Cox PH regression analysis that adequate bilinguals perform significantly better than inadequate bilinguals in terms of accuracy and response time in a flanker task (Fig.1). This further indicates the advantage of using the new method when analyzing behavioral data in cognitive tasks.

5. Conclusion

We used both the traditional and the Cox proportional hazard model to analyze the performance data of bilinguals in a flanker task. The traditional analysis showed that the length of time spent actively using the L2 in immersion settings was a more reliable predictor of the bilingual advantage than the overall bilingual level. While the results of Cox analysis indicated that the behavioral performance got better with the increased overall bilingual level.

Acknowledgment

We are grateful to everyone who helped us write this article. We thank the assistance of Professor Hou, who has offered us invaluable suggestions in the studies.

Supported by Inner Mongolia Natural Science Foundation [2020LH03013], the Fundamental Research Funds for the Inner Mongolia Normal University [2022JBQN117], High-level Talent Research Start-up Fund for the Inner Mongolia Normal University [2015YJRC012], National Natural Science Foundation of China [31960184].

References

[1] Tao, L., et al., Bilingualism and domain-general cognitive functions from a neural perspective: A systematic review. Neurosci Biobehav Rev, 2021. 125: p. 264-295.

[2] Calabria, M., et al., Neural basis of bilingual language control. Annals of the New York Academy of Sciences, 2018. 1426(1): p. 221-235.

[3] Hannaway, N., B. Opitz, and P. Sauseng, Exploring the bilingual advantage: manipulations of similarity and second language immersion in a Stroop task. Cogn Neurosci, 2019. 10(1): p. 1-12.

[4] Bialystok, E., F.I.M. Craik, and G. Luk, Bilingualism: consequences for mind and brain. Trends in Cognitive Sciences, 2012. 16(4): p. 240-250.

[5] Anderson, J.A.E., et al., The language and social background questionnaire: Assessing degree of bilingualism in a diverse population. Behav Res Methods, 2018. 50(1): p. 250-263.

[6] De Cat, C., A. Gusnanto, and L. Serratrice, Identifying a Threshold for the Executive Function Advantage in Bilingual Children. Studies in Second Language Acquisition, 2017. 40(1): p. 119-151.

[7] DeLuca, V. and C. Pliatsikas, Bilingualism and the brain. OpenNeuro. [Dataset], 2022.

[8] DeLuca, V., et al., Duration and extent of bilingual experience modulate neurocognitive outcomes. Neuroimage, 2020. 204: p. 116222.

Research on Image Encryption Algorithm Based on One-dimensional Logistic and Two-dimensional Hénon Chaotic System

Xionghui Wu[1], Rui Dong[1], Kuo Yang[1*] and Wenli Zhang[2*]
[1] PLA, Beijing 100043, China
[2] Taiyuan University of Technology, Taiyuan 030024, China
[*]yangkuo1985@163.com(Yang), zhangwenli0374@link.tyut.edu.cn (Zhang)

Abstract

Aiming at the problems of high leakage risk and serious information loss that may exist in the encryption and decryption of digital images, an image encryption and decryption algorithm based on chaos theory is proposed. The algorithm introduces both one-dimensional Logistic mapping and two-dimensional Hénon mapping. In the encryption process, the digital image is dislocation encrypted by Logistic mapping and diffusion encrypted by Hénon mapping, and finally the cipher image is obtained. Compared with other similar algorithms, this algorithm has the advantages of low algorithm complexity and high operation efficiency. The simulation experiment is carried out on the MATLAB platform, and the results show that the algorithm can meet the test standard in the field of digital image encryption.

Keywords: Image Encryption, Chaotic System, Logistic Mapping, Hénon Mapping.

1 INTRODUCTION

With the rapid development of society and the extensive progress of information technology, digital images have become one of the important carriers for people to obtain information. At the same time, the serious challenges in the field of information security are increasingly highlighted, such as military secret images, satellite monitoring images, medical diagnostic images and so on. Therefore, a certain degree of security and privacy protection for classified information is becoming more and more important [1], and the security protection of image data always occupies an important position in the field of information security, and image encryption and decryption technology is an important method to implement image information protection, and it is also an effective way to solve the security transmission of images in the network.

In order to provide some level of security for plaintext images, the most straightforward method is to encrypt them [2]. Image encryption means that the content of the image is changed and the decryptor can view the original image content only if he has the correct decryption key [3]. It is necessary to design an image encryption algorithm with high security and good encryption effect. In 1963, the meteorologist Lorenz [4] published his famous article "The Butterfly Effect" to address the non-periodic and unpredictable characteristics of weather changes, which led to the birth of This gave birth to the modern chaos theory. Chaos is a state of chaos and disorder, and chaotic systems have many characteristics, such as boundedness, internal randomness, ergodicity, sensitivity to initial values, and long-term unpredictability [5], which coincide well with the basic requirements of cryptography. link, making it uniquely advantageous in the field of image encryption [6].

2 BASIC KNOWLEDGE

2.1 Chaos Theory

The beginning of chaos theory dates back to a series of studies by French mathematician Jules Henri Poincaré on planetary motion in the solar system in the early 20th century [7]. Poincaré combines dynamics with topology and points out that the interaction between stars can produce very complex behaviors, and some solutions of deterministic equations are unpredictable. In the 1960s, the American meteorologist Edward Norton Lorenz discovered a physical phenomenon while studying atmospheric motion and gave a description in the form of the Lorenz equation, stating that there is some connection between acyclicity and unpredictability, and that when the initial value of the equation varies within a particular range, the result becomes unpredictable and the system enters a chaotic state [8], opening a new chapter in the study of modern chaos theory. In 1989, R. Matthews [9] first used Logistic map to generate chaotic sequences and used them in data encryption, which opened a new journey of chaotic theory in cryptography.

International Conference on Computer Graphics, Artificial Intelligence, and Data Processing (ICCAID 2022)
edited by Ruishi Liang, Jing Wang, Proc. of SPIE Vol. 12604, 1260421
© 2023 SPIE · 0277-786X · doi: 10.1117/12.2674585

Typical Chaotic Systems. Two typical chaotic systems used in this paper are described below.

One-Dimensional Logistic Chaotic Mapping. One-dimensional Logistic chaotic mapping is a discrete chaotic system, which is often applied to dislocation operations in the encryption process of digital images because of its simple structure and efficient implementation [10]. The bifurcation diagram of the Logistic mapping is shown in Figure 1.

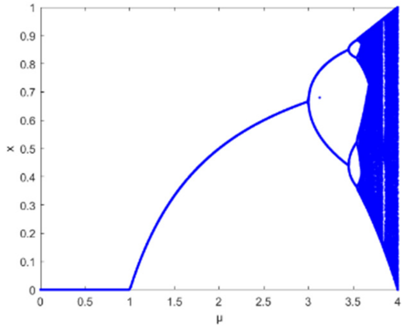


Figure 1. Bifurcation diagram of Logistic mapping.

Two-Dimensional Hénon Chaotic Mapping. Compared with the one-dimensional chaotic mapping, the two-dimensional Hénon chaotic mapping has more complex structure and better chaotic performance. The bifurcation diagrams of x-component and y-component of the Hénon mapping are shown in Figure 2.

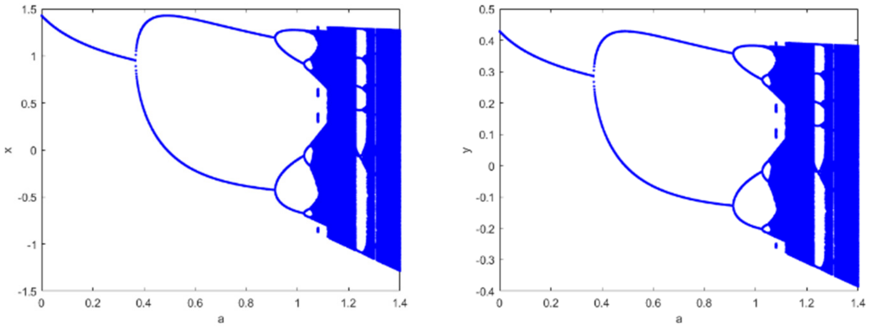


Figure 2. Bifurcation diagram of x-component and y-component of Hénon mapping.

2.2 Cryptography

Cryptography has been born for a long time, and with the development of the times, cryptography has penetrated into various aspects of people's daily life through its integration with many fields such as mathematics, computer science, biology, and information and communication engineering [11].

Cryptographic Foundations. In the late 1940s, Claude Elwood Shannon, an American mathematician and founder of information theory, published a landmark paper "Communication Theory of Secrecy Systems", which laid the theoretical foundation for modern cryptography [12]. A complete cryptographic communication system contains five components: plaintext, key, ciphertext, encryption algorithm and decryption algorithm. The schematic diagram of cryptographic communication system is shown in Figure 3.

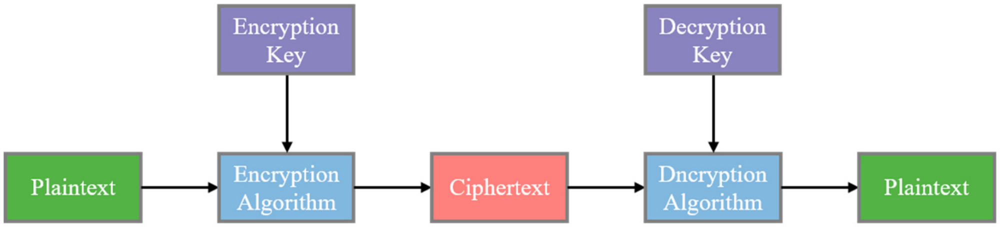


Figure 3. Schematic diagram of cryptographic communication system.

Digital Image Encryption Basics. In computer systems, digital images exist in the form of pixel value matrix. Therefore, the encryption of images can first transform the matrix data into one-dimensional data flow, and then change the position of each pixel or the size of each pixel value, and then transform it into the matrix form, and finally obtain the encrypted ciphertext image.

In the encryption process, if only change the pixel position, this encryption method is called scrambling encryption; if only change the size of the pixel value, this encryption is called diffusion encryption; if the two are both, it is called hybrid encryption [13]. At present, hybrid encryption is mainly used in the mainstream digital image encryption algorithms.

3 IMAGE ENCRYPTION AND DECRYPTION ALGORITHMS

3.1 Encryption Process

The steps to implement the digital image encryption algorithm are as follows.

Step 1: Read Plaintext Image. For an original grayscale image of size m×n, the i-th row and j-th column elements of the pixel value matrix (where $0 \leq i \leq m$ and $0 \leq j \leq n$) are represented by P_{ij}, and the size of P_{ij} value indicates the size of the pixel grayscale value. Since the pixel grayscale values are in the range of [0,255], the grayscale value of each pixel can be expressed as an eight-bit binary number, which results in a two-dimensional binary number matrix P_{Bin}, with the number of rows m and columns 8n.

Step 2: Dislocation Encryption. Two sets of pseudo-random sequences L_H and L_V are generated using the one-dimensional Logistic chaos system, where $L_H = \{L_{H1}, L_{H2}, L_{H3}, \ldots, L_{Hm}\}$ and $L_V = \{L_{V1}, L_{V2}, L_{V3}, \ldots, L_{V8n}\}$. The L_H and L_V are used to respectively dislocation encrypted the rows and columns of the matrix P_{Bin} obtained in step 1 to obtain the dislocation encryption matrix $P_{Bin\text{-}Z}$, and then converts $P_{Bin\text{-}Z}$ from a two-dimensional matrix to a one-dimensional sequence $P_{Bin\text{-}Z\text{-}1D}$ with length m×n.

Step 3: Diffusion Encryption. Two sets of pseudo-random sequences H_{I} and H_{II} are generated using the two-dimensional Hénon chaotic system, where $H_{\mathrm{I}}=\{H_{\mathrm{I}1}, H_{\mathrm{I}2}, H_{\mathrm{I}3}, \ldots, H_{\mathrm{I}mn}\}$ and $H_{\mathrm{II}}=\{H_{\mathrm{II}1}, H_{\mathrm{II}2}, H_{\mathrm{II}3}, \ldots, H_{\mathrm{II}mn}\}$. The H_{I} and H_{II} are processed using formula (1) as follows.

$$H_k = floor(mod(H_k \times 10^{14}), 256) \tag{1}$$

where k = 1, 2.

The processed H_{I} and H_{II} are converted into binary sequences $H_{Bin\,\mathrm{I}}$ and $H_{Bin\,\mathrm{II}}$; then the one-dimensional sequence $P_{Bin\text{-}Z\text{-}1D}$ and $H_{Bin\,\mathrm{I}}$ obtained in Step 2 are bitwise dissociated to obtain E_{I}, and bitwise dissociation between E_{I} and $H_{Bin\,\mathrm{II}}$ to obtain E_{II}.

Step 4: Generate Ciphertext Image. The E_{II} obtained in step 3 is converted from a one-dimensional binary sequence to a two-dimensional binary matrix, and then from a two-dimensional binary matrix to a two-dimensional decimal matrix E. Finally, the E is the encrypted image.

3.2 Decryption Process

The decryption process is the inverse operation of the encryption process, and the steps of the digital image decryption algorithm are as follows.

Step 1: Read Ciphertext Image. For a ciphertext image E of size m×n, the E is a two-dimensional matrix composed of pixel values, first convert the decimal two-dimensional matrix into a binary two-dimensional matrix, and finally convert the binary two-dimensional matrix into a binary one-dimensional sequence $E_{Bin\text{-}1D}$.

Step 2: Diffusion Decryption. Two sets of pseudo-random sequences H_{I} and H_{II} are generated using the two-dimensional Hénon chaotic system, where $H_{\mathrm{I}}=\{H_{\mathrm{I}1}, H_{\mathrm{I}2}, H_{\mathrm{I}3}, \ldots, H_{\mathrm{I}mn}\}$ and $H_{\mathrm{II}}=\{H_{\mathrm{II}1}, H_{\mathrm{II}2}, H_{\mathrm{II}3}, \ldots, H_{\mathrm{II}mn}\}$, and H_{I} and H_{II} are processed using formula (1), and convert the processed H_{I} and H_{II} into binary sequences $H_{Bin\,\mathrm{I}}$ and $H_{Bin\,\mathrm{II}}$. The one-dimensional sequence $E_{Bin\text{-}1D}$ and $H_{Bin\,\mathrm{II}}$ obtained in step 1 are bitwise dissociated to obtain E_{II}, and the bitwise dissociation between E_{II} and $H_{Bin\,\mathrm{I}}$ to obtained E_{I}. Finally, the one-dimensional sequence E_{I} is converted into a two-dimensional inverse diffusion matrix $E_{Bin\text{-}k}$.

Step 3: Dislocation Decryption. Two sets of pseudo-random sequences L_H and L_V are generated using the one-dimensional Logistic chaos system, where $L_H = \{L_{H1}, L_{H2}, L_{H3}, \ldots, L_{Hm}\}$ and $L_V = \{L_{V1}, L_{V2}, L_{V3}, \ldots, L_{V8n}\}$. The L_H and L_V are used to inverse the rows and columns of the matrix $E_{Bin\text{-}k}$ obtained in Step 2, and the inverse scrambling matrix $E_{Bin\text{-}k}$ is obtained.

Step 4: Generate Plaintext Image. The matrix $E_{Bin\text{-}Z}$ obtained in step 3 is a binary pixel value matrix, and $E_{Bin\text{-}Z}$ is converted from binary to decimal to obtain a decimal pixel value matrix P, and P is the plaintext image.

The key scheme of grayscale image encryption and decryption algorithm is shown in Table 1, and the flow chart of grayscale image encryption and decryption is shown in Figure 4.

Table 1. Key scheme of grayscale image encryption and decryption algorithm.

one-dimensional Logistic chaotic mapping				two-dimensional Hénon chaotic mapping			
μ_H	x_H	μ_V	x_V	x	y	a	b
3.9992	0.03	3.9989	0.03	0.75	0.32	1.39	0.3

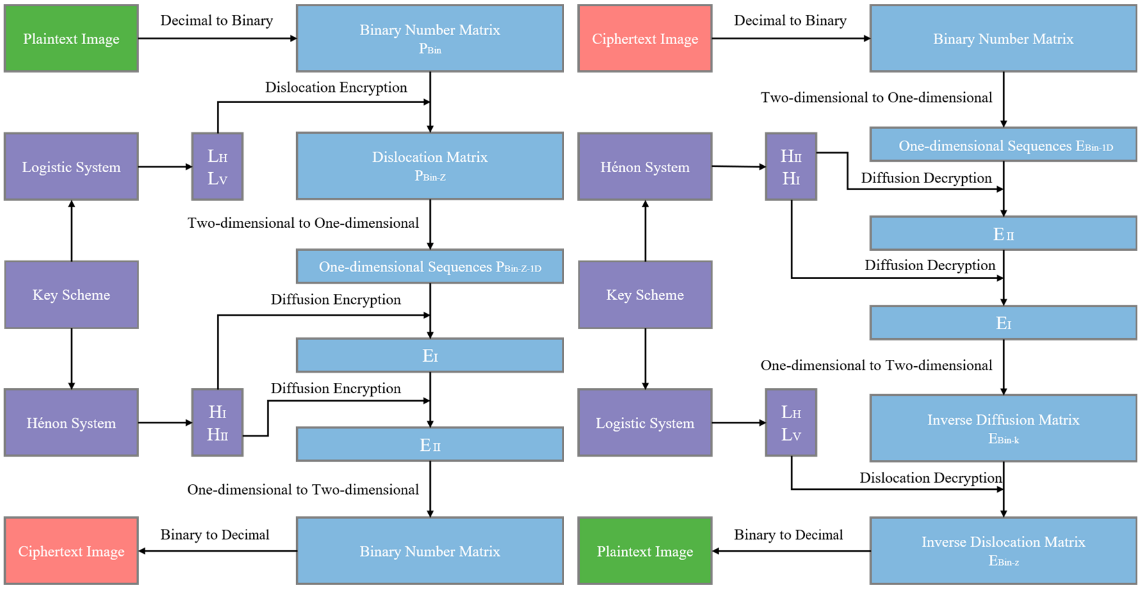
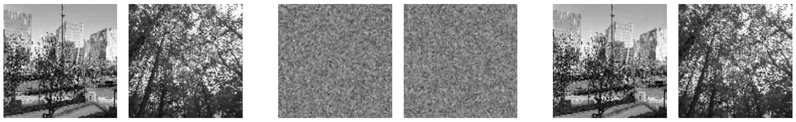


Figure 4. Flow chart of grayscale image encryption and decryption.

4 EXPERIMENTAL RESULTS AND ANALYSIS

4.1 Encryption and Decryption Effects

In order to verify the effect of the encryption and decryption algorithm proposed in this paper, the classical Lena (512×512) and Baboon (512×512) gray images are used as the processing objects, and the key scheme shown in table 1 is used. The encryption and decryption effect of grayscale images are shown in Figure 5.



a) Plaintext Images b) Encrypted Images c) Decrypted Images

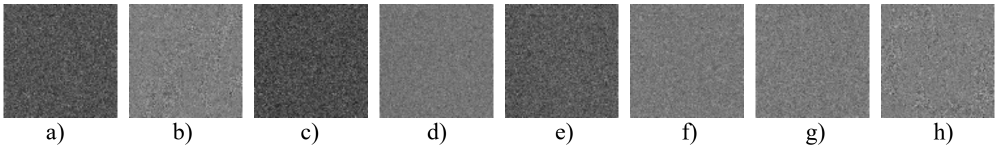
Figure 5. Encryption and decryption effect of grayscale images.

4.2 Key Space Analysis

For an encryption algorithm, it is effective against brute force cracking when the size of its key space exceeds 2^{100} [14]. In the image encryption algorithm proposed in this paper, the initial value of the key consists of a total of 8 values of μ_H, x_H, μ_V and x_V in the dislocation encryption phase, and x, y, a and b in the diffusion encryption phase, respectively, and according to the calculation accuracy of 64-bit double precision numbers of about 10^{-15}, the key space size of this encryption algorithm is calculated to be about 10^{120}, which is much larger than the required 2^{100}. therefore, the image encryption algorithm has a large enough key space to effectively resist brute force cracking.

4.3 Key Sensitivity Analysis

Key sensitivity is extremely important for a good encryption algorithm. In this subsection, Lena image is chosen as the test object to test the key sensitivity of the encryption algorithm. The 8 initial values of keys μ_H, x_H, μ_V, x_V, x, y, a and b are added respectively with $\Delta\mu_H = 10^{-15}$, $\Delta x_H = 10^{-15}$, $\Delta\mu_V = 10^{-15}$, $\Delta x_V = 10^{-15}$, $\Delta x = 10^{-15}$, $\Delta y = 10^{-15}$, $\Delta a = 10^{-15}$ and $\Delta b = 10^{-15}$, and the key scheme is tested by making a small modification to one of the keys at a time. The decryption effect on the encrypted image is shown in Figure 6. It can be seen from the observation that any minor modification cannot correctly decrypt the original image. Therefore, the encryption algorithm proposed in this paper has good key sensitivity and can resist violent attacks.



a) μ_H=3.9992+$\Delta\mu_H$, b) x_H=0.03+Δx_H, c) μ_V=3.9989+$\Delta\mu_V$, d) x_V=0.03+Δx_V,
e) x=0.75+Δx, f) y=0.32+Δy, g) a=1.39+Δa, h) b=0.3+Δb.

Figure 6. Image decryption effect of wrong key.

5 CONCLUSION

In this paper, an image encryption and decryption algorithm based on chaos theory is designed and implemented. The algorithm has low complexity and high operation efficiency. The simulation experiment is carried out by using MATLAB software platform to test the encryption and decryption effect of the algorithm. Through performance analysis, it can be seen that the algorithm can meet the test standard in the field of image encryption, which has a certain reference value in the research field of image encryption based on chaos theory.

REFERENCES

[1] Zhang Y.: Chaotic Digital Image Encryption. 1nd edn. Tsinghua University Press, Beijing (2016).

[2] Li C. Q.: Cracking a Hierarchical Chaotic Image Encryption Algorithm Based on Permutation. Signal Processing 118, 203-210 (2016).

[3] Chengqing Li, F., Yun Zhang, S., Eric Yong Xie. T.: When an Attacker Meets a Cipher-image in 2018: A Year in Review. Journal of Information Security and Applications 48, 1-9 (2019).

[4] Lorenz E. N.: Deterministic Nonperiodic Flow. Journal of the Atmospheric Sciences 20, 130-141 (1963).

[5] Morris W. Hirsch, F., Stephen Smale, S., Robert L. Devaney, T.: Differential Equations, Dynamical Systems, and an Introduction to Chaos. 3nd edn. Academic Press, Salt Lake City (2013).

[6] Mohamed Zakariya Talhaoui, F., Xing-yuan Wang, S., Mohamed Amine Midoun, T.: Fast Image Encryption Algorithm with High Security Level Using the Bülban Chaotic Map. Journal of Real-Time Image Processing 18, 85-90 (2020).

[7] Chen Y. C.: The Cryptanalysis and Design of Chaos Based Image Encryption Scheme. Doctor, Guangzhou University (2021).

[8] Shih-Yu Li, F., Zhangming Ge, S.: A Novel Study of Parity and Attractor in the Time Reversed Lorentz System. Physics Letters A 373(44), 4053-4059 (2009).

[9] Matthews R.: On the Derivation of Chaotic Encryption Algorithm. Cryptologia 13(1), 29-42 (1989).

[10] E. Xu, F., Liangshan Shao, S., Guanghui Cao, S., et al.: A New Method of Information Encryption. 2009 ISECS International Colloquium on Computing, Communication, Control, and Management 4, 583-586 (2009).

[11] Li Z. C.: Cryptography. 1nd edn. Publishing House of Electronics Industry, Beijing (2019).

[12] Shannon C. E.: Communication Theory of Secrecy Systems. Bell System Technical Journal 28(4), 656-715 (1949).

Proc. of SPIE Vol. 12604 1260401-547

[13] Fangzheng Zhao, F., Chenghai Li, S, Chen Liu, T, et al.: Analysis of the Effects of Scrambling and Diffusion of Logistic Chaotic Map on Image Encryption. In: 11th International Conference on Digital Image Processing (ICDIP 2019), pp. 768-777. SPIE Press, Guangzhou (2019).

[14] Alvrez G., F., Li S., S.: Some Basic Cryptographic Requirements for Chaos-Based Cryptosystems. International Journal of Bifurcation Chaos 16(8), 2129-2151 (2006).

Application of Deep Neural Network in Cost Estimation of Hydropower Projects

Xin Qiu[1*], Meiru Li[2], Peiyu Li[3], Yang Jiang[4], Li Peng[5]
[1] Chengdu Engineering Corporation Limited, Chengdu, Sichuan, China
[2] Chengdu Engineering Corporation Limited, Chengdu, Sichuan, China
[3] Chengdu Engineering Corporation Limited, Chengdu, Sichuan, China
[4] Chengdu Engineering Corporation Limited, Chengdu, Sichuan, China
[5] Chengdu Engineering Corporation Limited, Chengdu, Sichuan, China
[*]Corresponding author's e-mail: 2021001@chidi.com.cn

Abstract

Investment estimation is an essential part of hydropower projects. This paper proposes a learning rate control-enabled deep learning neural network model that can be optimized for different data sizes, especially when small. Then, a DNN model with learning rate optimization is constructed based on the existing hydropower project data in China; finally, the practicality and reliability of the learning rate control enabled example calculations to verify the DNN model. According to the results, the learning rate control-enabled DNN model accurately predicts outcomes. Therefore, it can achieve accurate, fast, and adequate investment estimation for large-scale and middle-scale hydropower projects.

Keywords: Hydropower, Cost Estimation, Deep Neural Network

1. Introduction

1.1. Background

Total investment in hydropower projects is remarkably relevant to investors in financing and managing projects. However, hydropower projects are different from other traditional construction projects. Because they have many influencing factors, such as complex composition, significant changes over time, wide distribution of construction sites, and complex and changing construction conditions, and they are often influenced by topography, hydrology, meteorology, and the natural resources of the construction site. Meanwhile, the entire project construction period requires strong collaboration between multiple government departments and related industries, resulting in a long construction cycle and a vast investment scale for hydropower projects. These inherent peculiarities make it challenging to accurately determine the total project investment during the pre-construction project. An effective project investment estimation method is the basis for project establishment and implementation. It is not only for the project decision with a controlled investment after the start of construction but also for the final accounting for the investment. Without an effective investment estimation model, it can easily cause a severe waste of funds. Therefore, studying the investment estimation of hydropower project construction is crucial and necessary.

In this paper, we analyzed the main influencing factors of total project investment. We constructed the prediction model of hydropower projects with a relatively small dataset scale to get the dynamic prediction value of total project investment, which is a deep learning neural network with learning rate control.

1.2. Research review

Accurate cost estimation is critical to the management of the project, such as budgeting, planning, monitoring, and construction, to ensure compliance with the client's available budget, time, and outstanding work [1]. It is shown that project decision-making and design are between 30% and 75% likely to influence project investment, while the likelihood of construction is between 5% and 25%. [2]. Ansar et al. [3] found that lots of dams suffered cost overruns. Therefore, an accurate cost estimate at an early stage is beneficial for the entire project cycle.

The traditional methods of investment estimation for hydropower projects mainly include the production capacity index estimation method, capital turnover rate method, comprehensive index investment estimation method, fixed amount measurement method, etc. The production capacity index estimation method, also known as the index estimation method, refers to estimating the investment amount of similar proposed and completed projects based on the investment amount

International Conference on Computer Graphics, Artificial Intelligence, and Data Processing (ICCAID 2022)
edited by Ruishi Liang, Jing Wang, Proc. of SPIE Vol. 12604, 1260422
© 2023 SPIE · 0277-786X · doi: 10.1117/12.2674628

and production capacity. As for the comprehensive index investment estimation method, the construction project is divided into construction works, equipment installation works, equipment purchase costs, and other capital construction costs or unit works. Then, the total investment amount is measured based on the specific investment estimation index, the cost items, or unit works investment estimation. Quota measurement measures the total investment amount based on pre-approved quotas and related budget base figures. Traditional methods are often either less accurate or complicated to calculate and operate, making them difficult to apply to today's construction project estimates.

Classical statistical model analysis methods can also be applied to hydropower project investment estimation, such as multi-primary linear regression, time series analysis, factor analysis method, fuzzy mathematical estimation method, etc. [4]. Lu Yeqi [5] proposed a regression model to predict the investment amount of a hydropower project with the help of the Project Cost Base Database System led by GREEI. Awojobi and Jenkins [6] applied the reference class forecasting (RCF) method to the investment estimation of the Bujagali Dam in Uganda and improved forecasting accuracy in the planning stage of hydropower projects. Ansar et al. [7] applied multilevel statistical techniques to large dams and fitted simple models for hydroelectric dam cost prediction and cost overruns.

Recent studies on the application of neural networks in hydropower projects are generally on hydropower plant operation and maintenance [8] and hydroelectric energy generation prediction [9][10], but there are fewer studies on its application to engineering cost and investment estimation, and the existing studies related to investment estimation are mainly focused on traditional civil engineering projects. Tarek Hegazy performed a parametric cost estimation for a highway project using a neural network approach [11]. Rafiei M H investigated a three-layer back-propagation neural network model using the SOFTMAX activation function for the complexity of construction projects, taking into account various factors such as materials, labor, equipment, project location, type, construction duration, and schedule, as well as relevant physical and financial variables and validated it for 372 building cost data [12]. Kim G H used construction cost data from 530 residential buildings in Korea to train a neural network to evaluate construction costs through genetic algorithms and traditional BP neural networks [13]. In a limited number of studies applying neural networks for estimating the cost of hydropower, Qiang M S and Song X S [14] first used a neural network algorithm for investment estimation of hydropower projects and achieved good results. Ren H and Zhou Q M [15] applied the BP neural network improved by the momentum method and learning adaptive adjustment strategy to project cost and main project quantity estimation. Tang Q [16] et al. combined typical hydropower projects, extracted engineering features for the complex characteristics of hydropower projects, and applied genetic algorithm and neural network theory to establish a genetic neural network model for investment prediction of hydropower projects, and the model prediction accuracy reached 90%.

In a comprehensive view, the existing research on hydropower project investment estimation uses a more traditional model approach, which is challenging to achieve a rapid estimation of complex hydropower project investment. In addition, the existing neural network modeling studies have scarcely involved optimizable and adjustable neural network models for small data sets that are difficult to collect in practice.

1.3. Research objectives

In the above context, this paper proposes a learning rate-control enabled deep neural network investment estimation model, which can be better applied to predictions with different sample sizes, especially when small. The purpose is to apply to the investment planning and site selection stage of hydropower projects to improve the accuracy and efficiency of total investment estimation and support project decision-making.

2. Methodology

2.1. Deep learning

The deep learning technology that has emerged in recent years has been widely used in the solution of complex classification, recognition, and prediction problems with promising results and efficiency due to its multi-layer nonlinear mapping network layer, which enables it to obtain a high capability of fitting complex functions [17]. Deep learning can be seen as an evolution of neural networks. A deep neural network possesses at least one hidden layer. Like shallow neural networks, deep neural networks can provide modeling for complex nonlinear systems. However, the extra layers provide a higher level of abstraction for the model, thus improving the model's capabilities. Deep neural networks are generally trained with back-propagation algorithms for neural layers. The weight update can be solved by stochastic gradient descent using the following equation:

$$\Delta w_{ij}(t+1) = \Delta w_{ij}(t) + \eta \frac{\partial C}{\partial w_{ij}} \qquad (1)$$

where η is the learning rate, and C is the cost function. This function is related to the type of learning (e.g., supervised learning, unsupervised learning, reinforcement learning) and the activation function.

Deep learning is currently widely used and has been applied in computer vision, speech recognition, natural language processing, audio recognition, bioinformatics, material inspection, etc.: they produce results comparable to and sometimes even surpassing the performance of human experts [18][19][20][21]. However, the current engineering research is mainly for project management in the construction process of infrastructure projects [22], as well as the prediction of engineering structural safety investigation[23][24], but less for the investment application analysis of hydropower engineering construction projects.

2.2. Learning rate control

In machine learning, the learning rate is a tuning parameter in the optimization algorithm that determines the step size of each iteration while moving toward the minimum value of the loss function. It will test each training epoch's conditions using an Early Stopping callback. Training is automatically stopped if there is no improvement after several training epochs. To achieve faster convergence, prevent oscillations, and get into undesired local minima, the learning rate is often varied during training according to the learning rate schedule or using adaptive learning rates.

Learning rate control can effectively reduce overfitting when DNN neural network models are trained with insufficient data.

3. Model Building

3.1. Acquisition of data

This study mainly adopts the data of constructed power stations of Chinese projects, from which 136 hydropower projects with total investment data were initially selected, and the engineering characteristics of these 136 hydropower projects were collected and analyzed. One hundred ten hydropower projects with more complete engineering characteristics were selected through a lot of data collation work, and these engineering characteristics are latitude and longitude, installed capacity, average power generation capacity, normal storage level, reservoir capacity, average power generation capacity, dam type, dam height, region, basin, time, seismic fortification intensity, etc.

3.2. Statistical analysis and pre-processing

3.2.1. *Statistical analysis.* According to the statistical analysis, the sample of hydropower plants selected in this paper are mainly distributed in seven major basins in China, namely: Yellow River Basin (7), Inner River Region Basin (2), Southeast Region Basin (2), Hailuan River Basin (1), Yangtze River Basin (72), Pearl River Basin (7) and Cross-over International Basins of Yunnan, Tibet and Xinjiang in Western China (19).

The hydropower resource of the Yangtze River basin is abundant and relatively more exploitable, followed by the cross-sectional international basins in the western and southern regions, following the actual distribution of hydropower plant construction in China (see Table 1). As shown in Table 2, the dam types of hydropower plants are mostly gravity dams and rockfill dams. The seismic fortification intensity of hydropower plants is generally magnitude 7 (56) or 8 (29) and can reach magnitude 9 (4) in high seismic intensity zones, and hydropower plants built before 2016 can be unprotected for areas with seismic intensity less than magnitude 6 (see Table 3).

Table 1. Numbers of Projects in Different Basins

Basins	Yellow River	Yangtze River	Pearl River	Inner River Region	Hailuan River	Southeast Region Basin	Cross-over International Basins of Yunnan, Tibet, and Xinjiang in Western China
Numbers	7	72	7	2	1	2	19

Table 2. Numbers of Different Dam Types

Dam type	Rockfill dam	Arch dam	Gravity dam	Others
Number of samples	35	10	35	30

Table 3. Distribution of Seismic Fortification Intensity Resistance Levels

Seismic Fortification Intensity Resistance Levels	VI	VII	VIII	IX
Number of samples	25	56	29	4

Statistical analysis of the sample quantitative engineering characteristic parameters was conducted, and the results are shown in Table 4. The investment per kilowatt of electric energy (yuan/KWh) of China's hydropower plant construction projects ranges from 0.51–7.69 yuan/KWh, with an average value of 1.96 yuan/kWh. The average power generation capacity ranges from 0.31 to 64 billion kWh. The maximum total installed capacity is 16,000 MW, the minimum is 10.8 MW, and the average is 931.42 MW. The normal storage level ranges from 51 to 4383 m, with an average level of 1243 m. The maximum reservoir capacity is 38.6 billion m^3 and the average value of 17.88 m^3. The highest dam height is 314 m, and the lowest is 16 m. Thus, the sample projects selected for the study in this paper have comprehensive coverage and significant differences in engineering characteristic values.

Table 4. Engineering Character Parameter Statistical Analysis

Parameters	Maximum	Minimum	Average
Total investment (Million CNY)	126020	193	8723.44
Average power generation (TWh)	64	0.031	3.887
Investment per unit KW(CNY/KWh)	7.69	0.51	1.96
Total installed capacity (MW)	16000	10.8	931.42
Normal storage level (m)	4383	51	1243.44
Reservoir capacity (km^3)	38.6	0.000058881	1.788
Dam height (m)	314	16	95.07

3.2.2. Normalization of engineering characteristics indicators. In this study, 12 engineering features are used: longitude, latitude, average power generation, total installed capacity, impounded level, reservoir capacity, dam height, dam type, seismic fortification intensity resistance, basin, province, and construction start time. Due to the nonlinear characteristics of the classification indexes, without knowing the weights of each engineering characteristic, the classification indexes are flattened to Boolean values to ensure that the DNN neural network can converge well. In this report, the data normalization function is established by the mean and standard deviation of the data set, and the formula is shown as follows.

$$x_i^* = (x_i - \overline{x_\iota})/s_i \tag{2}$$

where xi represents the original value of the original series; xi represents the mean of the original index series xi; si is the standard deviation of the original series xi.

Quantitative indicators such as latitude and longitude, average power generation, total installed capacity, impounded level, reservoir capacity, dam height, seismic intensity, etc. are subject to pre-processing data work such as unified units and checking abnormal values before being standardized for calculation and analysis. Qualitative indicators such as dam type, river basin where the project is built, and the province where the project is located are Booleanized, with 4 quantified codes for dam type, 7 quantified codes for the river basin, and 13 quantified codes for the province expanded into 24 Boolean indicators.

3.3. Model applicability analysis

The goal of a complex nonlinear system, such as engineering investment estimation for hydropower projects, is to predict a continuous predictive value from multiple numerical or non-numerical indicators.

After plotting the existing data with total investment as the vertical axis and each parameter as the horizontal axis, the graph shows that total investment presents an approximately linear relationship with average power generation and total installed capacity, and some parameters show exponential correlation. However, most are not simply linearly correlated and have a deeper, higher-dimensional connection.

The training sample characteristics also indicate that total investment estimation does not apply to the traditional multiple linear regression method, and the deep connection between the data needs to be explored. Applying DNN to the prediction of investment estimation amount of hydropower projects can give engineering-friendly learning algorithms for complex nonlinear mapping between hydropower project characteristics and investment without the constraints of complex nonlinear relationships. The model predicts a continuous total investment by nine numerical indicators: latitude and longitude, average power generation, total installed capacity, impounded level, reservoir capacity, dam height, seismic fortification intensity, and construction start time, as well as three qualitative indicators: basin, province, and the dam type. The DNN model has a minor prediction error with sufficient samples compared with the traditional neural network algorithm. The model itself is scalable and allows the model to be trained by adding more training samples to improve the prediction accuracy further. It is also insensitive to extreme values and can exclude individual extreme samples' influence.

3.4. Model development and training

3.4.1. Dataset preparation. In order to avoid overfitting, 20% of the samples are randomly selected as test samples, and the rest are used as training samples.

3.4.2. Neural network structure. This model uses normalized key engineering feature indicators as input vectors and total investment as output vectors to establish a DNN-based investment estimation model for hydropower projects.

The general idea is to set the relevant parameters according to the number of input variables and training samples and input the training dataset into the model for training. The model structuring flowchart is shown in Figure 1.

Input layer: 12 engineering characteristics of longitude, latitude, average power generation, total installed capacity, normal storage level, reservoir capacity, dam height, dam type, seismic intensity resistance, basin, province, and construction start time are selected as the input of the model, and the number of input nodes of the neural network is 33 after standardization and Booleanization.

Hidden layer: This model tries to adopt different scales of neural network structure, starting from a large neural network with 10 layers of 512 neurons per layer and gradually optimizing it, finally determining the number of neurons in the hidden layer as 6 layers with 64 neurons per layer, and the activation function is a linear rectifier function (ReLU) with a total of 23401 parameters, of which 23108 are trainable parameters, the normalized parameters are 66 (33 mean, 33 standard deviations), and the output parameter is 1.

Output layer: The output variable is the total investment, and the number of nodes in the output layer is 1.

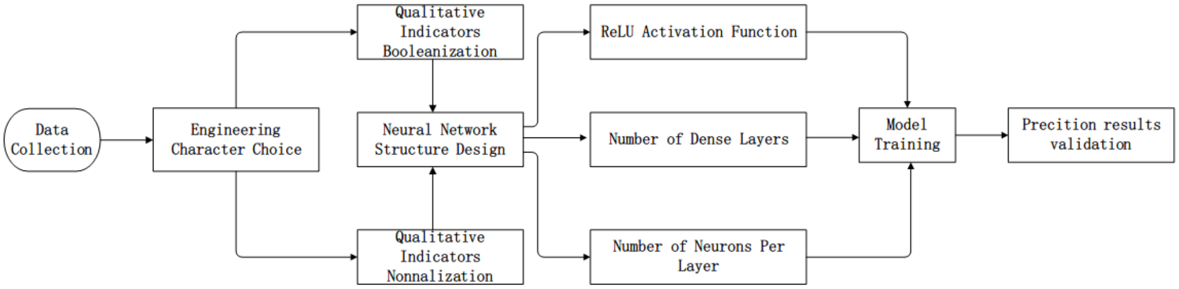

Data Collection
Engineering Character Choice
Qualitative Indicators Booleanization
Neural Network Structure Design
Qualitative Indicators Nonnalization
ReLU Activation Function
Number of Dense Layers
Number of Neurons Per Layer
Model Training
Precition results validation

Figure 1. Modeling Flowchart

3.4.3. Learning rate control and early-stopping. Because of this study's relatively small number of samples, this model is trained with both learning rate control and early stopping to improve prediction accuracy and reduce overfitting.

In machine learning and statistics, the learning rate is a tuning parameter in the optimization algorithm, which determines the step length of each iteration of model training while moving toward the minimum value of the loss function, and learning rate control can effectively reduce the model overfitting.

The early-stopping is to stop training early by selecting a node in the iterative process of model training in order to intentionally reduce the model fitting because further training after the node will better fit the training sample but also increase the general error.

In this model, the initial learning rate is set to 0.001, and there are learning rate decays every 100 training iterations, in which the decay rate is 1. Training is terminated after 800 iterations, so the final learning rate is 0.000125.

4. Results

4.1. Error analysis of prediction results

Different prediction models are evaluated by comparing the prediction values of the training samples with validation samples. Mean absolute error (MAE), root mean square error (RMSE), and mean absolute percentage error (MAPE) are selected for this evaluation.

In this study, there are two directions of model optimization, one is to optimize the neural network structure itself, i.e., the number of neurons and the depth of the neural network; another is to optimize the neural network training process, mainly by adding learning rate control to optimize the step length of each iteration of neural network training.

For the three models with different optimization levels, the average error percentage (MAPE) varies with the number of training epochs as shown in the figure below. According to this figure, the result of the original model has a higher error, and the LOSS function, which is MAPE in this case, has a wide fluctuation during the training process, which is attributed to the excessive training parameters of the neural network and the complexity of the problem. After optimizing to a smaller neural network, the training curve fluctuation is reduced significantly, indicating that the neural network capacity is controlled at a reasonable level. After adding the learning rate control, the MAPE fluctuation varies smoothly, the validation error is substantially reduced, and the overfitting problem is significantly solved, indicating that the learning rate control effectively optimizes the training neural networks, especially for small-scale data sets (see Figure 2).

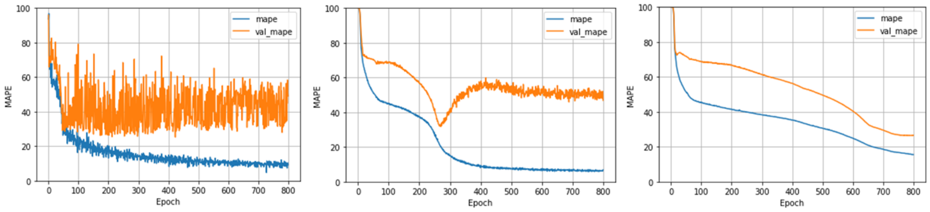


Figure 2. Comparison of Training Curves of Different Models

As seen from Table 5, when the amount of predicted data is small, the MAPE difference between the training and testing samples of the large-scale neural network model is large, which is over 55%. While the difference between the MAPE of the validation sample and the training sample is around 30% after optimizing the model into a small-scale neural network model. After the learning rate control is enabled, the difference between its test sample and training sample MAPE is reduced to about 10%.

This indicates that the large-scale neural network model is more likely to lead to overfitting in the case of a small data volume. In comparison, the prediction accuracy of the validation samples with learning rate control enabled is significantly improved, and its MAPE value is reduced from 62.12% to 26.57%, which is less than 30%.

Table 5. Model Prediction Error Matrix

Error indicators	Not optimized		Optimized neural structure		Learning rate control enabled	
	Training sample	Test sample	Training sample	Test sample	Training sample	Test sample
MAE	16278.29	725902.81	94427.34	659248.56	245459.39	295359.41
RMSE	61860.11	1700184.0	280726.12	1383926.5	795684.63	502609.91
MAPE	1.70%	39.19%	6.55%	46.92%	15.45%	26.57%

In summary, with the optimized model, the 22 test sample prediction results have a mean absolute percentage error of 26.57%, and the mean value of the error percentages is 0.99%. By calculating the standard deviation of error percentages with a result of 35.36%, t the distribution pattern of the model prediction results could be fitted to a normal distribution curve with a 95% confidence interval of 0.99% ± 14.78% (see Figure 3).

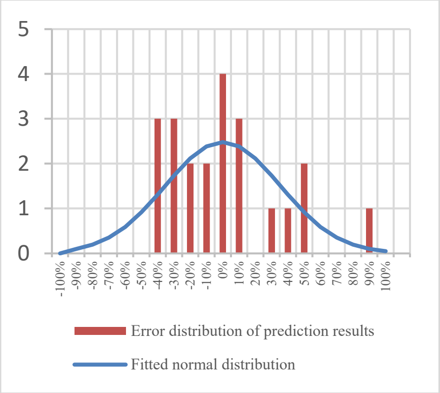


Figure 3. Distribution of Error Percentages

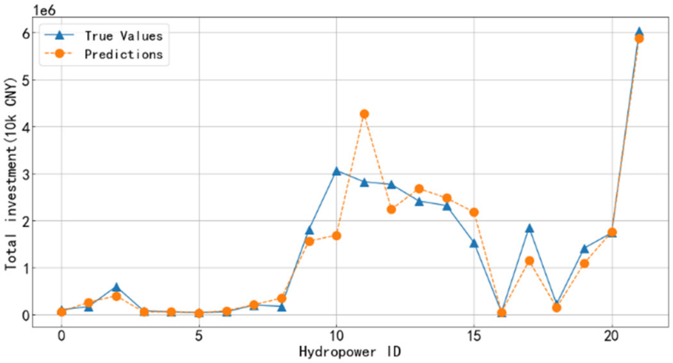


Figure 4. Comparison of Predicted Results and Actual Investment Value

According to the prediction results and the actual investment value comparison figure, the distribution of the predicted value shows a specific pattern that is distributed near the actual value (see Figure 4). The prediction points with more significant errors are mainly distributed in large hydropower projects with investments higher than 15 billion, but the actual error percentage does not increase significantly. This indicates that the choice of MAPE as the LOSS function in the model training process is more advantageous than simply choosing MSE because it avoids the excessive influence of large hydropower projects with high investment in the training dataset on the model weights during training.

4.2. Case study

The Nan'ou River is the largest first-class tributary of the Mekong River in Laos, with a river basin area of 25,634 km^2, a length of 475 km, and a natural drop of 430 m. The Nan'ou River Hydropower Station invested by China Power Construction Group in Laos has a total installed capacity of 1272 MW, a total of 21 turbine generating units, a total reservoir capacity of 8 billion cubic meters, an average head of 25 meters, an average dam height of 75.25 meters, an average dam length of 525 meters, and a dam type of gravity.

The engineering characteristics parameters of the Nan'ou River hydropower project are preprocessed and input into the learning rate control-enabled DNN model for prediction. The average value of unit kW prediction is obtained as 2034.56 USD/kW, meaning that the project's total investment can be obtained as 2.587 billion USD. In comparison, the actual investment of the project is 2.733 billion USD, which could be converted to 2.911 billion USD in 2018, with an estimation error of -11.86%. Therefore, it can satisfy the empirical requirement of ±30% of the investment estimation error in the project planning stage.

5. Conclusion

This paper proposes a learning rate control-enabled DNN model that can effectively solve the overfitting problem of the neural network model in small data sizes. The results show that this model improves the prediction accuracy compared with the traditional estimation method, especially in the watershed planning and selection stage. It provides decision-makers with a more reliable basis for investment forecast analysis. Furthermore, it achieves a quicker and more convenient way for investors to understand the total investment amount of the project.

Given the limited amount of data we have collected so far, every project sample has several vital parameters to describe its rough outline without much detailed information. While we still get a relatively accurate result compared to other existing methods.

In the future, with more data from more hydropower projects, the prediction accuracy of this model could be further improved.

References

[1] Kim G H, Yoon J E, An S H, et al.(2004) Neural network model incorporating a genetic algorithm in estimating construction costs. Building and Environment, 39(11): 1333–1340.

[2] Chen H.(2021) An enquiry into Investment Control of Engineering Projects.E3S Web of Conferences. EDP Sciences, 276: 02005.

[3] Ansar A, Flyvbjerg B, Budzier A, et al.(2014) Should we build more large dams? The actual costs of hydropower megaproject development. Energy policy, 69: 43–56.

[4] Kouskoulas V, Koehn E. (1974) Predesign Cost-Estimation Function for Buildings. Journal of the Construction Division, 100:589–604.

[5] Lu Y Q.(2019) Research on the Method of the Investment Prediction of Hydropower Project-based on Date Base. Hydropower and Pumped Storage,5(3):112–115.

[6] Awojobi O, Jenkins G P.(2016) Managing the cost overrun risks of hydroelectric dams: An application of reference class forecasting techniques. Renewable and Sustainable Energy Reviews, 63: 19–32.

[7] Ansar A, Flyvbjerg B, Budzier A, et al. (2014) Should we build more large dams? The actual costs of hydropower megaproject development. Energy policy, 69: 43–56.

[8] Feng Z, Niu W, Zhang R, et al. (2019) Operation rule derivation of hydropower reservoir by k-means clustering method and extreme learning machine based on particle swarm optimization. Journal of Hydrology, 576: 229–238.

[9] Li L, Yao F, Huang Y, et al. (2019)Hydropower generation forecasting via deep neural network//2019 6th International Conference on Information Science and Control Engineering (ICISCE). IEEE, 324–328.

[10] ALRAYESS H, GHARBIA S, BEDEN N, et al. (2018) Using machine learning techniques and deep learning in forecasting the hydroelectric power generation in almus dam turkey[J]. SAFETY, 72.

[11] Hegazy T, Ayed A. (1998) Neural network model for parametric cost estimation of highway projects. Journal of construction engineering and management, 124(3): 210–218.

[12] Rafiei M H, Adeli H.(2018) Novel machine-learning model for estimating construction costs considering economic variables and indexes. Journal of construction engineering and management, 144(12): 04018106.

[13] Kim G H, Yoon J E, An S H, et al. (2004) Neural network model incorporating a genetic algorithm in estimating construction costs. Building and Environment, 39(11): 1333–1340.

[14] Qiang M S, Song X S.(2002)Application of Neural Networks(NNs) in Quick Estimating for Hydroelectric Engineering Projects. Journal of Hydroelectric Engineering, 54–62.

[15] Ren H, Zhou Q M.(2005) Application of Neural Network For Quick Estimation of Engineering Construction Cost and Main Quantities. China Civil Engineering Journal, 38(8):135–138.

[16] Tang Q, Chen X, Liao Y Y, et al.(2013) Construction of high side slope deformation prediction genetic neural network model. Journal of Southwest University for Nationalities Natural Science Edition, 06:942–947.

[17] Wang H, Luo P, Zhang J. (2018)A New Appraisal Model for Urban Land Benchmark Price based on Deep Learning. China Land Science,32(9):59–65.

[18] Bengio, Yoshua; LeCun, Yann; Hinton, Geoffrey (2015). Deep Learning. Nature. 521 (7553): 436–444.

[19] Hu, J.; Niu, H.; Carrasco, J.; Lennox, B.; Arvin, F. (2020). Voronoi-Based Multi-Robot Autonomous Exploration in Unknown Environments via Deep Reinforcement Learning. IEEE Transactions on Vehicular Technology. 69 (12): 14413–14423.

[20] Ciresan, D.; Meier, U.; Schmidhuber, J. (2012). Multi-column deep neural networks for image classification. 2012 IEEE Conference on Computer Vision and Pattern Recognition. 3642–3649.

[21] Krizhevsky, Alex; Sutskever, Ilya; Hinton, Geoffry (2012). ImageNet Classification with Deep Convolutional Neural Networks. NIPS 2012: Neural Information Processing Systems, Lake Tahoe, Nevada.

[22] Zhao H. (2022) Application of Deep Learning in Civil Engineering Management. Computational Intelligence and Neuroscience, 2022.

[23] Kumar S S, Abraham D M. (2019)A deep learning-based automated structural defect detection system for sewer pipelines//Computing in Civil Engineering 2019: Smart Cities, Sustainability, and Resilience. Reston, VA: American Society of Civil Engineers, 226–233.

[24] Li Y, Bao T, Gao Z, et al. (2021)A new dam structural response estimation paradigm powered by deep learning and transfer learning techniques. Structural Health Monitoring, 2021: 14759217211009780.p

Design of medium and long-term power market electricity purchase and sale strategy considering time-sharing power consumption deviation assessment

Xu XIE[1a*], Zhe ZHANG[1], Bingqi LIU[2b*], Le YU[1], Yanbin LI[1], Xi CHEN[1]
[1]North China Branch of State Grid Corporation of China, State Grid Corporation, Beijing 100031, China
[2]Smart Grid Department, Beijing QU Creative Technology Co., Ltd., Beijing 100031, China
[a*]121316924@qq.com, [b*]liu_bq@qctc.com.cn

ABSTRACT

In order to improve the effectiveness of the medium and long-term market power purchase and sale strategy and enhance the trading efficiency of the electricity selling company, a medium - and long-term market power purchase and sale strategy considering the time-sharing power consumption deviation assessment is proposed. In the medium - and long-term market model of electricity, it will suffer from bias assessment when the actual electricity consumption of electricity users deviates from the amount of electricity traded in the medium - and long-term market. The purchasing cost of the power selling company is modeled and the assessment cost of time-sharing deviation of the power users is considered. On this basis, a model for evaluating the utility of electricity sales in the medium - and long-term market is proposed. If the benefit of purchasing and selling electricity exceeds its expected income, the purchasing and selling electricity strategy is effective; Otherwise, you need to adjust your trading strategy. Finally, a numerical example is constructed based on the actual data of a provincial power grid to verify the effectiveness of the proposed strategy. The results of the example show that if the time-sharing deviation assessment is not considered, the electricity purchasing and selling strategy of the electricity selling company may lead to higher transaction risk.

Keywords-Medium and long-term electricity market; Electricity purchase and sale strategy; Time-sharing power consumption deviation; Assessment of deviation

1 INTRODUCTION

Our country is carrying out a new round of power system reform, and the opening of power selling side market is the important content of the epicycle power system reform. With the continuous promotion of the liberalization policy of the power selling side market, more and more power users choose to purchase power through the power selling company. In recent years, it can be seen from the situation of electricity market transactions that the power selling companies bear a great risk of price fluctuation in the agent purchasing of electricity. This gives rise to the medium - and long-term market power strategy design.

The goal of the design of medium - and long-term electricity purchase and sale strategies of electricity selling companies is to maximize the transaction returns while effectively preventing transaction risks by evaluating the expected returns of medium - and long-term market transactions [1-2]. Literature [3] puts forward a power purchase and sale strategy considering the needs of sensitive users. By introducing high-reliability power sale service, it can provide high-reliability power supply service for sensitive users and improve the transaction income of the power selling company. Literature [4-5] studies the design of retail packages, which aims to strive for the market scale and increase the revenue from electricity sales on the basis of maximizing the consumption habits of users. Literature [6-7] studied the power sale strategy problem under the participation of user demand response, and respectively proposed the power purchase and sale strategy considering the deep interaction between interruptible load and flexible load. Literature [8] further studied the optimal dispatching mode of power selling companies under the combination of multiple virtual power plants, and proposed the optimal purchasing and selling power strategy scheme on this basis.

It should be noted that the above studies are all based on day-ahead trading strategy design, and relatively few studies have been conducted on mid - and long-term market trading. Studies in literature [9-10] show that electricity selling companies will face more uncertainties under medium - and long-term market transactions, including electricity demand prediction on the power side and power generation capacity prediction on the power side. The above uncertainties will

International Conference on Computer Graphics, Artificial Intelligence, and Data Processing (ICCAID 2022)
edited by Ruishi Liang, Jing Wang, Proc. of SPIE Vol. 12604, 1260423
© 2023 SPIE · 0277-786X · doi: 10.1117/12.2674859

lead to the actual benefits of electricity selling companies and medium - and long-term market transactions are different. In particular, the uncertainty of electricity demand of power users is large. In the medium and long-term market time-sharing trading mode, the power consumption deviation of each period will lead to the deviation assessment of the power selling company [10]. The above deviation assessment has become the main influencing factor of the trading benefit of the electricity selling company.

Therefore, this paper will study the medium - and long-term market strategy of electricity purchase and sale considering time-sharing deviation assessment. Firstly, a power purchase cost evaluation model considering the assessment of time-sharing power consumption deviation is proposed. According to the expected power consumption deviation of power users, the medium and long-term transaction expected power purchase cost is established. Then, the medium and long term market electricity purchase and sale transaction benefit evaluation model considering the assessment of time-sharing electricity consumption deviation is constructed. If the expected transaction income is higher than the expected income, the electricity purchase and sale strategy is reasonable. Finally, a numerical example is constructed based on the actual data of a provincial power grid to verify the effectiveness of the proposed method.

2 POWER PURCHASE COST ASSESSMENT CONSIDERING TIME-SHARING DEVIATION ASSESSMENT

Since China started the electric power system reform in 2015, it has established the long-term market trading system characterized by time-sharing transaction. Under the time-sharing medium and long-term market trading system, the power selling company and the power users sign the time-sharing power selling contract, and determine the electricity consumption and selling price of each time period. And signed with the power generation enterprises time - period power purchase contract, to determine the purchase of electricity and price. However, if the actual time-segment electricity consumption deviates from the medium and long-term transaction electricity, the electricity selling company will undertake the electricity consumption deviation assessment. According to the above analysis, the medium - and long-term electricity purchase cost of the electricity selling company considering time-sharing electricity consumption deviation assessment includes medium - and long-term transaction electricity purchase cost and expected electricity consumption deviation assessment, which can be expressed as:

$$C^C = C^L + C^A \tag{1}$$

where C^C is the medium and long-term electricity purchase cost of the electricity selling company considering the assessment of time-sharing electricity consumption deviation. C^L and C^A are respectively the medium and long-term transaction electricity purchase cost and the expected electricity consumption deviation assessment cost.

Among them, the medium and long-term transaction power purchase cost of the power selling company is the sum of the power purchase cost at different periods, which can be expressed as:

$$C^L = \sum_{t=1}^{N^T} E_t^B p_t^B \tag{2}$$

where N^T is the number of time segments in the medium - and long-term market trading mode. E_t^B and p_t^B are respectively the amount of electricity purchased and the price of electricity purchased at the time of the electricity selling company.

Deviation assessment is related to the market model. For the power grids of provinces and regions that have started electricity spot market trading, the power consumption deviation of each time period will be settled at the price of the electricity spot market. For the provincial power grid that has not started the electricity spot market transaction, the power consumption deviation of each period is calculated statistically in the form of the deviation assessment cost. Considering that the current power spot market is still in its infancy, this paper focuses on the deviation assessment model. In this mode, the electricity deviation assessment price is related to the deviation range, which can be expressed as:

$$p_t^A = \begin{cases} p^{A,1} & \left| {\Delta E_t^B}/{E_t^B} \right| \geq \gamma^1 \\ 0 & \left| {\Delta E_t^B}/{E_t^B} \right| < \gamma^1 \end{cases} \tag{3}$$

where p_t^A is the electricity deviation assessment price. ΔE_t^B is the power distribution company acting as the power user at different times. γ^1 is the power consumption deviation proportion limit. $p^{A,1}$ is the assessment price after the electricity consumption deviation exceeds the specified value.

The electricity deviation cost is the sum of the product of the deviation assessment price and the deviation electric quantity in each period, which can be expressed as:

$$C^A = \sum_{t=1}^{N^T} \Delta E_t^B p_t^A \tag{4}$$

According to Equations (1)-(4), it can be seen that the key to evaluate the medium and long-term electricity purchase cost of a power selling company considering time-sharing electricity consumption deviation assessment is to accurately evaluate the deviated electricity quantity in each time period. Reference [11-12] can be used to calculate the probability of electric quantity with different deviation based on the historical data of power users. When the distribution of deviated electricity quantity is known, the medium - and long-term electricity purchase cost of the electricity selling company considering the assessment of time-sharing electricity quantity deviation is the expectation of the assessment cost of deviated electricity quantity in different scenarios, which can be expressed as:

$$C^C = \sum_{t=1}^{N^T} E_t^B p_t^B + \sum_{s=1}^{N^S} \rho_s \sum_{t=1}^{N^T} \Delta E_{s,t}^B p_t^A \tag{5}$$

where N^S is the scene number of deviated electric quantity. ρ_s is the probability of the deviation charge. $\Delta E_{s,t}^B$ is the power deviation of time segments in the scenario s .

3 POWER PURCHASE AND SALE STRATEGY DESIGN

This section will first put forward the purchase and sale of electricity strategy evaluation model. The model will integrate the cost of electricity purchase and the revenue of electricity sale and calculate the comprehensive revenue of electricity purchase and sale considering the assessment of the power consumption deviation in different periods. On this basis, this paper will design the power purchase and sale strategy evaluation process for the power sale company to evaluate and analyze the effectiveness of its power purchase and sale strategy [12].

3.1 Comprehensive income evaluation model of electricity purchase and sale

The revenue from electricity sales in the medium and long term market transactions of the electricity selling company is the sum of the product of electricity sold and electricity sold price at each time period, which can be expressed as:

$$Q^S = \sum_{t=1}^{N^T} E_t^S p_t^S \tag{6}$$

where Q^S is the revenue of electricity sales from medium - and long-term market transactions. E_t^S and p_t^S are the amount of electricity sold by the company and the price of electricity sold by the company. The medium and long-term market comprehensive benefit of the electricity selling company is the difference between the electricity selling revenue and the electricity purchasing cost assessed by considering the time-of-use deviation, which can be expressed as:

$$Q^C = Q^S - C^C \tag{7}$$

Proc. of SPIE Vol. 12604 1260401-559

where Q^C is the medium and long-term market comprehensive benefit of the electricity selling company. The electricity sold by the electricity selling company is equal to the sum of the electricity purchased in the medium and long-term market and the electricity consumption deviation without considering the loss, which can be expressed as:

$$E_t^S = E_t^B + \Delta E_t^B \tag{8}$$

If the comprehensive loss is considered, the above relation can be expressed as:

$$\lambda^C E_t^S = E_t^B + \Delta E_t^B \tag{9}$$

where λ^C is the comprehensive loss coefficient, which is generally 1.09.

3.2 Evaluation and analysis of electricity purchase and sale strategies

On the basis of the above comprehensive income evaluation model of electricity purchase and sale, the evaluation strategy of electricity purchase and sale strategy proposed in this paper is shown in fig. 1. The key points of its implementation include:

(1) Calculation of comprehensive income from electricity purchase and sale. According to the comprehensive income evaluation model of electricity purchase and sale shown in Equation (7) of this paper, the comprehensive income from electricity purchase and sale considering the assessment of power consumption deviation in different periods is calculated.

(2) Assessment of comprehensive income from electricity purchase and sale. If comprehensive income from electricity purchase and sale meets the requirements of expected income, it will be passed; otherwise, electricity purchase and sale strategy optimization shall be carried out. The determination condition can be expressed as follows:

$$Q^C \geq Q^{C,set} \tag{10}$$

where $Q^{C,set}$ is the limit value of expected comprehensive income from electricity purchase and sale.

(3) Optimization of electricity purchase and sale strategy. If the comprehensive revenue from electricity purchase and sale does not meet the requirements of expected revenue, the expected revenue of medium and long-term transactions of each power user will be evaluated one by one, and the power user with low expected revenue will be canceled first until the expected revenue requirements are met.

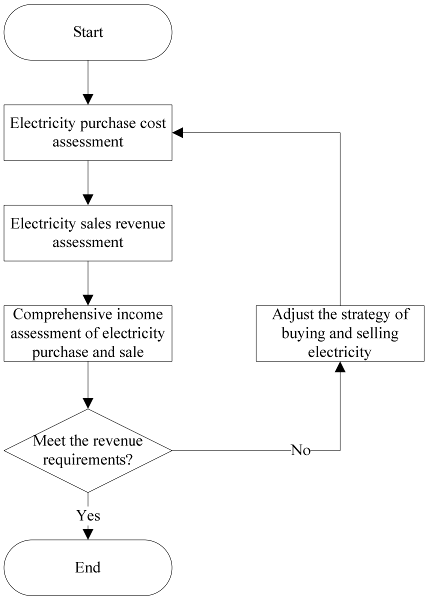

Start
Electricity purchase cost assessment
Electricity sales revenue assessment
Comprehensive income assessment of electricity purchase and sale
Adjust the strategy of buying and selling electricity
Meet the revenue requirements?
No
Yes
End

Fig. 1 Electricity purchase and sale strategy evaluation process

4 CASE STUDY

4.1 Basic data

In this paper, a numerical example is constructed based on the actual data of a provincial power grid to verify the effectiveness of the proposed algorithm. In the example, the running day is divided into five periods, which are 00:00-06:00 and 06:00-13 respectively. 00, 13:00-16:00, 16:00-21:00 and 21:00-24:00. The price of electricity purchased by the power selling company at different time periods is 0.35, 0.65, 0.55, 0.70 and 0.40 yuan/KWH respectively.

As shown in Table 1, the power selling company can choose to act as the agent of power users in different periods. The five power users used different amounts of electricity, with power user 1 using the highest amount of electricity at 450 MWH and power user 4 using the lowest amount of electricity at only 360 MWH. However, the expected power consumption deviation and its occurrence probability of the five power users are quite different, as shown in Table 2. The probability of deviation of power user 5 is higher, the probability of occurrence is 50% when the deviation amplitude is 15%, 30% when the deviation amplitude is 5%, and only 20% when there is no deviation. In order to simplify the analysis, the comprehensive loss coefficient between purchased electricity and sold electricity is set as 1 in the calculation example, that is, no loss.

Table 1. Different power users use power on a time-sharing basis

user	Time period 1	Time period 2	Time period 3	Time period 4	Time period 5
1	50	100	70	150	80
2	40	80	80	120	60
3	60	90	80	90	60
4	60	80	80	80	60
5	40	120	70	120	50

table 2. Expected power consumption deviation of different power users

ueser	Deviation amplitude 1	Probability of occurrence 1	Deviation amplitude 2	Probability of occurrence 2	Deviation amplitude 3	Probability of occurrence 3
1	15%	5%	5%	40%	0	55%
2	15%	5%	5%	45%	0	50%
3	15%	5%	5%	50%	0	45%
4	15%	10%	5%	20%	0	70%
5	15%	50%	5%	30%	0	20%

4.2 Data Analysis

In order to verify the effectiveness of the proposed algorithm, the trading returns of the following two power buying and selling strategies will be compared.

Strategy 1 is the electricity purchase and sale strategy without considering the deviation assessment;

Strategy 2 is the electricity purchase and sale strategy considering bias assessment proposed in this paper.

Table 3 shows the income analysis results under the two power purchase and sale strategies of the power selling company. Under purchasing and selling strategy 1, the price of electricity sold by the power selling company to users is higher than that purchased by the power generation company. Therefore, the power selling company will act as the agent for the above 5 users, whose expected constituent cost, electricity selling income and comprehensive income from purchasing and selling electricity are in turn. The comprehensive income from electricity purchase and sale obtained in strategy 1 is taken as the expected limit value of comprehensive income from electricity purchase and sale in strategy 2. If the power selling company acts as the agent of the above 5 power users, its comprehensive income from electricity purchase and sale will be 850,000 Yuan, which is lower than the expected limit value of comprehensive income from electricity purchase and sale. The detailed analysis of the expected comprehensive income of each power user agent shows that the expected deviation assessment cost of power user 5 is relatively high, which may lead to the higher time-sharing deviation assessment for the power user agent of the power selling company. After comprehensive comparative analysis, the electricity selling company will only represent the electricity selling business from power users 1 to 4.

table 3. Comparative analysis of the benefits of electricity strategy trading

strategy	Power purchase cost	Sell electricity income	Assessment of deviation	Comprehensive income from electricity purchase and sale
1	350	450	0	100
2	280	410	20	110

The fundamental reason for the above differences in purchasing and selling strategies lies in the different probability of electricity deviation among different power users. In the medium - and long-term market transaction deviation assessment model, different power users have different expected deviation assessment costs. The deviation assessment risk borne by the power selling company acting power users is not the same.

5 CONCLUSION

In order to increase the expected revenue of medium and long term electricity market transaction, this paper proposes a power purchase and sale strategy considering time-sharing deviation assessment. This strategy fully considers the power consumption deviation of different power users and the risk of purchasing power by proxy, and can avoid the deviation assessment cost that may be borne by a large number of high-risk power users by the power selling company.

In the next step, the risk guidance mechanism of power selling companies can be further considered, and the power buying and selling strategies of power users facing different transaction risks can be established.

ACKNOWLEDGMENT

This paper is supported by the Science and Technology Project of the State Grid (Medium and Long term Power Transaction Mechanism of inter-province in North China Region and Beijing-Tianjin-Tang Power Grid adapted to the new power system, and consultation and research on the power grid enterprises' purchasing and selling transmission business strategy, No. : SGNC0000BJJS2100349).

REFERENCES

[1] J. YANG, J. ZHAO, F. WEN and et al. (2017) Key Business Framework and Purchase/Sale Decision-making for Electricity Retailers. Automation of Electric Power Systems, 41: 10-18+20-23.

[2] Y. REN, G. LI, M. ZHOU. (2017) A Survey on Electricity Procurement Decision-making and Risk Management for Electricity Sale Company. Modern Electric Power, 34: 79–85.

[3] H. DONG, J. CHEN, Q. JIA, et al. (2022) Power quality gradation and power purchase and sale strategy considering sensitive users'demand. Electric Power Automation Equipment, 42: 201–209.

[4] Y. PENG, J. LIU, J. LIU, (2022) Electricity Purchasing and Selling Strategies for Electricity Retailers Considering Multiple Types of Retail Packages in Two Level Electricity Market. Power System Technology, 46: 944–957.

[5] L. TANG, J. LIU, Y. YANG, et al, (2019) Study on Strategies of Electricity Procurement and Sale of Power Retailer With Multiple Retail Contract Modes Based on Information Gap Decision Theory. Power System Technology, 43: 1978–1988.

[6] P. ZOU, C. LI, J. GUO, et al, (2017) Optimal Marketing Strategy of Distribution and Retail Companies Considering Interruptible Load. Southern Power System Technology, 11: 71-77.

[7] S. YE, J. WEI, H. RUAN, et al, (2021) Distributional Robust Electricity Trading Strategy of Virtual Power Plant with Deep Interaction of Diversified Flexible Loads. Electric Power Construction, 42: 59-66.

[8] X. DOU, J. WANG, F. YE, et al, (2020) Optimal Dispatching and Purchase-sale Decision Making of Electricity Retailers Considering Virtual Power Plant Combination Strategies. Power System Technology, 44:2078-2086.

[9] S. LUO, C. JIANG, X. WANG, et al, (2019) Power Trading Strategy and Risk Assessment of Electricity Retailing Company Under Power System Reform. Power System Technology, 43: 944-953.

[10] L. WANG, L. ZHANG, F. ZHANG, et al, (2018) Decision-making and Risk Assessment of Purchasing and Selling Business for Electricity Retailers. Automation of Electric Power Systems, 42: 47-54+143.

[11] X. DOU, J. WANG, P. SHAO, et al, (2019) Purchase-sale Strategy of Power Retailers Considering User Contribution Degree. Automation of Electric Power Systems, 43: 2752-2760.

[12] X. DOU, P. ZHANG, J. LI, et al, (2020) Analysis of Power Purchase-sale Strategy of Power Retail Companies With Different Establishment Qualifications. Proceedings of the CSEE, 40: 181-187.

Mechatronic Intelligent Control Technology based on PLC Technology

Lina Guo*, Dayu Fang, Feng Tian
Applied Electronics Department, shandong institute commerce and technology, jinan 250103, shandong, China
*Corresponding author's e-mail: gln20080592@163.com

Abstract

Up to now, PLC output power is getting higher and higher, creating more favorable conditions for PLC control coil. With the maturity of sensing technology, the level of manipulator assembly operation is enhanced. The application of electromechanical intelligent(EI) control technology based on PLC technology is studied and analyzed in this paper. It takes YL-235A optical EI equipment as the research object, introduces its basic working process, introduces the hardware composition, electrical control composition, software required for work and use requirements of YL-235A optical EI training equipment in detail, and finally discusses the overall working process and requirements of the equipment in detail, and discusses the application of EI intelligent control technology of PLC technology.

Keywords. PLC Technology, Electromechanical Integration, Intelligent Control Technology, Application Analysis

1. Introduction

According to the development status of mechatronics and automation technology, the control system based on the combination of "PLC sensor hydraulic components" still occupies a key position; The research and development of adaptive control hydraulic components can truly integrate electronic technology, so that the hydraulic technology will transition from "switch control" to "feedback control", and the accuracy will be more guaranteed. The key is to effectively combine it with flexible manufacturing systems and cells, which can promote the huge development of mechanical manufacturing systems, completely get rid of human needs, and truly achieve the goal of complete automation. The application of EI control technology based on PLC technology is studied and analyzed in this paper.

Many scholars at home and abroad have studied and analyzed the application of EI control technology based on PLC technology. As the equipment of the skill contest, the equipment for EI equipment assembly and debugging undertakes the important task of testing the teaching results. In various competitions, the smaller the defects of the equipment itself, the stronger the ability of competitors can be tested [1]. If the contestants waste unnecessary competition time due to their own equipment defects, the real strength of the contestants will be difficult to reflect, which is a pity for the individual contestants, instructors and even participating schools. As time goes by, the ability of higher levels will be limited. Therefore, the function of the device itself is particularly important [2].

Relying on YL-235A optical EI, this paper introduces the composition and structure of the software and hardware of the equipment in detail, analyzes the hardware structure from two aspects of mechanical structure and electrical control composition, and divides the mechanical structure into three parts: disk, manipulator, material conveying and sorting mechanism, Each part has carried on the simple elaboration to the corresponding basic workflow. The hardware composition, software use, operation process and work requirements of the equipment. A systematic and detailed introduction of the equipment will help to understand the composition and working status of the equipment in detail, and help to understand the reasons for the improvement suggestions. It will play a great role in laying the groundwork and promoting the subsequent demonstration [3-4].

2. Research on Mechatronic Intelligent Control

2.1. Assign Touch Screen

The basic structure of the positioning system consists of PLC, which sends pulses to control the stepping motor and servo motor; HMI man-machine interface: real-time display and detection of working status.

International Conference on Computer Graphics, Artificial Intelligence, and Data Processing (ICCAID 2022)
edited by Ruishi Liang, Jing Wang, Proc. of SPIE Vol. 12604, 1260424
© 2023 SPIE · 0277-786X · doi: 10.1117/12.2674656

The input/output address touch screen of the relevant PLC can display the data by monitoring the transformation of the data register D in the PLC control program. Table 1 shows the input/output address allocation of PLC related to touch screen.

Table 1. Input/output address assignment

Enter address			Output address		
Serial number	name	address	Serial number	name	address
1	Start button	XO	1	20Hz high-speed operation	Y20
2	Stop button	X1		Data register	
3	Chute I in place detection sensor	X12	1	Number of metal workpieces	DO
4	Material trough II in place detection sensor	X14	2	Number of white workpieces	D1

2.2. Create Touch Screen Configuration Screen Project

To create a configuration screen project, first create a new project, establish the connection between and, set the station number and serial port parameters. Then create a configuration screen, create switching elements, and create a numerical display element. At this time, you can follow the steps below [5].

Determine the basic properties of the digital display element In the configuration window, click the numerical display icon in the left component window, drag it into the configuration window, and then the numerical display element properties dialog box will pop up, click the basic properties icon, and set the address type of the numerical display element.

Confirm that the number of the display element is switched to the digital page, set the data type and integer digit, save the file, connect the computer first, download the program to, then connect the computer with the touch screen, and download the project to the touch screen. Then refer to the electrical schematic diagram for electrical wiring; Connect the communication with the touch screen; Set the parameters of the frequency converter according to the control requirements; Confirm that there is no error in parameter setting and connection, and power on. When the communication is normal, the belt conveyor can be started and stopped through the touch screen, and the quantity of different materials can be monitored remotely [6-7].

2.3. Basic Introduction of Electromechanical Equipment Assembly and Commissioning

2.3.1. Disc. In actual production, to send materials or stored workpieces stored in the silo to the processing location, a feeding device is often required. The feeding device in the YL-235A opto EI training device consists of a disc (with a material outlet), a paddle, a micro DC motor, a disc fixing bracket and related fasteners, which are also called a storage disc, a feeding disc or a feeding disc. The materials are directly put into the disk manually. A micro DC motor is installed under the disk chassis. The paddle in the disk is installed on the shaft of the DC motor, and the micro DC motor drives the paddle to rotate [8]. The micro DC motor is equipped with a gearbox, so the rotation speed of the paddle is relatively slow.

2.3.2. Outlet detection sensor. The sensor represents a device that can intuitively feel the measured object and convert it into usable signal output according to certain rules [9]. The outlet detection sensor uses a diffuse reflective photoelectric switch, and the optical receiver and optical transmitter are installed at the same position to form an integrated structure. When the optical transmitter operates normally, it always keeps the state of transmitting and detecting light. If there is no object near the position in front of the switch, the light will not be reflected and sent to the receiver at this time, and the position close to the switch will remain normal without any other action; On the contrary, there is an object near the front of the switch. At this time, the light is blocked by the object and reflects light of a certain intensity. The receiver makes the switch act under the diffuse reflection of the light and adjusts the output state [10-11].

2.3.3. Warning light. Generally, different signs will be marked on the electromechanical equipment to remind the operator of the equipment operation status, and the warning light will send out alarm signals according to the user's setting requirements. There are many types of warning lights, which can be divided into green, red, multi-color and yellow according to their colors; The warning lights can also be divided into long light type and flashing type according to the luminous state [12].

2.4. YL-235A Optical Electromechanical Integration Equipment

YL-235A opto electromechanical integration training and assessment equipment mainly includes electromechanical equipment and facilities, aluminum alloy guide rail training platform, touch screen module unit, PLC module unit, button module unit, frequency converter module unit, simulation production equipment training module, power module unit, multiple sensors and terminal strip. The system uses a dismountable and open structure. The training situation can realize the assembly of mechanical parts, and the training device can be assembled according to the current mechanical part. It can also add other mechanical parts to become other training equipment after assembly, improve the flexibility of equipment operation, so as to meet the assembly requirements of the competition and the actual teaching stage, and can simulate the electromechanical integration production state.

The three-phase AC asynchronous motor of the belt conveyor in the YL-235A optical mechanical electrical integration training device uses a frequency converter to control speed regulation, and each section of the frequency converter is led out to the frequency converter module panel. The general working parameters of frequency converter in YL-235A optical electromechanical integration training device are set as three speed settings, and the parameters involved generally include upper and lower frequency, acceleration and deceleration time, etc.

3. EI Design Based on PLC Technology

3.1. Mechanical Gripper Reconstruction Design

3.1.1. Problems and cause analysis of pneumatic mechanical gripper

Grasping is not allowed due to the detection range of the material detection sensor. After installation, the pneumatic mechanical gripper needs to conduct hardware detection before use to ensure that the pneumatic mechanical gripper can accurately grasp the materials on the feeding table. Because the material detection sensor on the feeding table has a certain detection range, which makes the material on the feeding table not stay at a fixed position every time, but a material detection area with a certain margin. In any position within the material detection area, as long as the material is detected, the material will immediately stop according to the program settings and wait for the gripper to grasp, while the swinging position of the manipulator is only fixed each time, The position of the gripper to grasp the material is also uniquely fixed, and it will not make appropriate fine adjustment with the small offset of the material position. This led the contestants to adjust the position and range of the material detection sensor to a small fixed range as much as possible when installing and debugging the material detection sensor.

3.1.2. Analysis of measures affecting inaccurate grasping of pneumatic mechanical gripper

The material detection range is controlled by the material detection sensor. Even if the material detection range of the material detection sensor is adjusted to the minimum, the error between it and the material grasped by the mechanical gripper cannot be completely eliminated. The adjustment of the material detection sensor can not be adjusted any more, but can only be considered on the manipulator. At present, the mechanical structure of the manipulator gripper is double-claw structure, in which the gripper is straight, and the gripper material is cylindrical material. The straight two claw structure has the defect of insufficient natural fit when grasping cylindrical materials. See Figure 1 for the specific capture.

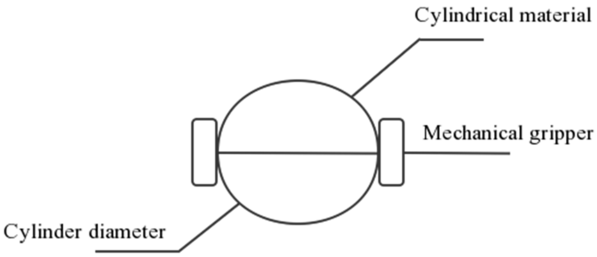


Figure 1. Schematic Diagram of Material Grasping Position of Mechanical Claw

Based on the analysis of the above factors, if the two claw structure of the mechanical hand is replaced by the three claw structure, and the grasping surface is made into an arc, the awkward situation of the only fixed position can be eliminated. Program control requirements: all requirements for the corresponding position of the actuator shall be met by using the pneumatic transmission method, including the requirements for the movement track, the movement sequence, the movement time and the movement speed. The command can also be sent to the actuator according to the control system information.

The main components of the actuator include wrist, hand, column and arm, and some of them are also equipped with walking mechanism.

Hand: the translational type is used in a small range in the market. The reason for this phenomenon is that the structure of this type is complex. The force transmission mechanism generates the clamping force under the action of fingers to achieve the clamping and lowering of objects.

Arm: It is mainly used to support hands, objects and wrists. It can drive fingers to grasp an object and move the goods to the designated area according to preset requirements.

Base: it is the basic part of the manipulator. The base is installed with the drive system and multiple components of the manipulator's operating mechanism. The function that can be realized is connection and support. The operating system of the pneumatic manipulator is shown in Figure 2.

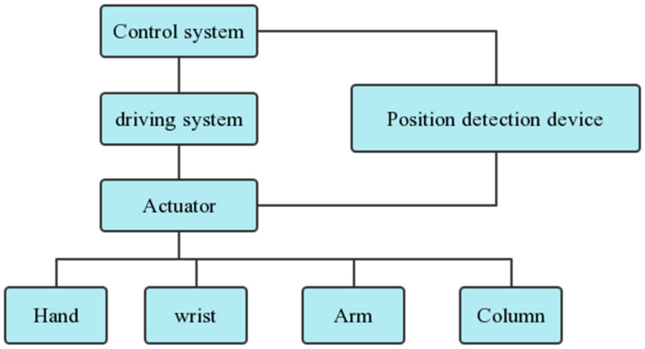


Figure 2. Pneumatic manipulator operating system

3.2. Design of Manipulator Hardware Transformation Measures

3.2.1. Hardware composition of mechanical gripper

At present, the mechanical gripper used in YL-235A optical EI equipment is a two claw internal supporting mechanical gripper. This kind of gripper mechanism is basically similar to the chuck in function. It can clamp cylindrical or centering parts. Its characteristic is that the rack rotates when driven by the cylinder, and then drives the gear and uniformly distributes three or four centering sliders around the circumference to complete the centripetal movement. At this time, the gripper shows clamping and loosening actions.

3.2.2. Calculation and analysis of mechanical gripper

The gripper hand is designed according to the clamping force exerted by the finger on the workpiece. At the same time, the direction, action point and size of the gripper should be studied and calculated. Generally, when designing the clamping force of the gripper, the inertia force or inertia moment caused by the gravity of the workpiece should be avoided to the maximum extent to ensure that the workpiece can always maintain a high reliability clamping state. The following formula is used to calculate the force of finger clamping workpiece:

$$F_N \geq K_1 \bullet K_2 \bullet K_3 \bullet G \tag{1}$$

In the above formula, K1 represents the safety factor. According to the requirements of the design manipulator and the process requirements, the value range of the safety factor is 1.2 to 2.0, and 1.5 is selected here; K2 represents the condition coefficient of the workpiece. When setting the workpiece coefficient, the influence of inertial force shall be analyzed, and for acceleration calculation, G represents the weight of the grasped workpiece; K3 represents the azimuth coefficient, which is determined by referring to the workpiece shape, finger shape, and the position between the workpiece and fingers when setting the azimuth coefficient. The clamping force F is calculated by the above formula.

The clamping force and driving force formulas are as follows:

$$F = \frac{Fc}{2b\sin\alpha} \tag{2}$$

In the formula, b represents the distance between the pins; C represents the distance between roller and pin shaft; A represents the tilt angle of the wedge.

According to the calculation, it can be inferred that the three claw manipulator is suitable for replacing the manipulator part of YL-235A optical EI equipment and replacing the two claw mechanism to work. In order to further ensure the stability of its working operation, it is necessary to add a force sensor on the manipulator to ensure that the manipulator claw can accurately grasp the workpiece without error.

4. Mechatronic Intelligent Control Technology based on PLC Technology

4.1. PLC Module

There are many types of devices and elements included in automatic control. This paper adopts PLC programming mode. This method has strong functions, simple operation, complete hardware facilities, higher cost performance, strong anti-interference, and high reliability. In practical application, the workload required for designing, installing and debugging the system is small, so it is widely used in industrial control. The PLC used on YL-235A optical EI training device is Mitsubishi FX2N48-MR, which is installed on a small module, connected to the output terminal, input terminal, external power connection line terminal and internal DC24V power supply, and finally output to the module panel jack.

On the left side of the panel, a three row jack is arranged in parallel to connect the PLC output terminal, and two jacks are arranged below the left side to connect the PLC power supply and the PLC power switch. Two sockets are set on the upper right side to lead out the DC24V power supply inside the PLC. The two rows of sockets arranged in parallel on the right side lead out sockets to the PLC input terminal. Two rows of switches are also set on the right side to provide the required input signals to the PLC input terminal. Generally, the external wiring input terminal of PLC module adopts the sink point wiring mode, and the output terminal wiring is set according to the load requirements, and the grouping wiring mode is adopted.

The basic operation process of PLC software is:

To create a new project, select Create New Project in the menu bar, open the Create New Project dialog box, and set the project name option when saving the program; Input the program. After the new project is created, the ladder diagram can be input in the programming area; Download the program. After the ladder diagram is compiled, click PLC Write in the online menu to download the control program.

4.2. Touch Screen Programming Software

The embedded configuration software simulation environment and configuration environment are similar to a tool software, which can run on a PC and design user configuration functions and objectives according to different processing methods selected by users in the configuration project. The MCGS embedded configuration software can be combined with a variety of hardware devices to develop more devices for collecting field data, processing data and controlling the system. Configure different types of intelligent instruments in a flexible way, and develop a variety of special equipment. The composition of MCGS embedded version software configuration is shown in Figure 3. The user system is generated based on MCGS embedded version.

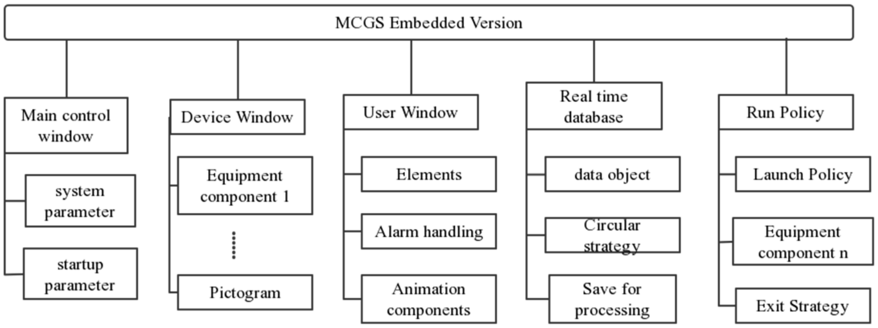


Figure 3. MCGS embedded version configuration user system composition

The basic operation process of MCGS software is to create a new project, create a project name, and create a blank project. Determine the connection and communication mode between PLC and touch screen. Create the configuration project screen. Save the file. Set the parameters of the frequency converter according to the control requirements. Conduct communication connection test. Power on after confirming that there is no error in parameter setting and connection.

4.2.1. Complete the design and connection of part processing line circuit

Allocation of PLC input and output address: determine the number of input points, determine the output required to complete the part processing line according to the working process and pneumatic system diagram, list the PLC input and output address allocation table, draw the electrical control schematic diagram of the part processing line according to the allocation, and finally complete the electrical installation process.

4.2.2. Preparation of PLC control program

Programming idea: take action control method as the basic idea, and cooperate with experience method to complete the overall program. Function analysis: including the realization of self inspection, main functions and reset functions of the equipment. It is required to press the self inspection button for self inspection during each power on, and this function can be executed by using the power on initialization pulse of PLC. The equipment self inspection shall pay special attention to the action sequence and protection function during self inspection. The reset function of the device requires unconditional reset as long as the device is not in a cyclic state. The basic action of the equipment can only be operated after the self inspection is normal, and the reset action is effective when it is no longer in cyclic operation.

Self check function and normal operation. The main function program of the system: This part mainly includes the processing of the beginning of the main function of the system, the manipulator action control, the control of the belt conveyor and the realization of the sorting process. Preparation of reset program: Before the reset process, the functions of other parts of the system that are running must be clear, so as to avoid the confusion of simultaneous action of two processes. Indicator lamp programming.

4.3. Workflow of Optical EI Device

The operator clicks Open on the touch screen. First, the device needs to be reset. After resetting to the specified position, the PLC turns on the feeding motor. At this time, the photoelectric sensor for material detection is started for detection; If the material detection photoelectric sensor detects no material after several seconds of operation of the feeder, it indicates that no material is placed on the feeder, and the operation must be stopped and an alarm must be sent to the staff; The detection shows that there is material, and sends the material position and signal to the PLC. Under the action of the PLC, the manipulator drives the arm to lower and grab the object, then raises and retracts the arm, turns the arm to the right to the far right, and then extends the arm. To open the conveyor belt to continue transmitting materials, the manipulator also needs to reset to start the next process.

5.Conclusions

Relying on YL-235A optical EI, this paper studies and analyzes the application of PLC technology in EI control technology. At present, it is still in the initial stage. In order to promote the promotion of practical applications, there is still a lot of work to be done. A lot of work needs to be further supplemented and improved. In addition, some aspects may be considered unsound. For example, although the design selection of electrical components can also meet the requirements of the system, the actual design needs further refinement and improvement. In a word, this electromechanical precise positioning device and intelligent control system still need to be verified in specific working conditions for feedback improvement.

References

[1] Khairul Annuar Abdullah, Zuriati Yusof, Raja Mohd Tariqi B. Raja Lope Ahmad, Muhammad Fairuz Abd. Rauf, Zuraidy Adnan, Wan Azlan Wan Hassan, Riza Sulaiman: Algebraic models based on trigonometric and Cramer's rules for computing inverse kinematics of robotic arm. Int. J. Mechatronics Autom. 9(1): 1-11 (2022)

[2] Ali Al-Ghanimi, Abdal-Razak Shehab, Adnan Alamili: A tracking control design for linear motor using robust control integrated with online estimation technique. Int. J. Mechatronics Autom. 9(3): 151-159 (2022)

[3] Hang Cui, Jiaming Zhang, William R. Norris: A real-time embedded drive-by-wire control module for self-driving cars with ROS2. Int. J. Mechatronics Autom. 9(2): 61-71 (2022)

[4] Chiharu Ishii, Ryo Sugiyama, Takahiro Yamada: Proposal of guidelines for application of endoskeleton assist suit 'sustainable' to transfer assistance in nursing care. Int. J. Mechatronics Autom. 9(2): 81-89 (2022)

[5] Ken'ichi Koyanagi, Daisuke Takata, Takumi Tamamoto, Kentaro Noda, Takuya Tsukagoshi, Toru Oshima: Design and development of a 3D-printed balloon type actuator for a hybrid force-display glove. Int. J. Mechatronics Autom. 9(1): 47-59 (2022)

[6] Naoki Moriya, Hiroki Shigemune, Hideyuki Sawada: A robotic wheel locally transforming its diameters and the reinforcement learning for robust locomotion. Int. J. Mechatronics Autom. 9(1): 22-31 (2022)

[7] Fusaomi Nagata, Kei Furuta, Kohei Miki, Maki K. Habib, Keigo Watanabe: Implementation and evaluation of calibration-less visual feedback controller for a robot manipulator DOBOT on a sliding rail. Int. J. Mechatronics Autom. 9(3): 142-150 (2022)

[8] Abhilasha Singh, V. Kalaichelvi, R. Karthikeyan: Prototype design and performance analysis of genetic algorithm-based SLAM for indoor navigation using TETRIX Prizm mobile robot. Int. J. Mechatronics Autom. 9(1): 32-46 (2022)

[9] Nina Tajima, Koichiro Kato, Eriko Okada, Nobuto Matsuhira, Kanako Amano, Yuka Kato: Development of a walking-trajectory measurement system. Int. J. Mechatronics Autom. 9(3): 113-122 (2022)

[10] Yamato Umetani, Masahiko Minamoto, Shigeki Hori, Tetsuro Miyazaki, Kenji Kawashima: Estimating future forceps movement using deep learning for robotic camera control in laparoscopic surgery. Int. J. Mechatronics Autom. 9(2): 72-80 (2022)

[11] Mitsuhiro Yamano, Naoya Hanabata, Akira Okamoto, Toshihiko Yasuda, Yasutaka Nishioka, M. D. Nahin Islam Shiblee, Kazunari Yoshida, Hidemitsu Furukawa, Riichiro Tadakuma: Development and motion analysis of a light and many-joint robot finger using shape memory gel and tendon-driven mechanism with arc route. Int. J. Mechatronics Autom. 9(2): 99-111 (2022)

[12] Keita Abe, Yumeta Seki, Yu Kuwajima, Ayato Minaminosono, Shingo Maeda, Hiroki Shigemune:Low-Voltage Activation Based on Electrohydrodynamics in Positioning Systems for Untethered Robots. J. Robotics Mechatronics 34(2): 351-360 (2022)

Research on stability control of blast furnace coal injection based on LMI

Pengcheng Xiong[a b], Guimei Cui*[a b]
[a] School of Information Engineering, Inner Mongolia University of Science and Technology, Baotou 014010, Nei Mongol, China; [b] School of Information Engineering, Inner Mongolia University of Science and Technology, Baotou 014010, Nei Mongol, China
* Corresponding author: cguimei1@163.com

ABSTRACT

In the blast furnace (BF) ironmaking process, to solve the problem of uncertain thermal hysteresis in pulverized coal combustion system, the uncertain hysteresis is converted into the uncertainty of the coefficient matrix of the system state equation. Based on Lyapunov Stability Theory and Linear Matrix Inequalities (LMI) method, it is proved that the sufficient condition for stable control of pulverized coal combustion system through state feedback control is equivalent to the solution of a set of LMI. Compared with the traditional PID control mode, The simulation results show that this method can effectively realize the stable control of coal injection in blast furnace.

Keywords: BF ironmaking; Uncertain hysteresis; LMI; Stability

1. INTRODUCTION

The steel industry is the pillar industry of China's national economic development. The energy consumption of the steel industry is huge. The total energy consumption of China's steel industry accounts for 13.2% of the total national energy consumption, which is far higher than the average of 6.7% of the world's iron and steel industry according to the international energy deployment statistics [1]. Among them, the BF ironmaking is the link with the largest pollutant emission and energy consumption, so the important technical means of replacing coke with coal is of great significance for energy conservation, cost reduction and national economy improvement. According to relevant data, the world’s advanced level of pulverized coal injection is 180-220 kg/t, and the world-class level of pulverized coal injection is 220-240 kg/t. However, the average coal injection ratio of Chinese steel enterprises hovers at 140kg/t, China’s BF injection pulverized coal technology is lower than the world’s advanced level [2]. The main reason is that manual control model is adopted for BF coal injection at present, and PID controller is used for coal injection compensation to stabilize furnace temperature. However, there is a thermal lag phenomenon in the pulverized coal combustion process, and the furnace head dare not inject pulverized coal easily, which may lead to furnace temperature fluctuation or even out of control of the blast furnace. Therefore, it is significant to study the thermal hysteresis in the process of pulverized coal combustion to enhance international competitiveness and improve national economy.

In the actual production of BF, pulverized coal and high temperature hot air are injected into the BF from the tuyere, the pulverized coal whirl in front of the tuyere to form the tuyere raceway under high pressure [3]. Pulverized coal and coke are continuously combusted in the raceway, which is called pulverized coal combustion system. In the process of coal injection of the BF, the temperature of hearth is cooled first and then heated. The main reason is that the pulverized coal is decomposed to absorb heat during combustion, make the temperature of hearth decrease. Until the newly increased coal injection changes the gas volume and the concentration of reducing gas in the BF, which rises the temperature of hearth, which is called the phenomenon of thermal hysteresis [4]. However, the thermal hysteresis produced by pulverized coal combustion in the raceway is variable and uncertain, which reduces the stability of the pulverized coal combustion system, and the coal injection rate is proportional to the thermal hysteresis. At present, most of studies are about the predictive modeling to compensate the coal injection rate for the furnace temperature stability control [5-7], without considering the variable hysteresis problem in the pulverized coal combustion process.

Aiming at the uncertain factors with variable hysteresis in the pulverized coal combustion system, based on the continuous time model of the pulverized coal combustion system in reference [8], this paper transforms it into a linear discrete system with uncertain. By using Lyapunov theory, the sufficient conditions for stable control of pulverized coal combustion system are derived. Based on LMI method, the sufficient conditions for state feedback to make the system stable are equivalent to the feasible solution of LMI.

International Conference on Computer Graphics, Artificial Intelligence, and Data Processing (ICCAID 2022)
edited by Ruishi Liang, Jing Wang, Proc. of SPIE Vol. 12604, 1260425
© 2023 SPIE · 0277-786X · doi: 10.1117/12.2674537

2. PROBLEM STATEMENT

According to the reference [8], the continuous time model of pulverized coal combustion system is shown in the Equation (1).

$$\begin{cases} \dot{\boldsymbol{x}}(t) = \boldsymbol{A}\boldsymbol{x}(t) + \boldsymbol{B}\boldsymbol{u}(t - \tau) \\ \boldsymbol{y}(t) = \boldsymbol{C}\boldsymbol{x}(t) \end{cases} \tag{1}$$

Where $\boldsymbol{x}(t)$ are the temperature of tuyere raceway and the temperature of hearth (characterization of [Si] content in molten iron), respectively, $\boldsymbol{u}(t)$ represents amount of coal injection, τ is the thermal hysteresis, $\boldsymbol{A} \in \mathbf{R}^{n\times n}$, $\boldsymbol{B} \in \mathbf{R}^{n\times 1}$, $\boldsymbol{C} \in \mathbf{R}^{1\times n}$ are the coefficient matrix, the input coefficient matrix and output coefficient matrix, respectively. According to the identification of historical data of blast furnace, the maximum delay of PC combustion process is $\tau_{\max}$, The sampling period is T, and $T > \tau_{\max}$. In consideration of the influence of thermal hysteresis τ of pulverized coal combustion, the above equation is discretized, such as the discrete time model of pulverized coal combustion system in closed loop tuyere raceway of the Equation (2):

$$\begin{cases} \boldsymbol{X}(k + 1) = \boldsymbol{A}_0\boldsymbol{X}(k) + \boldsymbol{B}_{\mathrm{d}}(\tau)\boldsymbol{U}(k) + \boldsymbol{B}_{\mathrm{d1}}(\tau)\boldsymbol{U}(k - 1) \\ \boldsymbol{Y}(k) = \boldsymbol{C}\boldsymbol{X}(k) \end{cases} \tag{2}$$

Where $\boldsymbol{A}_0 = e^{AT}$, $\boldsymbol{B}_0(\tau) = \int_0^{T-\tau} e^{At}dt \cdot \boldsymbol{B}$, $\boldsymbol{B}_{\mathrm{d1}}(\tau) = \int_{T-\tau}^{T} e^{At}dt \cdot \boldsymbol{B}$.

The coefficient matrix in equation (2) $\boldsymbol{B}_0(\tau)$, $\boldsymbol{B}_{\mathrm{d}}(\tau)$ varies with τ. According to matrix theory, The sufficient and necessary condition for $\boldsymbol{A}^{n\times n}$ to be diagonalized is that $\boldsymbol{A}^{n\times n}$ has n characteristic roots $\lambda_1, \lambda_2, \cdots, \lambda_{\mathrm{n}}$ which are not 0 and are different from each other, then $\boldsymbol{A} = \boldsymbol{P}\mathrm{diag}(\lambda_1, \lambda_2, \cdots, \lambda_{\mathrm{n}})\boldsymbol{P}^{-1}$, Where $\boldsymbol{P}$ is a matrix composed of eigenvectors of matrix $\boldsymbol{A}$. Therefore, the Equation (2) is transformed into the Equation (3) of a linear discrete model with uncertain of hysteresis.

$$\begin{cases} \boldsymbol{X}(k + 1) = \boldsymbol{A}_0\boldsymbol{X}(k) + (\boldsymbol{B}_0(\tau) + \Delta\boldsymbol{B}_0(\tau))\boldsymbol{U}(k) + (\boldsymbol{B}_{\mathrm{d}}(\tau) + \Delta\boldsymbol{B}_{\mathrm{d}}(\tau))\boldsymbol{U}(k - 1) \\ \boldsymbol{Y}(k) = \boldsymbol{C}\boldsymbol{X}(k) \end{cases} \tag{3}$$

$$\begin{cases} \boldsymbol{B}_{\mathrm{d}}(\tau) = \boldsymbol{B}_0(\tau) + \Delta\boldsymbol{B}_0(\tau) \\ \boldsymbol{B}_{\mathrm{d1}}(\tau) = \boldsymbol{B}_{\mathrm{d}}(\tau) + \Delta\boldsymbol{B}_{\mathrm{d}}(\tau) \\ \Delta\boldsymbol{B}_0(\tau) = \boldsymbol{D}\boldsymbol{F}(\tau)\boldsymbol{E} \\ \Delta\boldsymbol{B}_{\mathrm{d}}(\tau) = -\boldsymbol{D}\boldsymbol{F}(\tau)\boldsymbol{E} \end{cases}$$

Where $\boldsymbol{B}_0(\tau) = \boldsymbol{P}\mathrm{diag}(-\frac{1}{\lambda_1}, -\frac{1}{\lambda_2}, \cdots, -\frac{1}{\lambda_{\mathrm{n}}})\boldsymbol{P}^{-1}\boldsymbol{B}$, $\boldsymbol{B}_{\mathrm{d}}(\tau) = \boldsymbol{P}\mathrm{diag}(\frac{1}{\lambda_1}\mathrm{e}^{\lambda_1 \mathrm{T}}, \frac{1}{\lambda_2}\mathrm{e}^{\lambda_2 \mathrm{T}}, \cdots, \frac{1}{\lambda_{\mathrm{n}}}\mathrm{e}^{\lambda_{\mathrm{n}} \mathrm{T}})\boldsymbol{P}^{-1}\boldsymbol{B}$

$\boldsymbol{D} = \boldsymbol{P}\mathrm{diag}(\frac{1}{\lambda_1}\mathrm{e}^{\lambda_1\alpha_1}, \frac{1}{\lambda_2}\mathrm{e}^{\lambda_2\alpha_2}, \cdots, \frac{1}{\lambda_{\mathrm{n}}}\mathrm{e}^{\lambda_{\mathrm{n}}\alpha_{\mathrm{n}}})$, $\boldsymbol{F}(\tau) = \mathrm{diag}(\mathrm{e}^{\lambda_1(T-\tau-\alpha_1)}, \mathrm{e}^{\lambda_2(T-\tau-\alpha_2)}, \cdots, \mathrm{e}^{\lambda_{\mathrm{n}}(T-\tau-\alpha_n)})$

$\boldsymbol{E} = \boldsymbol{P}^{-1}\boldsymbol{B}$, $\boldsymbol{D} \in \mathbf{R}^{n\times n}$, $\boldsymbol{E} \in \mathbf{R}^{n\times n}$, $\boldsymbol{F}(\tau) \in \mathbf{R}^{n\times n}$. $\alpha_1, \alpha_2, \cdots, \alpha_{\mathrm{n}}$ should be selected to meet the condition $\boldsymbol{F}^{\mathrm{T}}(\tau)\boldsymbol{F}(\tau) \le \boldsymbol{I}$.

According to above derivation, it can be obtained that the discrete model (2) of the pulverized coal combustion system can be transformed into a linear discrete model with uncertain (3). That is, the stability control problem of the pulverized coal combustion system can be transformed into the stability control problem of the uncertain discrete system with hysteresis.

3. STABILITY DESIGN

Design of state feedback controllers for linear discrete systems with uncertain as follows:

$$\boldsymbol{U}(k) = \boldsymbol{Kx}(k) \tag{4}$$

Substituting equation (4) into equation (3), the closed loop control system model of pulverized coal combustion in the tuyere raceway is obtained as follows:

$$\begin{cases} \boldsymbol{X}(k+1) = \left[\boldsymbol{A}_0 + \boldsymbol{B}_0(\tau)\boldsymbol{K} + \boldsymbol{DF}(\tau)\boldsymbol{EK}\right]\boldsymbol{X}(k) + \left[\boldsymbol{B}_\mathrm{d}(\tau) - \boldsymbol{DF}(\tau)\boldsymbol{E}\right]\boldsymbol{KX}(k-1) \\ \quad \boldsymbol{Y}(k) = \boldsymbol{CX}(k) \end{cases} \tag{5}$$

Where $\boldsymbol{G} = \boldsymbol{A}_0 + \boldsymbol{B}_0(\tau)\boldsymbol{K} + \boldsymbol{DF}(\tau)\boldsymbol{EK}$, $\boldsymbol{H} = \boldsymbol{B}_\mathrm{d}(\tau) - \boldsymbol{DF}(\tau)\boldsymbol{E}$.

Lemma 1: By using the control law shown in Equation (4), If the existence of positive definite symmetric matrix $\boldsymbol{X}$、$\boldsymbol{Y} \in \mathbf{R}^{n\times n}$, and feedback gain matrix $\boldsymbol{K} \in \mathbf{R}^{m\times n}$, as well as constant ζ , makes the following matrix inequality (6) hold, then $\boldsymbol{U}(k) = \boldsymbol{Kx}(k)$ is the stable control law of system (4).

$$\begin{pmatrix} \zeta\boldsymbol{DD}^\mathrm{T} - \boldsymbol{X} & \boldsymbol{A}_0 + \boldsymbol{B}_0\boldsymbol{K} & \boldsymbol{B}_\mathrm{d}\boldsymbol{K} & \boldsymbol{0} \\ (\boldsymbol{A}_0 + \boldsymbol{B}_0\boldsymbol{K})^\mathrm{T} & -\boldsymbol{X} + \boldsymbol{Y} & \boldsymbol{0} & (\boldsymbol{EK})^\mathrm{T} \\ (\boldsymbol{B}_\mathrm{d}\boldsymbol{K})^\mathrm{T} & \boldsymbol{0} & \boldsymbol{Y} & -\boldsymbol{EK} \\ \boldsymbol{0} & \boldsymbol{EK} & -\boldsymbol{EK} & -\zeta\boldsymbol{I} \end{pmatrix} < \boldsymbol{0} \tag{6}$$

Corollary 1[9]**:** If there is a matrix $\boldsymbol{M} \in \mathbf{R}^{m\times n}$ and positive definite symmetric matrix $\boldsymbol{Q}$、$\boldsymbol{R} \in \mathbf{R}^{n\times n}$, as well as constant ζ , makes the following matrix inequality (7) hold, then $\boldsymbol{U}(k) = \boldsymbol{Kx}(k)$ is the stable control law of system (4).

$$\begin{pmatrix} \zeta\boldsymbol{DD}^\mathrm{T} - \boldsymbol{Q} & \boldsymbol{A}_0\boldsymbol{Q} + \boldsymbol{B}_0\boldsymbol{M} & \boldsymbol{B}_\mathrm{d}\boldsymbol{M} & \boldsymbol{0} \\ (\boldsymbol{A}_0\boldsymbol{Q} + \boldsymbol{B}_0\boldsymbol{M})^\mathrm{T} & -\boldsymbol{Q} + \boldsymbol{R} & \boldsymbol{0} & (\boldsymbol{EM})^\mathrm{T} \\ (\boldsymbol{B}_\mathrm{d}\boldsymbol{M})^\mathrm{T} & \boldsymbol{0} & \boldsymbol{Y} & -\boldsymbol{EM} \\ \boldsymbol{0} & \boldsymbol{EM} & -\boldsymbol{EM} & -\zeta\boldsymbol{I} \end{pmatrix} < \boldsymbol{0} \tag{7}$$

Where $\boldsymbol{Q} = \boldsymbol{X}^{-1}$, $\boldsymbol{M} = \boldsymbol{KX}^{-1}$, $\boldsymbol{R} = \boldsymbol{X}^{-1}\boldsymbol{YX}^{-1}$. Cui Guimei et al[10].proved the corollary.

4. SIMULATION EXAMPLE

In this section, taking the 2500m^3 BF as an example. According to the identification of historical data of BF, the maximum delay of pulverized coal combustion process is 0.3 h. The continuous time model of pulverized coal combustion system is obtained as shown in Equation (8). sampling period T is 1h, all experiments were performed using the MATLAB 2016b toolbox.

$$\begin{cases} \dot{x}(t) = \begin{bmatrix} -0.094 & 0 \\ 0.016 & -0.043 \end{bmatrix} x(t) + \begin{bmatrix} 0.014 \\ 0 \end{bmatrix} u(t-\tau) \\ y(t) = \begin{bmatrix} 1 & 0 \end{bmatrix} x(t) \end{cases} \tag{8}$$

The coefficient matrix of the discrete time model is:

$$\boldsymbol{A}_0 = \begin{bmatrix} 0.9103 & 0 \\ 0.0149 & 0.9579 \end{bmatrix}, \boldsymbol{B}_0 = \begin{bmatrix} 0.0095 \\ 0.0001 \end{bmatrix}, \boldsymbol{B}_\mathrm{d} = \begin{bmatrix} 0.0039 \\ 0.0001 \end{bmatrix},$$

$$\boldsymbol{D} = \begin{bmatrix} 0 & -22.1071 \\ -10.5912 & 6.9355 \end{bmatrix}, \boldsymbol{E} = \begin{bmatrix} -0.014 \\ 0.014 \end{bmatrix}, \boldsymbol{F} = \begin{bmatrix} 0.9406 & 0 \\ 0 & 0.9741 \end{bmatrix}.$$

Using LMI toolbox to solve Equation (7), the feedback gain matrix of the model is $\boldsymbol{K} = 10^{-3} \times \left[-0.5765, -0.3652\right]$. The initial condition is $\boldsymbol{x}(0) = \left[1, 0\right]$, and an increment of coal injection volume with amplitude of 1t/h. The state response curve of the system is obtained through simulation in Simulink, as shown in Figure 2. Figure 1 shows that the PID controller is used as the controller, and the input is the coal injection amount after data preprocessing of a steel plant. Given an increment of amplitude of 1t/h, affected by thermal hysteresis of pulverized coal combustion, the temperature fluctuation amplitude in the raceway is large, which is not conducive to the stable and smooth operation of the blast furnace. It can be seen from Figure 2 that the feedback controller designed in this paper can ensure the stability of $\boldsymbol{x}_1(\mathrm{t})$ (the temperature of raceway) after increasing pulverized coal injection, and the temperature of hearth $\boldsymbol{x}_2(\mathrm{t})$ ([Si] content of molten iron) is stable at a higher temperature value due to the increase of coal injection.

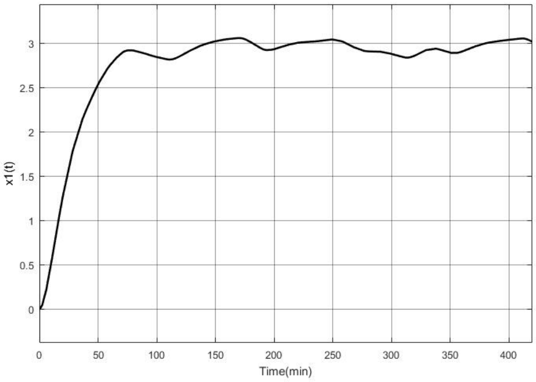


Figure 1. Temperature response curve of raceway region in PID controller

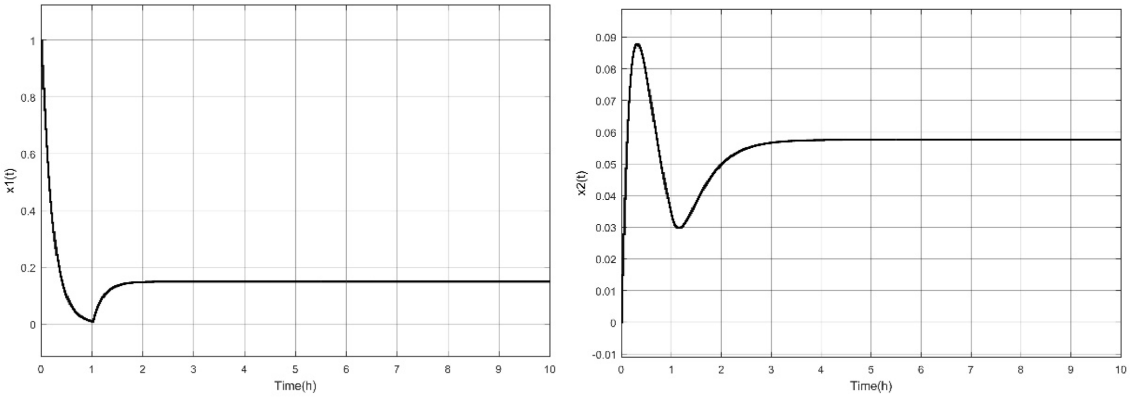


Figure 2. State feedback response curve of closed loop system

5. CONCLUSION

On the basis of the continuous time model of the pulverized coal combustion system, and the model with hysteresis uncertain after discretization, then the stability control problem of the pulverized coal combustion system is studied. It is deduced that the design of the stable static state feedback controller of the pulverized coal combustion system is equivalent to solving the LMI. Through simulation, it is verified that the system has good stability after adding pulverized coal.

Acknowledgment

This work was supported in part by the Natural Science Foundation of China (No.61763039).

REFERENCES

[1] Xing Yi, Cui Yongkang, Tian Jinglei. "Application status and prospect of low carbon technology in iron and steel industry," Chinese Journal of Engineering, Papers 44(04), 801-811 (2022).

[2] Cui Guimei, Yao Yanqing, Ma Xiang, Zhang Yong. "Fuzzy Comprehensive Evaluation Model of Pulverized Coal Digestibility in Blast Furnace Raceway Based on the Fusion of Subjective and Objective Evidence," Papers 44(04), 801-811 (2022).

[3] Xiao Weifeng, He Xinjie. "Numerical simulation of combustion behavior for semi-coke in raceway of blast furnace," China Metallurgy, Papers 32(5), 86-92 (2022).

[4] Cui Guimei, Chen Rong, Ma Xiang, Zhang Yong. "Decision-making optimization of coal injection volume based on evaluation of blast furnace condition," Control and Decision, Papers 32(11), 2803-2809 (2020).

[5] Cui Bo, Chen Wei, Wang Baoxiang. "Prediction of silicon content in hot metal of blast furnace based on grey correlation analysis and extreme learning machine," Metallurgical Industry Automation, Papers 46(1), 54-62 (2022).

[6] Guan, Xin. " Prediction of Blast Furnace Temperature Based on Improved Extreme Learning Machine," Lecture Notes in Electrical Engineering, Papers 634, 292-298 (2020).

[7] Jiao, Hongyuan; Zhang, Yingwei; Luo, Chaomin; Bi, Zhuming. "Collaborative Multiple Rank Regression for Temperature Prediction of Blast Furnace," IEEE Transactions on Instrumentation and Measurement, Papers 71, (2022).

[8] Cui Guimei, Li Shuqi Zhang Yunqiang, Ma Xiang. "Research on synergistic optimization control strategy of blast furnace pulverized coal injection based on T-S model," Journal of Iron and Steel Research, Papers 34(03), 222-230 (2022).

[9] Cui Guimei, Mu Zhichun, Li Xiaoli. "Guaranteed cost control of networked control systems," Chinese Journal of Engineering, Papers 28(6), 595 (2006).

[10] Cui Guimei, Li Xiaoli, Mu Zhichun. "Discrete fuzzy control of network closed loop system," Journal of Liaoning Technical University, Papers 26(2), 260-263 (2007).

Sentiment analysis of food safety Internet public opinion based on XLNet

Hu Wang, Chaofan Jiang *, Changbin Jiang and Di Li
School of Management, Wuhan University of Technology, Wuhan 430000, China;
* Corresponding author: jiangchaofan9812@163.com

ABSTRACT

Internet public opinion sentiment analysis is significant for managing and controlling food safety events. Since emotions can play a decisive role in behavior, netizens' emotions towards the food safety events will influence their expressions of opinions on the Internet, thereby influencing the development of public opinion on the events. However, few scholars have analyzed the sentiment of Internet public opinion regarding food safety. We employ XLNet, a dynamic text representation method, to build context-dependent word vectors for the distributed representation of Internet public opinion in order to better analyze Internet public opinion on food safety events according to its characteristics. Then, we input the word vectors into Convolutional Neural Networks (CNN) and Bi-directional Long Short-Term Memory (BiLSTM) layers for local semantic features and contextual semantic extraction. Additionally, we introduce an attention mechanism to assign different weights to the features based on their importance before conducting sentiment tendency analysis. The experimental results showed that the average accuracy and Fl values of the sentiment analysis model proposed in this study for Internet public opinion regarding food safety reached 94.12% and 94.61%, respectively, which achieved better results than comparable sentiment analysis models.

Keywords: sentiment analysis; food safety public opinion; XLNet; CNN; BiLSTM; attention mechanism

1. INTRODUCTION

According to the China Livelihood Survey General Report in 2019, food safety is respondents' primary concern regarding the social environment [1]. In addition, due to the ripple effect of food safety incidents, the public opinion generated by food safety incidents frequently influences the national policy system and other quality-of-life concerns, including healthcare, education, and housing. People typically use concerns about food safety as an outlet for their emotions regarding various social issues [2]. According to psychological study [3], emotions play a critical role in behavior. Due to the rapid growth of mobile internet terminal technology, online public opinion over food safety incidents has spread rapidly and broadly. Consequently, the emotional tendency of netizens toward public opinion on food safety significantly influences the development of public opinion on the events [4]. In particular, when netizens' sentiments are negative, they are more likely to engage in irrational behaviors [5], which has a significant and detrimental impact on the government's credibility and social order [6]. If netizens' emotions about food safety are not dredged, it may not only trigger new public opinion events but may also provoke more extreme group behaviors in real life. Therefore, this study proposes a method for analyzing the sentiment of food safety public opinion on the Internet, which can assist public opinion controllers in developing appropriate management and control tactics.

At present, many scholars have achieved significant progress in sentiment analysis of online public opinion, but there are still some limitations. Traditional methods of public opinion sentiment analysis can be grouped as sentiment lexicon-based methods and machine learning methods. The former relies heavily on the construction of sentiment dictionaries. However, there are no food safety sentiment dictionaries available to the public. The latter requires manual selection of text features, but the efficiency of the sentiment analysis model has largely influenced the construction of text features and training corpus, and it is also easy to disregard the grammatical and semantic information of the text. Deep learning approaches emerging in recent years can better compensate for the shortcomings of traditional sentiment analysis approaches. Specifically, pre-training techniques can better extract text features and retain more grammatical and semantic information, hence enhancing the performance of the sentiment analysis model [7]. The majority have extracted text features using Word2Vec word vector technology [8]. However, this static text representation method can only learn a shallow representation of the text [9], where the same word has the same meaning in different contexts. Therefore, it is unable to capture some deep-level information [10], and the performance enhancement of the sentiment analysis model is limited. Particularly, netizens frequently use slang, catchphrases, and ironic language when discussing food safety events.

International Conference on Computer Graphics, Artificial Intelligence, and Data Processing (ICCAID 2022)
edited by Ruishi Liang, Jing Wang, Proc. of SPIE Vol. 12604, 1260426
© 2023 SPIE · 0277-786X · doi: 10.1117/12.2674590

Therefore, this paper proposes the XLNet-CNN-BiLSTM-Att model for Internet public opinion sentiment analysis about food safety events in an effort to address the issues outlined above. We employ XLNet, a dynamic text representation method, to finish the initialization of the downstream model, which can learn both shallow and deep textual information. In the deployment of the subsequent model, we use the CNN that can better extract the local semantic features of food safety opinion more efficiently. Nonetheless, its capability to learn text context remains inadequate. To further improve the model, the BiLSTM [11], which can successfully extract the contextual semantic features of text, is applied. In addition, we introduce the attention mechanism to improve the attention of sentiment words in the text in order to increase the accuracy of sentiment analysis.The model utilizes the advantages of CNN and BiLSTM to model and represent sentences, appropriately handling local and global aspects of texts at different scales, so considerably enhancing the accuracy of sentiment analysis for online public opinion regarding food safety. Furthermore, different from the training and testing of previous sentiment analysis models, the training set data and test set data constructed in this study are not proportionally divided within the same data set. Instead, multiple network public opinion text data of the same type of food safety events are utilized as the model's training set data, while three network public opinion text data of the same type of food safety events are chosen as the test data. The universality of the model supports its practical implementation in the management and control of public opinion events regarding food safety by government regulator.

2. SENTIMENT ANALYSIS MODEL BASED ON XLNET AND CNN-BILSTM-ATT

The XLNet-CNN-BiLSTM-Att model consists of four main components, namely the XLNet layer, CNN layer, BiLSTM layer, and Attention layer. The model architecture is shown in Figure 1. The overall process of the model is as follows. First, the XLNet layer is used to obtain word vectors containing contextual semantic information. Second, then local semantic features are extracted using CNN. Then, BiLSTM extracts contextually relevant features. Finally, the attention mechanism is introduced to assign weights to the extracted features to highlight the key information. The model utilizes a softmax classifier for sentiment classification.

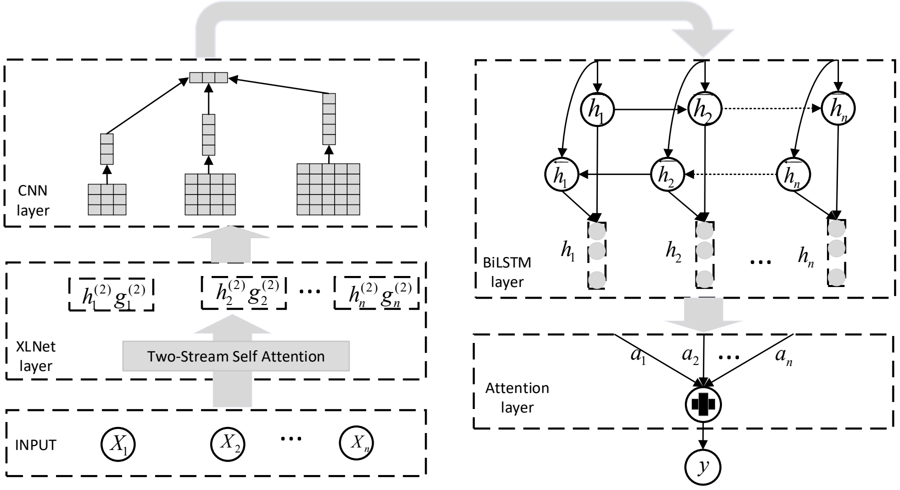


Figure 1. XLNet-CNN-BiLSTM-Att architecture

2.1 XLNet model

XLNet, a generalized autoregressive pre-training model, compensates for the shortcomings of the original autoregressive and autoencoding language models. The model reconstructs the input text in the form of permutations, putting a part of the following context into the preceding part of the text, and making full use of the contextual information to achieve a two-way prediction. Specifically, given a sequence $W=[w_1,w_2,...,w_T]$ of length T, there are $T!$ different combinations of sequences. Let Z_T be all possible combinations of permutations of the sequence W, $\boldsymbol{z}$ be one of the permutations, z_t be the t-*th* element and $\boldsymbol{z}_{<t}$ be the first $t-1$ element. The following function is used to find the maximum expected likelihood probability value of all factorial decomposition sequences.

$$\max_\theta \boldsymbol{E}_{z\sim \boldsymbol{Z}_T}\left[\sum_{t=1}^{T}\log \boldsymbol{p}_\theta\left(\boldsymbol{x}_{z_t} \mid \boldsymbol{X}_{z_{<t}}\right)\right] \tag{1}$$

$$p_\theta\left(X_{z_t}=x \mid X_{z_{<t}}\right)=\frac{\exp\left(e(x)^T g_\theta\left(X_{z_{<t}},z_t\right)\right)}{\sum_{x'}\exp\left(e\left(x'\right)^T g_\theta\left(X_{z_{<t}},z_t\right)\right)} \tag{2}$$

Since the original input text cannot be permuted during the fine-tuning phase, the order of the input text can only be changed within the Transformer during pre-training. XLNet achieves this through Attention masks. As XLNet disrupts the order of the sentences, information about the position of the token becomes important when predicting. At the same time, the content information of the token must also be covered up when predicting; otherwise, the input contains the content information to be predicted, and the model will not be able to learn knowledge. This means that XLNet needs to see the position of the token but not the content of the token. Therefore, XLNet uses a two-stream self-attentive mechanism, i.e., the content stream h_{z_t} and the query stream g_{z_t} . The computation process is shown in Figure 2.

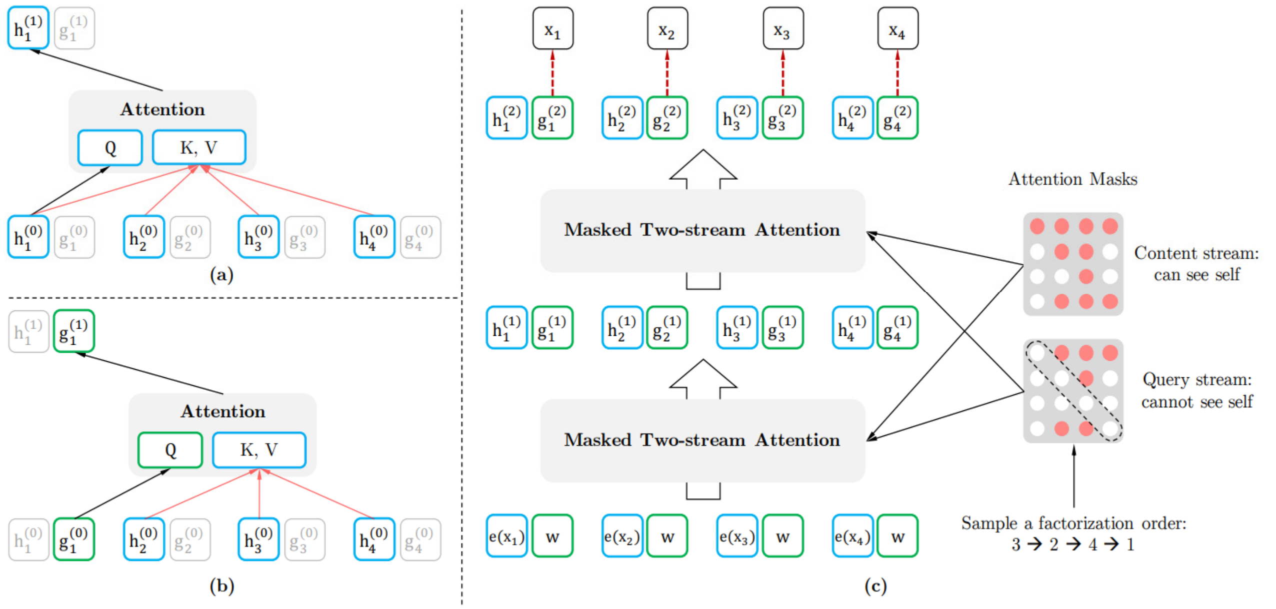

$h_1^{(1)}$ $g_1^{(1)}$
Attention
Q
K, V
$h_1^{(0)}$ $g_1^{(0)}$ $h_2^{(0)}$ $g_2^{(0)}$ $h_3^{(0)}$ $g_3^{(0)}$ $h_4^{(0)}$ $g_4^{(0)}$
(a)
$h_1^{(1)}$ $g_1^{(1)}$
Attention
Q
K, V
$h_1^{(0)}$ $g_1^{(0)}$ $h_2^{(0)}$ $g_2^{(0)}$ $h_3^{(0)}$ $g_3^{(0)}$ $h_4^{(0)}$ $g_4^{(0)}$
(b)
x_1 x_2 x_3 x_4
$h_1^{(2)}$ $g_1^{(2)}$ $h_2^{(2)}$ $g_2^{(2)}$ $h_3^{(2)}$ $g_3^{(2)}$ $h_4^{(2)}$ $g_4^{(2)}$
Masked Two-stream Attention
$h_1^{(1)}$ $g_1^{(1)}$ $h_2^{(1)}$ $g_2^{(1)}$ $h_3^{(1)}$ $g_3^{(1)}$ $h_4^{(1)}$ $g_4^{(1)}$
Masked Two-stream Attention
$e(x_1)$ w $e(x_2)$ w $e(x_3)$ w $e(x_4)$ w
Attention Masks
Content stream: can see self
Query stream: cannot see self
Sample a factorization order: 3 → 2 → 4 → 1
(c)

Figure 2. Dual-stream self-attention mechanism

In the process, (a) is content stream attention of $X_{z_t}: h_{z_t}^{(m)} = Attention\left(Q=h_{z_t}^{(m-1)}, KV=h_{z_{\le t}}^{(m-1)};\theta\right)$; (b) is the Query stream attention calculation that includes the contextual information $X_{z_{<t}}$ and the position z_t , but does not have access to content X_{z_t}: $g_{z_t}^{(m)} = Attention\left(Q=g_{z_t}^{(m-1)}, KV=h_{z_{\le t}}^{(m-1)};\theta\right)$; (c) is the overview of the permutation language modeling training with two-stream attention, where h and g are initialized to $e(x_i)$ and w, respectively. Then the Content mask and Query mask calculate the output of each layer in turn. On the far right is the mask matrix. Assuming that the arranged order is [3, 2, 4, 1], then in the Content mask matrix, the first word can use the information of all words; the second word can use the second word and the third word information. Since the Query mask cannot make use of information about itself, the diagonal lines are white dots.

XLNet draws on two of the most important technical points of Transformer-XL that is relative positional encoding and the segment recurrence mechanism, which allows the hidden layer information of the previous word to be used in the computation of the next word, thus obtaining longer distance contextual information. In other words, the XLNet pre-training model can fully learn contextual semantic information. In this layer, the input text sequence is converted into a word vector that can be recognized by the machine, and it is then used as the input of the CNN network.

2.2 CNN Networks

The output of the XLNet embedding layer is used as the input to the CNN network. The word vector of each word in the sentence is $x_i, x_i \in R^{t\times d}$. The sentence with length t is represented as $x_{1:n} = x_1 \oplus x_2 \oplus \ldots \oplus x_t$, where t is the number of words, d is the vector dimension, and $\oplus$ is the concatenation symbol. The purpose of the convolutional layer is to extract the semantic features of the sentence, by setting a filter w to complete the extraction of the features of j-*th* word of the input text sentence: $z_j = f(w \times x_{i:i+h-1} + b)$, where f is the activation function relu, h is the size of the convolutional kernel, $x_{i:i+h-1}$ is the sentence vector of words i to $i+h-1$, b is the bias term. The feature matrix $Z=[z_1, z_2, ..., z_{t-h+1}]$ is obtained after the convolutional layer. The pooling layer aims to obtain the optimal solution of the local value by downsampling the local feature matrix Z of the sentence obtained from the convolution layer using the MaxPooling technique. That is, it is used to acquire the most important features $\max\{Z\} = max(z_1, z_2, ..., z_{t-h+1})$. In this paper, three filters with different sizes are set up for feature acquisition, and the acquired feature matrixes are fully concatenated to form a serialized feature vector, which is used as the input matrix of the BiLSTM model. The CNN network architecture is shown below.

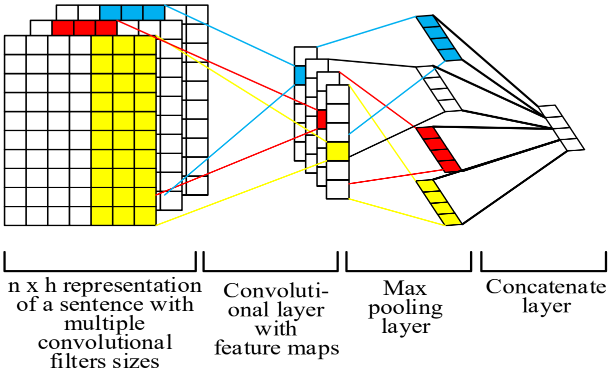

n x h representation of a sentence with multiple convolutional filters sizes
Convoluti-onal layer with feature maps
Max pooling layer
Concatenate layer

Figure 3. CNN network architecture

2.3 BiLSTM networks

The LSTM is a special recurrent neural network model consisting of three gates: forget gate, input gate, and output gate. In the LSTM network, the input first goes through the forget gate, discarding some information in the previous cell state; then, the input gate determines how much information is added to the cell state; and finally, the output gate determines the output information from the cell state. The LSTM architecture is shown below.

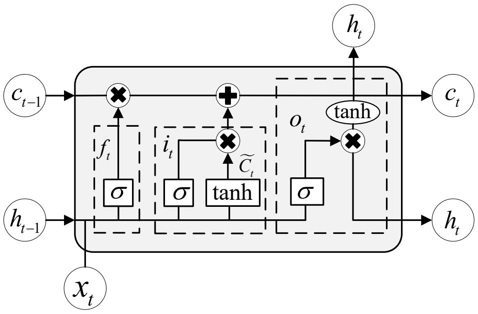


Figure 4. LSTM architecture

The calculation formula of each gate is as follows.

The forget gate determines what information is retained from the cell state of the last moment to the current moment C_t:

$$f_t = \sigma\left(W_f \cdot [h_{t-1}, x_t] + b_f\right) \tag{3}$$

The input gate determines how much information of the input x_t at the current moment is fed into the cell state C_t:

$$i_t = \sigma\left(W_i \cdot [h_{t-1}, x_t] + b_i\right) \tag{4}$$

$$\widetilde{C}_t = \tanh\left(W_C \cdot [h_{t-1}, x_t] + b_C\right) \tag{5}$$

$$C_t = f_t * C_{t-1} + i_t * \tilde{C}_t \tag{6}$$

The output gate controls the outputted information h_t of cell state C_t:

$$o_t = \sigma\left(W_o \cdot [h_{t-1}, x_t] + b_o\right) \tag{7}$$

$$h_t = o_t * \tanh\left(C_t\right) \tag{8}$$

The BiLSTM contains a forward LSTM and a backward LSTM, which can learn the contextual information of each word in the sequence, respectively. The network architecture is shown in Figure 5. The BiLSTM layer receives the CNN layer output feature vector as input. The forward LSTM processes the input vector in order (from X_1 to X_t) and calculates the forward hidden state sequence $(\overrightarrow{h_1}, \overrightarrow{h_2}, ..., \overrightarrow{h_t})$. The backward LSTM processes the sequence in reverse order (from X_t to X_1) to obtain a backward hidden state sequence $(\overleftarrow{h_1}, \overleftarrow{h_2}, ..., \overleftarrow{h_t})$. Then, the forward hidden state and the backward hidden state are concatenated to obtain the complete sequence. For example, the state of *i-th* word is $h_i = [\overrightarrow{h_i}; \overleftarrow{h_i}]$. Using this approach, bi-directional information is obtained for each word. In this way, the feature of the input text vector can be further extracted, and the output of the BiLSTM layer can be obtained.

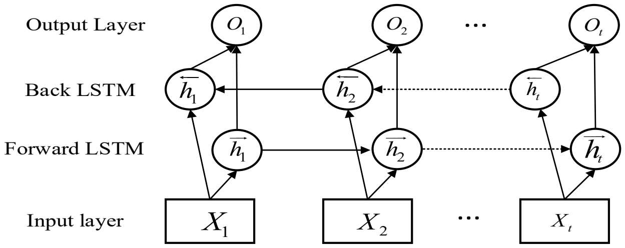


Figure 5. BiLSTM architecture

2.4 Attention mechanism

The attention mechanism simulates the characteristics of the human brain's attention, with the core idea that more attention is allocated to what is important and less to other parts. The mechanism was first proposed in the field of computer vision. However, after Bahdanau first integrated it into natural language processing, the attention model began to be widely used in various tasks. In a sentence, each word has a different impact on its final sentiment tendency. In order to enlarge the influence of the key parts on the final sentiment, the key parts need to be found and highlighted. The formula of the attention mechanism is as follows.

$$u_t = \text{sigmoid}\left(W_C h_t + b_w\right), a_t = \text{soft}\max\left(u_t\right), v = \sum a_t H_t \tag{9}$$

u_n is the hidden cell of the BiLSTM output h_n, which is randomly initialized and continuously learned during training. a_n is the attention vector, and v is the final output vector after the attention mechanism.

An attention mechanism is introduced into the BiLSTM network to extract the parts of the text that are relatively important for sentiment classification by increasing their weight in the final generated text features.

3. EXPERIMENT AND ANALYSIS

3.1 Datasets

In June 2020, four departments, including the State Administration for Market Regulation and the Ministry of Education of the People's Republic of China, jointly issued the "Campus Food Safety Protection Action Plan (2020-2022)", which clearly proposes to strictly prevent and control campus food safety risks and prevent the occurrence of major food safety accidents. Based on the background, this paper uses Python and XPath technology to customize a web crawler to crawl a totally of 86670 related blog posts from several food safety events, including “Mouldy and rotten meat found in a primary school canteen of Chengdu”, “Dead rats found in school meals in Xinzhou”, “Food poisoned in the school canteen in Bazhou”, “Food in the canteen of Weixin Middle School deteriorated” and “Handan school students have diarrhea”. After preprocessing such as deleting duplicates and objective invalid remarks, 41,420 annotated Weibo data with balanced

580

positive and negative sentiments are finally obtained as training data. At the same time, we crawled three other events, including “Mass vomiting of students in Henan”, “The kitchen of one Tianjin school is untidy”, and “Students vomited after dinner in Anhui”, that broke out on Weibo and had a wide range of effects, as test data. To ensure the validity of the data, the text content was first classified by the Baidu Sentiment Analysis API. Second, we manually annotated the text content and classified it as positive or negative. Finally, a more accurate positive and negative corpus is obtained by comparing and correcting the annotations obtained by the manual and the Baidu API. Table 1 shows the descriptive statistic of test datasets.

Table 1. Test event data

event name	event name abbreviation	positive	negative	total
Mass vomiting of students in Henan	Henan incident	2203	2152	4355
The kitchen of one Tianjin school is untidy	Tianjin Incident	2029	1999	4028
Students vomited after dinner in Anhui	Anhui Incident	2124	1929	4053

3.2 Experimental Setup

This study uses the jieba package in python to segment the training data and download the base version of the Chinese XLNet pre-training model released by the Harbin Institute of Technology Xunfei Joint Laboratory. The XLNet model that had been pre-trained on the general-purpose dataset was further pre-trained on our training dataset. The parameters of the XLNet model were fine-tuned using CNN, BiLSTM, and attention mechanisms, which is equivalent to transferring the model from the general-purpose domain to food safety sentiment analysis. After obtaining the word vectors generated by XLNet pre-training model, the microblog sentiment analysis model is trained in the environment with the deep learning framework Tensorflow 2.0.0 and its wrapper interface Keras 2.3.1. A grid search method is used to adjust the main parameters and obtain the ideal set of parameters. The values of the model parameters are shown in Table 2.

Table 2. Model parameters

parameter name	**parameter value**
Eembedding size	768
Convolution kernels num	128、128、128
convolution kernels length	3、4、5
Batch size	32
Sentence length	70
Learning rate	0.00005
Lstm units	128
Dropout rate	0.5
Validation split	0.3

3.3 Evaluation Criteria

The common metrics used to evaluate the performance of sentiment analysis models are Precision (P), Recall (R), F1-score(F1), Accuracy (Acc) and Loss rate (L).

$$P = \frac{T_P}{T_P + F_P} \times 100\% \text{，} \quad R = \frac{T_P}{T_P + F_N} \times 100\% \text{，} \quad F_1 = \frac{2 \times P \times R}{P \times R} \times 100\% \text{，} \quad Acc = \frac{T_P + T_N}{T_P + F_N + F_P + F_N} \times 100\% \tag{10}$$

$$L = -\left[\log \hat{y} + (1 - y)\log(1 - \hat{y})\right] \tag{11}$$

Where T_P is the number of positive samples identified as positive, T_N is the number of negative samples identified as negative, F_P is the number of negative samples identified as positive, and F_N is the number of positive samples identified as negative. Moreover, the y refers to the actual value and $\hat{y}$ is the predicted value in loss function. P refers to the proportion of correct positive predictions to all positive predictions. R is the proportion of correct predictions to be positive to all actual positives. Fl is a summed average of P and R. Acc is the proportion of all samples that were correctly predicted. The higher the F1 and Acc values, the better the model sentiment analysis. On the contrary, the lower the loss rate, the better the model.

3.4 Experimental Results

In this experiment, five models were chosen as experimental control models: Word2Vec-BiLSTM, BERT, XLNet, XLNet-CNN, and XLNet-BiLSTM. The sentiment analysis experiments were conducted on three test events, including Henan event, Tianjin event, and Anhui event.

Figure 6 depicts the dynamic changes in the accuracy and loss rate of the model training over the 10 epochs. According to the accuracy and loss rates of the validation set data during the training phase, we may conclude that the model proposed in this paper outperforms the other five models.

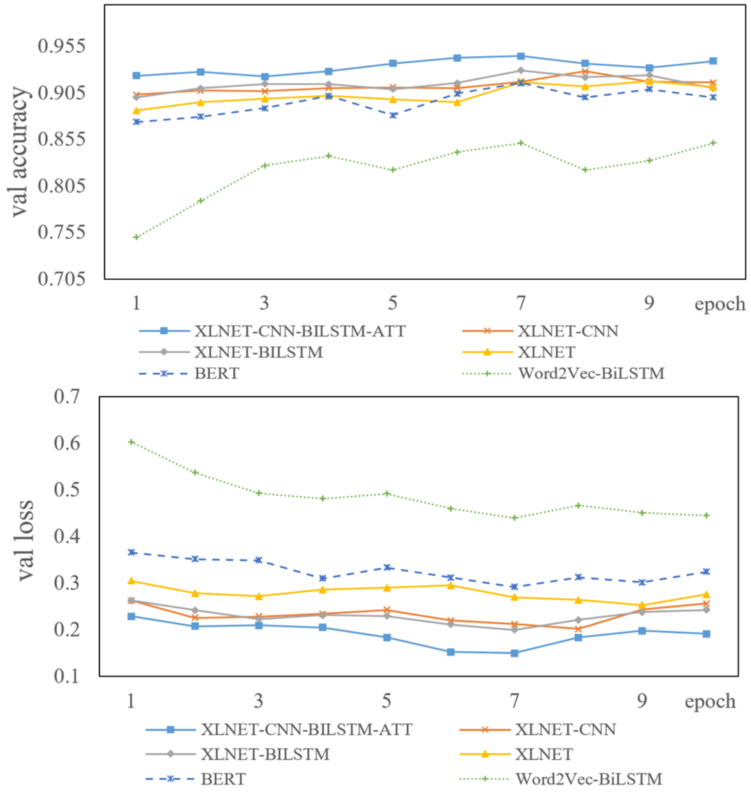


Figure 6. Variation of accuracy(front) and loss(behind) of the model during training

In this study, the optimal parameter values of XLNet-CNN-BiLSTM-Att were utilized to classify Internet users' sentiments, and the results are presented in Table 3. The average accuracy of the three test events is 94.12%, with the Anhui event having the highest accuracy at 94.75%. This demonstrates that the XLNet-CNN-BiLSTM-Att model proposed in this research is capable of achieving outstanding performance in Internet public opinion sentiment analysis. Besides, the model outperformed the other five models in terms of accuracy in all three events tested. This implies that the dynamic text representation approach is superior to the static text representation method Word2Vec for classification tasks.

Table 3. Models test result

Model	Evaluation indicators	Henan incident	Tianjin Incident	Anhui Incident
XLNet-CNN-BiLSTM-Att	Acc	93.87	93.76	94.75
	P	95.09	93.98	95.15
	R	94.86	94.32	94.76
	F1	94.97	94.15	94.73
XLNet-CNN	Acc	91.96	92.04	92.12
	P	92.41	92.33	92.21
	R	92.91	91.97	93.83
	F1	92.66	92.15	93.17
XLNet-BiLSTM	Acc	92.26	91.33	92.18
	P	93.91	92.79	91.76
	R	92.23	91.36	92.9
	F1	93.06	92.07	92.33
XLNet	Acc	90.97	90.63	91.13
	P	91.13	91.38	91.83
	R	90.79	90.79	90.41
	F1	90.96	91.08	91.11
BERT	Acc	89.12	89.23	90.47
	P	89.71	88.51	91.32
	R	89.96	89.67	90.59
	F1	89.83	89.09	90.95
Word2Vec-BiLSTM	Acc	82.46	81.96	82.1
	P	82.39	82.18	84.23
	R	83.66	81.93	83.14
	F1	83.02	82.05	83.68

4. CONCLUSION

The research offers an XLNet-CNN-BiLSTM-Att model for analyzing Internet public opinion sentiment regarding food safety. The XLNet approach is used to optimize the text representation of public opinion on food safety. In addition, we integrate CNN and BiLSTM to capture local features and contextual information of the text. Lastly, the attention mechanism is used to assign weights to several essential features. The model was compared with several different sentiment classification models, and the accuracy and loss rate of the validation set of the model were superior to those of other models during the model training process. The trained model was evaluated on three online opinion data on food safety events, and the results showed that, in comparison to other models, the proposed model could obtain a better Fl value and accuracy.

References

[1] Zhang, K.; Ge, Y.; Jin, S.; Zhu, X. China Livelihood Satisfaction Remains at a High Level:China Livelihood Survey 2019 General Report. J. Manag. World 2019, 35, 1-10.

[2] Kasperson, R.E.; Renn, O.; Slovic, P.; Brown, H.S.; Emel, J.; Goble, R.; Kasperson, J.X.; Ratick, S. The social amplification of risk: A conceptual framework. Risk Anal. 1988, 8, 177-187.

[3] Sun, S.-x.; Guan, Y.-h.; Qin, Y.; Zhang, L.; Fan, F. Social support and emotional-behavioral problems: Resilience as a mediator and moderator. Chin. J. Clin. Psychol. 2013, 21, 114-118.

[4] Choi, Y.; Lin, Y.-H. Consumer responses to Mattel product recalls posted on online bulletin boards: Exploring two types of emotion. J. Public Relat. Res. 2009, 21, 198-207.

[5] Wang, Q.; Bai, X.-J.; Guo, L.-J.; Shen, D.-L. The effect of suppressing negative emotion on economic decision-making. Acta Psychol. Sin. 2012.

[6] Wu, P.; Qiang, S.; Gao, Q. Modelling internet users' negative emotion based on soar model. Chin. J. Manag. Sci. 2018, 26, 126-138.

[7] Kiros, R.; Zhu, Y.; Salakhutdinov, R.R.; Zemel, R.; Urtasun, R.; Torralba, A.; Fidler, S. Skip-thought vectors. In Proceedings of the 28th International Conference on Neural Information Processing Systems, Montreal, Canada, 7–12 December 2015; pp. 3276–3284.

[8] Mikolov, T.; Chen, K.; Corrado, G.; Dean, J. Efficient estimation of word representations in vector space. arXiv 2013 arXiv:1301.3781.

[9] Deng, L.; Yu, D. Deep learning: methods and applications. Found. Trends Signal Process. 2014, 7, 197-387.

[10] Yang, X.; Macdonald, C.; Ounis, I. Using word embeddings in twitter election classification. Inform. Retrieval J. 2018, 21, 183-207.

[11] Mnih, A.; Hinton, G.E. A scalable hierarchical distributed language model. In Proceedings of the 21st International Conference on Neural Information Processing Systems, British Columbia, Canada, 8 - 10 December 2008; pp. 1081–1088.

Effect of different armature sizes on the performance of electromagnetic rail guns

Jingpu Bian *, Tao Shu, Taidou Qin, Shuai Zhang, Ruicong Zhu, Meng Qian
Air and Missile Defense College, Air Force Engineering University, Xi'an shan xi 710051, China
*Corresponding author's e-mail: 1844812330@qq.com

Abstract

In order to further investigate the influence of different structural parameters of the armature on the performance of the electromagnetic rail gun, the Ansys Maxwell simulation software is used to simulate and obtain the results of different diversion arc angle, width and centre circle radius of the armature. In order to further investigate the influence of different structural parameters of the armature on the performance of the electromagnetic rail gun, the Ansys Maxwell simulation software was used to obtain the magnetic induction strength and maximum electromagnetic thrust values for different diversion arc angles, widths and radii of the centre circle. The simulation results show that for the design of the rail gun structure, the armature width should be as small as possible within a reasonable range, the diversion arc angle should be large enough and the centre circle radius should be at a suitable position, which can effectively improve the electromagnetic thrust performance of the rail gun and obtain a larger electromagnetic shielding range.

Keywords: Electromagnetic railgun; Armature dimensions; Electromagnetic shielding; Electromagnetic thrust values

1. INTRODUCTION

The electromagnetic rail gun is generally composed of a pair of parallel rails and an armature that maintains sliding electrical contact with the rails, as well as reinforcement devices. The electromagnetic rail gun can make the speed range from 95m/s to 8km/s, while the traditional gunpowder launch technology can only meet the maximum speed of 1km/s[1]. Secondly, the launch efficiency is also much higher than the traditional gunpowder launch technology, with a theoretical efficiency of 50%, while the theoretical efficiency of traditional launch technology is only 4% to 6%. As the loading carrier, the armature plays a crucial role in the electromagnetic rail gun, therefore, the in-depth study of the armature size is a prerequisite for the performance improvement of the electromagnetic rail gun[2]. In this paper, the effects of armature diversion arc angle, armature width and center circle radius on the maximum thrust value and magnetic induction strength of electromagnetic rail gun are studied.

2. Current status and prospects of electromagnetic rail gun research

Electromagnetic rail guns are one of the main effective means of striking conventional weapons in the future battlefield environment, as they can increase range and shorten time to target compared to traditional weapons[3]. The electromagnetic launch technology meets the needs of the battlefield, with high kinetic energy and variable launch loads, short launch intervals, high sustained operational capability, high reliability, few operators and maintenance personnel, and low maintenance requirements. On the basis of the large-calibre long-range electromagnetic rail gun launch technology, consideration should be given to optimising the small calibre, high rate of fire and high precision to achieve enhanced tactical capabilities.Its rapid response characteristics are closer to the actual combat environment, modern highly informative and highly intelligent battlefield, and has broad military application prospects in medium- and close-range air defence and anti-missile, and distant precision strike[4]. The main application prospects are: counter-immediate space target threats, medium- and close-range air defence and anti-missile operations, and long-range precision strikes[5].

3. Introduction to the armature rail gun model

The electromagnetic railgun launcher model is shown in Figure 1. The angle of the drainage arc is A, the width of the armature is B, and the radius of the center circle is C. In this paper, the armature and the guide rail are configured in the same way. Use Ansys Maxwell simulation software to analyze the magnetic induction intensity on a specific path, and the path taken is shown in Figure 2. By changing the angle of the drainage arc to A, the width of the armature to B, and the radius of the center circle to C, three sets of independent comparative experiments were conducted to explore the impact

International Conference on Computer Graphics, Artificial Intelligence, and Data Processing (ICCAID 2022)
edited by Ruishi Liang, Jing Wang, Proc. of SPIE Vol. 12604, 1260427

on the performance of the electromagnetic railgun.

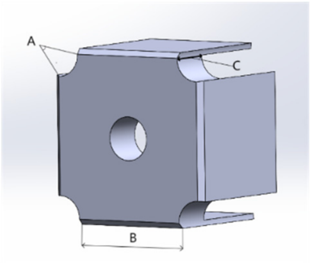
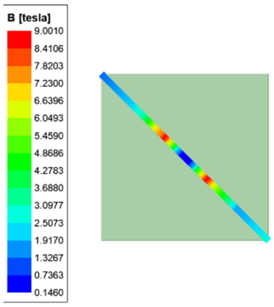
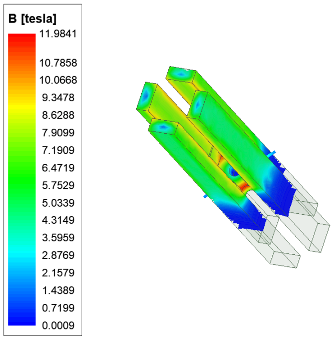


Figure 1 Armature model

To calculate the magnitude of the magnetic field force on the armature, the Biot-Savard law can be used. The Biot-Savard law can be described as the magnitude of the magnetic induction intensity dB generated by the current element Idl at a point P in space is proportional to the magnitude of the current element Idl[6], proportional to the sine of the position vector from where the current element Idl is to the point P and the angle between the current element Idl, and inversely proportional to the square of the distance from the current element Idl to the point P[7].





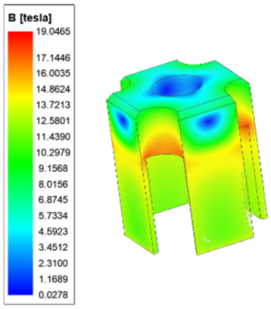


Figure 2 Magnetic field path Figure 3 Simulation of electromagnetic rail gun Figure 4Armature simulation diagram

4. Simulation analysis

4.1 Simulation analysis of different diversion arc angles

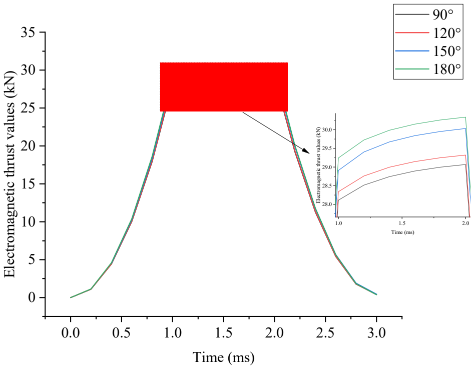


Figure 5 The variation of electromagnetic thrust

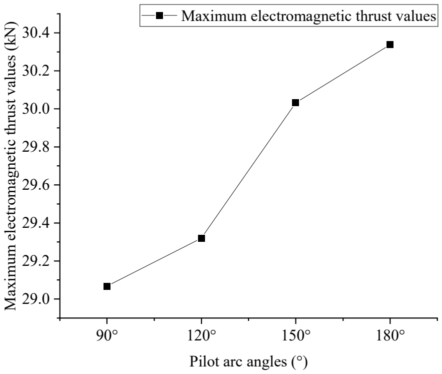


Figure 6 The variation of the maximum electromagnetic thrust

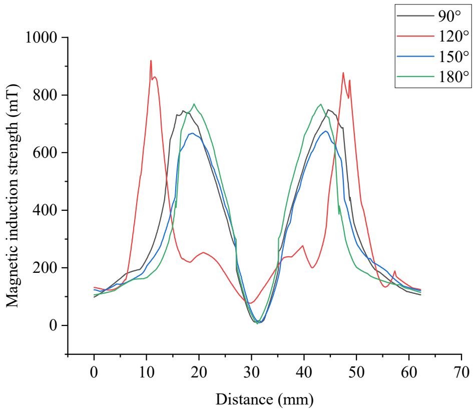


Figure 7 The variation of magnetic induction strength

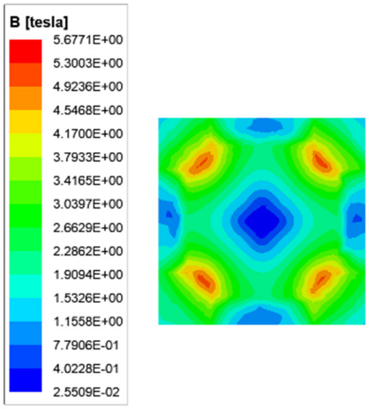


Figure 8 Simulation of magnetic induction at the front face of the armature

The electromagnetic thrust curve for armatures with different diversion arc angles is calculated over time using Ansys Maxwell simulation software, as shown in Figure 5. Figure 6 shows the maximum values of the electromagnetic thrust corresponding to different diversion arc angles.

(1) As shown in Figure 5, the electromagnetic thrust as a whole is a rapid rise, then a steady rise and then a decline in the trend, 90º and 120º electromagnetic thrust maximum value is approximately equal, 150º and 180º electromagnetic thrust maximum value is approximately equal, indicating that the influence of the diversion arc angle on the electromagnetic thrust is relatively small, at the same time through Figure 6 can visually show that the maximum value of electromagnetic thrust with the diversion arc angle increases and increases.

(2) It can be seen from Figure 8 that the magnetic field at the front face of the armature shows a regular distribution, with its magnetic induction intensity concentrated in the pilot arc part, from the inside of the track to the middle of the armature, the magnetic fields generated by the currents at both ends of the armature track cancel each other out in the armature firing region, and the magnetic induction intensity drops to zero; it can be seen from Figure 7 that for the selected path, there are certain regions where the magnetic induction intensity cancels each other out, where when the pilot arc The electromagnetic shielding area is greatest at an angle of 120º.

4.2 Simulation analysis of different widths

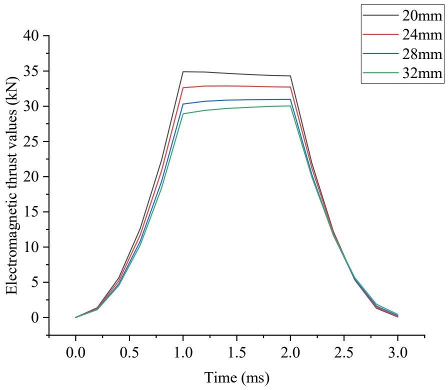


Figure 9 The variation of electromagnetic thrust

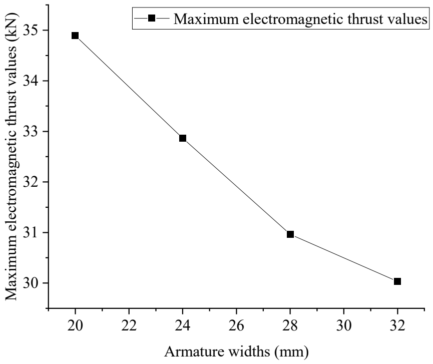
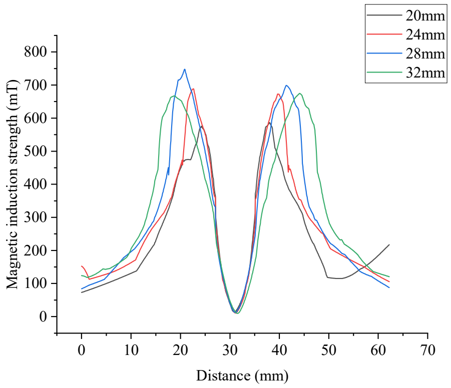


Figure 10 The variation of the maximum electromagnetic thrust



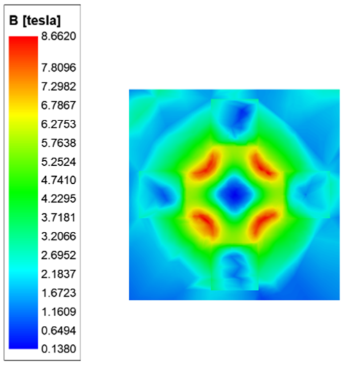


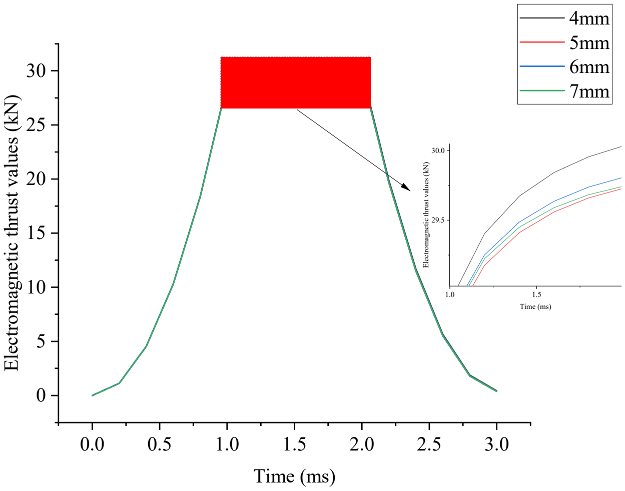
Figure 11 The variation of magnetic induction strength Figure 12 Simulation of magnetic induction at the front face of the armature

The electromagnetic thrust curves for armatures with different widths were calculated with the Ansys Maxwell simulation software as a function of time, as shown in Figure 9. Figure 10 shows the maximum values of electromagnetic thrust for different armature widths.

(1) As shown in Figure 9, there is a significant difference in the maximum values of electromagnetic thrust corresponding to different armature widths, indicating that the electromagnetic railgun thrust performance can be effectively improved by varying the armature width; it is evident from Figure 10 that the value of electromagnetic thrust decreases with the increase in armature width under constant excitation conditions.

(2) From Figure 9 and Figure 10, the effect of armature width on the electromagnetic shielding area is not too significant.

4.3 Simulation analysis for different centre circle radii



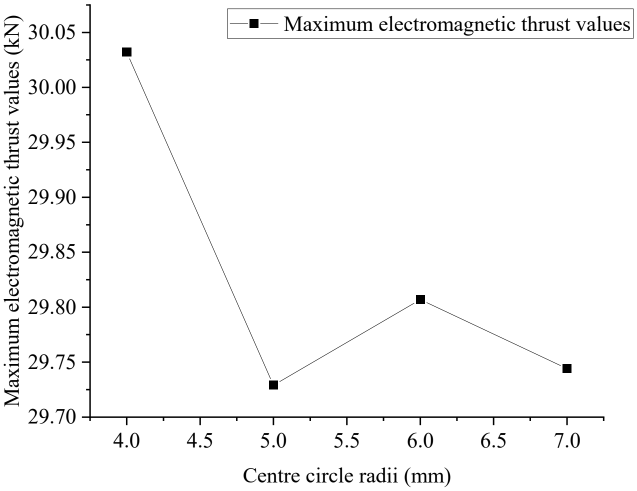
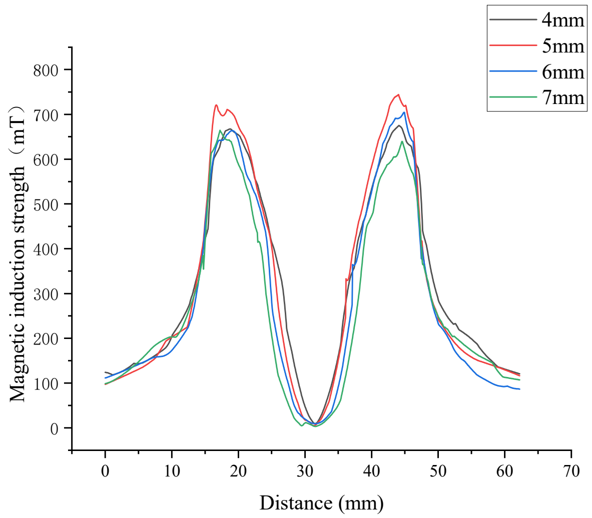


Figure 13 The variation of electromagnetic thrust Figure 14 The variation of the maximum electromagnetic thrust



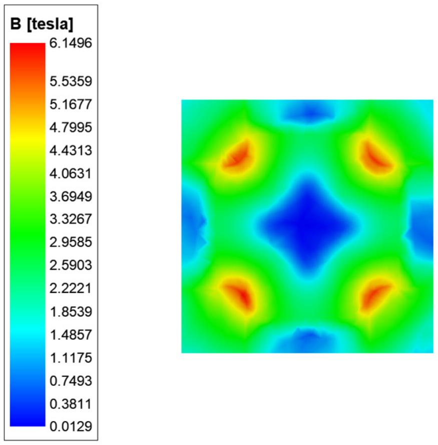


Figure 15 The variation of magnetic induction strength   Figure 16 Simulation of magnetic induction at the front face of the armature

The electromagnetic thrust curves for armatures with different centre circle radii were calculated with the Ansys Maxwell simulation software as a function of time, as shown in Figure 13. Figure 14 shows the maximum values of electromagnetic thrust corresponding to different centre circle radii[8]

(1) As shown in Figure 13, there is no significant difference in the maximum values of electromagnetic thrust corresponding to different centre circle radii, indicating that the centre circle radius has no significant effect on the thrust performance of the electromagnetic rail gun, while it is shown by Figure 14 that there is no significant relationship between the maximum values of electromagnetic thrust and centre circle radius[9].

(2) As can be seen from Figure 13, the electromagnetic shielding area on the armature increases as the radius of the centre circle increases. An increase in the radius of the centre circle provides a better electromagnetic shielding area, and a larger shielding area better meets the demand for magnetic fields for the launch[10].

5.CONCLUSION

This paper analyses and investigates the effect of armature structures with different pilot arc angles, armature widths and centre circle radii on the performance of electromagnetic rail guns and draws the following conclusions.

(1) The three parameters have different degrees of influence on the thrust performance of the electromagnetic railgun, with the armature width having the greatest influence, the diversion arc angle the second and the centre circle radius the least.

(2) The magnetic fields generated by the currents in the orbits cancel each other on the axis of the entire launch area, but the magnetic fields at the two adjacent orbits are strengthened and this characteristic is continued in the armature area as well. Due to the structural characteristics of the armature, which has its maximum value in the middle of the chosen path, the current density at this strong magnetic field also happens to be at its maximum value, so that the strong current in the armature interacts with the strong magnetic field to generate more electromagnetic thrust and the propulsion performance of the launcher is therefore mainly influenced by the electromagnetic characteristics in the main current path of the armature.

REFERENCES

[1] CHEN Qingrong, SHU Tao, DING Rixian. Structural optimization design of a new quadrupole orbital electromagnetic transmitter[J]. Journal of Weaponry Equipment Engineering,2019,40(10):30-35.

[2] Liu M, Shu T, Xue XP. New quadrupole orbital electromagnetic launcher[J]. Firepower and command and control,2019,44(03):23-27.

[3] CHEN Qingrong, SHU Tao, DING Rixian, XUE Xinpeng, LIU Ming. Comparative analysis of electromagnetic rail gun structure improvement and performance simulation[J]. Journal of Ballistic Arrows and Guidance, 2019,39(04):9-14+18. DOI: 10.15892/j.cnki.djzdxb.2019.04.003.

[4] Li Tengda, Feng Gang, Liu Shaowei, Shi Jianming, Fan Chengli. Simulation analysis of electromagnetic characteristics of copper-based composite quadrupole rail [J]. Systems Engineering and Electronics Technology, 2021, 43(11):3054-3063.

[5] Ren Shida, Feng Gang, Liu Shaowei, Li Tengda, Wei Dongbin. Analysis of armature structure of four-track electromagnetic launcher based on contact characteristics[J]. Journal of Ballistics,2022,34(02):59-64+92.

[6] Li Tengda, Feng Gang, Liu Shaowei, Shi Jianming, Fan Chengli. Analysis of the initial armature-rail contact characteristics of a four-track electromagnetic transmitter [J]. Weapon Materials Science and Engineering, 2022,45(02): 36-42. DOI: 10.14024/j.cnki.1004-244x.20211026.002.

[7] Ma Weiming, Lu Junyong. Electromagnetic emission technology[J]. Journal of the National University of Defense Technology, 2016,38(06):1-5.

[8] Li Jun, Yan Ping, Yuan Weiqun. Development and status of electromagnetic rail gun firing technology[J]. High Voltage Technology,2014,40(04): 1052-1064. DOI: 10.13336/j.1003-6520.hve.2014.04.014.

[9] Lv Q Ao, Lei B, Li Ziyuan, Chi S. Review of military applications of electromagnetic rail guns[J]. Journal of Artillery Firing and Control,2009(01):92-96. doi: 10.19323/j.issn.1673-6524.2009.01.024.

[10] Gugang, Xiang Yang, Zhang Jiange. Current status of international research on electromagnetic emission technology[J]. Ship Science and Technology,2007(S1):156-158.

Research on Artificial Intelligence Product Design Method Based on Product Semantics

Zonghua Zhu
Jingchu University of Technology

Abstract

With the development of Internet of Things, cloud computing, big data and other information technologies, the design objects and design methods have changed greatly. In order to meet the development direction of the times, this paper studies and analyzes the product design method of artificial intelligence based on product semantics. Firstly, this paper analyzes the design method and theoretical basis of product semantics, and then makes an in-depth research and analysis on the formation of artificial intelligence technology, including detailed discussion and planning of behaviorism, connectionism and symbolism of AI. Finally, it expounds the method of using artificial intelligence technology in product semantics design. By making full use of network technology, big data and artificial intelligence technology, a user model which is more in line with the user's thinking is established, and the application and development of artificial intelligence in product semantics are promoted.

Keywords: product semantics, artificial intelligence, product design, method research

1. Introduction

With the emergence of artificial intelligence, the design object has changed, and the design object begins to develop in the direction of "tangible" functional interaction and "intangible" information communication media, which is no longer limited to the design of physical hardware. In the era of artificial intelligence, how to fully apply artificial intelligence technology to product design on the premise of product semantics is a problem that everyone in this field and industry will face. This paper discusses the above problems and gives the corresponding solutions at the end of the article. Through understanding and learning the framework of artificial intelligence, and using big data technology to collect users' information, a relatively perfect user model is finally constructed to promote the use of artificial intelligence in product design. Research on product semantic-based design method of artificial intelligence products is not only an important medium to evaluate the design results of artificial intelligence products, but also a methodology combining design practice with design theory.

2. The meaning of product semantics and the concept of product design

Like product language, it has symbolic characteristics and is used to communicate and convey meaning. In fact, product semantic design is to examine and think about design by using linguistic signs. The semantic design of products is based on the basic theory of design symbols, and the concepts of meaning, communication, context and rhetoric are expounded. Several cases are analyzed, and the design practice process is displayed and explored. Through the correlation of symbols and rhetorical methods in linguistics, divergent thinking and association are carried out, so as to find the best way to convey the predetermined concepts and meanings of products. The concept of "product semantics" was first put forward in an article written by American Association of Industrial Designers in 1984, which distinguished the relationship between things and things, people and things, and environment and things. Product design is to solve the relationship between people and things in the product system China, such as product structure, material, operation comfort and other issues. With the continuous development of network technology, big data and artificial intelligence in today's society, the field of product design has been affected to some extent. Object-oriented product design based on artificial intelligence is no longer only limited to physical components, but also has great changes in design objects and design techniques. More and more product designs begin to incorporate the concept of system interaction. With the continuous development of science and technology in China, the design methods of products are constantly evolving. As the core of artificial intelligence technology, deep learning plays an important role in the research of artificial intelligence product design methods.

International Conference on Computer Graphics, Artificial Intelligence, and Data Processing (ICCAID 2022)
edited by Ruishi Liang, Jing Wang, Proc. of SPIE Vol. 12604, 1260428

3. The design method of product semantics

3.1 Mapping from meaning to behavior

Meaning can absorb, analyze and interpret the logical behavior and sensory perception of machines and people. It is a structured space, but it is not suitable for direct product design. As we all know, the communication between people, between machines and between machines is impossible to form a sharing mechanism, and it is impossible to share with others, so the meaning also belongs to a personalized framework. To some extent, meaning is objective, because it comes into being because of the communication between people. Different people have great differences in their personality characteristics, and different people have different experiences and relationships between things, so the formed meaning is diversified and unrestricted. The direction of a person's work is often determined according to the meaning. The mapping process from meaning to behavior can not only provide product designers with information of different users, but also make the designed products meet the needs of users, thus forming a logical causal relationship. In order to map meaning to behavior smoothly, it takes two big steps. First, we need to imitate people's actions and behaviors, then use meaning to construct the interaction between people and things into a whole, and finally complete the mapping from meaning to behavior.

3.2 Establish a structure from perception to meaning

Perception is an active and purposeful search activity, not a aimless glance. Learning is a constructive process. Merleau-Ponty turned the perceptual subject from consciousness to body, and the perceptual object was led from physical object, personal feeling and sensory material to intentional activity of human beings, thus leading perception into a new direction. Perception can directly capture and acquire external things, which is a working process without psychological processing. For example, people can directly smell different smells with their noses, see objects directly with their eyes, and taste them directly with their mouths. Manifestation is the core concept of perception theory, which can connect people's perceptions and clearly show the relationship between people and objects. However, the interaction between people and objects can't be directly driven by perception, so we need to use meaning to build a framework between perception and meaning, so that the interaction can be driven.

4. Discussion and overview of artificial intelligence technology

4.1 Overview and development of artificial intelligence technology

Artificial intelligence (AI) is a new science and technology that researches and develops theories, methods, technologies and application systems used to simulate, extend and expand human intelligence. At the same time, it is also a branch of computer science and technology. The research in this field includes robotics, language recognition, image recognition, natural language processing and expert systems. The development of AI has gone through three levels: perceptual intelligence, cognitive intelligence and decision-making intelligence, and it is widely used in many fields, as shown in the figure 1. Since the birth of artificial intelligence, its theory and technology have gradually matured and improved, and its application fields are constantly expanding. Therefore, artificial intelligence is no longer just a small branch of computer science and technology, but it has begun to involve mathematics, psychology, philosophy, neurology and other fields. Researchers have divided the paradigms of artificial intelligence into three categories according to different research directions, and through studying the three paradigms, they can fully master the artificial intelligence technology.

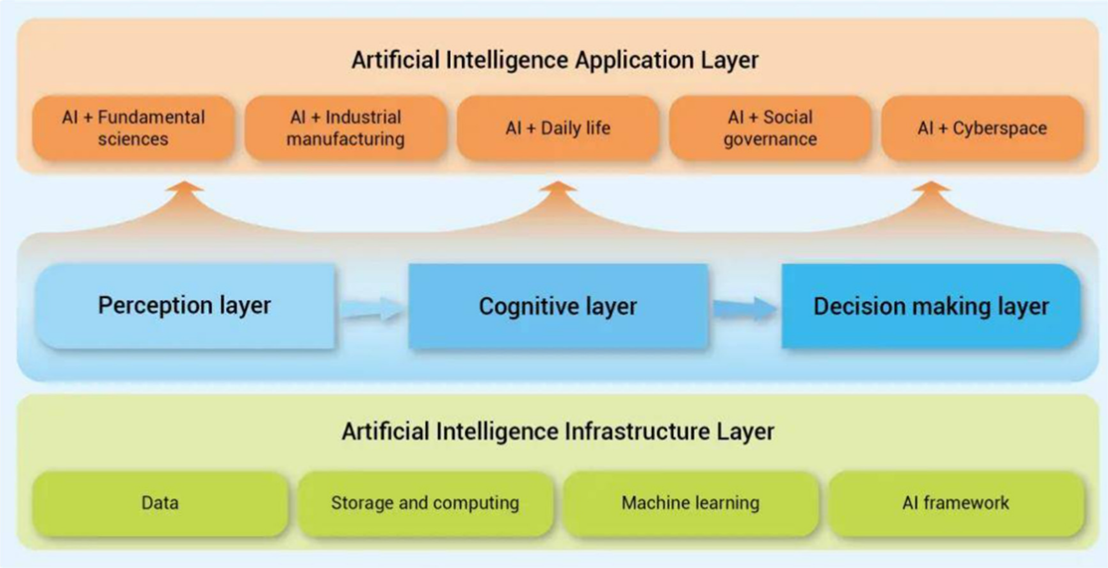


Figure 1 Three levels of AI development

4.2 Paradigm research of connectionism

Connectionism is inspired by brain science. It attributes human intelligence to the high-level activities of the human brain, and emphasizes that the generation of intelligence is the result of a large number of simple units through complex interconnection and parallel operation. Its representative achievement is the brain model created by physiologist Mccullough and mathematical logician Pitts in 1943, namely MP model. As shown in Figure 2, it creates a new way to imitate the structure and function of human brain with electronic devices. It starts with neurons and then studies neural network model and brain model, which opens up another development path of artificial intelligence. Connectionism holds that "neuron" is the basic structure of artificial intelligence and plays an important role in the development of artificial intelligence. Using neural network toolbox to design a network is like building blocks, which can greatly simplify the task of model building. The neural network contains many nodes, parameters and layers, but there are not many core parts and components. Of course, these core components are not independent, and they are closely related to each other and other components of the neural network. After linking the above components and layers, we can build a neural network instance by using nn toolbox in Python. In 2006, Jeff of the University of Toronto in Canada put forward "deep learning" based on the existing "neural network". Deep learning is the core technology of artificial intelligence, which can optimize the data results of neural network layer and at the same time accelerate the learning speed. At present, deep learning technology has been applied to intelligent products such as iris recognition, face recognition, and sound and image processing. Especially in the application of face recognition, it distinguishes individual organisms by their own biological characteristics. Biometrics research includes face, fingerprint, palmprint, iris, retina, voice, etc. Target detection is the most important technology in face recognition image monitoring. OverFeat is the first method to detect targets using deep learning and has made great progress. It was proposed by new york University in 2013, and a convolution neural network was proposed. In addition, a general processing function-—softmax loss function is applied in the face recognition technology, which is the initial face recognition function, and its function is shown in formula 1.

$$\mathrm{Softmax}(y)*k=p*k=e*\cos\theta*k*\sum e\cos\theta j \qquad (1)$$

Figure 2 MP model

4.3 Behaviorism based on "perception-action"

Artificial intelligence behaviorism came into being through the research of control system and the influence of biological evolution theory. The core of behaviorism is to improve the thinking mode of machine learning by studying and imitating people's behaviors, psychological activities and neural thinking, so as to improve the efficiency of machine work learning. Behaviorism focuses on "perception" and "movement", and it needs constant training in complex real environment in the process of learning. Compared with connectionism, behaviorism is more difficult in the process of learning and training, so it has certain challenges. From the 1940s to 1950s, a group of artificial intelligence workers appeared all over the world. At this stage, the driving force for them to study artificial intelligence was the emergence of cybernetics, which also laid the foundation for the emergence of behaviorism in the later period. In 1950, biological cybernetics and engineering cybernetics were put forward by some researchers. After that, these researchers began to want to control people's behaviors and psychological activities through computers. However, behaviorist paradigm research shows that the reason why machines can perform some intelligent activities and operations is not from computer network technology, nor from some formal structures. Behaviorist paradigm research holds that the work of machines can capture signals and work given by external objects and respond directly to the captured signals. Therefore, thinking is the prerequisite for agents to respond to external stimuli. Whether behaviorism or connectionism, their paradigm research aims at solving the possible problems in the development of artificial intelligence, so as to ensure the better development of artificial intelligence.

4.4 Research paradigm of symbolism

Symbolism belongs to the category of modern artificial intelligence, which is an intelligent simulation method based on logical reasoning to simulate people's intelligent behavior. Symbolism holds that "intelligence" is the most basic symbolic public office, which can store and analyze information symbols collected by intelligent agents. At the same time, like behaviorism, symbolism holds that "thinking" is the prerequisite for intelligent development. Symbolism is a section of mathematical logic in the field of artificial intelligence. At the end of 19th century, mathematical logic began to get rapid development and progress. With the continuous evolution and progress of information technology, followed by the emergence of computer technology, big data technology and cloud computing technology, the logical deduction system was realized.

5 Application of artificial intelligence in design method based on product semantics

5.1 Application of "deep learning" in human intelligence

Deep learning is the core of artificial intelligence, and the core of deep learning is "machine learning". In fact, machine learning is to learn some computer data, then predict and judge other data, analyze the data by using algorithms, learn from them, then make decisions and predictions on the data, be good at the messy user data, and integrate and analyze the analyzed data. It is the main purpose of product semantic design to establish a mapping relationship between product physical form and users, and whether the constructed mapping relationship can be fully perceived by users can be used as a good evaluation factor. Deep learning is a discipline and technology based on neural network algorithm and simulating the iterative pattern of human brain. In 2014, Google applied artificial intelligence and machine learning to home manufacturing, face recognition, iris recognition and other fields. Face recognition technology can distinguish different faces through target detection technology in deep learning. In 2018, the design laboratory of MIT developed deep learning shoes, and used deep learning technology to make shoes with "temperature". This shoe is mainly composed of circuit layer, microbial layer and electronic component layer. Using PyTorch to generate shoes from the border, this process mainly involves the definition of imported data, preprocessing, training the main function of the model, etc., then generating colored shoes to shoe sketches, then mapping them to colored shoes, and finally generating shoes. The design of this shoe uses the working principle of DiscoGAN, which can recommend shoes to users and match shoes for users according to their clothes. Figure 3 shows the architecture of DiscoGAN, which is composed of two GAN models. Finally, the microcontroller converts the data into digital data and transmits it to the terminal for processing. Finally, the personalized shoes with algorithm model are established.

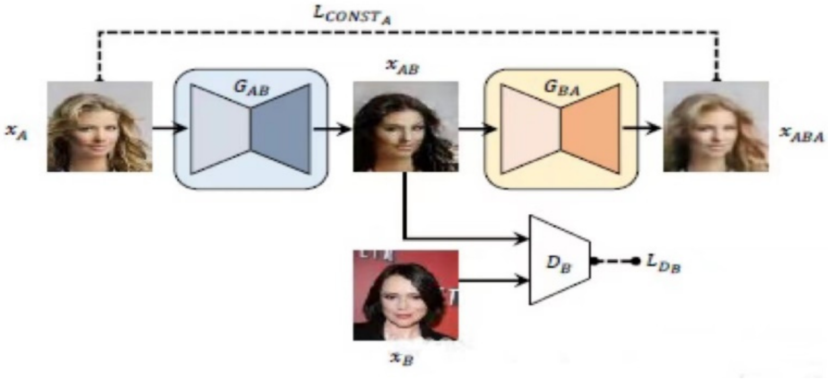


Figure 3 The architecture of DiscoGAN

5.2 Application of "data" in user portrait extraction

In recent years, with the rapid development of cloud computing, big data, Internet of Things and other network technologies, people's lifestyles and working styles have been greatly affected, especially in the application of product semantics, which provides new opportunities and new thinking for the development of product semantics in the big environment. Integrating small data into big data through Internet of Things technology is a big environment in the process of communication and evolution between people and objects. For the traditional product design, product design mainly focuses on in-depth research of product materials and packaging modules. However, in the intelligent environment, product design is not only the connection between people, people and objects, objects and objects, but also the research and exploration of factors such as space dimension and time of product design, so as to promote the design and development of product service system. User portrait is the result of collecting, sorting and analyzing user information, and it is also an important condition for product design. "Data" includes big data and small data, and the combination of big data and small data is mainly used in the extraction of user portraits. Investigating users' needs, observing users' psychology, conducting questionnaire survey on users, etc. are all the tasks that product semantic design needs to complete before the design. Users' information collected by workers is mainly analyzed and processed by "small data" to the perception system and behavioral logic. However, small data has certain drawbacks, for example, there is a certain limit on the number of user portraits extracted from data, and it is impossible to collect subconscious user portraits. Therefore, small data can only integrate some user portraits, which is not universal. Then, with the rapid development of Internet, big data and other technologies, big data technology has been widely used in the extraction of user portraits. Big data can obtain portraits of a large number of users, and can use data analysis, data induction and other technologies to process and analyze user portrait data, thus realizing the collection of user group data. "Small data" can be aggregated into "big data" by using the Internet of Everything. Therefore, the application of artificial intelligence in product semantic design can make product design more suitable for users' needs.

6. Conclusion

With the rapid development of information technology, big data, Internet of Things and other network technologies, each of us is a tiny object in big data. Through artificial intelligence, we can combine product design with user's needs, simulate human's perception and brain activity, map the collected user information into deep learning, and then apply machine learning to formulate product design that fits users. From this, we can see that the application of artificial intelligence in product design has injected new vitality into product design.

Reference

[1] Wu Suyan. Research on Key Technologies of Remote Design of Mechanical Products Based on Artificial Intelligence[D]. Zhengzhou University, 2003.

[2] Gao Zhenbin. Product Design Method Based on Artificial Intelligence[J]. Electro-Mechanical Engineering, 2000(05):42-44.

[3] Ran Bei. Research on Product Design Method of Artificial Intelligence Based on Product Semantics[J]. Design, 2021, 34(12):3.

[4] Liu Jie. Research on Artificial Intelligence Design Method of Appearance Color Matching Based on Product Function[J]. Journal of Changchun Normal University (Natural Science Edition), 2021(005):040.

[5] Lai Chaoan. Research on the Theory and System of Conceptual Design of Human-Computer Collaborative Innovation Based on Knowledge Fusion[D]. South China University of Technology, 2003.

[6] Wang Runfang, Jiang Bo. Research on Automatic Generation Method of Electronic Software Data Based on Artificial Intelligence Algorithm[J]. SP, 2020(4):1.

[7] Tang Zhichuan, Wang Dongling, Xia Dan, et al. "Artificial Intelligence+Design" --A New Exploration of the Teaching Practice of Product Design Courses for Design Majors[J]. Art & Design, 2020(1):4.

[8] Wang Lingxiao, Ban Ningqiu. Research on the Influence of Artificial Intelligence on Product Design[J]. China Southern Agricultural Machinery, 2020, 51(6):2.

[9] Liu Jingzheng. Research on Design Method of Electrical Products Based on Artificial Intelligence Technology[D]. Hebei University of Technology, 2009.

[10] Wang Shanshan, Lu Xiaoyi. Fusion Platform of Nanjing Art Resources and Cultural and Creative Products Based on Artificial Intelligence Aesthetics[J]. Heilongjiang Textile, 2020(1):4.

A Study on Vehicle Bottom Safety Detection Technology Based on Image Feature Matching Algorithm

Wei Gao*, Maoying Li, Chen Chen
Automotive Data of China (Tianjin) Co., Ltd, Room 12-17, New City Center B1, No. 3, Wanhui Road, Zhongbei Town, Xiqing District, Tianjin, PRC.
* Corresponding author: gaowei2020@catarc.ac.cn

ABSTRACT

In recent years, the vehicle safety detection system is a common information management method of vehicle safety detection in various ports and forts. However, it is generally unable to accurately identify by the system, and still needs manual re-inspection. In this paper, the image algorithm is comprehensively utilized after the functions of automatic identification, display storage and foreign object comparison. A large number of image processing, target recognition, depth learning and other technologies are used to realize the accurate identification of abnormal objects. According to the difference between the normal image and the current image in the database, the vehicle bottom image will be detected for the presence of abnormal objects. From the intelligent point of view, the processing level and accuracy of the detection system can be greatly improved to more than 99.1% and the human intervention and foreign object investment can be reduced to the greatest extent. From the perspective of image processing methods, this paper introduces a new safety detection method of vehicle bottom image data identification.

Keywords: 3D image algorithm, Image matching technology, Target detection database annotation

1. INTRODUCTION

The vehicle bottom safety detection technology based on the image feature matching algorithm can solve the problem of limited obvious features and poor detail information among low-resolution images. At present, the vast majority of under car safety inspection system adopts a high-speed linear ARRAY CCD camera combined with a wide-angle lens and high brightness LED light source to scan and transfer digital signals to the processing system at high speed, and then complete the scanning through acquisition, processing, image combination. Because of the inconsistent speed, the deformation of the picture leads to the picture length and width information is also more difficult to restore, and it is very difficult to match using the template. Even if the accuracy of the library detection is about 70%, it can not reach the level of automatic detection and management. What is more, when dangerous foreign objects could be obscured by some structures of the chassis such as exhaust pipes or close to the edges, then the camera does not capture a clear outline at all and the miss detection rate is nearly 100%. A good image feature matching algorithm can solve this problem, which also requires matching probability as high as possible, negligible matching error, faster algorithm computation speed. These parameters can meet the real-time requirements of the application environment.

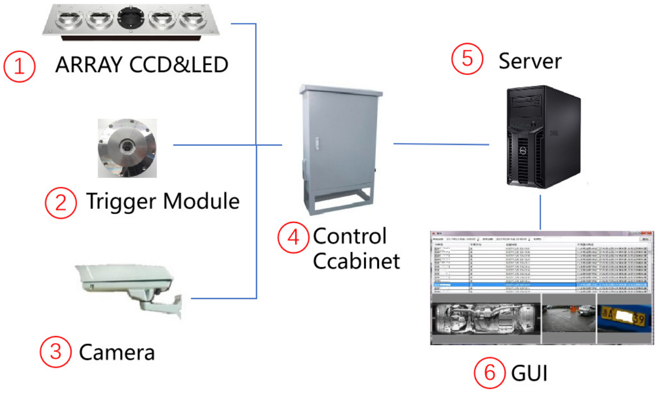


Figure 1 The system structure of vehicle bottom safety detection

International Conference on Computer Graphics, Artificial Intelligence, and Data Processing (ICCAID 2022)
edited by Ruishi Liang, Jing Wang, Proc. of SPIE Vol. 12604, 1260429
© 2023 SPIE · 0277-786X · doi: 10.1117/12.2674545

In this paper, The vehicle bottom safety detection technology based on the image feature matching algorithm can solve the problem of limited obvious features and poor detail information among low-resolution images. From the perspective of 3D image processing methods, it will improve the vehicle bottom image database security detection algorithm technology. At the same time, The comparison method of YOLOv4 and RetinaNet is used to improve the detection accuracy after data pre-processing. This method of SsD does not be used, which is often used in practice actually but is slightly worse than YOLOv4 in terms of performance. Of course, because SsD is not upgraded in recent years, they are used as the baseline for many new methods such as RFB-Net, DSSD, etc [7]. About the problem of some foreign object categories with small data volume, this paper proposals a new way. It can achieve the purpose of data enhancement base on image matching techniques using the hazardous materials database. The robustness of system detection is enhanced combined with 3D object deep learning. In addition, we also use our method to develop automatic tagging functions for training database-based data generation of 3D depth images. Online learning can improve model accuracy by adding modified detection result data to original chassis data set, which can greatly improve the accuracy of detection and fundamentally solve the problem of high missing detection rate caused by obscuring, etc.

2. ESTABLISHMENT OF IMAGE DATABASE

2.1 Establishment of Image Database

The vehicle bottom safety detection scheme based on binocular depth estimation needs to establish a perfect vehicle bottom image database in advance. The vehicle bottom images in the database will be used for depth map matching technology to detect the existence of abnormal objects based on the difference between the image in the database and the current target. Furthermore, The vehicle bottom image database needs to be established under the following strict requirements to acquire high-quality images.

a. Make sure the changes in external natural conditions would not bring negative influence such as light intensity, weather conditions, etc .

b. Acquire image data under different lighting conditions (early morning, midday, night, etc.).

c. Acquire image data under different weather conditions (rainy, snowy, cloudy, etc.).

d. The resolution of the acquired images remains essentially the same.

e. During the acquisition process, the direction of the vehicle remains consistent. It should be ensured that the angle change and the distance change between the vehicle and the camera should keep within a certain range.

f. Image acquisition for different vehicle models: the number of models should be more than 1700, and the number of images should be more than 50 per model.

g. It is recommended to repeat the acquisition using different vehicles for one model.

h. Acquire images at different positions and angles.

i. Using the same equipment or the same type of equipment to collect images, the equipment used for data collection should be consistent with the equipment used for inspection, and the equipment used for actual inspection should also ensure that the images meet the above requirements when generating pictures.

j. If there are foreign objects, the location and type need to be marked manually.

2.2 Image Retrieval

During the whole system working process, the first step is the need to find the corresponding model template image from the image database for the next step-foreign object detection. It is recommended to search the database images with image matching techniques in this step [3]. The original image is taken by the binocular depth camera and then that depth map are also used to improve the robustness of system. There are three matching methods: template matching, feature matching, and deep learning feature matching. This paper combines three matching methods that and follows the below the three steps.

Firstly, it should perform visual feature extraction for all the existing images in the database with traditional feature methods such as SIFT descriptors.

Secondly, this is an important process to find out the identical feature and the area around features. Because the image can not distinguish the main feature for recognition, for instance, colour, texture and so on. The system needs to identify the identical feature and describe the area around features among other images. SIFT (full name is Scale Invariant Feature Transform), is a descriptor used in image processing. This algorithm can detect small area features in an image by finding the features and their descriptors regarding size and orientation, it can obtain very stable image features and perform image feature matching [5]. SIFT as the feature extractor, can change the rotation angle, image brightness or shooting angle to get the best detection characteristics. The system can get the perfect detection characteristics with the image classification algorithm:

$$L(x,y,z,e)=G(x,y,z,e)*I(x,y,z) \quad (1)$$

whiere $G(x,y,z,e)$ is a scale-variant Gaussian function, (x, y,z) is spatial coordinates and e is scale coordinate.

The following briefly describes the SIFT feature extraction process. SIFT is an initialization operation with a clear purpose-firstly it simulates the key point localization of the image data, then uses multi-scale features and re-specifies the orientation parameters, while finally generating key point descriptors. What is more, we can further expand the SIFT features in the form of a spatial pyramid and then do further refinement analysis.

The system algorithm also requires visual depth feature extraction for all images in the database. Since better performance is obtained with depth features, this paper combines the depth features to improve the robustness of the model. While in the case of a large amount of depth features can be provided from image data, we collect a large amount of labelled data and uses a per-trained model on the network as feature extractors [1].

① Fusion of multiple features will be performed.

② The proposed stitching or multi-modal fusion method is used to conduct the experiments separately.

③ The data captured by multiple cameras (CCD arrays) with different angles and positions and image feature fusion is performed.

④ Extract the image features of the vehicle to be retrieved, and use the same feature extraction algorithm as above.

⑤ The features extracted from the vehicle images to be retrieved are matched with the features in the database to do the matching calculation, and finally the most similar rather than the most identical vehicle images are taken as the matching to derive the results, or the top K most similar vehicles can be extracted.

Last but not least, there are many matching calculation methods, including: the error sum of squares algorithm, normalized product correlation algorithm, sequential similarity detection algorithm, Hadamard transform algorithm, and so on in the above five steps. Separate matching combined with fusion matching is used and integrated learning is performed for the images from multiple cameras [4].

2.3 Foreign object detection

Foreign object detection must be based on an image matching algorithm, which is a process of comparing two images with similar features and following that pixel-wise alignment and discrimination are performed. The images obtained by matching are often sourced from the same or similar natural or unnatural scenes and objects. In the way, we name this kinds of images with the same or similar semantic information as match-able conformation requirements [2]. It uses image matching algorithm to detect foreign objects, which requires the following steps:

In beginning, it uses the SIFT algorithm to extract the feature points and descriptors of the image, and the effect shows in Figure 2. We pick up feature descriptions to represent the image with a matching algorithm to regional match or search among the same or similar features. In this study, we work in the same way with the FLANN algorithm. As shown in Figure 3, we find those highly similar data feature segments according to the above process. To obtain the best matching descriptor pairs, we correspond the coordinates of key points of the paired images one by one. After that, it can gain the final position with a separate correspondence matrix to solve the transformation.

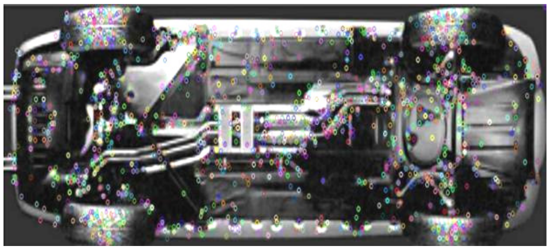


Figure 2 SIFT feature points in a template image

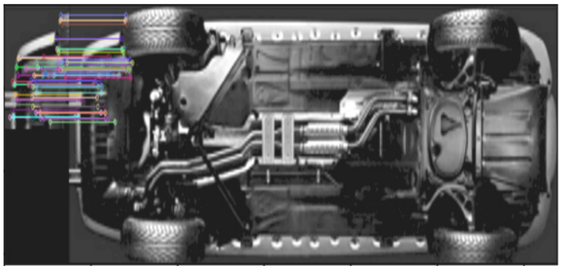


Figure 3 Feature Point in a template image

Figure 4 shows a detecting image which has added a knife entrapped in the car. The system should get the corresponding template image from the database according to the model retrieval module described in Section 2.2, and the template image is shown in Figure 5. After that, the testing image is cropped in a grid-like manner to obtain n sub-images of equal size, as shown in Figure 6. As an individual image, these sub-images need to be matched with the template image retrieved from the image database again, and if the sub-image was matched with a template image, it means that there is no foreign object in this position. On the contrary, if a sub-image fails to match the same regional part of the template image. It means that there is a foreign object in the sub-image. The matching result shows in Figure 7, where the green squares are marked as correct matching ones which means there is no foreign object that is present in these squares, and the orange squares are marked as incorrect matching ones, indicating the presence of foreign objects.

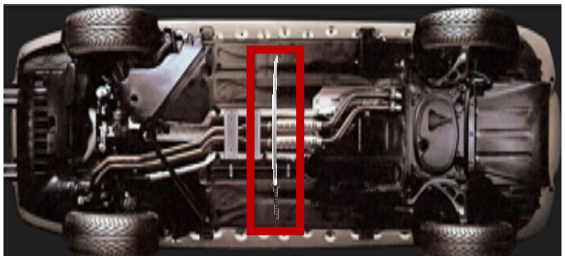


Figure 4 Detecting image with knife

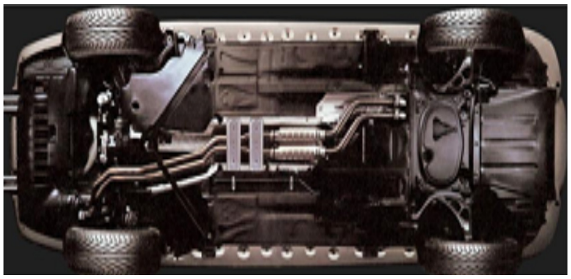


Figure 5 Template image in the database

The system has found the approximate position of the foreign object and can move the position of the detection frame and change its size based on this position after the above process. So it generates more sub-plots to get the more accurate position by repeating the above operation.

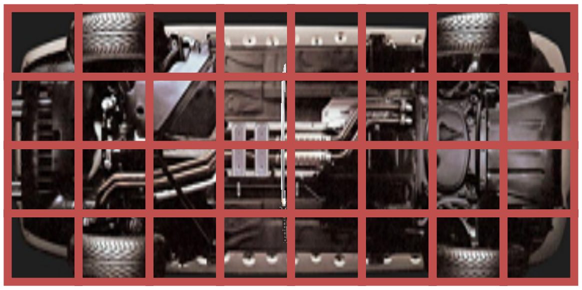


Figure 6 Testing image after grid division

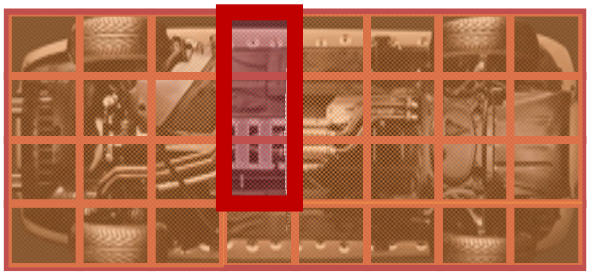


Figure 7 Matching result of testing image

2.4 Foreign object recognition and online learning

When a potential foreign object is detected and gain its location from a image, the system crops down the object part from that and identifies it. Since there is a lot of uncertainty types of hazardous objects, foreign object recognition is a challenging task for image classification.

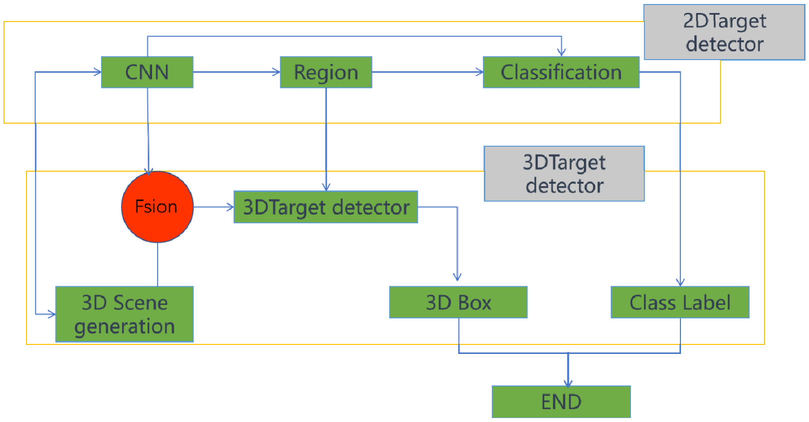


Figure 8 foreign object recognition

There are several steps In the process of foreign object identification are required:

Step 1: We build a dataset of hazardous items for this study. First of all, it can be collected from the Internet. In fact everyone knows that the potential hazardous items are unknown in practical situations and the hard part is that image data of all hazardous items cannot be collected.

Step 2: A pre-trained image classification model is used as the base model, which can classify some types of items already in the dataset.

Step 3: The base model is modified using an algorithm of open-set identification so that it can identify some unknown hazardous items into unknown classes.

Step 4: In the application, such unknown items can be continuously collected and labelled by repeatedly training the model, so that the model will be able to gradually identify as many hazardous items as it could.

For the above steps, the new data can be added to the existing dataset by online learning, so that the system can achieve the effect of incremental updates. If a image that be detected below the threshold in step 2 is considered as new one in the database and would be added to the existing database. The new classes in the open-set in step 4 can be considered as new hazards items and can be added to the hazards items database, which is benefit of the identification model updates. Meanwhile, online learning improves model accuracy by enriching the corrected inspection result data into the original vehicle chassis dataset.

3. DATABASE TRAINING BASED ON TARGET DETECTION

3.1 Building separate dataset

Since this solution is different from the regular target detection system in terms of identification types, it is recommended to build a separate dataset in advance. Therefore, an image-based target detection program needs the pre-establish the database with hazardous objects, which can be achieved in several ways. It is intentional to place some hazardous materials under the car, obviously considering the different combinations of hazardous objects, different models, different placement, different packaging, etc. It can simulate generating data, using the existing data and changing the position of foreign objects, combinations as well as other forms to generate more data.

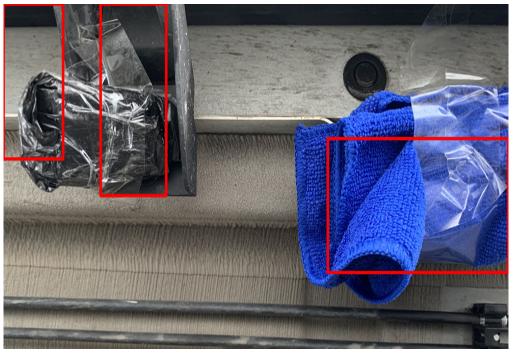


Figure 9 Image with foreign objects

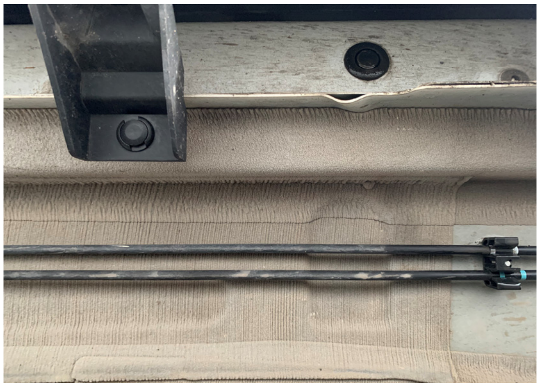


Figure 10 The template in the database

3.2 Target detection database annotation

Based on the existing dataset, the foreign object locations and species can be labelled in several ways. For example, manually labelling the location and type of foreign objects. Manually annotate a small amount of data, train the target detection model, manually review the detection combined with the review, and put the corrected data into the training set for retraining. The annotated object category is a comparison of the training image with the 5050 image test. An approach that takes open-set target detection, such as the YOLO method can improve the target detection with new images, the model is trained against an existing training set and then tested on the samples.

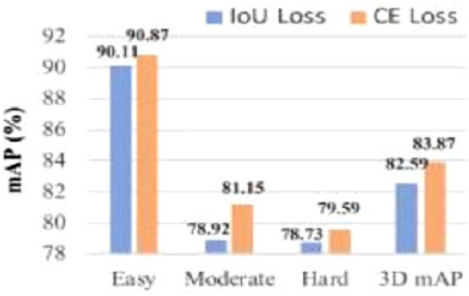


Figure 11 Result on the KITTI validation dataset

3.3 Target detection and evaluation indicators

In the target detection phase, it usually uses mean Average Precision (mAP) as a "litmus test" to check the detection capability of the model. In this study, we use mAP as one of the important indicators and Frames per second (fps) as the second important indicator to check the detection speed of the model. Thus, the final performance of the model is measured by these two indicators. Depth images are generated by the binocular camera and used to improve the robustness of detection by combining the image information in the database after the database is built and trained.

4. BINOCULAR DEPTH ESTIMATION

4.1 Foreign object detection

The foreign object detection will be based on a binocular depth estimation algorithm, and the 3D information of the scene will be obtained by processing the images that were taken by binocular cameras. In the same way, the result will be represented as a depth map. Foreign object detection based on binocular estimation has the following three steps:

1. It should calibrate and calibrate the binocular camera and then adjust the internal and external parameters of the camera, which are used to calculate the depth of the object.

2. The depth map obtained by depth calculation of the two images after correction, and the similar colours indicate that they are at the same depth.

3. We use the computed depth map to compare with the retrieved template depth map in the database. So that objects with significant bumps relative to the template depth map can be detected.

4.2 Foreign object identification

The database collects the different models after a long time of training and labels the type of these. In other words, the system collects different positions, angles, and packing hazards data with 3D target monitoring data. On the other hand, the 3D labelled information of the target goes up, as shown in Figure 12.

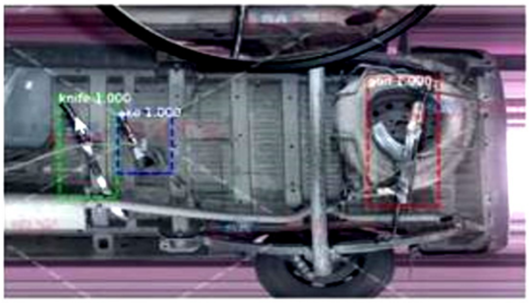


Figure 12 Annotation diagram in the database

Combining 3D target monitoring data, we currently have 3712 frames for the training Kitti dataset, 3769 validation frames and 7158 test frames, and 10335 frames for sun-rgbd dataset (type: simple, medium and complex), training images and 5050 images for testing. What is more, those were made to include annotations. The angle of inclination of the image in the horizontal direction is ±15°. For some minority categories with a small amount of data, data enhancement is performed with image matching technology from the hazardous materials database. This solution identifies ten categories of hazardous foreign objects (guns, explosives, grenades, knives, axes, sticks, plastic bags, broken concrete asphalt, paper products, leaves) that can be labelled.

Table 1. Training peak performance

Dataset	**Average identification time**	**Audio**
KITTI (>200k 3D labelled objects)	≤1s	98.90%
CityScapes (5000 fine labelled objects and 20000 roughness labelled objects)	1.3s	99.10%
nuScenes (14,000 multiple bounding boxes)	1.3s	98.80%
Waymo (25,000,000 3D multiple bounding boxes)	≤1s	99.40%

5. CONCLUSION

This paper will image processing methods from the point of view of the improved undercarriage image database of undercarriage safety detection technology. The use of YOLO to take an open set of target detection methods, while data pre-processing, for the existing training set for model training. The performance of detecting foreign objects can improve the accuracy of detection to about 99.1% in the database. For some categories with small data volume, data enhancement is performed using image matching techniques in the hazardous materials database. For the case of targets with more occluded foreign objects, 3D target depth learning is recommended to enhance the robustness of the system detection. In addition, the use of online learning to enhance the model accuracy by adding the corrected detection result data to the original vehicle chassis data set greatly improves the accuracy of detection and fundamentally solves the problem of the high rate of missed detection caused by occlusion, etc. The scheme proposed in this paper completely solves the problem of manual rechecking and improves the artificial intelligence of the system.

This algorithm solution essentially detects foreign objects by sequentially performing image processing operations, such as template matching and target detection on the target image based on a pre-established image database. The solution proposes a 3D target detection with a template matching method, where the accuracy and leakage rate can reach a level that is free from manual or only requires simply review. The dataset through online learning after manual correction by the inspector is more authentic and reliable with the detection requirements in real situations

REFERENCES

[1] Zhu S. S, Jia X. Y, Li Z. C., "Exploiting global and local label correlations for multi-label classification," Acta Electronica Sinica, 2345-2351 (2020).

[2] Bodla N, Singh B, Chellappa R, et al., "Soft-NMS-Improving Object Detection with One Line of Code," 2017 IEEE International Conference on Computer Vision, 5562-5570 (2017).

[3] Zhuang Q. W, Zhong Y, Zhang M. L., "Feature-induced labeling information enrichment for multi-label learning," Proceedings of the 32nd AAAI Conference on Artificial Intelligence, 4446-4453 (2018).

[4] Li S. T., "Image stitching system based on local feature extraction," Electronic World, 46-47 (2018).

[5] Guo Z. C, Dang J. W, Wang Y. P, et al., "Background modeling method based on multi-feature fusion," Opto-Electronic Engineering, 180206 (2018).

[6] Zhu X.Z, Xiong Y.W, Dai J. F, et al., "Deep Feature Flow for Video Recognition," Computer Vision and Pattern Recognition,2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR),2350-2358(2017).

[7] Bochkovskiy, A, Wang, C. Y, Liao, H. Y. M., "YOLOv4: Optimal Speed and Accuracy of Object Detection," arXiv preprint arXiv, 10934 (2020).

Multiple Vehicle Detection and Tracking using Improved YOLOv5 and Strong SORT

Yinan Zhang[1a*], Tong Zhang[1b], Zhichao Huang[1c]
[1]Department of Electrical Engineering, Guilin University Of Electronic Technology,
Guilin 541004, China
[a*]1205440049@qq.com, [b]253067155@qq.com, [c]644863316@qq.com

ABSTRACT

Multiple object tracking (MOT) is an important subject in applications of computer vision As a subtask of object detection and tracking, vehicle tracking has important research significance. This paper proposes a vehicle tracking and detection technology which is based on improved YOLOv5 and Strong SORT. The YOLOv5 combined with the CBAM attention mechanism work as the detector of Strong SORT in the tracking process, this arrangement decreases computational time. Experiments proved that this proposed algorithm can effectively deal with the problems of object occlusion, target loss, and ID switch. The trained model is easy to deploy for an embedded device, which makes it a very good candidate for a real-time surveillance system.

Keyword: tracking; detection; YOLOv5; Strong SORT; CBAM

1. INTRODUCTION

Multi-target tracking is an critical topic of computer vision, which is used in autonomous driving, violation detection and so on. The main task of the multi-target tracking algorithm is to correlate the motion object detected in the video and obtain the motion track of each moving target. In the video frame, the vehicle position has been obtained using the target detection algorithm, but the relationship between vehicles in the video sequence cannot be determined. Using the multi-target tracking algorithm, the vehicle information in the video sequence can be correlated to obtain the overall motion track of the vehicle. In the practical application of the road scene, most of the surveillance cameras are installed on both sides of the road, so vehicles often block each other in the process of video monitoring. When the target vehicle reappears in the video after being blocked, it easily identifies the target vehicle as a new target, causing the loss of the original target ID, which brings some difficulties to the target track tracking. Therefore, this paper proposes a simple and efficient tracking algorithm, using improved Yolov5 and Strong SORT for target tracking.

2. RELATED WORK

Before the advent of deep learning and correlated filter tracking algorithms, classical algorithms were used in the field of target tracking. At present, most algorithms used image edge features and probability density as tracking criteria to increase the direction of target search along probability gradient, such as particle filter [1], mean shift [2] and Kalman filter [3].

Fukunage has proposed mean shift algorithm in 1975， and Yizong Cheng expanded the algorithm later. He proposed two methods: one is to define kernel function and the other is to increase weight coefficient. The algorithm is not very computationally intensive and can fully achieve real-time tracking when the target area is known. The kernel function histogram model is used, which is insensitive to edge occlusion, target rotation, deformation and background motion.

The particle filter framework is a Monte Carlo method with an importance sampling layer idea. The basic idea behind this method is to use a cluster of samples (or particles) to approximate the posterior probability distribution of the network, and then uses this approximate representation to approximate the state of the non-linear system. With this idea, a particle filter can handle any kind of likelihood in the filtering process, unlike the Kalman filter which is only able to cope with the probability of a linear Gaussian distribution. Which is also a major advantage of particle filtering.

Kalman filter adopted the concept of state space,thus changing the general description of the filtering problem, instead of requiring a direct second-order characteristic or spectral density function of the signal process, this input-output relationship is described by an equation of state,thus making the signal process under study not only a smooth scalar In

International Conference on Computer Graphics, Artificial Intelligence, and Data Processing (ICCAID 2022)
edited by Ruishi Liang, Jing Wang, Proc. of SPIE Vol. 12604, 126042A
© 2023 SPIE · 0277-786X · doi: 10.1117/12.2674583

this way, the studied signal process can include not only the smooth scalar stochastic process, but also the non-smooth vector stochastic process. This overcomes two major drawbacks of classical filtering theory. So it is successfully applied in space technology, navigation and positioning of aircraft, fire control systems, etc. It is an efficient recursive filter that can obtain the optimal state estimation for linear stochastic dynamic systems and the best filter for discrete linear systems.

With the growth of deep learning, classical tracking methods have been abandoned, as classical methods are unable to be applied to complex ID tracking and detection, and the accuracy and robustness are outperformed by current algorithms.

To address these issues, this paper uses two current algorithms, YOLOv5 and Strong SORT [4], to improve the accuracy of vehicle tracking and perform multi-vehicle tracking in videos

3. METHODOLOGY

This section will explain our enhancement of the structure of the YOLOv5 network. In order to increase the feature learning ability of DarkNet-53, we introduce the CBAM (Convolutional Block Attention Module) [5] attention mechanism into the backbone of YOLOv5. We then combined YOLOv5 with Strong SORT in order to realize the vehicle tracking system.

3.1 YOLOv5 with CBAM Attention Module

Attention mechanisms in the neural network are resource allocation schemes that allocate computational resources to larger tasks and deal with information overload in the presence of limited computational resources. We introduce attention mechanism to make the algorithm is able to focus on the most critical information through the inputs, decrease the attention to other information, and filter out useless information. These can reduce the overload of information and improve the accuracy and performance of the task.

The CBAM algorithm consists CAM (channel attention module) and SAM (spatial attention module). Attention characteristics in the dimensions and channels are fused, respectively. This method not only saves the time and decreases the parameters, but also make sure that it can be a pluggable module to be inserted into other algorithms. In this case, we added spatial attention module and channel attention module to the bottom of each activation layer, as shown in Figure 1. And the CBAM structure diagram can be viewed in Figure 2.

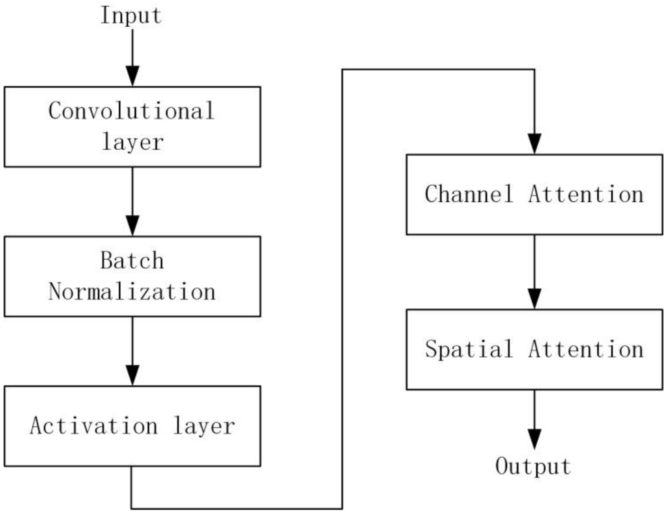


Fig. 1 Convolutional layer with attention module.

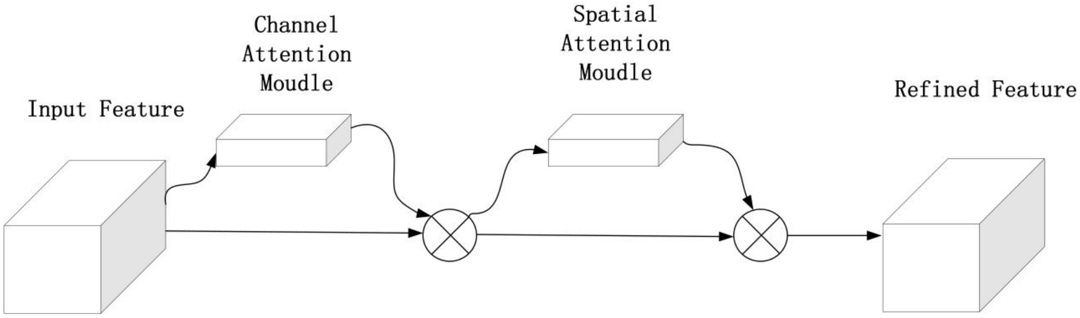


Fig. 2 CBAM structure diagram.

3.2 Strong SORT

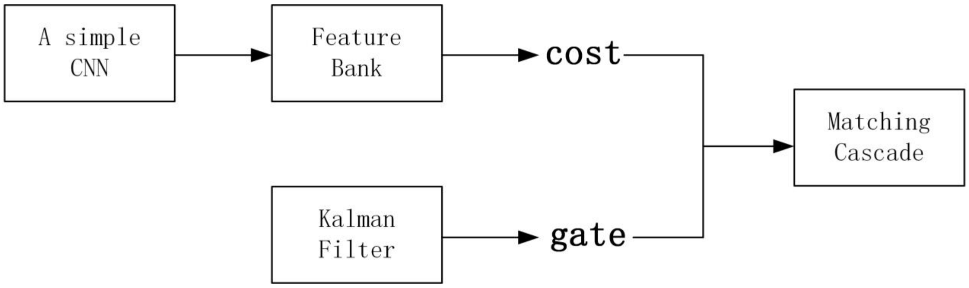


Fig. 3 Deep SORT Framework.

Strong SORT is an improved version of Deep Sort [6]. The improvements over Deep SORT lie mainly in two aspects, as shown in the bottom half of Figure. 4. Compared with Deep SORT Framework in Figure. 3, it uses BOT [7] and ResNeSt50 [8] to replace the original simple CNN neural network. ResNeSt50 as the backbone can extract more discriminative features. Then, it replaces the feature bank with an exponential moving average (EMA) strategy. The EMA update strategy not only improves the quality of matching but also reduces the time consumption.

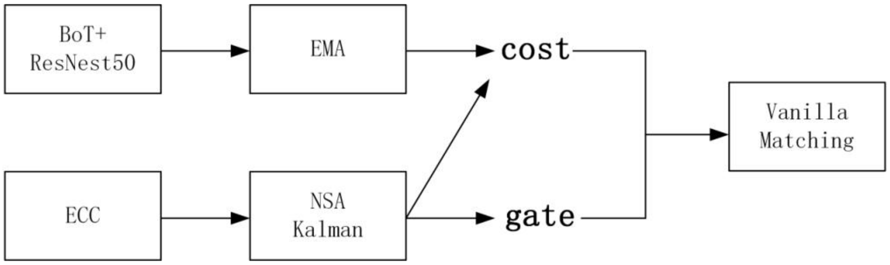


Fig. 4 Strong SORT Framework.

As for the motion branch, it adopts ECC for camera motion compensation. In addition, the original Kalman filter is vulnerable to poor quality detection and ignores information about the scales of the detection noise. To solve this problem, it uses the NSA Kalman filter which proposes a formula to calculate the noise covariance $\tilde{R}_k$:

$$\tilde{R}_k = (1 - c_k) R_k \tag{1}$$

c_k is confidence score and R_k is preset constant measurement of noise covariance.

In addition, it uses both Mahalanobis Distance and Cosine Distance instead of only using Cosine Distance. Cost matrix C is weighted sum of Mahalanobis Distance $d^{(1)}$ and Cosine Distance $d^{(2)}$ as follows:

$$C = \lambda d^{(1)} + (1 - \lambda) d^{(2)} \tag{2}$$

Where weight factor λ is set to 0. 02, another problem is that the matching cascade limits the performance as the tracker becomes better. So, the matching cascade is replaced with a vanilla global linear assignment.

4. EXPERIMENTS AND RESULTS

We used UA-DETRAC as the dataset and trained 100 epochs in the training phase in the UA-DETRAC dataset. We compared the network under different conditions and used the most common evaluation indicators for the boundary box regression tasks: mAP0.5 and mAP0.5:0.95 as the measurement of the experiment. Strong SORT Tracking FPS is the Strong SORT processing performance with YOLO models. The experiment result is shown in Table I.

In Table 1, the original YOLOv5 backbone has got 80.2% performance on mAP0.5 using an evaluation metric of IoU, and the mAP0.5:0.95 achieved a performance of 60.4%. The proposed YOLOv5 with CBAM can improve 1.1%mAP0.5 and 0.6%mAP0.5:0.95 in detection. And the Strong SORT tracking performance is almost the same. This proves that YOLOv5 embedded in CBAM has better performance under the stricter standard of IOU. On the other hand, we used the latest YOLOv7 and YOLOv7 with CBAM for another comparison. The mAP0.5 and mAP0.5:0.95 are slightly lower than YOLOv5 with CBAM, but Strong SORT Tracking FPS is much lower than YOLOv5 with CBAM. This proves that YOLOv5 with CBAM has a faster tracking speed in Strong SORT tracking than YOLOv7 with CBAM.

Table 1. Vehicle tracking results of different network structures.

Network Structure	***mAP0.5/%***	***mAP0.5:0.95/%***	***Strong SORT Tracking FPS***
YOLOv5	80.2	60.4	13.32
YOLOv5+CBAM	81.3	61.0	13.25
YOLOv7	79.3	59.4	2.58
YOLOv7+CBAM	79.6	60.1	2.56

In Strong SORT vehicle tracking, the position of the vehicle has been obtained by using YOLOv5 with the CBAM target detection algorithm, as shown in Figure. 5. Then, the Strong SORT tracking algorithm can associate the vehicle information in the video sequence, and obtain the overall vehicle motion track through the Kalman filter and matching cascade, as shown in Figure. 6 and Figure. 7.

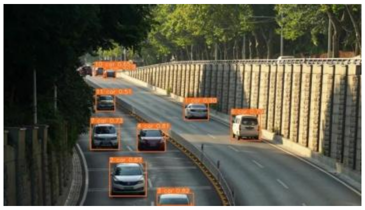


Fig. 5 Results of Target Detection.

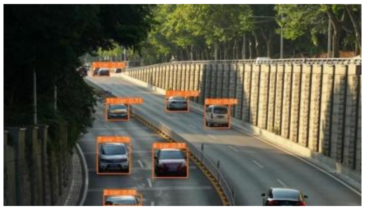


Fig. 6 Results from the first tracking.

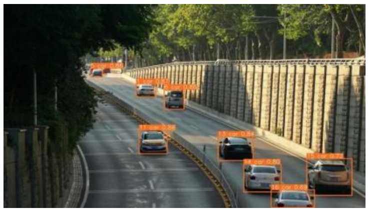


Figure. 7 Results after continuous tracking

5. Conclusion

We use the CBAM attention mechanism as the backbone of YOLOv5 to improve the performance, adds CBAM to the back of each activation layer, and uses YOLOv5 with CBAM as the detector of Strong SORT for vehicle tracking. It achieved good mAP performance and high speed in vehicle tracking.

However, the proposed approach still has some limitations. For the scenario in which IoU is larger than 0.5, it still needs further improvement in its performance. Therefore, in future work, we will consider extending the vehicle dataset and considering more improvements to the network structure.

REFERENCES

[1] Merwe R, Doucet A, Freitas N D, et al. (2001) The unscented particle filter. In: Advances in neural information processing systems.

[2] D. Comaniciu, P. Meer. (2002) Mean shift: a robust approach toward feature space analysis. In: IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 24, no. 5. pp. 603-619.

[3] Kalman R E. (1960) A New Approach to Linear Filtering and Prediction Problems. In: Trans of the ASME Journal of Basic Engineering.

[4] Du Y, Song Y, Yang B, et al. (2022) Strong SORT: Make Deep SORT Great Again. arXiv e-prints.

[5] Woo S, Park J, Lee J Y, et al. (2018) CBAM: Convolutional Block Attention Module. In: European Conference on Computer Vision.

[6] Wojke. N, Bewley A, Paulus D. (2017) Simple Online and Realtime Tracking with a Deep Association Metric. In: 2017 IEEE International Conference on Image Processing (ICIP).

[7] Luo H, Jiang W, Gu Y, et al. (2020) A Strong Baseline and Batch Normalization Neck for Deep Person Re-Identification. In: IEEE Transactions on Multimedia.

[8] Zhang H, Wu C, Zhang Z, et al. (2020) ResNet: Split-Attention Networks.

Research on Employment Quality of Wuhan Business University Based on AHP and BP Neural Network

Quanli Wang
School of General Education, Wuhan business University, Wuhan City, Hubei Province, China，
430000, 2023236027@qq.com

Abstract

In order to further improve the linkage mechanism between employment and enrollment plan, talent training, and specialty adjustment, this paper conducts a quantitative analysis and deepening of the *Report on the Employment Quality of Undergraduate Graduates of Wuhan Business University.* This paper uses the Analytic Hierarchy Process to build the evaluation index system of the employment quality of college graduates, focusing on the evaluation criteria, evaluation indicators and calculation methods to calculate the index weight. At the same time, we train the model of BP Neural Network method, which improved the accuracy of the model, and obtained a more comprehensive, objective, independent and measurable employment quality monitoring and evaluation results.

Keywords: Employment quality, Wuhan Business University, AHP, BP Neural Network

1. Introduction

Employment is the foundation of people’s livelihood [1]. Since the implementation of the college enrolment expansion policy, China's higher education has moved from elite education to mass education. The employment problem of college students has gradually become the focus of social attention and the hot issue in the field of higher education research. At present, the society and universities pay more attention to the employment of college students on the quantitative indicator of “employment rate”. The definition of “employment rate” is clear and the calculation is simple, which can explain the market supply and demand relationship of some colleges and universities and some professional graduates at a certain level, and can also reflect the talent training situation of colleges and universities. However, a single concept of “employment rate” cannot reflect the employment level, employment structure, job matching and other complex employment status of graduates, nor can it comprehensively and objectively reflect the quality of education and teaching in colleges and universities [2]. Therefore, it is extremely important to explore and establish a comprehensive, scientific and reasonable “employment quality” evaluation system for college graduates.

Wuhan Business University can date back to the Wuhan Service School founded in 1963. In 1985, with the approval of the former National Education Commission, Wuhan Institute of Business Services was established at the junior college level [3]. In 2004, with the approval of the People's Government of Hubei Province, Wuhan Institute of Economic Management Cadres, Wuhan Institute of Staff Finance and Economics and Wuhan University of Staff and Workers were merged into Wuhan Institute of Business Services. In 2021, the Enrollment and Employment Office of Wuhan University of Commerce prepared the *Report on the Employment Quality of Undergraduate Graduates of Wuhan Business University*, which comprehensively described the implementation rate, employment flow and further education of our undergraduate graduates. However, the above-mentioned report focuses more on the collection of basic information, lacking comprehensive evaluation of employment quality and in-depth learning training. This paper takes "employment quality" as the goal, comprehensively considers the employment level, labor relations, personal development and other factors of graduates, and constructs an evaluation index system for employment quality of college graduates. Through AHP, the evaluation values calculated by MATLAB are used as training samples and test samples to train the BP neural network, and more accurate evaluation results are obtained. The example proves that this method can comprehensively, objectively and scientifically evaluate the employment quality of college graduates [4].

2. Analytic Hierarchy Process and BP Neural Network

2.1 Analytic Hierarchy Process

In the 1970s, the American scholar Saaty. T.L. proposed the Analytic Hierarchy Process (AHP) [5]. According to the nature of the problem and the expected overall goal, the analytic hierarchy process decomposes the problem into several

International Conference on Computer Graphics, Artificial Intelligence, and Data Processing (ICCAID 2022)
edited by Ruishi Liang, Jing Wang, Proc. of SPIE Vol. 12604, 126042B
© 2023 SPIE · 0277-786X · doi: 10.1117/12.2674659

different constituent factors, compares them in pairs, aggregates them according to their relevance at different levels, and finally creates a multi-level analysis model, which provides a theoretical basis for selecting the optimal solution. Its operating principle is to treat all complex problems in a unified way, take them as a whole system, sort out and analyze all relevant factors in the system in detail, sort out the relationship between different factors, and then invite industry experts to score the importance of these influencing factors, calculate the weight of each factor and rank it. Finally, the analysis and planning are carried out according to the weight results, and a more scientific and objective decision is made on this basis. When applying the analytic hierarchy process to analyze and make decisions on problems, the first step is to hierarchize the problems to be solved, so as to create a hierarchical and clear structural model [6]. Through the use of AHP, the qualitative problem can be transformed into a quantitative analysis problem, because an important part of it is to build a judgment matrix. We get the maximum eigenvalue corresponding to judgment λ_{max} of matrix C to calculate the feature vector, and obtains the ranking weight value of the relative importance of each corresponding indicator at the same level to an indicator at the upper level through normalization calculation. This calculation process is called hierarchical single ranking. The judgment matrix is the basis for the final weight calculation, so the matrix must be generally consistent. It is necessary to check the consistency of the filled data to avoid data distortion.

2.2 BP Neural Network

Neural network is a mathematical calculation model designed to imitate the function of human brain. It has high nonlinear processing ability, good fault tolerance and associative memory functions, and super adaptive learning ability. It is widely used in pattern recognition, data mining, intelligent control, combination optimization and other fields [7]. BP neural network has become one of the most widely used neural network models due to its simple structure, convenient use and excellent function approximation ability. BP neural network is the most widely used artificial neural network at present. The basic idea of BP neural network is the forward calculation of signals and the reverse transmission of errors. At present, the three-layer BP neural network is widely used. The three-layer network has strong enough functions. Therefore, this paper adopts a three-layer BP neural network that includes three parts: the input layer, the hidden layer, and the output layer. Compared with other prediction models, BP neural network model has significant characteristics, strong nonlinear mapping ability, high self-learning and adaptive ability, ability to apply existing learning achievements to new knowledge, and strong fault tolerance. The model mainly includes one input layer and one output layer, including one or more hidden layers. Each layer has a certain number of neurons. For example, the number of neurons in the input layer corresponds to the number of each factor of the target. The number of hidden layers is obtained by a specific calculation formula, and the number of neurons in the output layer corresponds to the number of prediction targets. There are weights and thresholds for neurons between layers, and the specific calculation formula is referred to the literature [8]. The neural network can be used for prediction because the model can constantly change its connected weights and topology through the learning of samples, so that the output of the network gradually approaches the set target output. When the model reaches the target output after learning, the model can be used for subsequent prediction [9].

3. Determination of indicator system of employment quality based on AHP

3.1 Indicators determination

Through questionnaire and expert discussion, combined with years of practical experience in the employment of college graduates, we selected five major first level indicators, including job search process analysis, salary level, employment suitability and career development, job satisfaction, and employer evaluation, as the main factors to evaluate the employment status of college graduates, which is shown in Figure 1.

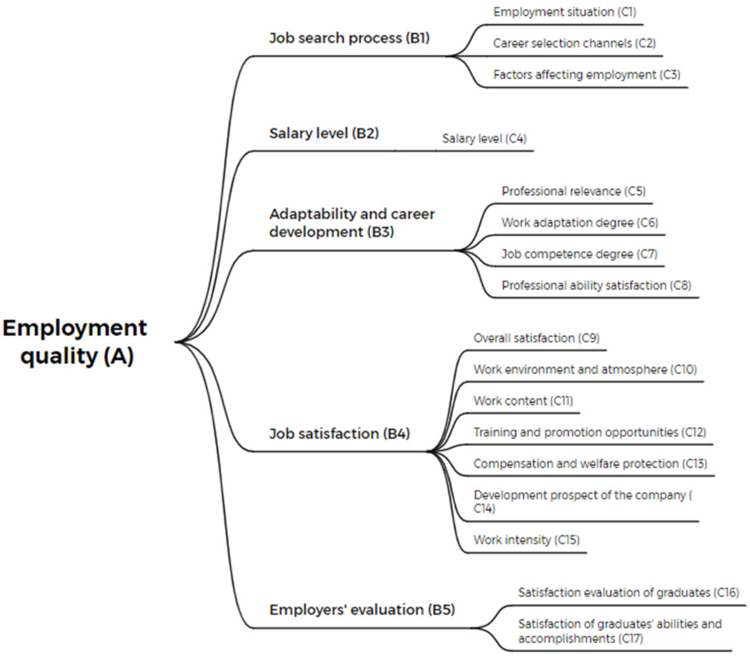

Employment quality (A)
Job search process (B1)
Employment situation (C1)
Career selection channels (C2)
Factors affecting employment (C3)
Salary level (B2)
Salary level (C4)
Adaptability and career development (B3)
Professional relevance (C5)
Work adaptation degree (C6)
Job competence degree (C7)
Professional ability satisfaction (C8)
Job satisfaction (B4)
Overall satisfaction (C9)
Work environment and atmosphere (C10)
Work content (C11)
Training and promotion opportunities (C12)
Compensation and welfare protection (C13)
Development prospect of the company (C14)
Work intensity (C15)
Employers' evaluation (B5)
Satisfaction evaluation of graduates (C16)
Satisfaction of graduates' abilities and accomplishments (C17)

Figure 1. Indicator system of employment quality

Job search process: to understand the evaluation of graduates on the employment situation of their majors in the process of job selection, 54.68% of them believe that their majors have strong advantages and competitiveness in the process of job seeking, and that the general proportion of employment opportunities in their majors is 35.45%, while only 9.87% of them believe that the employment opportunities in their majors are relatively few and very few. It can be seen that graduates' professional skills and soft skills are relatively important for current jobs. It can be seen that graduates' professional skills and soft skills are relatively important for current jobs. In view of the lack of job seeking skills and narrow access to professional counterpart information faced by graduates in the process of job selection, the university has actively integrated effective resources and provided targeted services to help graduates find jobs successfully by using the internet+ employment model.

Salary level: The salary level of graduates refers to the salary level given by the employer. This research investigates and calculates the current monthly income level of graduates whose graduation destination is domestic employment based on their pre-tax monthly income (including wages, benefits that can be converted into cash, etc.).

Adaptability and career development: The evaluation dimensions of the correlation between graduates' employment positions and their majors include very relevant, relatively relevant, average, relatively unrelated, and very unrelated. The degree of professional correlation is the proportion of the number of people who choose "very relevant", "relatively relevant", and "average" to the total number of people in this question. In addition, give 1-5 points to the feedback of the graduates, and calculate the average value. The evaluation dimensions of graduates' adaptability to the current job and their satisfaction with their comprehensive quality and professional skills to the current job needs are divided into five grades. The comprehensive quality and professional skills of the graduates match the social needs. 97.83% of the graduates believe that their ability and quality meet the current job needs at an average level or above, with an average score of 3.99 (5-point system), and tend to be "relatively satisfied".

Job satisfaction: Graduates' satisfaction with the current work unit is mainly composed of three aspects: first, the graduates' evaluation of the actual employment situation, including their satisfaction with the current work remuneration and welfare security, work environment and atmosphere, work content and work intensity; The second is the expected evaluation of the future development, including the satisfaction with the training or promotion opportunities and the development prospects of the unit (post); The third is the overall evaluation of the employment situation, that is, the overall satisfaction with the current job. Satisfaction evaluation dimensions include very satisfied, relatively satisfied, average, relatively dissatisfied and very dissatisfied. Job satisfaction refers to the proportion of the number of people who choose "very satisfied", "relatively satisfied" and "average" to the total number of people in this question.

Employers' evaluation: Putting the main body of quality measurement of school graduates on the employers can more truly reflect the quality of graduates, and further more comprehensively reflect the problems in the process of talent training of schools. Therefore, the establishment of an external evaluation system for the quality of graduates is of positive significance to the improvement and perfection of the school's talent training model. As far as the current job needs are concerned, the employer's evaluation of the satisfaction of the graduates' professional abilities is above 4.20 (5-point system), which is at the "relatively satisfied" level.

3.2 Weight determination

We use the expert interview method and expert questionnaire survey form, and at the same time refer to some literature, use the discrete scale of 1 to 9 to judge the relative neutrality of each indicator, and take the average value of the judgment value to construct a judgment matrix B [11]. Matrix B represents the relative importance of Ci and Cj relative to employment quality (A). For the N monitoring indicators of each layer, it is necessary to establish a n * n paired judgment matrix, as shown in Table 2. Usually b_{ij}=k (k=1,2,..., 9), then b_{ji}=1/k. Secondly, each element b_{ij} represents the relative comparison between the abscissa index Ci and the ordinate index Cj, and the result of their importance is expressed by 1~9 or its reciprocal. Based on the above model construction principle, on the basis of expert discussion and comprehensive analysis of the questionnaire, first compare each factor in pairs, and establish the evaluation index judgment matrix of the graduates' employment quality results, as shown in Table 1.

Table 1. Judgement matrix of B-A

Indicator	B1	B2	B3	B4	B5
B1	1	3	5	6	2
B2	1/3	1	3	2	3
B3	1/5	1/3	1	3	1
B4	1/6	1/2	1/3	1	1/3
B5	1/2	1/3	1	3	1

The maximum eigenvalue and consistency test index value of the corresponding judgment matrix are obtained through MATLAB software. The solution method is: input [V, D]=eig (A) in MATLAB software to obtain the eigenvector value ω_0 and λ_{max}.

The complete code is:

```
clc;clear;

% [V,D] = eig(A)

A = [ 1      3       5       6       2
 1/31        3       2       3
 1/5 1/3     1       3       1
 1/6 1/2      1/3    1        1/3
 1/2 1/3     1       3       1;

[V,D] = eig(A)
```

We have the result ω_0=0.1813 and λ_{max}= 4.0235.

The feature vector is normalized into the weight set of each index. After the weight is calculated, consistency test shall be carried out to determine whether the result is good. Among them, the consistency indicator CI=(λ_{max}-n)/(n-1). In addition, the consistency of the judgment matrix also has randomness, which can be represented by the average random consistency index RI. The value of RI is related to the size of matrix dimension. The average random one-time index of 1~9 order repeated calculations 1000 times obtained by querying relevant literature [12]. After querying the RI indicators of the corresponding dimensions of the matrix, calculate CR as the indicator to measure the consistency of the judgment matrix, that is, CR=CI/RI. It is generally believed that when CR $<$ 0.1, the consistency of the judgment matrix is acceptable, otherwise it should be corrected appropriately. The weights of B1, B2, B3, B4 and B5 relative to A are 0.1571, 0.2841, 0.2546, 0.1996 and 0.1046 respectively.

Similarly, we can calculate the following weights.

Table 2. Weights of C1, C2 and C3 relative to B1

Indicator	C1	C2	C3
Weight	0.38	0.25	0.37

Table 3. Weights of C5, C6, C7 and C8 relative to B3

Indicator	C5	C6	C7	C8
Weight	0.18	0.34	0.27	0.21

Table 4. Weights of from C9 to C15 relative to B4

Indicator	C9	C10	C11	C12	C13	C14	C15
Weight	0.11	0.11	0.08	0.23	0.15	0.19	0.13

Table 5. Weights of C16 and C17 relative to B5

Indicator	C16	C17
Weight	0.58	0.42

After getting the weight of each index relative to the criterion layer, use MATLAB software for consistency test simulation, and it can be considered that the weight value obtained through these five groups of comparison matrices has credibility.

Table 6. Indicator weights of employment quality based on AHP using MATLAB

Indicator	Weight	C9	0.022
C1	0.060	C10	0.022
C2	0.039	C11	0.016
C3	0.058	C12	0.046
C4	0.284	C13	0.030
C5	0.046	C14	0.038
C6	0.087	C15	0.026
C7	0.069	C16	0.061
C8	0.053	C17	0.044

4. Comprehensive evaluation of employment quality based on BP neural network

4.1 Structural design

The basic idea of applying BP neural network to the employment quality evaluation of graduates of Wuhan Business University is as follows. First, determine the input and output vector, take the normalized value of each evaluation index score as the input vector of the neural network, and the calculated employment quality comprehensive score value is normalized as the output vector of the neural network. Second, design the structure of BP neural network, including the number of network layers, the number of neurons in each layer, the selection of conversion function, etc. Third, train the network with enough samples, Different input matrix can get different output results [13]. Fourth, compare the real output results with the expected output results. If there is an error between the two and it exceeds the predetermined error range, the weights and thresholds between neurons at each layer of the neural network need to be modified until the specified error requirements are met. After the above training and testing, the neural network can be used as an effective evaluation tool to make accurate and effective comprehensive evaluation on other similar samples outside the training sample set. Theoretically, it has been proved that a three-layer BP network containing a hidden layer can approximate a nonlinear system with arbitrary accuracy without limiting the number of hidden layer nodes. Therefore, this paper selects a three-layer BP neural network with only one hidden layer [14].

The input layer of the neural network plays a buffer role, which is responsible for receiving external input signals. Therefore, the number of nodes in the input layer depends on the dimension of the input item. There are 17 evaluation indicators for the employment quality of college graduates. The output of the neural network is the evaluation result of the employment quality of college graduates, which is a single output, so the number of nodes in the output layer. The transformation function adopted by BP neural network must be differentiable everywhere, and the S-type function just meets this requirement, and it has a good nonlinear mapping ability. In addition, after the data of employment quality evaluation index is normalized, the input value and output value are between 0 and 1, so we have adopted logarithmic S-type function for the neuron transformation function of hidden layer and output layer [15].

4.2 Network training

In this paper, the input layer data and output layer data should be standardized first. After normalizing the comprehensive scoring value data of 500 graduates' employment quality, the first 400 data were trained using MATLAB 7.10 neural network toolbox. After 1000 times of learning, they converged to the allowable range. After the training, the trained BP network was tested with the remaining 100 test samples, and the corresponding comprehensive evaluation results of college graduates' employment quality were obtained, with an error of less than 5%, indicating that the network has certain reliability. By using the network model, input the evaluation index value of the graduates, and the employment quality score of the graduates can be automatically given. Since the weight of the function is randomly initialized, each training result is different. In the training process, three functions are randomly used to train 10 times each, and the one with the smallest error between the test group and all training groups is found as the training function. Through training, the training degree of function is ideal, but because the randomness of training directly affects the fitting degree of the function, it is preliminarily determined that the training number is large enough or the training cluster is large enough, and the curve fitting effect is better. Using MATLAB software, after the training program is completed, the figure shown in Figure 13 is obtained, which proves that the whole calculation process of the neural network model is over. Through the selection of training, the smaller mean square error result output can be selected. Results The analysis shows that the training effect is ideal, there is no data with a particularly large degree of dispersion, and the training process is acceptable.

5. Summary and prospect

Based on AHP and BP neural network, the comprehensive evaluation system of employment quality of graduates of Wuhan Business University has acquired the knowledge and experience of evaluation experts by learning from the existing sample models. When it is necessary to evaluate the employment quality of college graduates, as long as the trained BP neural network is input with the corresponding index data matrix, the BP network will reproduce the knowledge and experience of experts and make immediate response. It does not require human intervention, the evaluation accuracy is high, and the evaluation results are objective, fair and scientific. With the increasing attention paid by all sectors of society to the employment quality of college graduates, the comprehensive evaluation system of employment quality of college graduates based on AHP and BP neural network will be greatly applied.

Acknowledgement

This research is the result of the Ministry of Education Industry University Cooperation Collaborative Education Project (No.202101062037).

References

[1] Gevaert J, Van Aerden K, De Moortel D, et al. Employment quality as a health determinant: Empirical evidence for the waged and self-employed[J]. Work and Occupations, 2021, 48(2): 146-183.

[2] Peckham T, Fujishiro K, Hajat A, et al. Evaluating employment quality as a determinant of health in a changing labor market[J]. RSF: The Russell Sage Foundation Journal of the Social Sciences, 2019, 5(4): 258-281.

[3] LaBriola J. Post-prison employment quality and future criminal justice contact[J]. RSF: The Russell Sage Foundation Journal of the Social Sciences, 2020, 6(1): 154-172.

[4] Vanroelen C. Employment quality: An overlooked determinant of workers' health and well-being?[J]. Annals of work exposures and health, 2019, 63(6): 619-623.

[5] Mukhtar U, Zhong Z, Tian B, et al. Does rural–urban migration improve employment quality and household welfare? Evidence from Pakistan[J]. Sustainability, 2018, 10(11): 4281.

[6] Jia C, Zuo J, Lu W. Influence of entrepreneurship education on employment quality and employment willingness[J]. International Journal of Emerging Technologies in Learning (iJET), 2021, 16(16): 65-78.

[7] Xue K, Xu D, Liu S. Social network influences on non-agricultural employment quality for part-time peasants: A case study of Sichuan province, China[J]. Sustainability, 2019, 11(15): 4134.

[8] Dust S B, Rode J C, Arthaud ‐ Day M L, et al. Managing the self ‐ esteem, employment gaps, and employment quality process: The role of facilitation ‐ and understanding ‐ based emotional intelligence[J]. Journal of Organizational Behavior, 2018, 39(5): 680-693.

[9] Lu J, Li X, Wei Y, et al. Research on the Evaluation Index System and Countermeasures of Chongqing Graduates' Employment Quality[C]//3rd International Seminar on Education Innovation and Economic Management (SEIEM 2018). Atlantis Press, 2019: 161-163.

[10] Deng Q. Study on Improvement in Employment Quality of College Students[C]//2020 6th International Conference on Social Science and Higher Education (ICSSHE 2020). Atlantis Press, 2020: 904-912.

[11] Yu D. Employment quality index for the South African labour market[J]. Development Southern Africa, 2020, 37(2): 276-294.

[12] Davidson R, Pacek A, Radcliff B. Public Sector Employment, Quality of Government, and Well-Being: A Global Analysis[J]. International Area Studies Review, 2021, 24(3): 193-204.

[13] Yunna L, Peishan Y, Zhaoyang Z. Research on Employment Quality Development of Talent Ecosystem in Tobacco Industry[J]. Tobacco Regulatory Science, 2021, 7(5): 858-866.

[14] Travaglianti F, Babic A, Hansez I. Relationships between employment quality and intention to quit: Focus on PhD candidates as traditional workers[J]. Studies in Continuing Education, 2018, 40(1): 115-131.

[15] Bloom J, Dorsett P, McLennan V. Investigating employment following spinal cord injury: outcomes, methods, and population demographics[J]. Disability and rehabilitation, 2019, 41(20): 2359-2368.

Size and Defect Detection of Valve Based on Computer Vision

Xiaobo Yan[a], Le Feng[b], Mian Tan[b*], Yuan Yang[b], Yu Zhang[c], Lin Wang[a]
[a]Guizhou Key Laboratory of Pattern Recognition and Intelligent System; [b]Guizhou Minzu University, Guiyang 550025, China; [c]Guiyang Fusheng Intelligent Technology Co., Ltd.
[*] Corresponding author: tanmian@gzmu.edu.cn

ABSTRACT

Engine valve is the core component of the engine, and its quality determines the performance of the engine. In industrial production quality inspection, it is necessary to detect the size of the valve and whether there are defects on the surface. Usually, the quality of the valve is determined by comparing the image of the valve surface with the standard image. However, the existing surface defect detection technology cannot detect the curved surface device. In order to solve this problem, this paper designs a valve size and defect detection method based on computer vision. The experimental results show that the method can quickly and accurately detect the rod diameter, groove radius and surface defects of the valve. The method is practical, robust and real-time.

Keywords: Engine valve, rod diameter and groove radius, defect detection, computer vision

1. INTRODUCTION

Engine valve is an important component of the engine, which function is to input air into the engine and discharge exhaust gas after combustion. The valve is shown in Figure 1. The machining accuracy of valve is directly related to the overall performance and service life of automobile engine. When the gap between the valve rod and the pipeline increases, it may cause the valve leakage, the valve overheating may cause the valve stuck and cannot run. Therefore, the accurate detection of valve quality is the key to ensure the stable operation of the engine.

The detection of valve quality includes the size of the device and the smoothness of the valve surface. Due to the small size of the engine valve, manual detection is not only inefficient, but easy to cause valve scratches. Therefore, designing an automatic detection method to detect valve size and surface defects can greatly improve the production efficiency of the valve.

Figure 1. Engine valve.

Computer vision has made remarkable achievements in many fields. Compared with the traditional manual detection, machine vision detection can detect objects without contacting objects, avoiding damage in the detection process. Reasonable design of machine vision detection algorithm can greatly improve the detection efficiency and accuracy.

The state of engine valve plays an important role in engine performance, and many scholars have made contributions to valve breakdown detection. In order to solve the problems of compact structure, strong noise, instability and nonlinearity of diesel engine vibration signal, Zhao et al. [1] proposed a method based on improved wavelet packet-Mel frequency and convolution neural network. Zhao [2] designed a micro displacement detection platform based on machine vision technology to detect whether the micro displacement of the spool of the automobile solenoid valve after pressurization meets the industry standards. Chen et al. [3, 4] proposed a stacked autoencoder to diagnose whether the diesel engine valve is faulty, this method can adaptively extract features from signals, which solves the tediousness of traditional artificial feature construction. Mohammed [5] used yellow transform to detect engine faults, by acoustic testing to determine which engine valve combustion chamber problems, and early detection of failure. Huang [6] proposed a

International Conference on Computer Graphics, Artificial Intelligence, and Data Processing (ICCAID 2022)
edited by Ruishi Liang, Jing Wang, Proc. of SPIE Vol. 12604, 126042C
© 2023 SPIE · 0277-786X · doi: 10.1117/12.2674674

mathematical model for scattering data optimization. Kang et al. [7] developed an augmented reality application that can detect the screw assembly status of automobile valves and display the position and sequence of screw assembly. In terms of dimensional inspection, Fu et al. [8] describe several methods for measuring multiple axial dimensions of valves. Huang [9] used Prewitt operator image detection technology to detect part size. Gong [10] and others use visual inspection equipment to detect engine valve size images. However, with the improvement of engine valve dimensional accuracy requirements, these methods can no longer meet the current accuracy requirements, and the existing method is to detect in the case of a very standard image taken. It is still a problem to be solved that how to accurately detect the size information of the valve when the shooting image has environmental interference and angle tilt.

Edge detection algorithm is an important algorithm in the size detection of industrial components. He [11] proposed an edge detection method combining LOG operator and Canny operator, this method alleviates the problem of noise suppression and insufficient edge detection ability of traditional Canny operator. Gu [12] proposed an improved oblique edge diameter detection algorithm to alleviate the problems of slow speed and poor stability of Prewitt operator in edge detection. Hough transform [13] is a feature extraction technology in image processing, which is usually used to detect lines in images [14]. Affine transformation is widely used in image processing. Lin [15] designed an automatic image matching method based on affine transformation.

Aiming at the problem that the existing technology can only detect the valve in a very standard shooting environment and the detection accuracy is not high, we design a valve size and defect detection method based on computer vision. This method can accurately detect the size and surface defects of the valve in the presence of interference pixels and angle tilt in the shooting image. This method can be applied to the size and defect detection of most other industrial devices.

We summarize the main work of this paper as follows:

a. Aiming at the problem that it is difficult to accurately extract image edge due to environmental interference in the process of shooting images, an edge detection algorithm based on horizontal filtering is proposed.

b. In order to solve the problem of inclination in image shooting, an image correction based on Hough line is propose, the key points are extracted by horizontal and vertical projections, and the diameter of valve components and groove radius are calculated.

c. Moreover, a method based on edge stitching is proposed, which can detect the surface defects of the valve efficiently and accurately, and display the surface defect information through the thermal diagram of the edge coordinates.

2. PROPOSED METHOD

In this paper, a rapid and accurate detection method of valve size information without manual measurement is proposed, to solve the problem of engine valve which is difficult to detect manually in industrial production. Therefore, we first outline the method structure process in section 2.1. Then, the image correction and denoising are shown in section 2.2, and the measurement process of valve diameter and valve groove radius are given in section 2.3. Finally, in section 2.4, the process of detecting valve surface defects is shown.

2.1 Overview of proposed methods

Figure 2 shows the frame diagram of the proposed method, which is mainly composed of image preprocessing, valve diameter, groove radius measurement and device surface defect detection. The purpose of image preprocessing is to correct the tilted image and filter out the irrelevant pixels in the image to avoid interference with detection, including image graying, binarization, edge detection and Hough line detection. After image preprocessing, valve diameter and the lowest points of groove are obtained by horizontal projection of the image, then the two points of groove edge are obtained by vertical projection of the image. The circular radius of the three key points of the groove is obtained by circular fitting, which is the radius of the device groove. Finally, each edge coordinate point is traversed from top to bottom for the ith frame image, and the coordinate $p(x, y)$ of the point is recorded. Simultaneously, it is converted to a pixel $I(i, y) = y$ of a matrix image M, where i is the frame number of the image, x is the abscissa of the coordinate point, and y-axis is regarded as the pixel value of the point. The thermal diagram was visualized to detect product surface defects.

Proc. of SPIE Vol. 12604 1260401-618

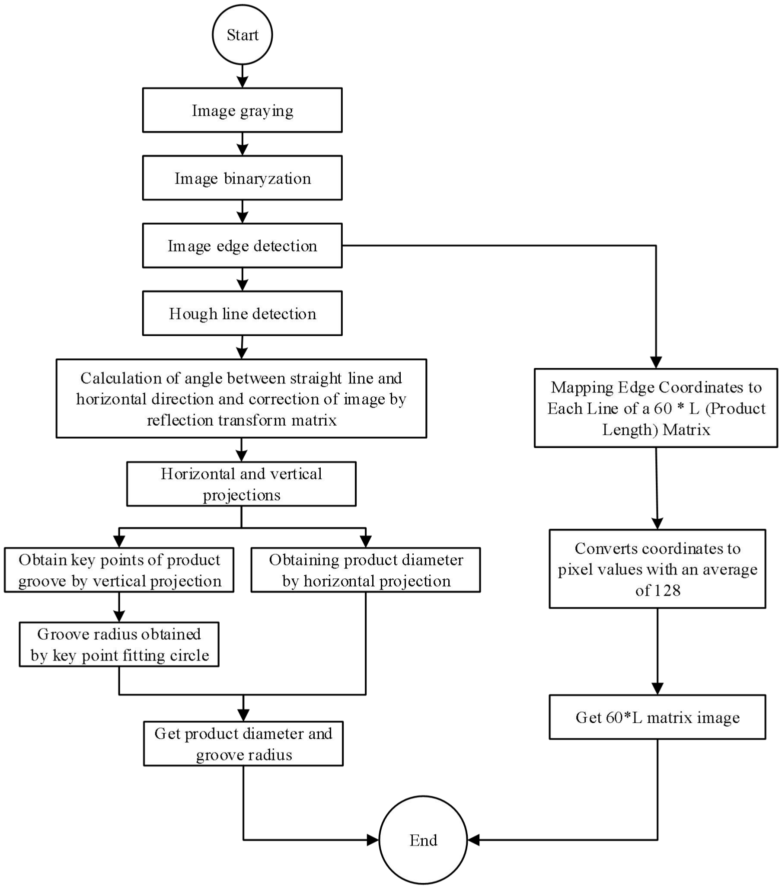

Start
Image graying
Image binaryzation
Image edge detection
Hough line detection
Calculation of angle between straight line and horizontal direction and correction of image by reflection transform matrix
Horizontal and vertical projections
Obtain key points of product groove by vertical projection
Obtaining product diameter by horizontal projection
Groove radius obtained by key point fitting circle
Get product diameter and groove radius
Mapping Edge Coordinates to Each Line of a 60 * L (Product Length) Matrix
Converts coordinates to pixel values with an average of 128
Get 60*L matrix image
End

Figure 2. Frame diagram of valve size and defect detection method based on computer vision.

2.2 Image correction and denoising

In the detection process, due to the factors such as the angle tilt and dirty points of the input image, the detection accuracy will be seriously affected. Therefore, it is necessary to preprocess the image to filter out some dirty points and correct the tilted image.

2.2.1 Image graying and binarization

In order to facilitate image processing, the valve image is first converted into gray image, and the image is divided into object part and background part after binarization. After processing, the pixels in the range of each pixel value [0, 127] are mapped to 0, and the pixels in the range of pixel value [128, 255] are mapped to 255. After binarization, 255 represents the background, and 0 represents the object part. Formula 2-1 shows the process.

$$\begin{bmatrix} 0 & 8 & 189 & 208 \\ 1 & 35 & 205 & 253 \\ 5 & 10 & 214 & 246 \\ 5 & 17 & 53 & 251 \end{bmatrix} \rightarrow \begin{bmatrix} 0 & 0 & 255 & 255 \\ 0 & 0 & 255 & 255 \\ 0 & 0 & 255 & 255 \\ 0 & 0 & 0 & 255 \end{bmatrix} \tag{1}$$

2.2.2 Image edge detection algorithm

The purpose of edge detection is to extract image straight line information conveniently, and further complete the level correction, we mainly uses Canny algorithm for edge detection.

(1) In order to reduce the influence of noise on edge detection results, the image must be filtered to prevent false detection caused by noise.

(2) In order to eliminate the misunderstanding of vertical continuous points into edges, we use the 2-2, Sobel horizontal operator G_y filter to smooth the image.

(3) In order to detect effective straight lines, only 200 to 600 pixels of straight lines are detected by parameter specification.

$$G_x = \begin{bmatrix} -1 & 0 & +1 \\ -2 & 0 & +2 \\ -1 & 0 & +1 \end{bmatrix} * A, \quad G_y = \begin{bmatrix} +1 & +2 & +1 \\ 0 & 0 & 0 \\ -1 & -2 & -1 \end{bmatrix} * A \tag{2}$$

Hough line detection is a feature extraction technology in image processing, which is usually used to detect geometric lines in images. Multiple linear coordinate points can be obtained by Hough detection, and the longest line is taken as the edge line of the device in the experiment. Moreover, the formula 2-3 is used to calculate the tilt angle between the line segment and the horizontal direction, and the affine transformation matrix T is calculated by triangular rotation, as shown in Figure 3.

$$\alpha = \frac{\frac{L_{13}}{L_{12}} * 180}{\pi} \tag{3}$$

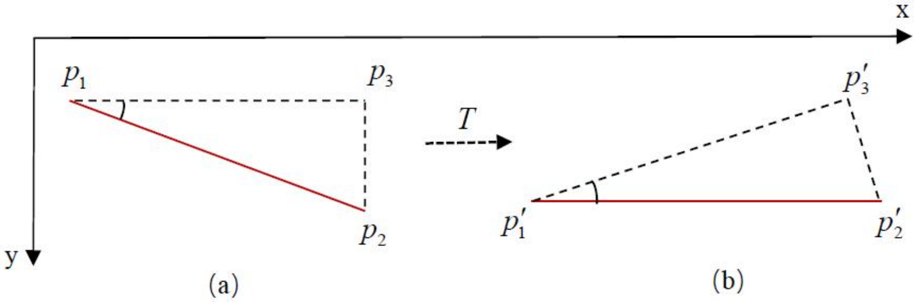


Figure 3. The affine transformation process.

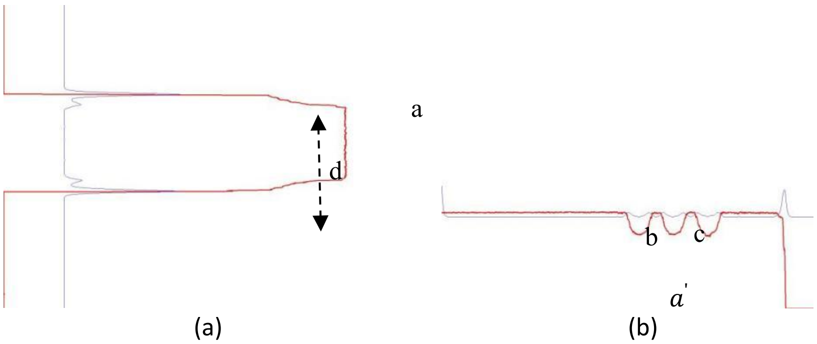
Here L_{13} denotes the length of point p_1 to point p_3, after obtaining the matrix T, the image is rotated by the affine transformation matrix T, and the affine transformation matrix T is like 2-4. Here θ represents the angle of clockwise rotation, and the image is corrected by the affine transformation matrix.

$$T = \begin{bmatrix} \cos\theta & -\sin\theta & 0 \\ \sin\theta & \cos\theta & 0 \\ 0 & 0 & 1 \end{bmatrix} \tag{4}$$

2.3 Detection of valve diameter and groove radius

2.3.1 Horizontal and vertical projections

The specific method of horizontal projection is to count the pixels whose pixel value is 0, and find the distance between the two points corresponding to the curve rising trend in the horizontal projection histogram as the valve diameter d. Meanwhile, the lowest a key point ordinate of the groove is obtained at the projection inflection point, that is, the a' key point in Figure 4(b). Similarly, on the vertical projection, the pixels with pixel value of 0 are vertically counted, and the coordinates of two key points b and c of the valve groove are obtained on the vertical projection. According to the fitting circle of the key points, the radius can be obtained. Figure 4 shows the horizontal and vertical projection processes. The red curve is the histogram of the horizontal and vertical direction of the image. The purple curve is obtained by calculating the second derivative of the histogram, and the peak value is the position of the maximum change rate of the histogram.



(a) (b)

Figure 4. Horizontal and vertical projection.

2.4 Surface defect detection

In the process of defect detection, the rate of the camera is 60f/s, corresponding to 60 device images per second. We propose an innovative image edge information stitching algorithm, which transforms the edge ordinate of each frame into a row in the matrix**.** Among them, the 60-frame edge information corresponds to a 60*L matrix M, where L is the length of the device, such as a coordinate point $p(x, y)$ at the edge of the i-frame image, mapped to the corresponding position $I(i, y) = y$ in the matrix M, as shown in formula 2-6.

$$M = \begin{bmatrix} y_{(1,1)} & \cdots & y_{(1,j)} & \cdots & y_{(1,L)} \\ \vdots & & & & \vdots \\ y_{(i,1)} & \cdots & y_{(i,j)} & \cdots & y_{(i,1)} \\ \vdots & & & & \vdots \\ y_{(60,1)} & \cdots & y_{(60,j)} & \cdots & y_{(60,L)} \end{bmatrix} \tag{5}$$

In order to show the product surface defects more clearly and intuitively, we map the matrix M to a matrix M' with an average of 128 per row, and then use the thermal diagram visualization to reflect the defects. The specific mapping process is shown in formula 2–7.

$$M' = M - \begin{bmatrix} m_1 \\ \vdots \\ m_{60} \end{bmatrix} + 128 \tag{6}$$

Where, m_i denotes the crowed number of the i row. When drawing M' into a thermal diagram, we use the COLORMAP_JET color corresponding to ColormapTypes provided in OpenCV. The deeper the blue is, the deeper the depression on the device surface is, and the deeper the red is, the more prominent the device surface is.

3. EXPERIMENTAL ENVIRONMENT AND RESULTS

3.1 Experimental equipment

To see the formats available with this manuscript, go to the Format menu and choose "Styles and Formatting". To view which style is being used in any part of this document, place your cursor on the line and look in the Styles and Formatting display. The experimental environment and equipment used in this paper are shown in Table 1. The industrial camera used in this paper is Basler acA1280-60gc GigE, 60 frames per second. The lens type used is XF-5MDT03X11, the background light is blue, and the resolution of the photograph is 16um, as shown in Figure 4 and Figure 5. In the detection, the turntable is needed to drive the valve to rotate to obtain the image of each frame.

Table 1. Experimental environment.

software and hardware	configure
operating system	Microsoft Windows 10
CPU	Intel(R) i7-9770 3.4GHz
RAM	16GB
development environment	VC++
Development tools	Visual Studio 2019
Industrial cameras and lenses	Basler acA1280-60gc GigE XF-5MDT03X11

Figure 5. Industrial cameras and lenses used.

Figure 6. Matching camera lens.

3.2 Analysis and discussion of experimental results

3.2.1 Results analysis of image correction and denoising

The original image captured by the camera is shown in Figure 7. It can be seen that the product in the original image has a certain inclination angle and the edge is blurred.

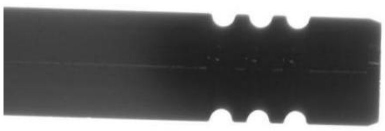
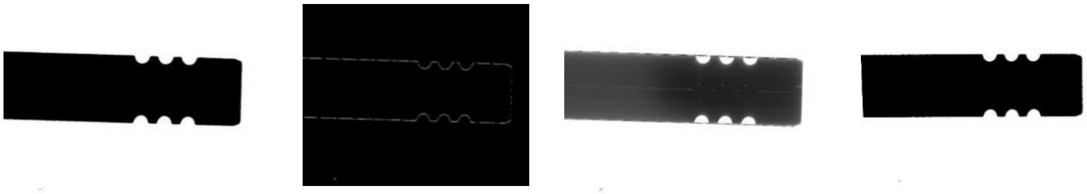


Figure 7. Original image.

Figure 8 shows the results of the original image after binarization, edge detection, Hough line detection and affine transformation. It can be seen that after binarization, the edge contour of the image is clearer, as shown in Figure 8(a). After edge detection, the edge of the image is clearly displayed, such as Figure 8(b), and the interference pixels are filtered out. The edge of the device is represented by a straight line through Hough line detection, as shown in Figure (c). Since the original image is tilted, the line detected by Hough line is still tilted. The image corrected by affine transformation is shown in Figure 8(d). It can be seen that the image has becomes horizontal after affine transformation.



a. Binarization b. Edge detection c. Hough line detection d. Affine transformation

Figure 8. Results of image binarization, edge detection, Hough line detection and affine transformation.

3.2.2 Analysis of valve diameter and groove radius detection results

The results of valve diameter and groove radius are shown in Figure 9. The diameter is 325, and the groove radius is 37. The length unit here is represented by pixels, which needs to be converted into actual diameter and groove radius according to the resolution of the camera.

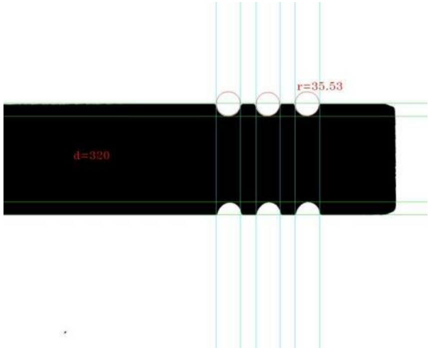


Figure 9. Detection results of valve rod diameter and groove radius.

3.2.3 Analysis of valve surface defect detection results

The detection results of valve surface defects are shown in Figure 10. Each small point in the figure represents the defects in the corresponding position. The dark blue point corresponds to the deep position of the depression, and the dark red point corresponds to the prominent position. The three dark blue positions represent the three groove positions of the valve. The color of most positions in the figure is uniform, only a few points represent several defects of the valve.

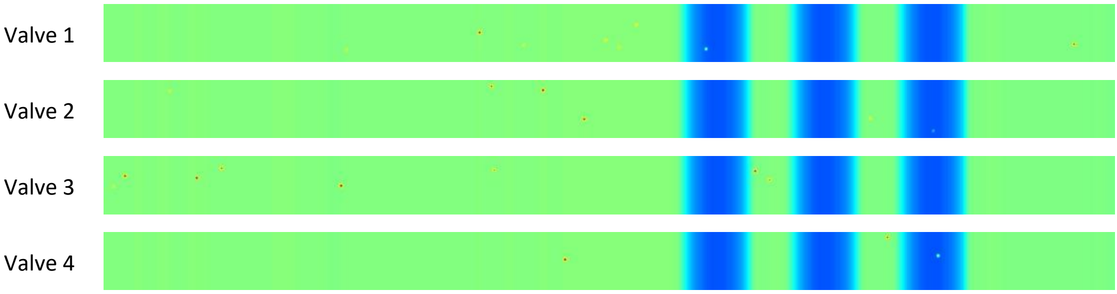


Figure 10. Thermal diagram of surface defect detection.

4. CONCLUSION

In this paper, an automatic detection method is designed for the size of the detection valve and the surface smoothness in industrial quality inspection. The edge of the device image is detected by image preprocessing, and the accurate device size is obtained by horizontal and vertical projection of the corrected image. In order to detect the defects of the product surface, we propose an edge information matrix image detection method, and takes the detection of rod diameter, groove radius and surface defects of engine valve in industrial production as an example. The experimental results show that the proposed method can accurately detect the diameter of the valve and the size of the groove radius, and can effectively detect the defects of the valve surface.

Acknowledgements

This work supported in part by the growth project of young scientific and technological talents in general colleges and universities in Guizhou Province (no.GZMUZK2022YB117), Guizhou Province Higher Education Teaching Content and Curriculum System Reform Project (no.2020053), the Natural Science Foundation of Guizhou Province (no. QKHJCZK2022YB195), the Natural Science Foundation of Guizhou Minzu University (no. GZMUZK2021YB24), the Youth Science and Technology Talents Cultivating Object of Guizhou Province (no. QJHKY2021104).

REFERENCES

[1] H. Zhao, Z. Mao, K. Chen, and Z. Jiang, "An intelligent fault diagnosis method for a diesel engine valve based on improved wavelet packet-Mel frequency and convolutional neural network," 2019 International Conference on Sensing, Diagnostics, Prognostics, and Control (SDPC), 354-359 (2019).

[2] S. Zhao, T. Huang, and S. Lv, "Design and Realization of Valve Core Micro-Displacement Detection System Based on Machine Vision for Solenoid Valve," 2018 2nd IEEE Advanced Information Management, Communicates, Electronic and Automation Control Conference (IMCEC), 967-970 (2018).

[3] K. Chen, Z. Mao, H. Zhao, Z. Jiang, and J. J. S. Zhang, "A variational stacked autoencoder with harmony search optimizer for valve train fault diagnosis of diesel engine," Sensors, 223 (2019).

[4] K. Chen, Z. Mao, H. Zhao, and J. Zhang, "Valve fault diagnosis of internal combustion engine based on an improved stacked autoencoder," 2019 International Conference on Sensing, Diagnostics, Prognostics, and Control (SDPC), 295-300 (2019)

[5] A. A. J. A. A. Mohammed, "Performance analysis of variable valve timing engine to detect some engine faults by using Hilbert Huang transform," Applied Acoustics, 108775 (2022).

[6] H. Huang, F. Feng, S. Huang, L. Chen, and Z. Hao, "Microscale Searching Algorithm for Coupling Matrix Optimization of Automated Microwave Filter Tuning," IEEE Transactions on Cybernetics, (2022).

[7] M. H. Kang, "Detecting the screw-assembly state of a valve-body using the AR method," Journal of the Korea Academia-Industrial cooperation Society, 24-30 (2021).

[8] W. L. Fu, J. S. Lan, D. P. Lin, J. S. Mo, J. S. Xiao and S. L. Lian, "Design and Application Analysis of Comprehensive Inspection Tool for Engine Valve Axial Dimension," electromechanical engineering technology, 94-96 (2006).

[9] J. X. Huang, D. Xu, P. Y. Jiang, W. K. Huang and Lin., "Research on a high precision image detection method for shaft sleeve parts size," Optics & Optoelectronic Technology, 85-87 (2008).

[10] C. Gong, "The Method and Realization of High-Precision Size Measurement Based on Machine Vision," Guangdong University of Technology (2014).

[11] Q. He and Y, "Algorithm of Edge Detection Based on LOG and Canny Operator," Computer Engineering, 210-212 (2011)

[12] Y. C. Gu, D. Xu, and J, "New high-precision taper surface edge detection algorithm based on prewitt operator," Computer Engineering and Applications, 201-203 (2013).

[13] R. O. Duda and P. E. J. C. o. t. A. Hart, "Use of Hough Transformation to Detect Lines and Curves in Pictures," CACM, 11-15 (1972).

[14] T. Nguyen, X. Pham, and J. Jeon, "An improvement of the Standard Hough Transform to detect line segments," IEEE International Conference on Industrial Technology, 1-6 (2008).

[15] H. Lin, P. Du, W. Zhao, L. Zhang, and H. Sun, "Image registration based on corner detection and affine transformation," 2010 3rd International Congress on Image and Signal Processing, 2184-2188 (2010).

Real option analysis in traffic network design with intelligent algorithm

Xinxin Yu[a], Shu Cui[a], Yu Wang[a], Yi Liu[a], Xing Yang[a], Di Wu[a], Heling Liu[b], Peng Zhang[a*]
[a]Transport Planning and Research Institute, Ministry of Transport, Beijing, PRC 01057802855;
[b]Beijing 101 middle school, Beijing PRC, 01051633235
[*]Corresponding author: 865404846@qq.com

ABSTRACT

In order to ensure the effective use of funds, the government must carry out reasonable traffic planning, in which traffic network design is one of the core contents of traffic planning. The traditional transportation planning has unreasonable factors due to deterministic assumptions. This paper assumes that the demand is a random variable, and then considers the time factor. With the cost recovery and link update as the constraint conditions, the real option is used to solve the problem of the flexibility value of the optimization strategy. The optimization model under uncertainty considering the time factor is given, and the real option is solved by using the LSM method. Genetic algorithm with Monte Carlo is used for network design. The example analysis shows that time factor has a significant impact on network construction decision-making, and real options can effectively describe the flexibility of network construction decision-making.

Keywords: real option, traffic network design, genetic algorithm, Monte Carlo simulation , demand uncertainty

1. INTRODUCTION

In order to alleviate traffic congestion, increasing the intensity of traffic infrastructure investment is an important means to solve the contradiction between supply and demand of transportation system. Construction funds are always relatively insufficient in the actual construction process. In order to improve the efficiency and efficiency of traffic investment and utilization, it is necessary to give a method to optimize the traffic network under limited capital investment. Traditional traffic planning only provides a static construction scheme under given conditions, and does not consider the possibility of planning scheme adjustment after future external environment changes. There is flexibility in network construction, which is not considered in the traditional network design.

For the flexibility of network construction, Morlok and Chang (2004) pointed out that the flexibility of network construction refers to decision maker can change and adjust construction strategies according to the actual situation of system operation. They also point out that the actual changes include changes in requirements, changes in network performance, changes in network fees or other important resources. [1] Kauffman and Kumar (2008) studied the value of real options contained in the network design problem under uncertainty. [2] Zhao, Sundararajan and Tseng (2004) studied the real option problem in highway network construction, and proposed to use the least squares Monte Carlo simulation method to solve the real option value. [3] Chow and Regan (2011) proposed a real option value calculation method in network design, and gave an example. [4]

This paper explores the use of dynamic network optimization strategy with real options under the uncertainty of demand. It is assumed that decision maker can dynamically change the planning scheme according to the changes in internal and external conditions in the future, so as to improve the flexibility and adaptability of traffic planning scheme, and provide scientific, systematic and effective decision-making schemes for planning and decision-making departments. It can make the capital investment obtain the best investment income, and make the speed of traffic construction meet the needs of social and economic development.

2. TRAFFIC NETWORK DESIGN MODEL

2.1 Real option

According to the general situation of network construction, network construction includes two kinds of real options: (1) decision maker can defer the construction time according to the actual situation, which is the deferral option; (2) Decision maker can change the construction plan according to the actual situation, which is flexible option. It is assumed

International Conference on Computer Graphics, Artificial Intelligence, and Data Processing (ICCAID 2022)
edited by Ruishi Liang, Jing Wang, Proc. of SPIE Vol. 12604, 126042D
© 2023 SPIE · 0277-786X · doi: 10.1117/12.2674516

that when the construction sequence is determined, the decision maker can only change the progress of the implementation plan according to the future development rather than the construction plan. In this case, the objective function with deferral option can be given as [5]

$$Z_{t_n}(Q_{rs}^{t_n}) = \max\{\pi_{t_n}(\phi(Q_{rs}^{t_n}, y_{at_0})), (1+r)^{-\Delta t} E[Z_{t_{n+1}}(Q_{rs}^{t_{n+1}})]\} \tag{1}$$

where y_{at_0} is the construction sequence scheme given at the beginning of the period, $\phi(Q_{rs}^{t_n}, y_{at_0})$ represents the total cost of each period of the network under the construction sequence given at the beginning of the period, and π_{t_n} represents NPV. The specific expression is

$$\phi(Q_{rs}^{t_n}, y_{at_0}) = E[\sum_a X_a^{t_n} T_a^{t_n}(X_a^{t_n}, y_{at_0}\overline{C}_{at_n})] = \sum_\omega p^{\omega t}[\sum_a X_a^{\omega t} T_a^{\omega t}(X_a^{\omega t}, y_{at}\overline{C}_{at})]] \tag{2}$$

$$\pi_{t_n}(\phi(Q_{rs}^{t_n}, y_{at_0})) = \sum_t^T (1+r)^{-t}\vartheta(\phi(Q_{rs}^{t_n}, 0) - \phi(Q_{rs}^{t_n}, y_{at_0})) - \sum_t \frac{G_{at}}{(1+r)^{t-1}} \tag{3}$$

The first term of formula (3) is the present value of the total income in the period, and the second term is the present value of the investment cost. Q_{rs} is demand between OD pair *rs*, X_a is the traffic flow on link *a*, T_a is travel time on link *a*, G_{at} is construction costs of government in the *t*-year, r is requested rate of return, ω is any random event, the probability is p^{ω}.

Assuming that when the construction sequence is determined, the decision maker can change the scheme and event of the hypothetical sequence according to the actual situation, and in this case, the objective function based on flexible option can be given as

$$Z_{t_n}(Q_{rs}^{t_n}) = \max\{\pi_{t_n}(\phi(Q_{rs}^{t_n}, y_{at_n})), (1+r)^{-\Delta t} E[Z_{t_n+1}(Q_{rs}^{t_n+1})]\} \tag{4}$$

It is worth noting that the difference between formula (1) and formula (4) is that under the deferral option, the construction sequence y_{at_0} is designed at the beginning of the period, and under the flexible option, the construction sequence y_{at_n} can be changed at t_n according to the actual condition.

When there is only one stage, it is the net present value under the static strategy, and the objective function becomes

$$Z_{t_n}(Q_{rs}^{t_n}) = \max\{\pi_{t_n}(\phi(Q_{rs}^{t_n}, y_{at_0})) \tag{5}$$

2.2 Network design model with real option

It is assumed that OD demand is a random variable subject to a certain probability distribution, and the construction period *t* is given. The traffic network design model with deferral options is composed of the upper level model formulas (6) - (10) and the lower level model formulas (11) - (14). The upper level model determines the sequence of the new or extended link sets for decision maker, which maximizes the net present value of the total travel time considering the deferral option. The constraints include cost recovery condition (7) and link update constraint (9) - (10). The lower level model is the user equilibrium corresponding to each demand scenario under the network construction sequence determined by the upper level model.

$$\max \quad Z_{t_n}(Q_{rs}^{t_n}) = \max\{\pi_{t_n}(\phi(Q_{rs}^{t_n}, y_{at_0})), (1+r)^{-\Delta t} E[Z_{t_{n+1}}(Q_{rs}^{t_{n+1}})]\} \tag{6}$$

$$s.t. \quad \sum_t^{t_n} \frac{B_t}{(1+r)^{t-1}} - \sum_t^{t_n} \frac{G_{at}}{(1+r)^{t-1}} \geq 0, \quad \forall t_n \tag{7}$$

$$\overline{C}_{at} \geq 0, \quad \forall a \in \overline{A}, \forall t \tag{8}$$

$$y_{at} \in \{0,1\}, \quad \forall a \in \overline{A}, \quad \forall t \tag{9}$$

$$y_{at_1} \geq y_{at_2}, \quad \forall a \in \overline{A}, \quad \forall t_1 \geq t_2 \tag{10}$$

$$\min \quad T(\mathbf{x}) = \sum_a \int_0^{X_a} T_a\,(w, \overline{C}_a)\,\mathrm{d}w \tag{11}$$

$$s.t. \quad \sum_{k \in P_w} F_k^{rs} = Q_{rs}, \quad \forall r \in R, s \in S \tag{12}$$

$$F_k^{rs} \geq 0, \quad \forall r \in R, s \in S, k \in P_{rs} \tag{13}$$

$$X_a = \sum_{r \in R} \sum_{s \in S} \sum_{k \in P_{rs}} F_k^{rs} \delta_{a,k}^{rs}, \quad \forall a \tag{14}$$

Where F_k^{rs} represents the traffic flow on path *k*, δ_a^{rs} is the parameter, R_t represents revenue from road charges in the *t*-year, B_t represents government funding in the *t*-year. This paper assumes that travel time satisfies BPR function.

When the formula (6) in the network design model with deferral option is replaced by formula (15), the model becomes a network design model with flexible option.

$$\max \quad Z_{t_n}(Q_{rs}^{t_n}) = \max\{\pi_{t_n}(\phi(Q_{rs}^{t_n}, y_{at_n})), (1+r)^{-\Delta t} E[Z_{t_{n+1}}(Q_{rs}^{t_{n+1}})]\} \tag{15}$$

When the formula (6) in the network design model with deferral option is replaced by formula (5), the model becomes a network design model under the static strategy.

3. ALGORITHM

3.1 Solution algorithm of deferral option

For the model with deferral option, the least squares Monte Carlo method is used to solve the network optimization problem. Assuming that each path of Monte Carlo simulation is *ω* ∈ *P*, *P* is the set of all simulation paths.

Step 1 : For each path *ω* ∈ *P* and time state *t*, where $0 \leq n \leq T / \Delta t$, the paths of demand are generated according to the given distribution.

Step 2 : For a given $Q_{rs}^{t_0}$ and simulation results $Q_{rs}^{t_n}(\omega)$, $0 \leq n \leq T / \Delta t$, *ω* ∈ *P*, we use the solution method under the static strategy to calculate the y_{at_0} of the design model of the traffic network conforming to the static strategy, and use the solution method of the net present value function π_{t_n} to calculate $\pi_{t_n}(\phi(Q_{rs}^{t_n}, y_{at_0}))$.

Step 3：Starting from $n = T / \Delta t$, the least squares Monte Carlo method is used to solve the objective function.

Step 3.1：if $n = T / \Delta t$, then let $Z_{t_n}(Q_{rs}^{t_n}(\omega)) = \max(0, \pi_{t_n}(\phi(Q_{rs}^{t_n}, y_{at_0})), 0)$;

Step 3.2：Set the parameter $\theta(\omega, t_n)$ to mark the optimal decision, if $Z_{t_n}(Q_{rs}^{t_n}(\omega)) > 0$, then it means that the option needs to be executed, and then $\theta(\omega, t_n)$ = 1, otherwise $\theta(\omega, t_n)$ = 0;

Step 3.3：Let *n*=*n*-1. If *n*=0，turn to Step 4;

Step 3.4：The least square method with Hermite polynomial is used to estimate $\hat{Z}_{t_{n+1}}$ in the form of

$$H_i(x) = (-1)^i e^{x^2/2} \frac{d^i}{dx^i} e^{-x^2/2} \tag{16}$$

i+1th polynomial can be expressed as $H_{i+1}(x) = xH_i(x) - iH_{i-1}(x)$, and the objective function is obtained by regression which is

$$Z_{t_n}(Q_{rs}^{t_n}(\omega)) = \sum_{i=0}^{\Pi} \beta_i (-1)^i e^{x^2/2} \frac{d^i}{dx^i} e^{-x^2/2} \tag{17}$$

where β_i is the coefficient obtained by least square method, x represents the net present value $\pi_{t_n}(\phi(Q_{rs}^{t_n}(\omega), y_{at_0}))|_{\pi_{t_n}>0}$ on each simulated path;

Step 3.5：Use $\hat{Z}_{t_{n+1}}$ to represent the objective function value. If $(1+r)^{-\Delta t} E[\hat{Z}_{t_{n+1}}] > \pi_{t_n}(\phi(Q_{rs}^{t_n}(\omega), y_{at_0}))$ is satisfied, then the optimal decision is to delay, the option needs to be executed and $\theta(\omega, t_n) = 1$;

Step 3.6：if $n>0$，turn to Step 3.3. Otherwise, get the function value using

$$\max \quad Z_{t_n}(Q_{rs}^{t_n}) = \max\{\pi_{t_n}(\phi(Q_{rs}^{t_n}, y_{at_0})), (1+r)^{-\Delta t} E[Z_{t_{n+1}}(Q_{rs}^{t_{n+1}})]\} \tag{18}$$

Step 4：If $Z_{t_0} > 0$, then options are valuable. If $(1+r)^{-\Delta t} E[\hat{Z}_{t_l}] > \pi_{t_0}(\phi(Q_{rs}^{t_0}(\omega), y_{at_0}))$, then use the right to delay. If $(1+r)^{-\Delta t} E[\hat{Z}_{t_l}] \le \pi_{t_0}(\phi(Q_{rs}^{t_0}(\omega), y_{at_0}))$, then invest immediately.

For models with flexible option, just change y_{at_0} to y_{at_n} .

3.2 Genetic algorithm

In order to solve the discrete traffic network design with real options, this paper uses genetic algorithm with Monte Carlo simulation, the process is shown in Figure 1.

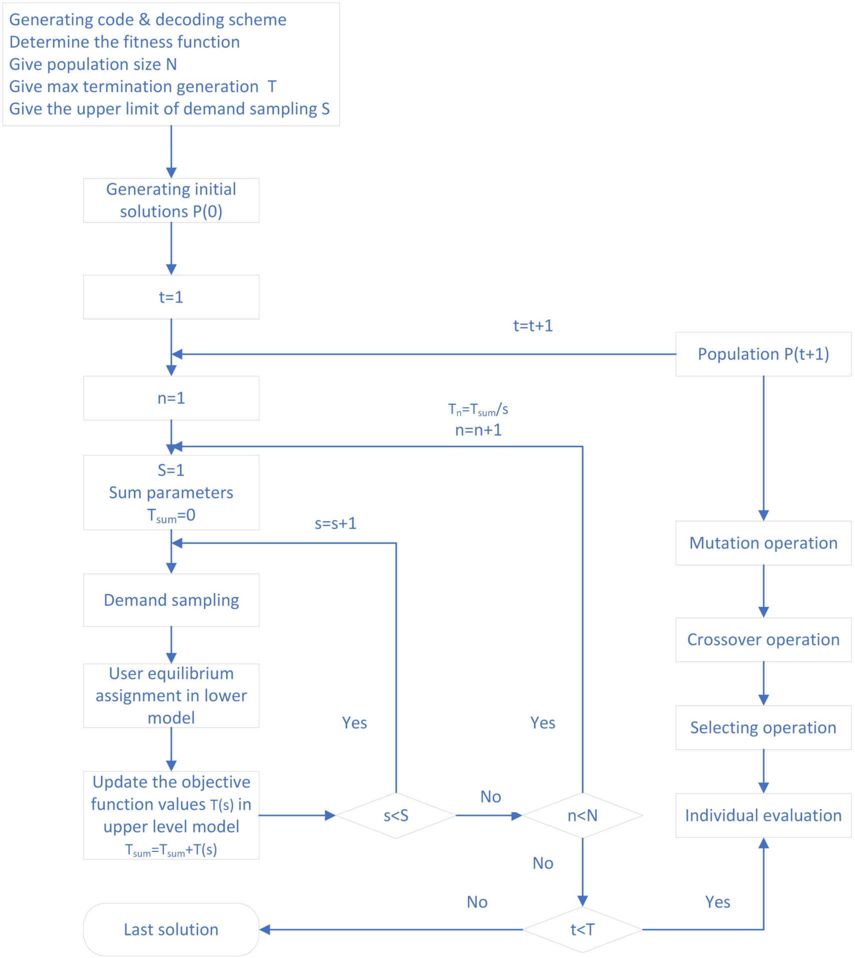

Generating code & decoding scheme
Determine the fitness function
Give population size N
Give max termination generation T
Give the upper limit of demand sampling S
Generating initial solutions P(0)
t=1
t=t+1
Population P(t+1)
n=1
$T_n=T_{sum}/s$
n=n+1
S=1
Sum parameters
$T_{sum}=0$
s=s+1
Mutation operation
Demand sampling
Crossover operation
User equilibrium assignment in lower model
Yes
Yes
Selecting operation
Update the objective function values T(s) in upper level model
$T_{sum}=T_{sum}+T(s)$
No
s<S
n<N
Individual evaluation
No
No
Yes
Last solution
t<T

Figure 1. Genetic algorithm with Monte Carlo method process.

4. CASE STUDY

In this paper, the classic Nguyen-Dupuis network is used for example analysis. As shown in Fig. 2, this network has 13 nodes, 25 links and 4 OD pairs.

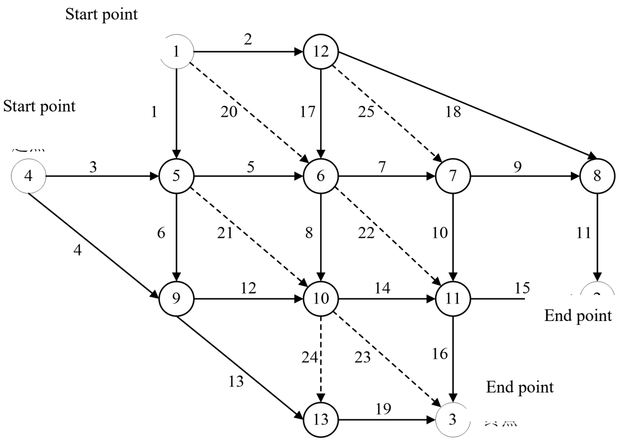


Figure 2. Nguyen-Dupuis network.

Assuming that there are four stages, the demand at time 0 is 250. Demand compliance formula (19).

$$\mathrm{d}\mathbf{Q}_{\mathbf{rs}} = (\boldsymbol{\mu}\mathrm{d}t)^T \mathbf{I}\mathbf{Q}_{\mathbf{rs}} + (\boldsymbol{\sigma}\mathrm{d}W)^T \mathbf{Q}_{\mathbf{rs}} \tag{19}$$

The parameters of genetic algorithm are as follows: evolutionary termination generation is 800, population size is 75, generation gap is 0.85, mutation probability is 0.16, crossover probability is 0.72. In the least squares Monte Carlo algorithm, the number of simulated paths is 150, and the polynomial order is 5. BPR function parameters *α* is 0.15, *β* is 4. The budget levels of the four stages are 200,125,200 and 125 respectively. The values of volatility *σ* are 0, 0.1, 0.2, 0.3, 0.4 and 0.5. *μ* is 0.2, *r* is 0. Fig. 3 shows the calculation results.

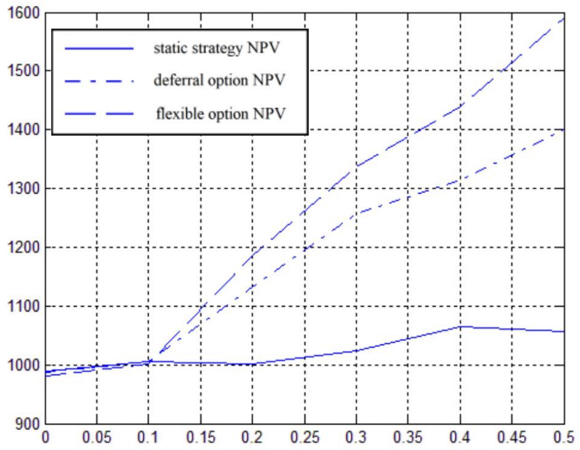


Figure 3. NPV with deferral option and flexible option.

5. CONCLUSION

In this paper, the multi-stage network design problem is considered. Taking cost recovery and link update as constraints, the problem of flexibility value of optimization strategy is solved by using real options, and the optimization model under uncertainty considering time dimension is given. The least squares Monte Carlo method is used to solve the real options. The Nguyen-Dupuis network example shows that: (1) when the volatility of demand is small, the value of the option is small. (2) The value of options increases rapidly with volatility. (3) The value of flexible option is greater than value of deferral option. According to the calculation results, the time factor has a significant impact on the network construction decision, and real options can effectively describe the flexibility of network construction decision and improve the rationality of decision.

REFERENCES

[1] Morlok, E. K. and Chang, D. J., "Measuring capacity flexibility of a transportation system," Transportation Research Part A, 38(6):405-420 (2004).

[2] Kauffman, R. J. and Kumar, A., "Network effects and embedded options: Decision-making under uncertainty for network technology investments," Information Technology and Management, 4 9(3):149-168 (2008).

[3] Zhao, T., Sundararajan, S. K. and Tseng, C. L., "Highway development decision-making under uncertainty: A real options approach," Journal of Infrastructure Systems, 10(1):23-32 (2004).

[4] Chow, J. and Regan, A., "Network-based real option models," Transportation Research Part B, 45:682-695 (2011).

[5] Bahrami S, Roorda M J., "Optimal traffic management policies for mixed human and automated traffic flows," Transportation Research Part A, 135: 130-143(2020).

[6] Soteropouls A, Berger M, Ciari F., "Impacts of automated vehicles on travel behaviour and land use: an international review of modelling studies," Transport Reviews, 39(1): 29-49(2019).

Design of intelligent home control system based on ZigBee technology

LinzhiZhou[1,2], Nanling Zhang*[3,4]
[1]CITI university of Mongolia, Mongolia; [2]Guangzhou Institute of Software, Department of Digital Media, Guangzhou, 510900, China
*[3]CITI university of Mongolia, Mongolia; [4]Guangdong Construction Polytechnic, Computer Application Technology, Guangzhou, 510900, China
*Corresponding author: 25883067@qq.com

Abstract

This paper designs a smart home control system based on ZigBee technology. The author analyzes the system design architecture, software and hardware implementation methods, software system construction, key processing flow and other aspects. This design uses home gateway (PC), pan Coordinator (network coordinator), home appliance sensor to build ZigBee smart home control system, and realizes communication with remote mobile phone terminal through GPRS wireless transceiver system. On the premise of ensuring energy saving, the system can also make the ZigBee wireless network stable during long-distance transmission and meet the remote monitoring needs of smart homes.

Key words: ZigBee; Smart home; Design; GPRS

1. The introduction

Smart home is the product of the rapid development of modern science and technology. [1]On the basis of traditional residential functions, the control system enables people to live in a safer, more convenient, more comfortable and more environmentally friendly home environment through intelligent control of household equipment, and the quality of people's life is improved. Therefore, the value of this paper mainly has two points. First, from the perspective of science and technology, modern smart homes resort to the Internet of things technology, which links the lighting system, air conditioning system, security system, protection system, electrical equipment, audio-visual equipment, curtain control, etc. in the home environment, and provides convenient functions such as remote control, alarm and monitoring. The Internet of things is connected with the existing Internet through various sensors, reflecting the new technology of the information technology industry. [2]Second, in terms of application field, the technology is applied to the field of human clothing, food, housing and transportation, which greatly improves the comfort and safety of human living environment, which is an important embodiment of the humanization of modern science and technology.

Based on the practical significance of the smart home control system, this research uses ZigBee CC2530 Internet of things development platform as the development tool to complete the overall hardware framework and software planning of the smart home control system. In terms of hardware, we will complete the design of the technical parameters of the microcontroller CC2530 and the whole set of peripheral circuits, and introduce in detail the characteristics and key points of the power supply circuit, sensor module circuit, network communication circuit and other hardware circuits. In terms of software, the ZigBee smart home control system is built with PC as the control center, pan Coordinator (network coordinator) and home appliance sensors, and then the control information is converted into control commands, which are transmitted to various home appliances through WiFi for corresponding control.

2. Overall scheme design of the system

2.1. ZigBee technology

The smart home control system adopts the currently popular ZigBee wireless self-organizing network. The main reason is that ZigBee communication technology has the characteristics of low power consumption, high reliability, low cost, small delay, large network capacity, good security, reasonable coverage, good compatibility and simple network formation.[3]GPRS network is a 2.5G mobile communication system, which uses packet switching technology. Its data transmission unit uses cmnet access to avoid the tedious application for fixed IP. It can be easily connected to the Internet by simply inserting a SIM card.[4]

International Conference on Computer Graphics, Artificial Intelligence, and Data Processing (ICCAID 2022)
edited by Ruishi Liang, Jing Wang, Proc. of SPIE Vol. 12604, 126042E
© 2023 SPIE · 0277-786X · doi: 10.1117/12.2674645

2.2. System design idea

Due to the close distance between the nodes, the smart home control system structure of ZigBee Technology (as shown in Fig. 1) is adopted in this paper. It does not need to expand the network coverage through routers, and only requires coordinators and sensor devices to meet the network construction. The coordinator is responsible for initiating and maintaining the network, and forwarding the collected information to the main gateway server, that is, the PC soft gateway. The server makes fusion decisions on the collected information according to the intelligent processing algorithm, and sends request information to the remote mobile phone terminal or response execution commands to the sensor according to the decision results.[5]

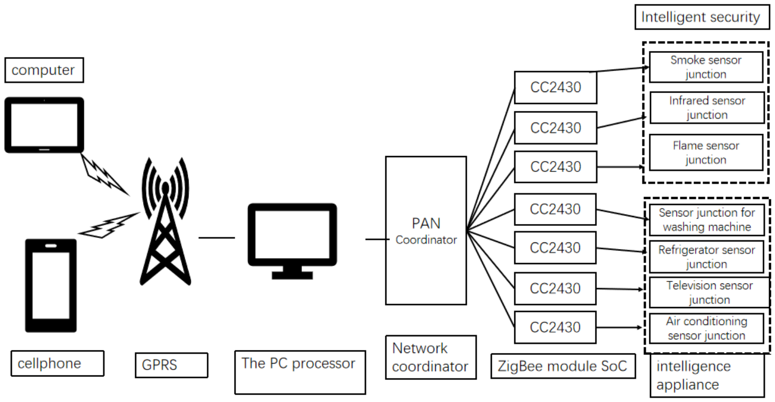


Figure 1. System structure diagram

The system workflow is as follows:

(1) Information collection. The home appliance information is transmitted to the ZigBee SOC module through the sensors of each module. This model uses the most common CC2430 chip, which integrates ZigBee RF front-end, memory and microcontroller. The CC2430 chip finally sends the information to the pan coordinator.

(2) Pan coordinator is responsible for networking and managing the sensors of each intelligent module. Pan coordinator mainly completes the tasks of allocating network IP, sending, interacting and ending information.

(3) PC processor. PC processor is the core component of this design scheme, which is responsible for the intelligent control of the whole system and also the core part of the model. The PC processor is mainly responsible for the analysis and processing of collected information, information data storage and remote control.

(4) Wireless transceiver system. This design completes the communication function with multiple user mobile terminals through the wireless transceiver system to realize the transmission of commands and the response feedback of smart home system information. This design remotely controls the home appliances through the mobile phone or PC, and the home gateway makes judgment and response to the control information. According to the event number and endpoint information in the data frame, it controls the corresponding ZigBee terminal equipment node and makes corresponding actions.

2.3 Controlling the network Structure

The core advantage of intelligent control is to bring convenience to users through intelligent management and control. The intelligent control process is as follows:

2.3.1. Smoke detection function and design

In the set home appliance monitoring area, once the system detects smoke, the system will automatically prompt and alarm to remind the user to deal with the alarm in time to prevent fire disaster.

2.3.2. Flame detection function and design

In the monitoring area set for household appliances, once the system detects the flame, the system will automatically prompt and alarm, and remind the user to deal with the alarm in time to prevent fire disaster.

2.3.3. Indoor temperature control and design

The built-in temperature sensor controls the air conditioner switch and adjusts the indoor temperature as required.

2.3.4. Combined control function and design

Through the design of a key control function to complete the control.

2.3.5 Other control functions and design

In case of fire or gas leakage, the system automatically closes the air valve and opens the window. Automatic window closing function when no one is home.

2.3.6 Intelligent terminal Status Monitoring Function and Design

Inquiry and inquiry are still the main ways for smart home control system to obtain monitoring information. During specific operation, the status of each module terminal in the whole smart home network can be understood by sending a query command to realize real-time monitoring of home appliances; You can also set the system to send status update information to users regularly to realize real-time update monitoring of status information.

3. Design of hardware system

The hardware of this system is composed of two parts: external network and internal network. The external network is composed of a wireless transceiver system and a PC processor. The two can communicate through RS232 serial port. The wireless transceiver system (GPRS data transmission module) uses sim900 module.

The intranet is composed of PC processor, pan Coordinator (network coordinator) and information collection (smart home sensor).[6]The information acquisition sensor communicates with the pan coordinator through the ZigBee wireless module. The sensors of the model communicate with the pan coordinator by sending and receiving information through the CC2430 chip (as shown in Fig. 2). The module is controlled by at command, which is convenient for system integration and software development.

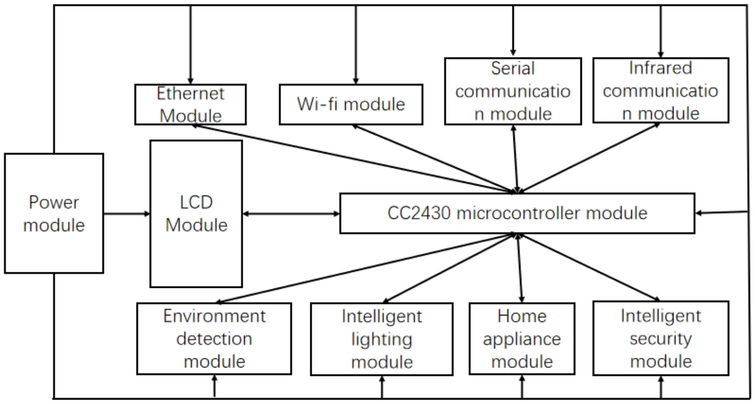

Ethernet Module
Wi-fi module
Serial communication module
Infrared communication module
Power module
LCD Module
CC2430 microcontroller module
Environment detection module
Intelligent lighting module
Home appliance module
Intelligent security module

Figure 2. Hardware module design framework of smart home control system

4. Design of software system

The software system needs to develop terminal control application program, intelligent gateway service program, coordinator program and programs of each distributed wireless sensor node. The working framework of the integrated system is shown in Figure 3 below:

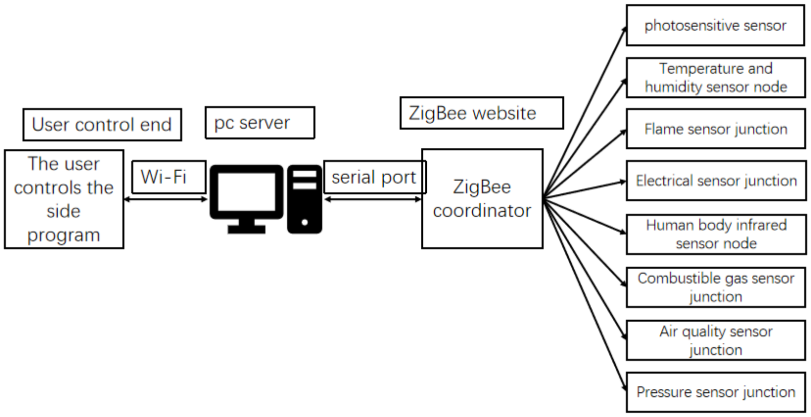

User control end
pc server
ZigBee website
The user controls the side program
Wi-Fi
serial port
ZigBee coordinator
photosensitive sensor
Temperature and humidity sensor node
Flame sensor junction
Electrical sensor junction
Human body infrared sensor node
Combustible gas sensor junction
Air quality sensor junction
Pressure sensor junction

Figure 3. Software design framework

4.1. Users control applications

The user control terminal runs on the IOT task platform, and this program can also run on Android mobile phones and other terminal devices. [7]The application program of the user control side generates the corresponding control command or the corresponding alarm information by receiving the input operation command, and generates the alarm signal (as shown in Fig. 4).

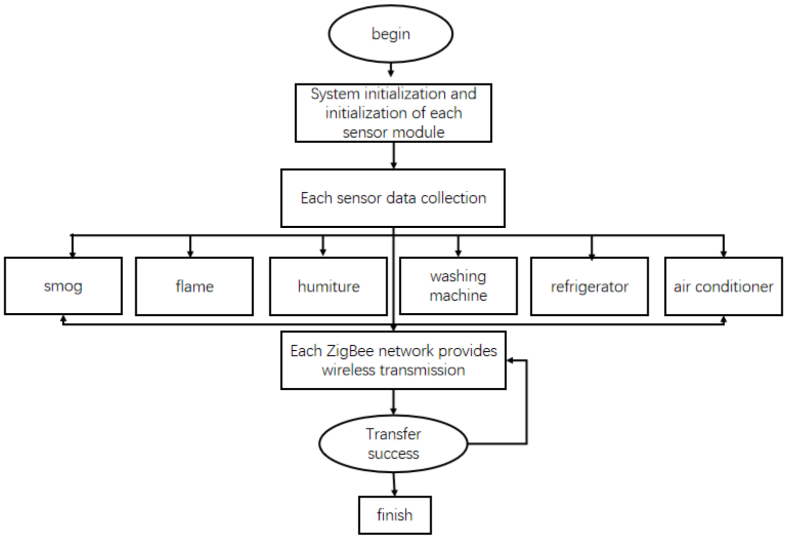

begin
System initialization and initialization of each sensor module
Each sensor data collection
smog
flame
humiture
washing machine
refrigerator
air conditioner
Each ZigBee network provides wireless transmission
Transfer success
finish

Figure 4.Flow chart of data acquisition and sending software at the control end

4.2. ZigBee Intelligent gateway service program

This program runs on the service layer of Android system and is a bridge connecting Android system and ZigBee wireless network. [8]It describes the entire data flow from ZigBee node to coordinator, coordinator to gateway, and gateway to Android client program ZigBee test software.

4.3. ZigBee Coordinator

ZigBee coordinator is a collective node of distributed wireless sensors, and then the terminal control application communicates with the main CPU system coordinator to control each sensor node.[9]

4.4. Wireless Sensor Node

The wireless sensor transmits the sensor data to the general ZigBee coordinator through CC2530 radio frequency, such as smoke, temperature and humidity, infrared and other sensor data.[10]

5. System test

In this paper, the remote monitoring of home appliances by APP is used for functional testing (as shown in Table 1), and the control of lighting equipment is taken as an example to conduct remote control testing of lighting equipment. The main test is the response of the remote server command to the terminal node, so as to verify the remote monitoring function of the system. When the user remotely logs in the smart home system, first select the living room option, and then click the switch on button. Assuming that the bulb is lit, it means that the switch on and off test is successful (as shown in Figure 5).

Table 1. APP functional test cases

Software module: APP function module			Version: V1.5		
Function Name: Status query					
test purpose	For APP terminal users, the smart home control system device status query, view the status of each terminal device added to the smart network				
precondition	All terminal devices have been added to the network				
abnormal					
Use case number	purpose	operating steps	input data	expected outcome	executive report
1	Example Query the air conditioner temperature status	Click living room Air Conditioning mode	No	Displays the current temperature of the air conditioner	OK
2	Querying an Air Conditioner Mode	Click the living room air conditioner and enter the mode page	No	Displays the current mode of the air conditioner	OK
3	Querying Air Conditioner Wind Speed	Click the living room air conditioner and enter the wind speed page	No	Displays the current air speed of the air conditione	OK
4	Example Query air conditioner options	Click the living room air conditioner and enter the Options page	No	Column Number Air conditioner option list	OK
5	Querying Area Information	Click on the area	No	Make a list of regions	OK
6	Query the status of smart lights	Click on living room Air conditioner and go to area, click on bedroom	No	View the current bedroom light status	OK
7	Querying TV Status	Click Living Room Air conditioner and go to the area and click Living room	No	Check the current living room TV	OK
8	Querying the Washing Machine Status	Click on the living room air conditioner and to go to the area, click on the balcony	No	Check the current balcony washing machine	OK

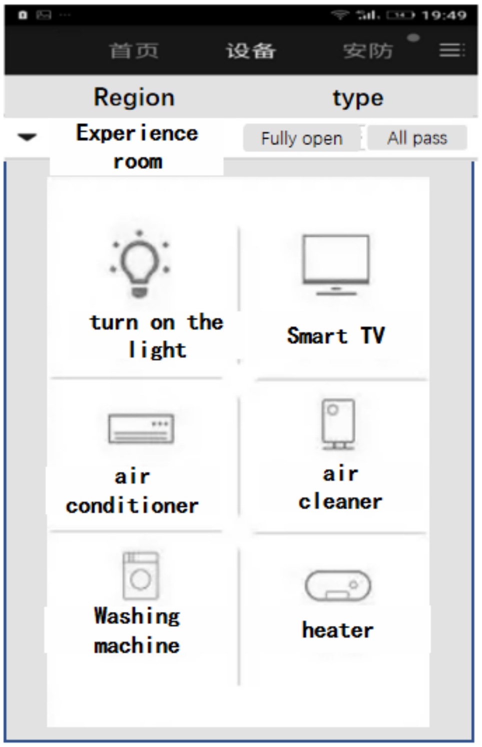


Figure5. monitoring interface of smart home system

6. Conclusion

In order to optimize the home control system, this paper presents the design concept of smart home system based on ZigBee and GPRS. The construction of ZigBee smart home control system mainly uses home gateway (PC), pan Coordinator (network coordinator) and home appliance sensor as basic equipment, and realizes communication with remote mobile phone terminals through GPRS wireless transceiver system. In the research process, this paper explains the system design architecture and the implementation methods of software and hardware, including the construction of software system and the key processing flow. Through the experimental test, the smart home control system based on ZigBee and GPRS technology developed in this paper not only shows its advantages in energy saving, but also has good stability, and can meet the remote monitoring needs of smart homes in reality.

References

[1] Xu Zhenfu. Research on Application of Zig Bee Technology in Smart Home Control System [D]. University of Chinese Academy of Sciences (School of Engineering Management and Information Technology),2014, (08) :7-8.

[2] ZHANG Xiong. Research and Design of Smart Home Control System based on Zig Bee technology [D]. Hangzhou Dianzi University,2015, (03) :8-9.

[3] Yao Guofeng, Zhuang Bin, Zhao Daming, Huo Xiaorui. Smart home system design based on ZigBee wireless technology [J]. Journal of modern electronic technology, 2016, 33 (22) 6:81-84. The DOI: 10.16652 / j.i SSN. 1004-373 - x. 2016.22.020.

[4] ZHANG Y W. Design of smart home control system based on Zigbee technology [D]. Dalian University of Technology,2018.

[5] WANG S G. Design and implementation of smart home control system based on ZigBee technology [D]. Henan university of science and technology, 2019. DOI: 10.27115 /, dc nki. Glygc. 2019.000079.

[6] You Kang. Design and Implementation of Smart Home Control System [D]. Hunan University,2018.

[7] Zhang Lianliu. Intelligent household design ecosystem balance study [D]. Wuhan university of technology, 2020. The DOI: 10.27381 /, dc nki. Gwlgu. 2020.001822.

[8] WANG Guodong. Design and implementation of smart home control System based on Bluetooth technology [D]. Inner Mongolia University,2019.

[9] MENG Chenxu. Design and implementation of smart Home Control System based on Android [D]. Shenzhen University,2016.

[10] GUO Rong. Design and Research of smart home control System based on Internet of Things technology [J]. Digital World,2017(05):141.

Occlusion Face Recognition Based on Improved Attention Mechanism

Mai Fu[1], Zhihui Wang[1*], Daoerji Fan[1], Huijuan Wu[1,2]
[1]Inner Mongolia University, Hohhot 010021, Inner Mongolia, China;
[2]Inner Mongolia University of technology, Hohhot 010321, Inner Mongolia, China
[*] Corresponding author: wzhbit2007@163.com

ABSTRACT

Due to the new crown and other epidemic diseases that make people wear masks to travel, the accuracy of the original face recognition system is affected. To address this challenge, a mask-wearing face recognition system based on an improved attention mechanism is proposed. First, Adding a maximum pooling operation to the CA (Coordinate Attention) attention module, then, placing attention module in the residual unit to form a feature extraction network. LResNet18E-IR is selected as the backbone network. Finally, the ArcFace loss and occlusion probability loss are combined to establish a multi-task network, which further promotes the accuracy of occluded face recognition. The results demonstrate that the system effectively increases the recognition accuracy of masked face and maintains almost the same accuracy as the original model on the unmasked dataset.

Keywords： Convolutional neural network; Masked face recognition; Attention mechanism; Multi-task

1. INTRODUCTION

The field of face recognition has developed rapidly in recent years, making face recognition one of the most reliable biometric identification technologies available[1]. But at present, occlusion is still the most important factor affecting the recognition accuracy, especially in today's social environment, due to the new crown epidemic, allergic rhinitis and some other infectious diseases, people always wear masks when they travel, which causes some areas of the face to be masked, affecting the accuracy of face recognition. The main problem facing current masked face recognition is divided into two sections. One is the lack of datasets, the other is that occlusion destroys the facial features of faces. Therefore, it is necessary to research face recognition algorithms that are not affected by occlusion.

Recently, main approaches on occlusion face recognition are occlusion repair, removal of occlusion, and combined masked and unmasked image co-training. The main methods for occlusion repair are generative adversarial networks and 3D reconstruction to repair the occluded part of a face. Xu[2] used generative adversarial networks to repair the occluded face and then combined convolutional networks to achieve the occluded face recognition. However, the face reconstructed by the occlusion restoration is synthetic, and its reliability depends on the data, network, and training process. Moreover, the occlusion removal process significantly increases the computation time. The mask removal methods remove the masked part and focus on the unmasked part and use the unmasked information for recognition. Li et al.[3] use the crop plus attention method to crop out the masked part and use CBAM [4] (Convolutional Block Attention Module) to focus the network on the unmasked part to achieve the masked face recognition; Vu HN[5]combines deep learning with local binary patterns to extract binary features using the eye, forehead, and periocular regions of the occluded face, then combines the features learned by RetinaFace[6] to jointly recognize the occluded face. These methods are designed for occluded faces and the accuracy degrades significantly when in an unmasked environment. Combined masked and unmasked methods use both masked and unmasked datasets to train together. Montero D et al.[7] introduced a multi-task approach to train the recognition of whether or not a mask is worn together with identity recognition, while training the masked and unmasked data together in equal proportions. The combined masked and unmasked approach fully considers both masked and unmasked environments and strives for outstanding performance in both environments.

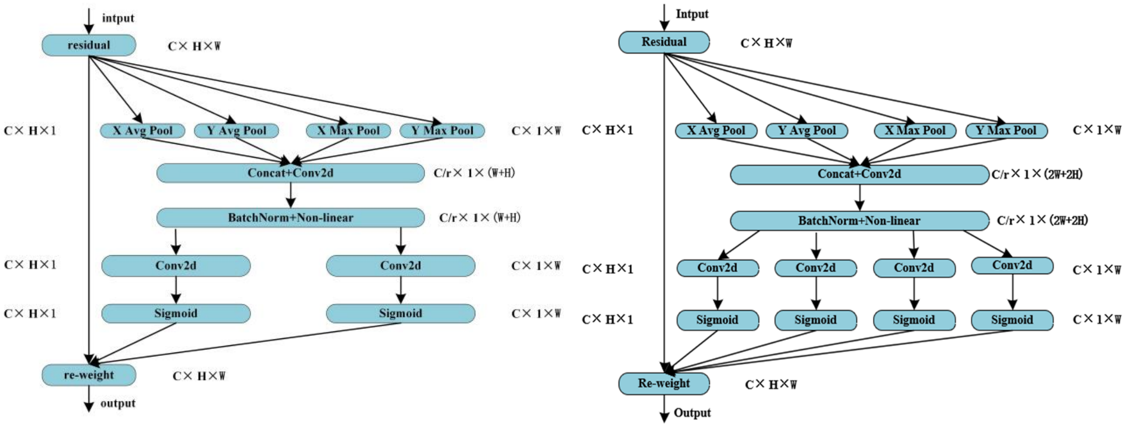
In this paper, we use the MaskTheFace[8] tool on MS1MV2[9] dataset to generate the dataset for training masked face recognition model and select LResNet18E-IR as backbone network. The maximum pooling is added to the CA attention to enrich the location information, and the attention mechanism is fused in the residual structure to construct a face feature extraction network, which is combined with ArcFace loss and mask wearing probability loss to form a multi-task network and applied to the mask face recognition task to improve the recognition accuracy of masked faces and unmasked faces.

International Conference on Computer Graphics, Artificial Intelligence, and Data Processing (ICCAID 2022)
edited by Ruishi Liang, Jing Wang, Proc. of SPIE Vol. 12604, 126042F
© 2023 SPIE · 0277-786X · doi: 10.1117/12.2674629

2. ALGORITHM

2.1Attention mechanism

Attention mechanisms have been shown to be effective in improving the recognition accuracy of networks. SE [10](Squeeze-and-excitation) attention mechanisms make full use of the relationships between channels to improve feature representation. CA[11] (Coordinate Attention) attention mechanism improves on the SE attention mechanism by using two one-dimensional global average pooling to decompose the original two-dimensional global average pooling to aggregate features along the horizontal and vertical directions, respectively, while obtaining long-term dependencies in both directions, reducing the damage of global average pooling on location information. Working with face recognition, the personal information is judged based on the distribution of the five senses of the face, so rich location information is required. We propose an improved version of the CA attention mechanism to facilitate face occlusion recognition. Two additional one-dimensional global maximum pooling (GMP) are introduced to capture the maximum value information along the horizontal and vertical directions respectively, capturing not only the global average value in a single direction, but also the global maximum value in a single direction, enriching the location information and thus enhancing the expression capability of specific features. The attention mechanism is named MCA and the architecture is displayed in Figure 1 (a). The effect of two-output and four-output on the recognition rate of the network was explored in the experiments. The four-output structure named CA1 and show in Figure 1 (b).

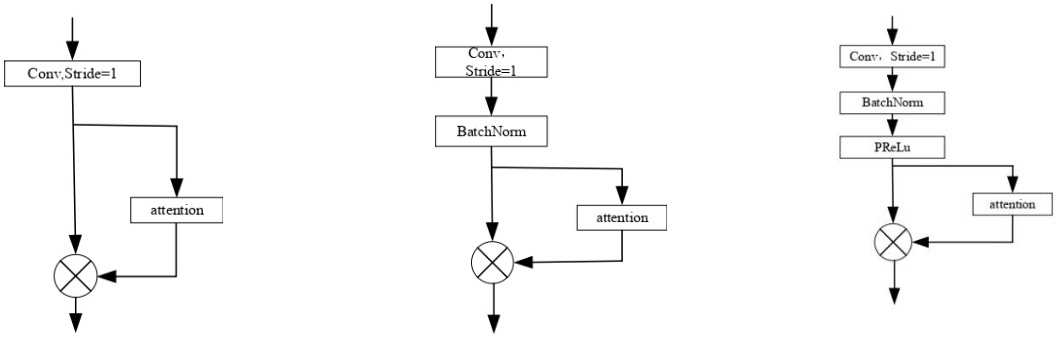


(a) (b)

Figure 1 attention mechanism

2.2Feature extraction module

To obtain facial features more efficaciously, an attention module is added to the original residual structure to form a face feature extraction module. Three feature extraction modules are investigated in the experiments, as shown in Figure 2, and the structure (b) is finally determined to be more sensitive to occlusion information and can effectively improve the recognition accuracy after experimental comparison.



(a) (b) (c)

Figure 2 Feature extraction structure

2.3Multi task network

Considering the existence of both masked and unmasked in real life, a multi-task network based on ArcFace loss[9] is designed for this purpose. Specifically, after the original dropout layer, a fully connected layer that outputs the probability score associated with whether a face is wearing a mask is used in parallel with the fully connected layer used to generate the feature vector as a way to force the network to know whether a face is masked or not, after which this information is used jointly with the generated face feature vector to generate the joint loss function. The ArcFace loss is generated as follow:

$$loss_{\mathrm{Arcface}} = \mathrm{crossEnt}\left(\mathrm{Softmax}\left(logits_{\mathrm{Arcface}}, labels_{\mathrm{ID}}\right)\right) \tag{1}$$

The equation to generate the masking loss ($loss_{\mathrm{Mask}}$) is as follows:

$$loss_{\mathrm{Mask}} = \mathrm{crossEnt}\left(\mathrm{Softmax}\left(logits_{\mathrm{Mask}}, labels_{\mathrm{ID}}\right)\right) \tag{2}$$

Where $\mathrm{crossEnt}$ is the cross-entropy loss, $\mathrm{Softmax}$ which is the activation function, $logits_{\mathrm{Arcface}}$ is the output of the fully connected layer that generates the face feature vector, and $logits_{\mathrm{Mask}}$ is the out of the newly attached fully connected layer, $labels_{\mathrm{ID}}$ means labeled. The joint loss is obtained by summing these two losses:

$$loss_{\mathrm{MTArcFace}} = loss_{\mathrm{Arcface}} + \lambda\, loss_{\mathrm{Mask}} \tag{3}$$

λ is the weight hyperparameter that balances the two losses. In our experiment, the value of λ is set to 0.001, $loss_{\mathrm{MTArcFace}}$ is a joint loss.

3. MASKED DATASETS

The MS1MV2 dataset was chosen as the training set, which is a cleaned MS-Celeb-1M dataset containing 5.8 million images and 85,000 identities. We selected 24,000 of these identities for training and used the MaskTheFace tool to simulate wearing a mask on all the images. Figure 3 shows a partial picture of the test dataset, which contains the simulated masking images. Four datasets were used as the test set, namely LFW[12] dataset,AgeDB_30[13] dataset, CFP_FP[14] dataset, and MFR[15] real mask-wearing dataset, and the corresponding simulated mask-wearing versions of the first three datasets were generated as the test set for the mask-wearing condition. We used MTCNN[16] network for face alignment and detection of labeled facial keypoints. Finally, the aligned face picture's size is 128×128 .

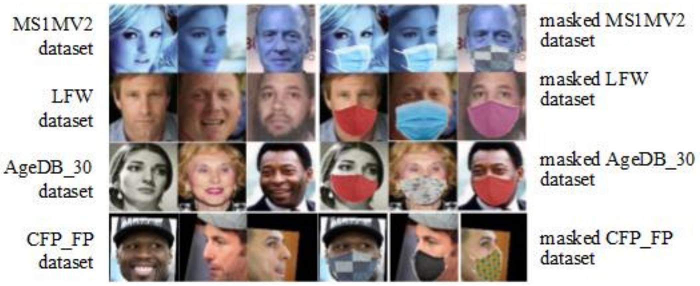


Figure 3 Original and simulated masked dataset

4. EXPERIMENTS

In the experiments, we use one NVIDIA GeForce RTX 3060 GPU to train the model, build a deep learning framework based on PyTorch version 1.10.1, the total batch size is set to 128, the step size is 650000. We use the SGD optimizer with momentum set to 0.9 and an initial learning rate of 0.0003, the learning rate is reduced using equal intervals, each time by 0.1, and the experiments use migration learning[17].

4.1Comparison of the recognition performance of three attention modules

We compared the proposed attention mechanisms. From Table 1, it can be found that the MCA attention mechanism has a higher improvement on face recognition accuracy compared to the CA attention mechanism and CA1 attention mechanism.

Table 1 Recognition accuracy under different attention mechanisms (%)

Technology	AgeDB_30	Masked AgeDB_30	MFR
Original model+CA	93.93	90.37	78.21
Original model+CA1	93.95	90.87	78.48
Original model+MCA	94.23	91.42	79.30

4.2Comparison of recognition performance of three feature extraction module structures

Compare the three proposed feature extraction modules, we can found from Table 2 that module (b) has better recognition accuracy in both masked and unmasked situations.

Table 2 Comparison of recognition performance of different feature extraction modules (%)

Feature extraction module	LFW	AgeDB_30	Masked LFW	MFR
Module a	99.28	93.66	97.9	78.16
Module b	98.98	94.23	98.15	79.30
Module c	99.20	93.61	98.10	78.28

4.3Performance comparison with other models

Table 3 shows the comparison with other face recognition models wearing masks, and from the data in the table, it can be found that the performance of our model is close to the model proposed by Montero D [7], but the model size is significantly smaller than the model proposed by Montero D.

Table 3 Comparison of recognition performance of different models (%)

Models	LFW	Masked LFW	CFP_FP	Masked CFP_FP	Size/Mb
Huang B[18]	99.01	97.08	93.58	86.07	—
Montero D[7]	99.45	98.92	92.27	88.43	167
Ours	98.98	98.15	91.63	89.53	98.20

5. Conclusion

A face recognition system based on improved attention mechanism is proposed to address the situation of poor recognition accuracy under occlusion conditions. The model takes into account both obscured and unobscured conditions, improves the recognition accuracy of obscured faces with minimal loss of unobscured recognition accuracy, and provides more convenient detection for the individuals to be recognized in practice, in the current epidemic, the risk of cross-infection from mask removal can be avoided. This work mainly focuses on masked faces, and the masked parts are known in advance. Future research will be devoted to the recognition of face pictures with uncertain masked locations, and further reduce the model size to make the arrangement of the model more convenient.

Acknowledgement

Inner Mongolia Natural Science Foundation Project (No. 2022MS06021)

REFERENCES

[1] Anot N, Singh K K. A review on biometrics and face recognition techniques[J]. International Journal of Advanced Research,2016,4(5):783-786.

[2] XU R H, CHENG J X, LI Z D, et al. Face recognition with occlusion based on cyclic generative adversarial networks [J]. Computer Engineering,2022,48(5):289-296,305.

[3] Li Y, Guo K, Lu Y, et al. Cropping and attention based approach for masked facerecognition [J]. Applied Intelligence, 2021,51(5):3012-3025.

[4] Woo S, Park J, Lee J Y, et al. Cbam: Convolutional block attention module[C]//Proceedings of the European conference on computer vision (ECCV). 2018: 3-19.

[5] Vu H N, Nguyen M H, Pham C. Masked face recognition with convolutional neural networks and local binary patterns[J]. Applied Intelligence, 2022, 52(5):5497-5512.

[6] Deng J, Guo J, Ververas E, et al. Retinaface: Single-shot multi-level face localisation in the wild[C] //Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. 2020: 5203- 5212.

[7] Montero D, Nieto M, Leskovsky P, et al. Boosting masked face recognition with multi-task arcface[J]. arXiv preprint arXiv:2104.09874, 2021.

[8] Anwar A, Raychowdhury A. Masked face recognition for secure authentication[J]. arXiv preprint arXiv:2008.11104, 2020.

[9] Deng J, Guo J, Xue N, et al. Arcface: Additive angular margin loss for deep face recognition[C]//Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. 2019: 4690-4699.

[10] Hu J, Shen L, Sun G. Squeeze-and-excitation networks[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. 2018: 7132-7141.

[11] Hou Q, Zhou D, Feng J. Coordinate attention for efficient mobile network design[C]//Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. 2021: 13713-13722.

[12] Huang G B, Mattar M, Berg T, et al. Labeled faces in the wild: A database forstudying face recognition in unconstrained environments[C]//Workshop on faces in'Real-Life'Images: detection, alignment, and recognition. 2008.

[13] Moschoglou S, Papaioannou A, Sagonas C, et al. Agedb: the first manually collected, in-the-wild age database[C]//proceedings of the IEEE conference on computer vision and pattern recognition workshops. 2017: 51-59.

[14] Sengupta S, Chen J C, Castillo C, et al. Frontal to profile face verification in the wild[C]//2016 IEEE winter conference on applications of computer vision (WACV). IEEE, 2016: 1-9.

[15] Wang Z, Wang G, Huang B, et al. Masked face recognition dataset and application[J]. arXiv preprint arXiv:2003.09093, 2020.

[16] Zhang K, Zhang Z, Li Z, et al. Joint face detection and alignment using multitask cascaded convolutional networks[J]. IEEE signal processing letters, 2016, 23(10): 1499-1503.

[17] Torrey L, Shavlik J. Transfer learning[M]//Handbook of research on machine learning applications and trends: algorithms, methods, and techniques. IGI global, 2010: 242-264.

[18] Huang B, Wang Z, Wang G, et al. When face recognition meets occlusion: A new benchmark[C]//ICASSP 2021-2021 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). IEEE, 2021: 4240-4244.