Retraction Notice

The Editor-in-Chief and the publisher have retracted this article, which was submitted as part of a guest-edited special section. An investigation uncovered evidence of systematic manipulation of the publication process, including compromised peer review. The Editor and publisher no longer have confidence in the results and conclusions of the article.

ZS does not agree with the retraction. CF, YL, and HZ either did not respond or could not be reached.

News image text classification algorithm with bidirectional encoder representations from transformers model

Zhan Shi,[a,*] Chongjun Fan,[a] Ying Li,[b] and Hongliu Zhang[a]
[a]University of Shanghai for Science and Technology, Business School, Shanghai, China
[b]Yonyou Network Technology Co., Ltd., Beijing, China

Abstract. With the development of Internet technology and the transformation of news media to informatization, news images, texts, and other information have exploded. The news image and text classification can effectively solve the disorder problem of news information. The early news image text classification is to establish artificial classifiers according to specific classification rules, but the classification error rate is high and the classification speed is slow. Later, machine learning technology replaced manual classification of news image texts. Although the classification efficiency has been greatly improved, the classification time is still a bit long. The bidirectional encoder representations from transformers (BERT) model uses transformer and encoder to pretrain news image text to improve classification efficiency. By comparing the differences between machine learning and BERT models in news image text classification, the experiments showed that the average precision, recall, and $F1$ values of the news image text classification algorithm using the BERT model were 96.6%, 95.7%, and 96.1%, respectively. All three evaluation criteria were about 5% more than the classification algorithm of the machine learning model. The classification speed of the news image text classification algorithm using the BERT model was 1.8 times that of the news image text classification algorithm based on support vector machine. Therefore, the news image text classification algorithm using the BERT model can improve the classification accuracy and efficiency of news image text. © *2022 SPIE and IS&T* [DOI: 10.1117/1.JEI.32.1.011217]

Keywords: news image text classification; bidirectional encoder representations from transformers model; transformer and encoder; machine learning; support vector machine.

Paper 220608SS received Jul. 1, 2022; accepted for publication Aug. 29, 2022; published online Sep. 13, 2022.

1 Introduction

With the rapid growth of information in the 21st century, traditional paper news can no longer adapt to the current era. The information transmission methods of news have also become diversified, and images and texts can convey more information. However, the influx of explosive news information leads to problems, such as difficulty in classifying news images and texts, slow classification speed, and low classification efficiency. Machine learning can already improve the classification of news images and texts by replacing manual classification, but it takes more time to build a feature classification model. The bidirectional encoder representations from transformers (BERT) model is a pretrained language model. Due to its internal abandonment of the decoder structure, it can improve the classification efficiency and classification accuracy of news image texts. Therefore, this paper has research significance.

With the explosive growth of news information, relevant researchers use classification algorithms to effectively classify news image text information. Among them, Ghosh's research showed that preprocessing news classification can improve the efficiency of news image text classification.[1] Elik and Ko[2] pointed out that the accuracy of classification of news images and texts by techniques, such as machine learning, can reach 96%. Tegegnie's research pointed out that the automatic classification of news images and texts depends on the classification of

*Address all correspondence to Zhan Shi, 211420079@st.usst.edu.cn

relevant features.[3] Endalie's research proposed a news image text feature classification method based on gain and chi-square mixture, which can improve the speed of classification.[4] Firdaus[5] improved the classification accuracy of news image text through Bayesian algorithm. Classification algorithms, such as machine learning, can improve the efficiency of classifying news images and texts, but it still takes a lot of time to build image and text feature models.

BERT is a pretrained classification model that does not require a decoder process. The BERT model is applied to news image text classification. Among them, Wan and Li[6] used BERT classifier to classify news image text information. Gao's research showed that the bidirectional encoding of BERT can improve the speed of classification.[7] Liu et al.[8] compared BERT with traditional classification through experiments and found that the performance of the BERT model far exceeded the traditional classification. Ji et al.[9] stated that BERT can solve the problem that machine learning is inflexible for news image text classification. Li's research showed that BERT model can improve the speed of news image text classification.[10] The BERT model applied to news image text can improve the accuracy of news image text classification, but it lacks comparison with other classification models.

The BERT model is a bidirectional encoding structure with the characteristics of efficient classification. In this paper, the BERT model is applied to the classification of news images and texts. The innovations of this paper include the following aspects: (1) research on the combination of BERT model and news image text and (2) comparison of the advantages and disadvantages of the news image text classification algorithm using the BERT model and the news image text classification algorithm based on machine learning.

2 News Image Text Classification Method Using BERT Model

News image text classification is to classify news information according to the specified image text characteristics so that the huge and complex news information becomes regular and followable.[11] The classification technology mainly includes machine learning and BERT model. The framework of news image text classification is shown in Fig. 1.

2.1 *Machine Learning Classification Methods*

Machine learning is a multidisciplinary intelligent research, which has a wide range of applications in the fields of text, images, and speech.[12] Machine learning has a high classification and prediction ability through the analysis of sample data. It can greatly improve the efficiency of news image text classification.

2.1.1 *Convolutional neural network*

Convolutional neural network (CNN) is an improvement on the basis of back propagation neural network. The CNN has the characteristics of deep learning and is widely used in image recognition, text recognition, etc.[13] The network topology of CNN is shown in Fig. 2.

As shown in Fig. 2, CNN consists of four parts: convolutional layer, subsampling layer, fully connected layer, and output layer. The CNN is a multilayer convolutional hierarchical computing model, which has strong learning ability and can handle complex nonlinear problems. It can classify news image text very well.[14]

The convolutional layer equation is expressed as

$$s_u^t = f\left(\sum_{i \in M_r} s_i^{t-1} * k_{iu}^t + g_u^t\right), \tag{1}$$

where s_u^t represents the u'th feature data of the t'th convolutional layer and $*$ stands for convolution operation.

The sampling dimensionality reduction process of the subsampling layer is expressed as

$$s_u^i = f(H_u^t h(s_u^{t-1}) + g_u^t), \tag{2}$$

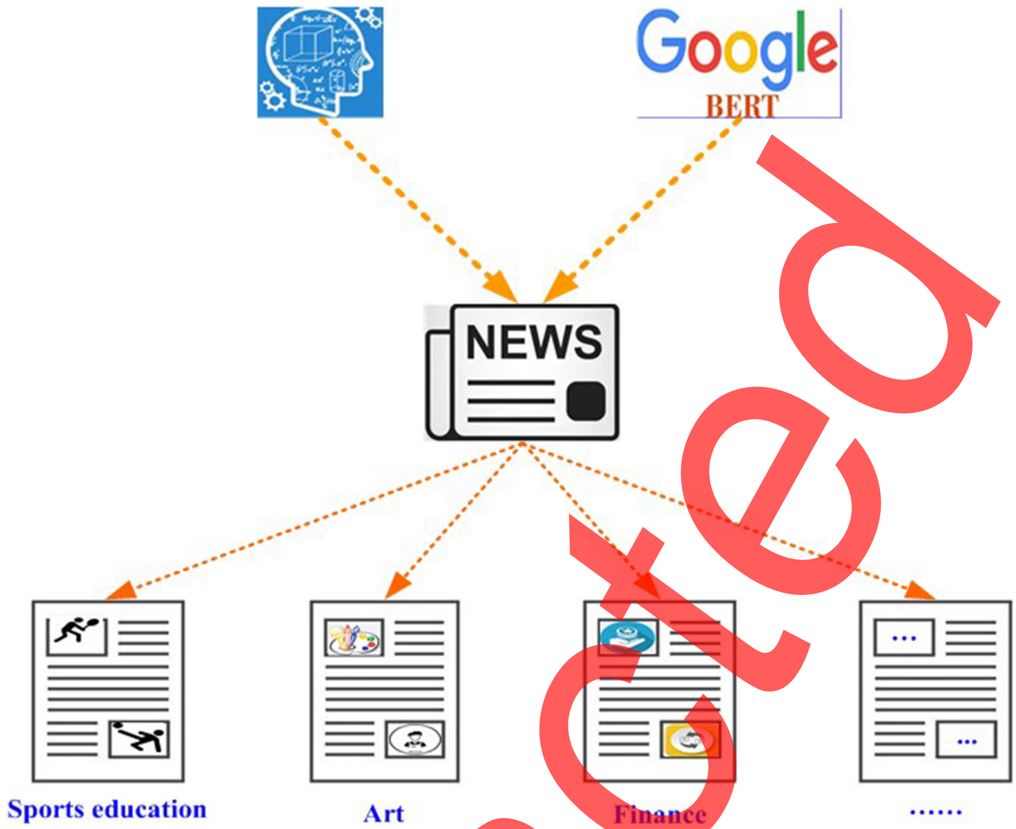

Fig. 1 Framework diagram of news image text classification.

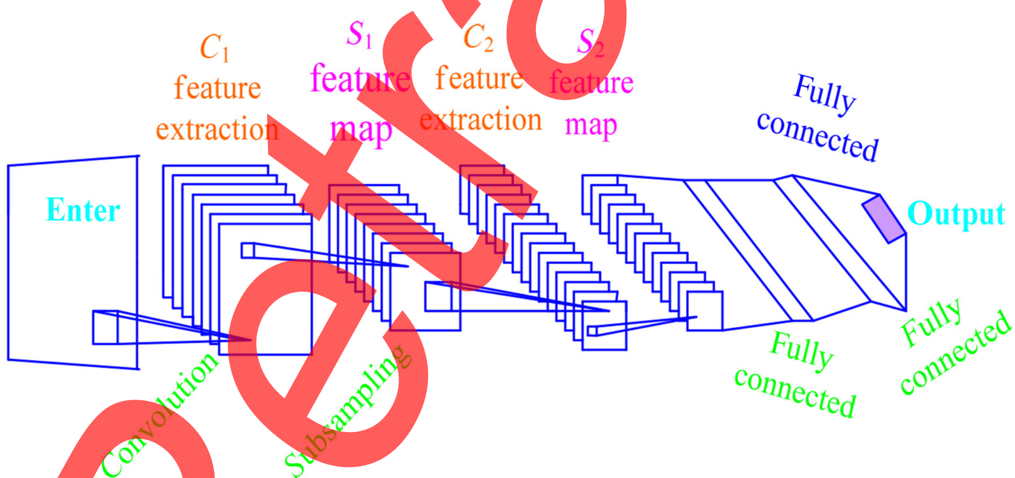

Fig. 2 CNN structure diagram.

where $h()$ represents the sampling dimension reduction function and H represents weighting parameter.

The reverse derivation of the fully connected layer is shown in the following equations:

$$\frac{\partial Q}{\partial w(t)} = \delta^{(t)}(s^{(t-1)})^T, \tag{3}$$

where t represents the current convolutional layer

$$\frac{\partial Q}{\partial b(t)} = \delta^{(t)}. \tag{4}$$

The transfer process from the convolutional layer $t-1$ to the t layer is shown in the following equation:

$$s_i^{(t)} = f\left(\sum_{n=1}^{N_{t-1}} B2(s_n^{(t-1)}, K_{\text{in}}^{(t)}) + g_u^t\right), \tag{5}$$

where B represents the convolution calculation process.

2.1.2 *Recurrent neural network*

Recurrent neural network structure. Recurrent neural network (RNN) is a network structure with memory function that can process sequence data.[15] There is a closed-loop mode in the hidden layer, which can contain the current and previous input information. The RNNs are great for handling image text classification problems. The structure of RNN is shown in Fig. 3.

As shown in Fig. 3, assuming that the input signal at time t is X_t, the weights between the three layers are A, B, and C, respectively, where C represents the weight between the hidden layers.[16] Then the expression of the hidden layer at time t is

$$Z_t = f(A \cdot X_t + C \cdot Z_{t-1}), \tag{6}$$

where Z_{t-1} represents the state of the hidden layer at time $t-1$ and f represents the activation function.

The expression of the output layer is

$$Y_t = f(B \cdot Z_t). \tag{7}$$

Long short-term memory network. Long short-term memory (LSTM) is an improvement of RNN. It can solve the long-term dependency problem of RNN and can effectively encode textual information.[17] The structure of LSTM is shown in Fig. 4.

The gate control information is entered into the memory state as shown in the following equations:

$$C_{ti} = \tanh(v_{xc} \cdot x_t + v_{hc} \cdot h_{t-1} + k_c), \tag{8}$$

$$I_t = \mu(v_{xi} \cdot x_t + v_{hi} \cdot h_{t-1} + v_{ci} \cdot c_{t-1} + k_t). \tag{9}$$

The working process of the forget gate is shown in the following equation:

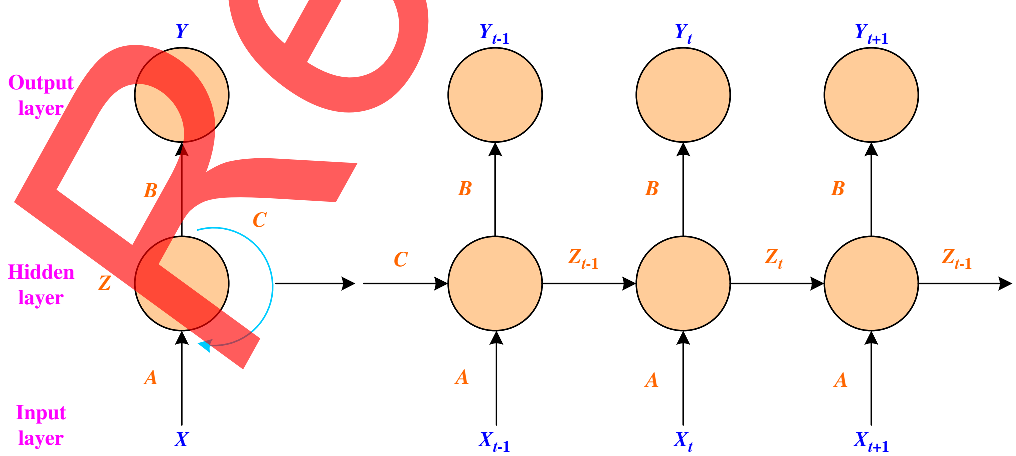

Fig. 3 RNN structure diagram.

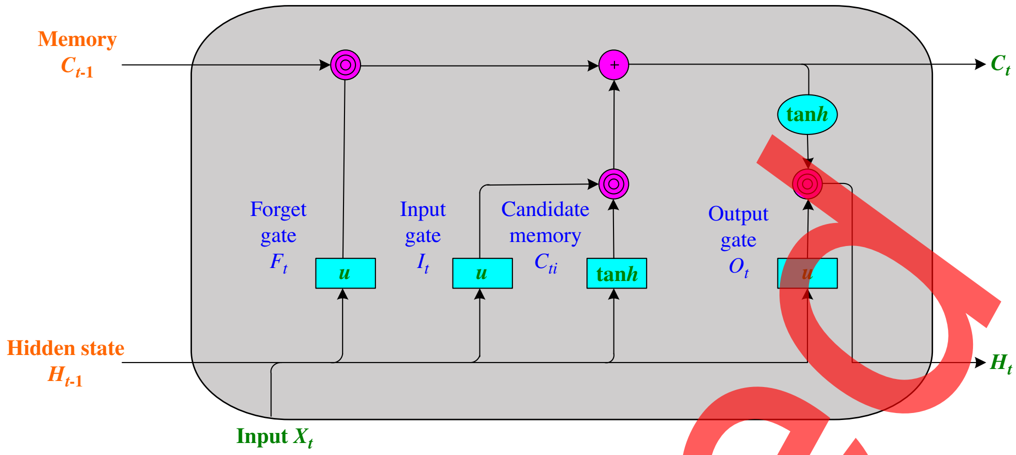

Fig. 4 LSTM structure diagram.

$$F_t = \mu(v_{xf} \cdot x_t + v_{hf} \cdot h_{t-1} + v_{cf} \cdot c_{t-1} + k_f). \tag{10}$$

The output gate works as shown in the following equation:

$$O_t = \mu(v_{xo} \cdot x_t + v_{ho} \cdot h_{t-1} + v_{co} \cdot c_{t-1} + k_o). \tag{11}$$

2.1.3 *Support vector machine*

Support vector machine (SVM) is a statistical method that can accurately classify processing. Its decision boundary is the maximum margin hyperplane for solving the learning samples.[18] The classification model of SVM is shown in Fig. 5.

Supposing that the sample data are $\{(p_1, q_1), (p_2, q_2), \ldots, (p_n, q_n)\}$, the SVM linear regression is expressed as

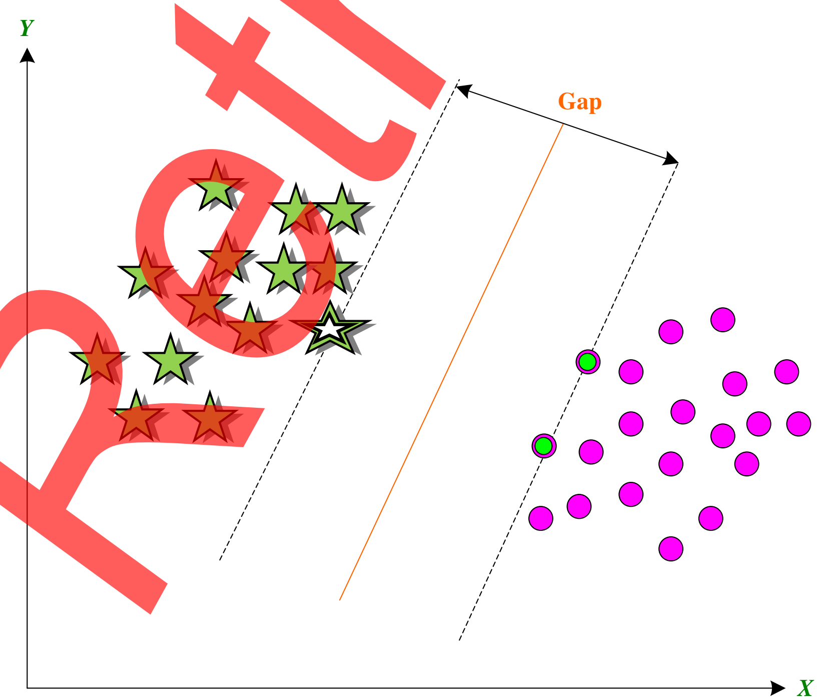

Fig. 5 SVM classification model diagram.

$$f(x) = k \cdot g(x) + b, \tag{12}$$

where $g(x)$ represents the mapping function.

The linear insensitive function is shown in the following equation:

$$R(f(x), y, d) = \begin{cases} |y - f(x)| - d, & |y - f(x)| > d \\ 0, & \text{other.} \end{cases} \tag{13}$$

2.2 *BERT Model*

BERT is a pretrained bidirectional transformer encoding method. Bidirectional structure enables efficient classification of news image text.[19] Its structure is shown in Fig. 6.

2.2.1 *Transformer structure*

Transformer is an attention mode consisting of encoder and decoder.[20] The structure is shown in Fig. 7.

The transformer calculation process is expressed as

$$A_i = \frac{e^{f(U,K_i)}}{\sum_{j=1}^{n} e^{f(U,K_j)}}. \tag{14}$$

The weighting process is shown in the following equation:

$$\sum_{i=1}^{n} A_i W_i. \tag{15}$$

In Eqs. (14) and (15), U, K, and W represent queries and key-value pairs.

The attention mechanism of transformer is shown in the following equation:

$$A(U, K, W) = f\left(\frac{UK^T}{\sqrt{h_k}}\right), \tag{16}$$

where $\sqrt{h_k}$ means reducing attention dependency.

The transformer attention structure is shown in Fig. 8.

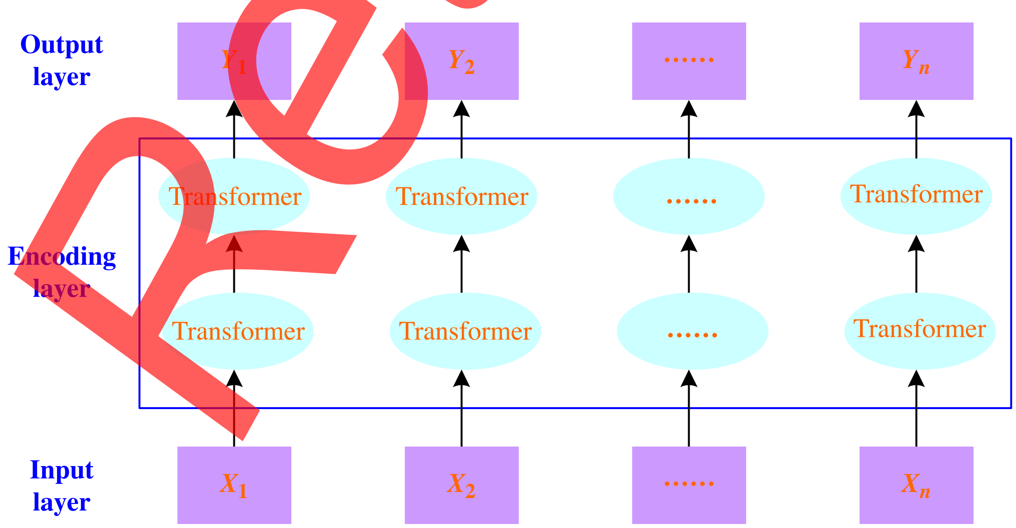

Fig. 6 BERT model diagram.

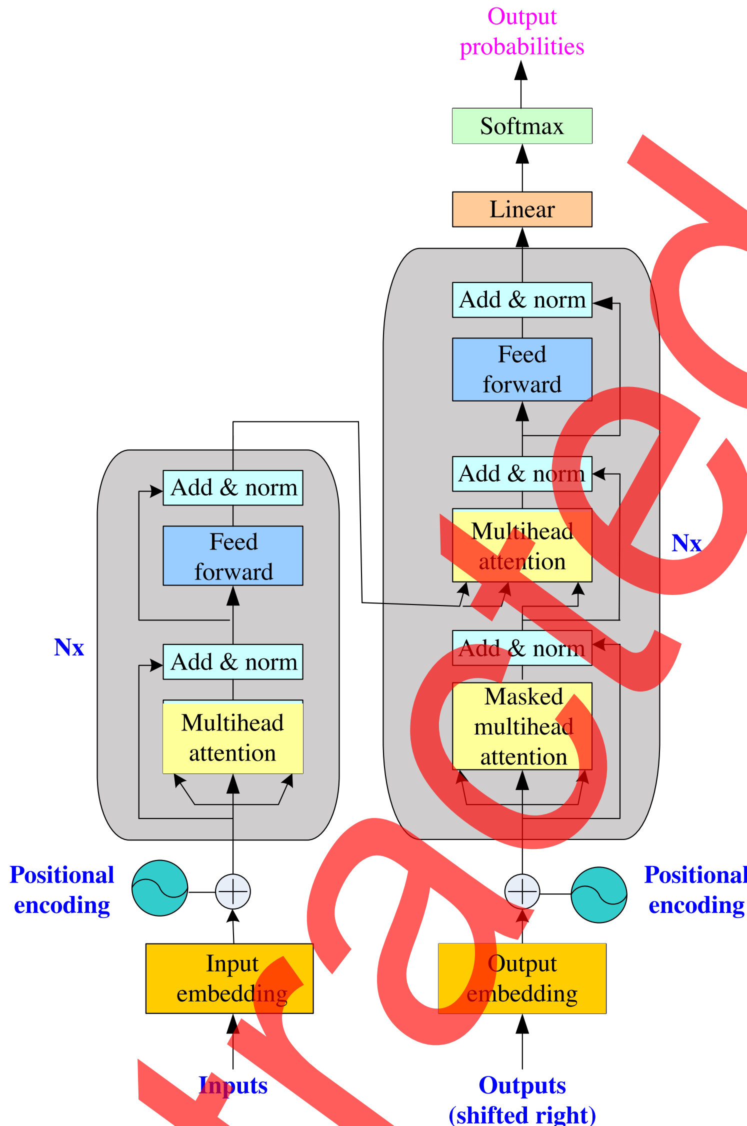

Fig. 7 Transformer structure diagram.

2.2.2 *Attention mechanism*

The attention mechanism is to map the U, K, and W features of the input information to the output as shown in the following equation:

$$A(U, K, W) = f(UK^T)W. \tag{17}$$

There are many ways to calculate the similarity of U and K as shown in the following equations:

$$G(U, K) = \frac{U^T K}{\|U\| \cdot \|K\|}, \tag{18}$$

$$G(U, K) = U^T K, \tag{19}$$

$$G(U, K) = R[U, K], \tag{20}$$

$$G(U, K) = W_a^T f(R_U + D_K), \tag{21}$$

where G represents the similarity.

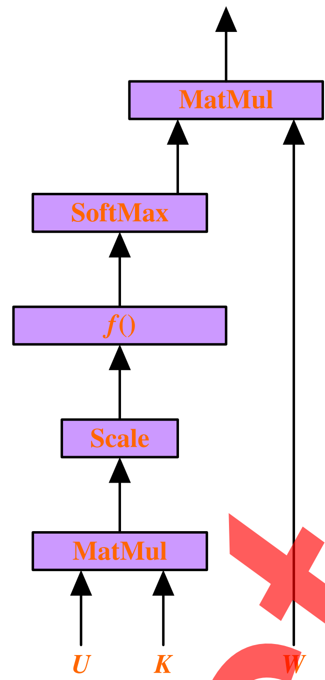

Fig. 8. Transformer attention structure diagram.

2.2.3 *BERT judgment mode*

There are many evaluation metrics for BERT, but the most commonly used are precision (P), recall (R), and $F1$ values. The higher the value of the indicator is, the higher the efficiency of classification will be. The BERT judgment is derived based on the binary classification, which is shown in Table 1.

The P value is expressed as

$$P = \frac{A}{A + C}. \tag{22}$$

The R value is expressed as

$$R = \frac{A}{A + N}. \tag{23}$$

The $F1$ value is expressed as

$$F1 = \frac{2PR}{P + R}. \tag{24}$$

Table 1 Binary classification table.

Classification type	Positive	Negative
True	A	B
False	C	D
Difference	T	T

Table 2 Statistical news data table.

News category	Quantity	Percentage (%)
Art	1450	14.5
Military	590	9.4
Finance	800	8.0
Agriculture	750	7.5
Music	1640	16.4
Sports	1290	12.9
Real estate	600	6.0
Entertainment	1430	14.3
Food	1100	11.0

Table 3 Classification of news popularity.

Classification	News type		
Top news	Music	Art	Entertainment
Medium news	Sports	Food	Military
Low hot news	Finance	Agriculture	Real estate

Fig. 9 Results of CNN-based news image text classification: (a) hot news, (b) medium hot news, and (c) low hot news.

3 Experimental Data on the Comparison of News Image Text Classification Algorithms Based on BERT Model and Machine Learning

3.1 *News Data Sources*

The operating system of this experiment is Windows 10, 10,000 pieces of Google news data were randomly selected, and statistics were made on the selected news data to observe the number and proportion of each type of news data. The statistical news data are shown in Table 2.

3.2 *Classification of News Data*

To better compare the news image text classification algorithm based on the BERT model and machine learning, the selected news data are classified, and the classification effect of news of different popularity is compared. The results of hot classification of news categories are shown in Table 3.

4 News Image Text Classification Results and Discussion

For the comparison of news image text classification algorithms based on BERT model, CNN, RNN, and SVM, the experiment analyzes the P, R, and $F1$ values of various algorithms in processing news images and texts from the perspective of three types of news: high, medium, and low.

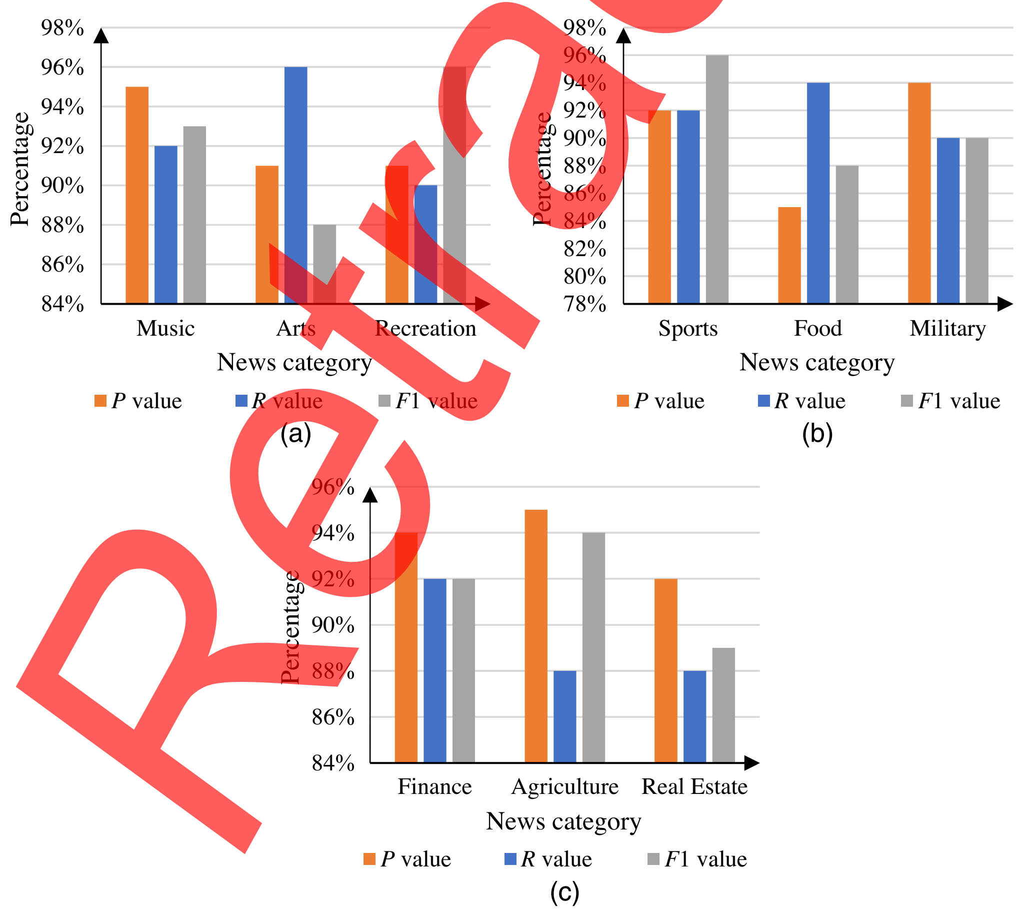

Fig. 10 Results of news image text classification based on RNN: (a) hot news, (b) medium hot news, and (c) low hot news.

4.1 *CNN-Based News Image Text Classification*

Image and text classification of three types of hot news is carried out with CNN, and the P, R, and $F1$ values were observed. The results of CNN-based news image text classification are shown in Fig. 9.

The data in Fig. 9 show that the indicators based on CNN news image and text classification perform almost the same under the three types of hot news, and the overall value is about 90%. The news with low popularity have less classification information, and the classification effect is also better.

4.2 *News Image Text Classification Based on RNN*

The image and text classification of the three types of news is carried out by the RNN method. The results of the RNN-based news image and text classification are shown in Fig. 10.

The data in Fig. 10 show that the averages of various indicators based on RNN news image and text classification in terms of low-interest news are 93.7%, 89.3%, and 91.7%. The data in high heat news and medium heat news are also similar.

4.3 *SVM-Based News Image Text Classification*

The three types of news are classified by SVM, and the result of SVM-based news image and text classification is shown in Fig. 11.

The data in Fig. 11 show that the average of each index based on SVM news image and text classification is better in terms of low-popularity news than in the two types of medium popularity and high popularity. The overall data of SVM are better than that of CNN and RNN.

Fig. 11 Results of news image text classification based on SVM: (a) hot news, (b) medium hot news, and (c) low hot news.

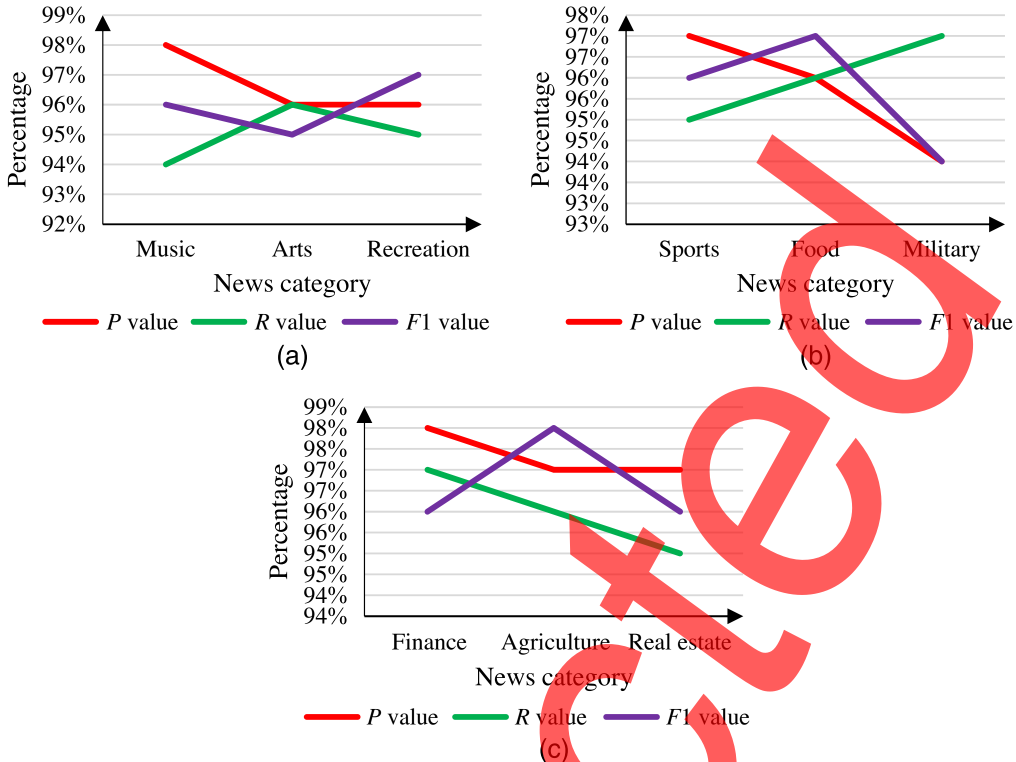

Fig. 12 Results of BERT-based news image text classification: (a) hot news, (b) medium hot news, and (c) low hot news.

4.4 *BERT-Based News Image Text Classification*

The image and text classification of the BERT model is used for three types of news, and the results of the BERT-based news image and text classification are shown in Fig. 12.

The data in Fig. 12 show that the P, R, and $F1$ values of BERT news image text classification are 96.6%, 95.7%, and 96.1%, respectively.

4.5 *Experiment Deconstruction*

Through the comparison of news image text classification algorithms based on BERT model, CNN, RNN, and SVM, the news image text classification algorithm based on BERT model is better than the machine learning method in terms of classification efficiency and accuracy. The specific average data are shown in Table 4.

Comparing the classification speed based on the BERT model and machine learning (SVM), the classification speed based on the BERT model is also much faster than the SVM algorithm. The classification speed of news images and texts of the two methods is shown in Fig. 13.

Table 4 Average data table of news image text classification algorithms.

Method	P value (%)	R value (%)	$F1$ value (%)
BERT model	96.6	95.7	96.1
CNN	88.0	80.9	89.7
RNN	92.3	92.7	92.3
SVM	93.6	94.0	93.7

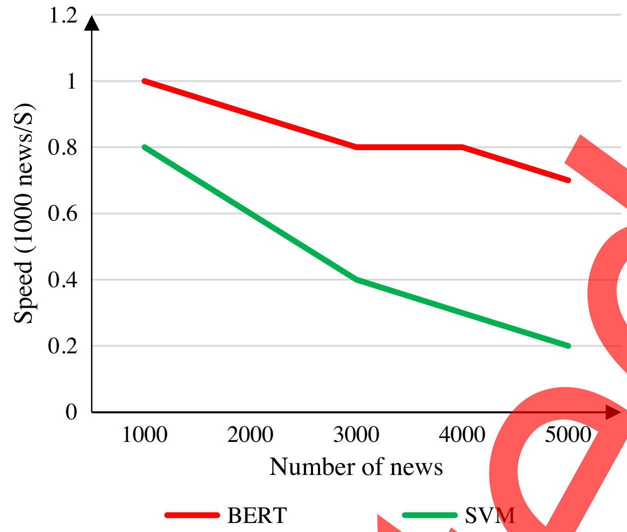

Fig. 13 Comparison of speed of news image text classification.

5 Conclusions

Through the comparison of news image text classification algorithms based on BERT model and machine learning (CNN, RNN, and SVM), the following conclusions are drawn: (1) the P, R, and $F1$ values of the news image text classification algorithm based on the BERT model are better than the news image text classification algorithm based on the machine learning method. The P, R, and $F1$ values of the BERT model classification are 96.6%, 95.7%, and 96.1%, respectively. (2) In terms of the speed of news image text classification, the average speed of the news image text classification algorithm based on the BERT model is 840 news per second, whereas the average speed of the news image text classification algorithm based on the SVM algorithm is 460 news per second. Therefore, the news image text classification algorithm using the BERT model is better than the news image text classification algorithm based on machine learning in terms of classification accuracy, classification efficiency, and classification speed. However, the model training of BERT is relatively complex, so improving the model training of BERT will be the direction of future research.

Acknowledgments

The author(s) received no financial support for the research, authorship, and/or publication of this article. The authors declare that there are no conflicts of interest regarding the publication of this article.

References

1. S. Ghosh et al., "Application of texture-based features for text non-text classification in printed document images with novel feature selection algorithm," *Soft Comput.* **26**(5), 891–909 (2022).
2. Z. Elik and B. C. Ko, "Classification of Turkish news text by TF-IDF, Word2vec and fasttext vector model methods," *Deu Muhendislik Fakultesi Fen ve Muhendislik* **23**(67), 121–127 (2021).
3. A. K. Tegegnie, A. N. Tarekegn, and T. A. Alemu, "MA comparative study of flat and hierarchical classification for amharic news text using SVM," *Int. J. Inf. Eng. Electron. Bus.* **9**(3), 36–42 (2017).
4. D. Endalie and G. Haile, "Hybrid feature selection for amharic news document classification," *Math. Probl. Eng.* **2021**(6), 5516262 (2021).

5. A. Firdaus, "Reputation scoring fake news using text mining," *ACMIT Proc.* **4**(1), 12–17 (2017).
6. C. X. Wan and B. Li, "Financial causal sentence recognition based on BERT-CNN text classification," *J. Supercomput.* **78**(5), 6503–6527 (2022).
7. W. Gao and H. Huang, "A gating context-aware text classification model with BERT and graph convolutional networks," *J. Intell. Fuzzy Syst.* **40**(3), 4331–4343 (2021).
8. N. Liu, Q. Wangand , and J. Ren, "Label-embedding bi-directional attentive model for multi-label text classification," *Neural Process. Lett.* **53**(9), 375–389 (2021).
9. H. Ji et al., "HSI-BERT: hyperspectral image classification using the bidirectional encoder representation from transformers," *IEEE Trans. Geosci. Remote Sens.* **58**(1), 165–178 (2019).
10. W. Li et al., "Joint image-text news topic detection and tracking by multimodal topic and-or graph," *IEEE Trans. Multimedia* **19**(2), 367–381 (2017).
11. B. Al-Shargabi, F. Olayah, and W. A. Romimah, "An experimental study for the effect of stop words elimination for Arabic text classification algorithms," *Int. J. Inf. Technol. Web Eng.* **6**(2), 68–75 (2011).
12. R. Janani and S. Vijayarani, "Text classification a comparative analysis of word embedding algorithms," *Int. J. Comput. Sci. Eng.* **7**(4), 818–822 (2019).
13. Z. Wang and Z. Lin, "Optimal feature selection for learning-based algorithms for sentiment classification," *Cogn. Comput.* **12**(1), 238–248 (2020).
14. C. Hu et al., "Low-dose CT with a residual encoder–decoder convolutional neural network (RED-CNN)," *IEEE Trans. Med. Imaging* **36**(99), 2524–2535 (2017).
15. S. Kruthiventi, K. Ayush, and R. V. Babu, "DeepFix: a fully convolutional neural network for predicting human eye fixations," *IEEE Trans. Image Process.* **26**(9), 4446–4456 (2017).
16. M. Inoue, S. Inoue, and T. Nishida, "Deep recurrent neural network for mobile human activity recognition with high throughput," *Artif. Life Rob.* **23**(2), 173–185 (2018).
17. H. Li et al., "A hybrid method coupling empirical mode decomposition and a long short-term memory network to predict missing measured signal data of SHM systems," *Struct. Health Monit.* **20**(4), 1778–1793 (2021).
18. Z. Yunlong and X. Yuanchang, "Travel mode choice modeling with support vector machines," *Transport. Res. Rec.* **2076**(1), 141–150 (2018).
19. K. Kim and E. Lee, "Simultaneous production of transformer insulating oil and value added glycerol carbonates from soybean oil by lipase-catalyzed transesterification in dimethyl carbonate," *Energies* **11**(1), 82 (2017).
20. Y. Kim et al., "Wheel transformer: a wheel-leg hybrid robot with passive transformable wheels," *IEEE Trans. Rob.* **30**(6), 1487–1498 (2014).

Zhan Shi received her master's degree from Dongbei University of Finance and Economics, China. Now, she studies in the Business School at the University of Shanghai for Science and Technology. Her research interests include machine learning, data mining, and text analytics.

Chongjun Fan received his postdoctoral degree from Shanghai Jiaotong University, China. Now, he works in the Business School at the University of Shanghai for Science and Technology as a professor. His research interests include management science and engineering (information management and electronic commerce) and systems engineering (information systems engineering).

Ying Li received her master's degree from Dongnan University, China. Now, she works in Yonyou Network Technology Co., Ltd. as an engineer. Her research interests include programming, machine learning, and data mining.

Hongliu Zhang received her bachelor's degree from Zhongbei University, China. Now, she studies in the Business School at the University of Shanghai for Science and Technology. Her research interests include machine learning, data mining, and systems engineering.