Security Monitoring and Early Warning of Negative Public Opinion on Social Networks Under Deep Learning

Haixiang He* and Shiqi Ma

School of Network Communication, Zhejiang Yuexiu University, Zhejiang 312000, China
E-mail: hedu714@163.com
* Corresponding Author

Abstract

With the continuous development of social networks, negative social network public opinion appears frequently, which is particularly important for its safety monitoring and early warning. Taking Sina microblog as an example, this paper crawled texts from the platform, used BERT to generate word vectors, combined the bidirectional gated recurrent unit (BiGRU) and attention mechanism to design an emotion tendency classification method, and realized the classification of positive and negative emotion texts. Then, TCN was used to predict the negative emotion text to realize public opinion safety monitoring and early warning. It was found that BERT had the best performance. Compared with other deep learning methods, BERT-BiGRUA had a P value of 0.9431, an R value of 0.9012, and an F1 value of 0.9217 in the classification of emotion tendency, which were all the best. In the prediction of negative emotion text, TCN obtained a smaller mean square error and a

higher R^2 than long short-term memory and other methods, showing a better prediction effect. The results verify the usability of the approach designed in this paper for practical safety monitoring and early warning of public opinion.

Keywords: Deep learning, social network, negative public opinion, time convolutional network.

1 Introduction

With the development and popularization of the Internet, the number of Internet users is growing rapidly, and the integration of information technology and traditional media has become higher. With the help of social networking platforms such as Sina microblog and Twitter [1], Internet users can express their opinions anytime, anywhere, conveniently, and quickly. In this case, negative social network public opinions will spread more vigorously and widely, which will challenge the security of the network and even society. Negative social network public opinion has a great influence, and if no active intervention is made within a certain period, large-scale negative public opinion events may occur, which is not conducive to the stable development of society. Therefore, the security monitoring and early warning of negative public opinion on social networks has a very important practical value [2]. Establishing a reasonable and reliable public opinion security monitoring and pre-warning system has become a highly concerning issue in the field of cybersecurity and even social security. With the development of machine learning, deep learning, and other technologies, more and more methods have been applied to public opinion monitoring and early warning [3]. Xu et al. [4] identified the main causes of public opinion fluctuations through the Granger causality test, realized the prediction of public opinion trends based on the gray prediction model, and proved the effectiveness of the method through the analysis of four typical emergencies. Karamouzas et al. [5] designed an automated public opinion monitoring mechanism based on a natural language processing algorithm, used it in the 2016/2020 US presidential election tweet dataset, and achieved a good result. Taking COVID-19 as an example, Wei et al. [6] used the moving average method to analyze the shift in focus of public opinion and used enhanced term frequency-inverse document frequency to extract keywords to establish a word co-occurrence network. The results showed that the focus of online public opinion changed significantly as time went on. Aldiansyah et al. [7] identified the best convolutional neural network

(CNN) model for Twitter sentiment analysis through parametric search and found it achieved an accuracy rate of 88.21% for the experimental dataset. Sina microblog is a widely used social platform at present and also an important carrier for generating public opinion [8]. Taking negative public opinion on the Sina microblog as an example, this paper designed a classification method of sentiment tendency of public opinion based on the deep learning method. The bidirectional gated recurrent unit (BiGRU) was combined with an attention mechanism to classify sentiment for texts. Then, based on this, a time convolutional network (TCN) was used to predict the negative sentiment text. The approach achieved the monitoring and early warning of negative public opinion on social networks. The effectiveness of the proposed method was verified through experimental analyses. This study provides a new and usable method for monitoring and early warning of public opinion. This study is conducive to the correct guidance of public opinion, reasonably regulating Internet users' emotions, and improving the government's ability to cope with negative public opinion, thereby achieving the safe and stable development of cyberspace. This article provides theoretical references for further applications of the deep learning method in the relevant field and offers some bases for deeper research.

2 BiGRU-based Emotional Tendency Classification Method

With the rapid development of social media, social network platforms have become an essential carrier for public opinion events [9]. For an unexpected negative social event or natural disaster, a large number of netizens participate in the discussion of the event on social network platforms, which may cause the spread of a large number of negative remarks in a very short period, resulting in the generation of negative social network public opinion. Negative public opinion events are characterized by suddenness, repetition, and dispersion, and the information disseminated is negative and harmful, which may bring adverse effects on relevant subjects and objects. Therefore, timely intervention is needed.

Take the Sina microblog as an example. Netizens post their views on an event on the platform in the form of a short text, which can be viewed and commented on by other netizens. Meanwhile, netizens can also participate in discussing related events through keywords and topics to form online public opinions. In this process, the text with negative emotions is the key to forming

negative public opinions on social networks. The emotional tendency of the microblog can be classified by deep learning and other methods.

This paper divides the emotional tendency into positive and negative categories. The Python crawler is used to collect microblogs related to research events on the Sina microblog platform. In order to avoid repetition, only the original microblog text is crawled as experimental data. Irrelevant data and information, such as expressions and unreadable code, are removed. Jieba word segmentation is used [10]. Moreover, modal words, pause words, and other data unrelated to the classification of emotional tendency are removed to reduce redundancy. A BERT language model [11] is used to generate word vectors, and then the deep learning approach is used to classify the pre-processed text.

In selecting the deep learning method, this paper uses an algorithm combining bidirectional gated recurrent unit (BiGRU) and the attention mechanism (BiGRUA). GRU is a variant of recurrent neural network (RNN) [12], which has a simpler structure, solves the long-term dependence problem of RNN, and has good performance in data classification and prediction [13]. GRU is mainly composed of reset gate r_t and update gate z_t. The operation process can be written as:

$$r_t = \sigma(W_r \cdot [h_{t-1}, x_t]), \tag{1}$$

$$z_t = \sigma(W_z \cdot [h_{t-1}, x_t]), \tag{2}$$

where σ refers to the sigmoid function, W_r and W_z are the weight matrix of r_t and z_t, h_{t-1} refers to the hidden state at moment $t-1$, and x_t is the input of the current node. Then, candidate hidden state $\tilde{h}_t$ is calculated:

$$\tilde{h}_t = \tanh(W_{\tilde{h}} \cdot [r_t * h_{t-1}, x_t]), \tag{3}$$

where W_h is the weight matrix and $*$ is the matrix product. Hidden state h_t at time t is written as:

$$h_t = (1 - z_t) * h_{t-1} + z_t * \tilde{h}_t. \tag{4}$$

BiGRU adopts the structure of forward layer and backward layer to learn information in two different directions, so as to improve the accuracy of classification. Hidden state h_t of the bidirectional layer at moment t can be written as:

$$h_t = [\vec{h}_t, \overleftarrow{h}_t], \tag{5}$$

where $\vec{h}_t$ and $\overleftarrow{h}_t$ are the input of the forward GRU and backward GRU at time t.

In view of the defect that BiGRU cannot judge which word is more important for the classification of emotional tendency, the attention mechanism [14] is added to BiGRU to classify different weights for words with different importance. h_τ obtained by BiGRU is used as input, and a new vector, u_t, is obtained through a multi-layer perceptron:

$$u_t = \tanh(W_w \cdot h_t + b_2), \tag{6}$$

where W_w and b_w are the weight and bias of attention.

Weight a_t is obtained by calculating with u_t and context vector u_v:

$$a_t = \frac{e^{u_x^\tau \cdot u_v}}{\sum_t e^{u_x^\tau \cdot u_v}}. \tag{7}$$

The final result of attention can be written as:

$$Z = \sum_t a_t h_t. \tag{8}$$

Finally, a softmax classifier is used to classify the sentiment tendency. The formula is:

$$Y = softmax(W_z \cdot Z + b_z), \tag{9}$$

where W_z and b_Z are the weight and bias, Y is the classification result.

3 Safety Monitoring and Early Warning Method Based on TCN for Public Opinion

In the development process of a public opinion event, the number of microblog posts with negative emotions can reflect the accumulation of negative emotions of netizens. Therefore, predicting the change in the number of negative emotional texts according to the time series can realize the safety monitoring and early warning of public opinions, and timely measures can be taken to appease the emotions of netizens and quell negative public opinions before the outbreak of negative emotional texts. A TCN is a method based on a CNN [15] that performs well in predicting time series data, especially in handling long-distance temporal dependencies. Compared to an RNN, a TCN avoids problems of gradient disappearance or explosion, suggesting a better stability. Predicting negative sentiment texts is also a prediction of time series data; therefore, this paper applies TCN to achieve safety monitoring and early warning of public opinion.

TCN combines dilated convolution with causal convolution to obtain dilated causal convolution. It is assumed that there is an input sequence, $X = [x_0, x_1, \ldots, x_{t-1}]$, the calculation formula of dilated causal convolution is written as:

$$F(X) = \sum_{i=0}^{k-1} f_i * x_{t-k-1+i}, \tag{10}$$

where f_i is the filter and k is the filter size. After dilated causal convolution, WeightNorm is used to regularize the network. Then, the rectified linear unit (ReLU) function is used for processing:

$$ReLU = max(0, x) = \begin{cases} 0, & x \le 0 \\ x, & x > 0. \end{cases} \tag{11}$$

Then, the dropout layer is also used to regularize the network and help it converge faster. In the TCN, the residual block structure (Figure 3.1) is used to enhance network performance. The residual block structure can avoid the gradient problem caused by increased network depth, mitigate network degradation, and thus improve the generalization ability of the network.

Finally, the output obtained from the previous layer is sent to the fully connected layer for dimensionality reduction and then the output layer outputs the final result:

$$Y = [y_{t-S+1}, y_{t-S+2}, \ldots, y_{t-1}, y_t].$$

4 Results and Analysis

4.1 Experimental Setup

Data were collected from the Sina microblog with the keyword "Heavy rain in Hebei in 2023". The retrieval time was from July 25, 2023, to August 12, 2023. Forty thousand data were selected for pre-processing and word segmentation and then manually labeled, with positive emotion marked as "1" and negative emotion marked as "0". The training and test sets were divided in a ratio of 8:2.

The experiment was carried out in a Windows 10 environment. The programming language was Python. The development platforms were TensorFlow and Keras. The grid search method was used to determine the parameters. For BERT-BiGRUA, the learning rate was 0.001, the epoch was 15, the batch size was 64, the dropout was 0.2, and an Adam optimizer was
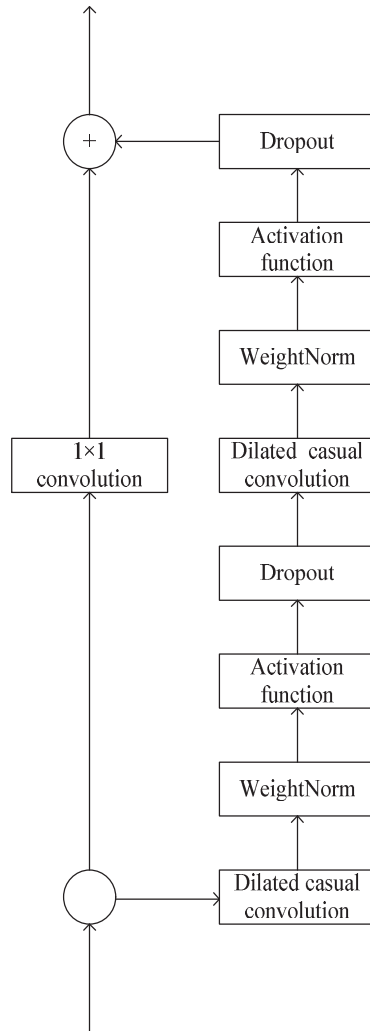
Figure 1   Residual block structure in TCN.

used. For TCN, the size of a convolution kernel was 2×3, the ReLU activation function was used, the epoch was 30, and the dropout was 0.3.

The performance of BERT-BiGRUA on emotion tendency classification for the collected microblog data was analyzed in Experiment 1. Experiment 2 was employed to analyze the performance of TCN in safety monitoring and early warning of public opinion. The data obtained after the emotion

tendency classification in experiment 1 was used to conduct experiments, and the number of positive and negative emotion texts in each hour was counted to generate an experimental dataset.

The evaluation of experiment 1 was based on the confusion matrix (Table 1).

Table 1   Confusion matrix

	Positive Emotion	Negative Emotion
Positive emotion	TP	FP
Negative emotion	FN	TN

TP: The number of positive emotions that are classified correctly
FP: The number of negative emotions that are classified as positive
FN: The number of positive emotions classified as negative
TN: The number of negative emotions that are classified correctly

(1) Precision: $P = \frac{TP}{TP+FP}$
(2) Recall rate: $R = \frac{TP}{TP+FN}$
(3) F1 value: $F_1 = \frac{2 \times P \times R}{P+R}$

The mean square error (MSE) and R^2 were used to evaluate experiment 2.

(1) MSE, used to determine the error between predicted value Y' and true value Y:

$$MSE = \frac{1}{N} \sum_{i=1}^{N} (Y' - Y)^2.$$

(2) R^2, used to measure the predictive power of the model, $R^2 = 1 - \frac{\sum_{i=1}^{N}(Y'-Y)^2}{\sum_{i=1}^{N}(Y'-\bar{Y})^2}$, where $\bar{Y}$ represents the sample average.

4.2 Emotional Tendency Classification Results

In order to determine BERT's performance as a word vector model, the BERT was compared with two other models: word-to-vector (Word2vec) [16] and Glove [17], both of which used BiGRUA for classification. Figure 2 shows the comparison results.

From Figure 2, it can be found that under the same classification algorithm, there was a gap in the effectiveness of the three word-vector models in the classification of emotional tendency. The classification effect of the
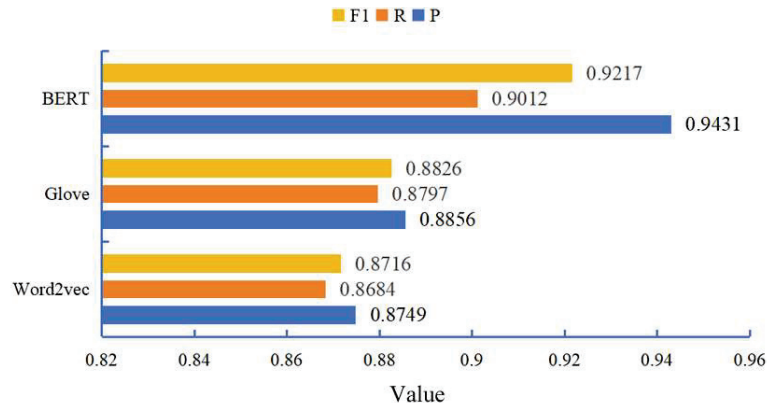
Figure 2   Comparison of different word vector models.

Word2vec model was the worst as its F1 value was only 0.8716, followed by the Glove model with an F1 value of 0.8826. The P value obtained by BERT combined with BiGRUA was 0.9431, the R value was 0.9012, and the F1 value was 0.9217, which increased by 0.0501 compared with the Word2vec model and 0.0391 compared with the Glove model. These results showed that BERT can obtain richer semantic features than Word2vec and Glove and improve the classification effect of emotional tendency. Word2vec does not change with the context, making it difficult to learn global statistical information for complex semantic features in natural language. Glove can utilize local context and consider global statistical information, resulting in better performance than Word2vec. However, it cannot solve the problem of polysemy. BERT is able to dynamically adjust word vector representations, thereby obtaining higher-quality word vectors. The results in Figure 2 also prove the reliability of selecting BERT as the word vector model.

In order to determine the performance of BiGRUA in the classification of emotional tendency, after using BERT to generate word vectors, it was compared with other deep learning methods:

(1) TextCNN [18],
(2) CNN-GRU [19]: a method that combines CNN and GRU,
(3) BiLSTM+Att [20]: a method combining BiLSTM and the attention mechanism.

The results of the comparison are displayed in Figure 3.

It can be seen that the TextCNN method performed poorly in emotional classification, with the P value, R value, and F1 value all below 0.9. The F1
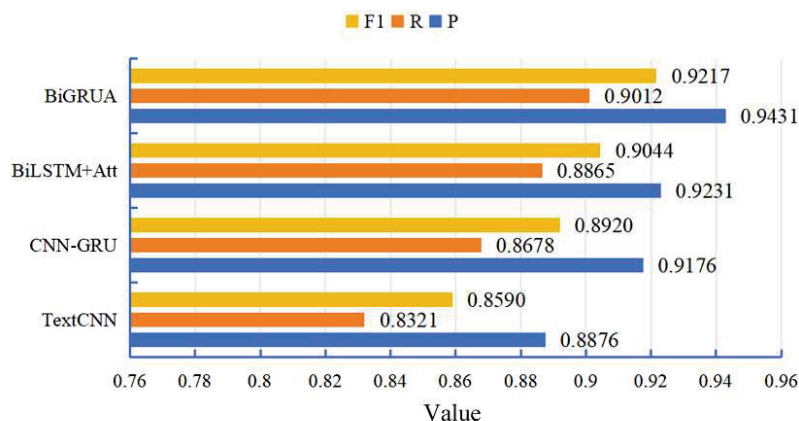
Figure 3 Comparison of different classification methods of affective tendencies.

value obtained by the CNN-GRU method was 0.033 higher than that of the TextCNN method, showing an improvement. The BiLSTM+Att and BiGRUA methods were both a combination of deep learning methods and the attention mechanism. The F1 value of the BiGRUA method was 0.9217, which increased by 0.0173 compared with the BiLSMT+Att method. The results showed that BiGRU is superior to BiLSTM in the classification of emotional tendency and can obtain a better classification effect. Although the TextCNN method has an advantage in extracting local features, it neglects the contextual semantic information before and after, resulting in poor classification performance. The GRU in CNN-GRU can only obtain semantic features from one direction, and its classification results are inferior to the BiLSTM+Att and BiGRUA methods. The results in Figure 3 further demonstrate the superiority of the designed BiGRUA in sentiment tendency classification.

4.3 Results of Safety Monitoring and Early Warning of Public Opinion

The selection of time steps (S) in TCN has a great impact on the predicted results. Different time steps were used for experiments, and the results are shown in Figure 4.

When the time step was 2, the MSE of TCN was 0.0051, and the R^2 was 0.8216. When the time step was 3, the MSE of TCN decreased to 0.0045, and R^2 increased to 0.8233. Then, as the time step continued to increase, the MSE increased again, and R^2 decreased. These results showed that the value
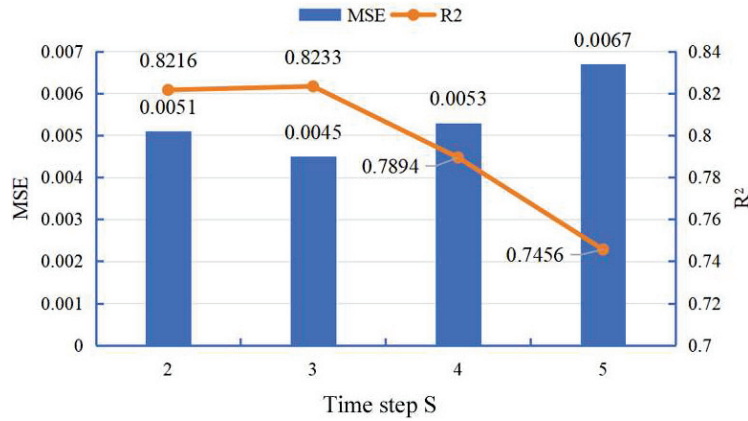
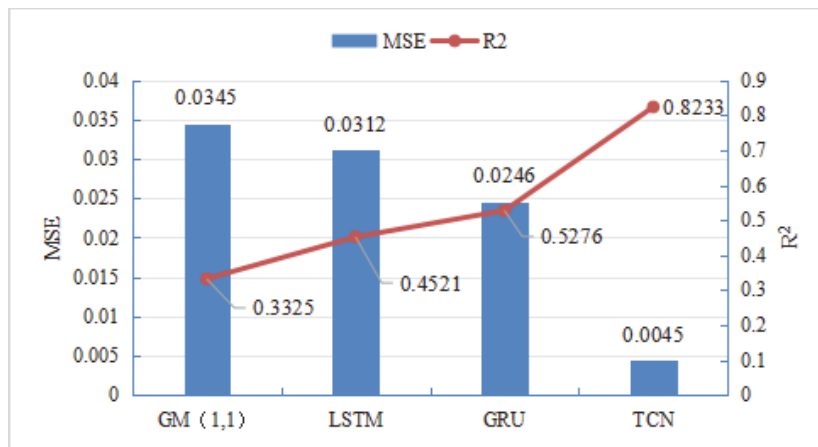**Figure 4**   Comparison of different time steps.

Figure 5   Comparison of different forecasting models.

of S was related to the prediction results. When the time step is 3, TCN can obtain the best effect on the prediction of negative emotion text; therefore, in the follow-up experiment, the time step was set as 3.

TCN was compared with other prediction models, including:

(1) grey prediction model GM (1,1) [21],
(2) LSTM,
(3) GRU.

Figure 5 presents the results obtained.

From Figure 5, it can be found that GM (1,1) had poor performance in predicting negative emotion text, with a high MSE of 0.0345 and a low R^2 of 0.3325, indicating that it is difficult to obtain effective results when applying this method to safety monitoring and early warning. In the comparison of several deep learning methods, the performance of LSTM and GRU was also inferior to that of TCN. The MSE and R^2 of LSTM was 0.0312 and 0.4521. The MSE and R^2 of GRU was 0.0246 and 0.5276. The MSE of TCN was 0.0045, which was reduced by 0.0267 compared to LSTM and 0.0201 compared to GRU. The R^2 of TCN was 0.8233, which was reduced by 0.3712 compared with LSTM and 0.2957 compared to GRU. Compared with other methods, TCN showed greater advantages in predicting time series data. Therefore, applying TCN in security monitoring and early warning of public opinion can achieve good results. Relevant departments can predict the number of negative emotional texts based on TCN. If the predicted results show that there will be more negative emotional texts, relevant measures should be taken in time to reduce the negative emotions of netizens. If the predicted results are within a controllable range, an early warning line is set according to the predicted results. If the actual number of negative emotional texts exceeds the early warning line, early warning and handling will be carried out immediately to avoid escalating the situation.

5 Conclusion

In this paper, by using the deep learning method, BERT-BiGRUA was designed to categorize the sentiment tendency of microblog posts in social network public opinions, and TCN was used to predict the number of negative sentiment texts to achieve safety monitoring and early warning. Through experimental analysis, it was found that the BERT model had a better effect than the Word2vec and Glove models. The F1 value of the designed BiGRUA in the classification of emotional tendency reached 0.9217, which was obviously superior to TextCNN and other methods. TCN also achieved good results in the prediction of negative emotional texts and can be applied in the actual safety monitoring and early warning of public opinion. However, there are also some limitations in this study. For instance, the data source is relatively limited as it only considers the social platform Sina microblog, and the research focuses solely on textual data, which may result in a biased perspective. Therefore, future work will take a more comprehensive approach to studying social network public opinion and incorporate analysis of social data from more modalities to explore further possibilities for optimizing

BiGRUA and TCN, thereby expanding the scope and effectiveness of public opinion security monitoring.

References

[1] A. Dandekar, and V. Narawade, 'Twitter Sentiment Analysis of Public Opinion on COVID-19 Vaccines', Comput. Vis. Robot., 2022:131–139, 2022. https://doi.org/10.1007/978-981-16-8225-4_10.

[2] J. Wang, Z. Liu, and F. Meng, 'Analysis on the Situation of Internet Public Opinion Research', 2020 12th International Conference on Computational Intelligence and Communication Networks (CICN), Bhimtal, India, 416–422, 2020. https://doi.org/10.1109/CICN49253.2020.9242578.

[3] V. Sathya, A. Venkataramanan, A. Tiwari, and D. D. Ps, 'Ascertaining Public Opinion Through Sentiment Analysis', 2019 3rd International Conference on Computing Methodologies and Communication (ICCMC), Erode, India, 1139–1143, 2019. https://doi.org/10.1109/ICCMC.2019.8819738.

[4] L. Xu, J. Qiu, and J. Zhai, 'Trend prediction model of online public opinion in emergencies based on fluctuation analysis', Nat. Hazards, 116:3301–3320, 2023. https://doi.org/10.1007/s11069-022-05808-8.

[5] D. Karamouzas, I. Mademlis, and I. Pitas, 'Public opinion monitoring through collective semantic analysis of tweets', Soc. Netw. Anal. Min., 12(1):91, 2022. https://doi.org/10.1007/s13278-022-00922-8.

[6] H. L. Wei, C. Hai, D. Shan, B. Lyu, and X. Wang, 'Text recognition and analysis of network public opinion focus events of a major epidemic: a case study of 'COVID-19' in Sina Microblogs', Multimed. Tools Appl., 82:25811–25827, 2023. https://doi.org/10.1007/s11042-023-14916-x.

[7] M. R. Aldiansyah, and P. S. Sasongko, 'Twitter Sentiment Analysis About Public Opinion on 4G Smartfren Network Services Using Convolutional Neural Network', 2019 3rd International Conference on Informatics and Computational Sciences (ICICoS), Semarang, Indonesia, 1–6, 2019. https://doi.org/10.1109/ICICoS48119.2019.8982429.

[8] D. Wei, C. Zhao, B. Zhang, J. Ma, Q. Zhu, X. Xu, and S. Wang, 'Analysis of Public Opinion on Weibo and Construction of Knowledge Gragh', 2020 19th International Symposium on Distributed Computing and Applications for Business Engineering and Science (DCABES), Xuzhou, China, 291–294, 2020. https://doi.org/10.1109/DCABES50732.2020.00082.

[9] D. Zhang, Q. Guo, Z. Ning, L. Zhang, and F. Wan, 'Visual Analysis of Public Opinion Based on Keywords', 2021 IEEE International Conference on Power Electronics, Computer Applications (ICPECA), Shenyang, China, 74–78, 2021. https://doi.org/10.1109/ICPECA51329. 2021.9362632.

[10] Q. N. Li, and T. H. Li, 'Research on the application of Naive Bayes and Support Vector Machine algorithm on exercises Classification', IOP Publishing, 2020:1–6. https://doi.org/10.1088/1742-6596/1437/1/012 071.

[11] A. Pardamean, and H. F. Pardede, 'Tuned bidirectional encoder representations from transformers for fake news detection', Indones. J. Electr. Eng. Comput. Sci., 22(3):1667–1671, 2021. https://doi.org/10.11591/I JEECS.V22.I3.PP1667-1671.

[12] Y. Yevnin, S. Chorev, I. Dukan, and Y. Toledo, 'Short-term wave forecasts using gated recurrent unit model', Ocean Eng., 268:1–8, 2023. https://doi.org/10.1016/j.oceaneng.2022.113389.

[13] M. Venkatanaresh, and I. Kullayamma, 'Deep learning based concurrent excited gated recurrent unit for crop recommendation based on soil and climatic conditions', Multimed. Tools Appl., 83(24):64109–64138, 2024. https://doi.org/10.1007/s11042-023-18004-y.

[14] H. Fan, Q. Dong, and N. Guo, 'Surface defect classification of hot-rolled steel strip based on mixed attention mechanism', Robot. Intell. Autom., 43(4):455–467, 2023. https://doi.org/10.1108/RIA-01-2023-0001.

[15] X. Chang, H. Cui, and M. Yang, 'Non-intrusive load decomposition algorithm based on temporal convolutional network and long short-term memory model combined with strongly correlated power sequences', Energy Rep., 9(Suppl.7):511–522, 2023. https://doi.org/10.1016/j.egyr .2023.04.163.

[16] S. Paliwal, A. K. Mishra, and M. N. N. Senthilkumar, 'XGBRS Framework Integrated with Word2Vec Sentiment Analysis for Augmented Drug Recommendation', Comput. Mater. Con., 72(3 Pt.2):5345–5362, 2022.

[17] D. Li, C. He, and M. Chen, 'Text sentiment analysis based on Glove model and United Network', J. Phys. Conf. Ser., 1748(3):1–6, 2021. ht tps://doi.org/10.1088/1742-6596/1748/3/032046.

[18] K. D. Angeli, S. Gao, I. Danciu, E. B. Durbin, X. C. Wu, A. Stroup, J. Doherty, S. Schwartz, C. Wiggins, M. Damesyn, L. Coyle, L. Penberthy, G. D. Tourassi, and H. J. Yoon, 'Class imbalance in out-of-distribution datasets: Improving the robustness of the TextCNN for the classification

of rare cancer types', J. Biomed. Inform., 125:103957–, 2022. https://doi.org/10.1016/j.jbi.2021.103957.

[19] A. Rozental, and D. Fleischer, 'Amobee at SemEval-2018 Task 1: GRU Neural Network with a CNN Attention Mechanism for Sentiment Classification', Int. Workshop Semant. Eval., 2018:218–225, 2018. https://doi.org/10.18653/v1/S18-1033.

[20] L. Zhou, Z. Zhang, L. Zhao, and P. Yang, 'Attention-based BiLSTM models for personality recognition from user-generated content', Inform. Sciences, 596:460–471, 2022.

[21] Y. Ge, and B. Wang, 'Predicting the impact of urban rail transit construction on employment attractiveness based on the GM(1,1) model: A case study of Guangzhou', IOP Conf. Ser.: Earth Environ. Sci., 647(1):1–7, 2021. https://doi.org/10.1088/1755-1315/647/1/012215.

Biographies

Haixiang He, born in January 1980, graduated from Assumption University of Thailand with a doctor's degree in December 2021. He is working at Zhejiang Yuexiu University as a professor. He is interested in network public opinion guidance and disposal.

Shiqi Ma, born in October 1996, is a doctoral candidate and major in communication innovation management. She is working at Zhejiang Yuexiu University as a teaching assistant. She is interested in network public opinion guidance and disposal.