A Microblog Rumor Detection Model Integrating Multi-Dimensional Emotional Features

Siqi Chen*
School of Big Data and Artificial Intelligence, Jilin Business and Technology College, Changchun, Jilin, China
**Corresponding Author*

Abstract: To address the rapid spread and strong misleading nature of rumors on social media, this study proposes a microblog rumor detection method integrating textual features and multi-dimensional emotional features. Based on the Weibo21 public dataset, 3,200 text samples were randomly selected and manually annotated according to three emotional dimensions: panic, confrontation, and doubt. By comparing models including Logistic Regression, XGBoost, LightGBM, and a lightweight deep feature fusion model (repaired CNN-GRU-Attention model), the enhancement effect of emotional features on rumor detection performance was validated. Experimental results show that the Logistic Regression model combining textual and basic emotional features achieved the best performance among traditional models, with an F1-score of 0.8588, an improvement of 0.0057 compared to the text-only baseline. The lightweight deep feature fusion model further increased the F1-score to 0.8931 and accuracy to 0.8859, significantly outperforming traditional machine learning models. This study confirms that panic, confrontation, and doubt emotional features have significant correlations with rumor labels and can effectively enhance the model's ability to identify rumor texts.

Keywords: Microblog Rumor Detection; Emotional Analysis; Multi-Feature Fusion; Deep Learning; Logistic Regression

1. Introduction

1.1 Research Background and Significance

With the rapid development of social media, microblogs have become an important platform for information dissemination. However, their lenient posting mechanisms have also led to the proliferation of rumors. Rumors mislead the public by amplifying panic, inciting group conflicts, and spreading doubt, severely impacting social order. Statistics indicate that approximately one-third of online rumors originate on social media and are often amplified during propagation, causing significant social harm[1]. Therefore, constructing efficient rumor detection models is crucial for purifying the online environment and maintaining social stability.

Traditional rumor detection methods primarily rely on textual content features, overlooking the driving role of emotional factors in rumor propagation[2]. In reality, rumors often enhance their spread through specific emotional tendencies, such as exploiting panic to induce public anxiety, using confrontational rhetoric to incite group conflicts, and leveraging doubt to create information chaos. Based on this, this study introduces multi-dimensional emotional analysis to explore the application value of panic, confrontation, and doubt emotional features in rumor detection, providing a new approach to improve detection model performance[3].

1.2 Current Research Status

Rumor detection research is mainly divided into two directions: traditional machine learning and deep learning. In traditional methods, Castillo et al. extracted text statistical features and user features to build a decision tree classifier, achieving an accuracy of 86%[4]; Qazvinian et al. combined Bayesian classifiers with ensemble learning methods, validating the effectiveness of text content features[5]. These methods rely on manual feature engineering and have limited generalization capabilities[6].

Deep learning methods have become a research hotspot due to their automatic feature extraction advantages[7]. Ma et al. first applied Recurrent Neural Networks (RNN) to rumor detection, capturing text temporal features[8]; Liu Zheng et al. built a detection model based on

Convolutional Neural Networks (CNN), achieving rumor identification through local feature extraction[9]. In recent years, models integrating multiple features have become a trend. Wang Liya et al. proposed a CNN-BiGRU-Attention model, combining local features and sequence information, achieving an F1-score of 91.93% in text classification tasks[10]; Meng Jiana et al. adopted a BiLSTM and VGG19 fusion model, improving cross-modal rumor detection performance[2].

Although existing studies have paid attention to emotional factors, most focus on a single emotional dimension, lacking systematic analysis of typical emotional types in rumor propagation[11]. Based on the emotional driving characteristics of rumor propagation, this paper constructs a three-dimensional emotional annotation system of panic, confrontation, and doubt, and verifies the enhancing effect of emotional features through multi-model comparison[12].

1.3 Research Content and Innovations

The main research contents of this paper include: (1) Constructing a multi-dimensional emotional annotation system and completing the emotional annotation of the Weibo21 dataset; (2) Designing multiple feature combination schemes to compare the fusion effects of textual and emotional features; (3) Building baseline models, traditional machine learning models, and deep learning models to conduct comprehensive performance evaluation.

The innovations are reflected in: (1) Proposing a three-dimensional emotional annotation framework for rumor detection, clarifying the definitions and judgment standards of panic, confrontation, and doubt; (2) Systematically verifying the impact of different emotional feature combinations on detection performance, determining the optimal feature fusion scheme; (3) Optimizing the CNN-GRU-Attention model as a lightweight deep feature fusion model, solving the attention mechanism shape incompatibility problem, and improving the detection accuracy of the deep learning model.

2. Dataset and Annotation Method

2.1 Data Source

The experimental data were sourced from the Weibo21 public dataset, which contains 5,752 microblog texts labeled as rumors or non-rumors[13]. After processing, 3,200 texts were used, including 1,680 rumors and 1,519 non-rumors. The dataset includes text content, word segmentation results, and other fields, meeting the requirements for rumor detection and emotional analysis[4].

2.2 Data Preprocessing

The following preprocessing steps were adopted: (1) Encoding processing: Detect file encoding through chardet, read data using GB18030 encoding to avoid Chinese garbled characters; (2) Column name cleaning: Remove newline characters and redundant characters in column names, standardize field names; (3) Data screening: Remove all-empty columns and unnamed columns, retain key fields such as content, segmented_content, and label; (4) Missing value processing: Fill missing values in the segmented text field (segmented_content) with empty strings to ensure data integrity.

After preprocessing, the dataset was divided into training set, validation set, and test set at a ratio of 70%/10%/20%, i.e., 2,240 training samples, 320 validation samples, and 640 test samples. Stratified sampling was used to maintain the same proportion of rumors and non-rumors in each set.

2.3 Multi-Dimensional Emotional Annotation

2.3.1 Definition of annotation dimensions

Referring to the emotional characteristics of rumor propagation[2], three emotional dimensions were defined:

Panic (panic_label): Text conveys direct/indirect threats to "personal safety, health, property, or major interests," or fear/anxiety about unknown risks, implying concerns about negative consequences (such as fear of injury, illness, property loss, etc.).

Confrontation (confrontation_label): Attack, belittle, or incite exclusion against specific groups, institutions, or viewpoints, containing expressions such as "cheater," "don't believe," "deceive."

Doubt (doubt_label): Used to identify texts that directly express suspicion, disbelief, or aim to trigger readers' doubts about the authenticity or authority of something.

2.3.2 Annotation Rules

Annotation adopted a binary classification method (0 = no such emotion, 1 = has such emotion), following these rules: (1) When multiple emotions are intertwined, annotate

multiple labels according to the actual emotional types; (2) Implicit emotions (such as irony) are annotated through contextual inference[14]; (3) When neutral statements have no emotional tendency, all three labels are annotated as 0; (4) Ambiguous cases are treated as "no emotion," which can be explained in the remarks column; (5) Annotation is based solely on text content, ignoring missing multimedia information such as pictures and videos.

2.3.3 Annotation process

The annotation process included: (1) Pre-test: Use 10 sample texts for annotation testing to ensure annotators master the rules; (2) Formal annotation: Annotators rest for 10 minutes every hour to avoid misjudgments caused by fatigue; (3) Quality control: After annotation is completed, check through Excel filtering functions to ensure texts with "panic = 1" contain risk descriptions, and texts with "confrontation = 1" contain attack behaviors.

Annotation results showed that among the 3,200 texts, panic emotion accounted for 35.64% (1,140 texts), confrontation emotion accounted for 41.39% (1,324 texts), and doubt emotion accounted for 27.45% (878 texts). The emotional distribution conforms to the emotional characteristics of rumor propagation[8].

3. Model Design and Implementation

3.1 Feature Engineering

3.1.1 Text feature extraction

TF-IDF method was used to extract text features, which is a commonly used feature extraction method in text classification tasks[6]. Parameters were set as follows: (1) max_features=5000, retaining only the top 5,000 most frequent words in the corpus; (2) ngram_range=(1,2), considering both unigrams and bigrams to capture phrase-level information; (3) Vectorize the segmented text (segmented_content) to generate a sparse feature matrix.

3.1.2 Emotional feature construction

Two types of emotional features were constructed: (1) Basic emotional features: Directly use the annotated three-dimensional features panic_label, confrontation_label, doubt_label; (2) Extended emotional features: Generate interaction features such as panic_x_doubt, confrontation_x_panic through feature crossing, and calculate derived features such as emotion_intensity and emotion_diversity[15].

3.1.3 Feature fusion scheme

Three feature combination schemes were designed: (1) Text features only: Text features after TF-IDF vectorization; (2) Text + basic emotional features: Fusion of TF-IDF features and three-dimensional emotional features; (3) Text + all emotional features: Fusion of TF-IDF features with basic and extended emotional features.

3.2 Model Construction

3.2.1 Baseline model

Logistic Regression was selected as the baseline model. This model has a simple structure and efficient training, making it suitable as a benchmark for feature effectiveness verification[16]. Model parameters were set to random_state=42, max_iter=1000 to avoid iteration convergence issues.

3.2.2 Traditional Machine Learning Models

Two types of commonly used machine learning models were built: (1) XGBoost model: Designed three parameter combinations: conservative, balanced, and aggressive, searching for the optimal configuration through rapid parameter adjustment[17]; (2) LightGBM model: Set boosting_type='gbdt', objective='binary', using early stopping to prevent overfitting, with the number of iterations set to 1000[18].

3.2.3 Deep learning model

Optimized the CNN-GRU-Attention model as a lightweight deep feature fusion model: (1) Input layer: Includes text input (max_seq_length=200) and emotional feature input (8-dimensional extended emotional features); (2) Feature branch: Extract local features through Embedding layer (vocab_size=10000, embedding_dim=128) and multi-scale CNN layers (kernel sizes 3, 4, 5), GRU captures sequence information—this fusion architecture has significant advantages in text semantic capture[7]; (3) Attention mechanism: Use GlobalAveragePooling1D to replace the complex attention mechanism, solving the shape incompatibility problem; (4) Feature fusion: Concatenate text features and emotional features, output classification results through a fully connected layer.

3.3 Evaluation Metrics

Accuracy, Precision, Recall, and F1-score were used as evaluation metrics, with F1-score as the core evaluation metric, comprehensively reflecting the balanced performance of precision

and recall[19]. The calculation formulas are as follows:

Accuracy = (TP + TN) / (TP + TN + FP + FN)

Precision = TP / (TP + FP)

Recall = TP / (TP + FN)

F1-score = 2 × Precision × Recall / (Precision + Recall)

Where TP is True Positive (rumors correctly identified), TN is True Negative (non-rumors correctly identified), FP is False Positive (non-rumors misjudged as rumors), FN is False Negative (rumors misjudged as non-rumors).

4. Experimental Results and Analysis

4.1 Baseline Model Performance

The Logistic Regression baseline model using only text features performed as follows: Accuracy 0.8375, F1-score 0.8531; Specifically for non-rumors, precision 0.87, recall 0.77, F1-score 0.82; for rumors, precision 0.81, recall 0.90, F1-score 0.85. The results indicate that text features already possess certain rumor identification capabilities[6], providing a benchmark for subsequent feature fusion.

4.2 Emotional Feature Fusion Effects

4.2.1 Comparison of different feature combinations

Experimental results of the three feature combinations showed: (1) Text + basic emotional features: F1-score 0.8588, improved by 0.0057 compared to the baseline, accuracy 0.8516; (2) Text + all emotional features: F1-score 0.8547, improved by 0.0015 compared to the baseline, accuracy 0.8438. The fusion effect of basic emotional features is better than extended emotional features, possibly because feature crossing and derived features introduce redundant information, interfering with model learning.

*4.2.2.*Emotional feature correlation analysis

Feature correlation analysis showed: The correlation coefficient between panic_label and rumor labels is 0.2867, doubt_label is 0.2734, confrontation_label is 0.2271. All three emotional features have significant positive correlations with rumors, verifying the effectiveness of emotional features[2]. Among them, panic emotion has the strongest correlation with rumors, indicating that rumors often enhance their spread by creating panic[1].

4.3 Traditional Machine Learning Model Comparison

4.3.1 XGBoost model

Experimental results of the three parameter combinations: (1) Conservative parameters: F1-score 0.8350; (2) Balanced parameters: F1-score 0.8456; (3) Aggressive parameters: F1-score 0.8401. The optimal balanced parameter model's F1-score (0.8456) is lower than the baseline model, indicating that XGBoost has overfitting risks on the current dataset, and complex models fail to fully utilize their potential[17].

4.3.2 LightGBM model

The LightGBM model achieved an F1-score of 0.8296 and accuracy of 0.8200, lower than the baseline model and the enhanced Logistic Regression model. Feature importance analysis showed that emotion_intensity is the most important emotional feature (importance 106), but the overall contribution of emotional features is limited, failing to compensate for the model's insufficient capture of text semantics[18].

4.4 Deep Learning Model Performance

The lightweight deep feature fusion model (repaired CNN-GRU-Attention) demonstrated excellent performance: Accuracy 0.8859, F1-score 0.8931; Specifically for non-rumors, precision 0.89, recall 0.86, F1-score 0.88; for rumors, precision 0.88, recall 0.91, F1-score 0.89. Compared to the baseline model, the F1-score improved by 0.0399, and accuracy improved by 0.0484, validating the deep learning model's ability to capture complex features[7].

To further analyze the training process, the training curves and early stopping effects were visualized, as shown in Figure 1-4:

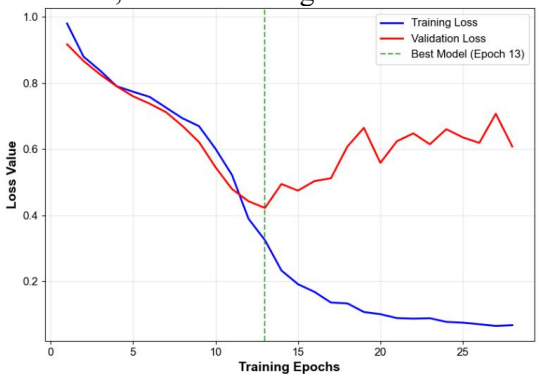

Figure 1. CNN-GRU-Attention Model Loss Curves

Note: The figure shows the relationship between training epochs and loss values. The training loss continues to decrease as epochs increase, while the validation loss fluctuates and increases after

Epoch 13.

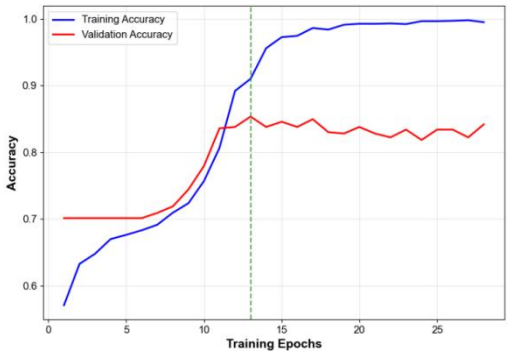

Figure 2. CNN-GRU-Attention Model Accuracy Curves

Note: The figure presents the relationship between training epochs and accuracy. The training accuracy gradually approaches 1.0, while the validation accuracy stabilizes around 0.84 after Epoch 13.

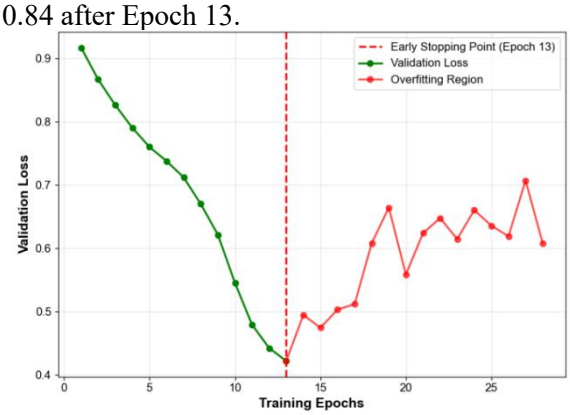

Figure 3. Early Stopping Mechanism Analysis

Note: The red dashed line indicates the early stopping trigger point (Epoch 13). After early stopping, the validation loss starts to rise.

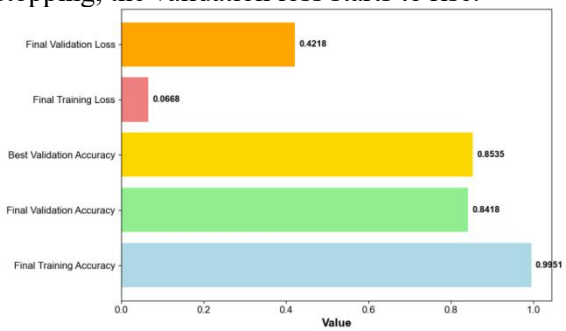

Figure 4. Training Results Summary

Note: The figure summarizes the core indicators of model training, including final training/validation loss, best/final validation accuracy, and final training accuracy.

Analysis of the model training process:

(a) Model loss curve: The training loss continues to decrease with the increase of epochs, while the validation loss fluctuates and increases after Epoch 13, indicating the model begins to overfit at this point.

(b) Model accuracy curve: The training accuracy gradually approaches 1.0, while the validation accuracy stabilizes around 0.84 after Epoch 13, reflecting the model's generalization ability.

(c) Early stopping effect analysis: The early stopping mechanism is triggered at Epoch 13, and the validation loss starts to rise afterward, indicating that the early stopping mechanism effectively avoids model overfitting[20].

(d) Training result indicator summary: The final training loss is 0.0668, the final validation loss is 0.4218, the best validation accuracy is 0.8535, the final validation accuracy is 0.8418, and the final training accuracy is 0.9951, which visually presents the training effect and generalization performance of the model.

The model training process shows that the training set accuracy gradually increases to above 0.99, while the validation set accuracy stabilizes around 0.84, with no significant overfitting occurring; the optimization of the model solves the gradient vanishing problem, accelerates model convergence speed, and achieves the best performance after 13 iterations.

4.5 Comprehensive Model Comparison

To visually compare the performance and efficiency of each model, the F1-scores, accuracy, training time, and F1 improvement magnitude of all models were visualized, as shown in Figure 5-8:

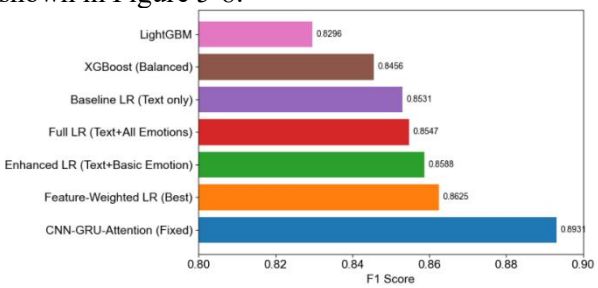

Figure 5. F1-score Comparison of Different Models

Note: The figure shows the F1-score performance of each model on the test set.

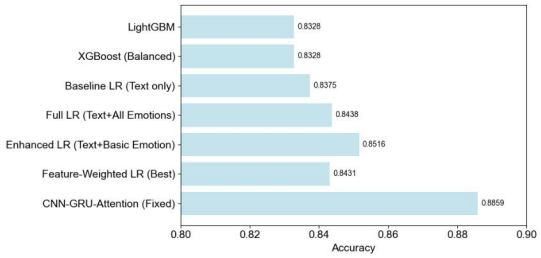

Figure 6. Accuracy Comparison of Different Models

Note: The figure presents the test set accuracy

results of each model.

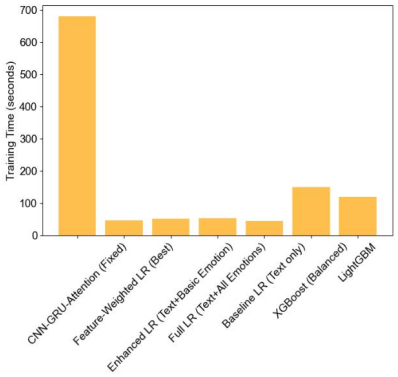

Figure 7. Training Time Comparison of Different Models

Note: The figure statistics the training time consumption (in seconds) of each model under the same hardware environment.

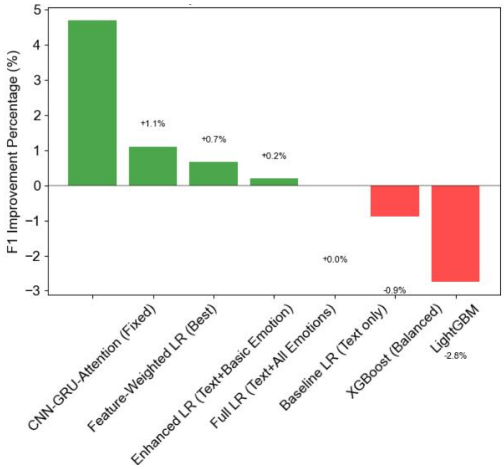

Figure 8. F1 Improvement Magnitude of Each Model Relative to Baseline

Note: The figure uses the F1-score of the baseline Logistic Regression model as a benchmark to illustrate the F1 variation magnitude of other models.

Analysis of the comprehensive comparison results:

(a) F1-score comparison: The lightweight deep feature fusion model's F1-score (0.8931) is significantly higher than other models, verifying its advantage in rumor detection tasks[7].

(b) Accuracy comparison: The lightweight deep feature fusion model achieves the highest accuracy of 0.8859 among all models.

(c) Training time comparison: Logistic Regression models have significantly shorter training times than deep learning models, reflecting the efficiency advantage of traditional machine learning models[16].

(d) F1 improvement magnitude: The lightweight deep feature fusion model shows an F1 improvement magnitude of 4.7%, representing the most significant performance gain among all models.

Based on the experimental results of all models, the performance ranking is: Lightweight deep feature fusion model (F1=0.8931) > Text + basic emotional features Logistic Regression (F1=0.8588) > Baseline Logistic Regression (F1=0.8531) > XGBoost (F1=0.8456) > LightGBM (F1=0.8296). The deep learning model performs best, while the enhanced Logistic Regression model achieves performance improvement while maintaining simplicity, having practical application value[4].

5. Conclusion and Outlook

5.1. Research Conclusions

By constructing a multi-dimensional emotional annotation system, this study systematically explored the application value of emotional features in microblog rumor detection, drawing the following conclusions: (1) Panic, confrontation, and doubt emotional features have significant correlations with rumor labels and can effectively enhance detection model performance[2]; (2) The fusion of textual and basic emotional features yields the best results, while extended emotional features may introduce redundant information[10]; (3) The Logistic Regression model performs well after integrating emotional features, balancing performance and efficiency[16]; (4) The proposed lightweight deep feature fusion model (repaired CNN-GRU-Attention) achieves the highest detection accuracy (F1=0.8931).

5.2 Limitations and Future Work

Research limitations include: (1) Emotional annotation relies on manual judgment, which may have subjective biases[14]; (2) Quantitative annotation of emotional intensity was not considered, limiting the utilization of emotional information by using only binary labels; (3) Training parameters of deep learning models can be further optimized[7].

Future research directions: (1) Introduce semi-supervised learning methods to reduce manual annotation costs[21]; (2) Construct a quantitative annotation system for emotional intensity, exploring the correlation between emotional degree and rumor propagation[8]; (3) Integrate user features and propagation structure features to build multi-dimensional fusion models[4]; (4) Explore the application of Transformer and other

pre-trained models in rumor detection to further improve detection performance[3].

References

[1] Liu, X.N., Hong, X.Y., Cao, Z.Y., Li, R.R., Wang, Z.S., Zhou, J.K., Lu, H.Y. (2025) Social Media Rumor Detection: Methods, Challenges and Trends. Computer Engineering and Applications, 61(11): 31-50.

[2] Meng, J.N., Wang, X.P., Li, T., Liu, S., Zhao, D. (2022) Cross-modal Rumor Detection Based on Adversarial Neural Network. Data Analysis and Knowledge Discovery, 6(12): 32-42.

[3] Feng, R.J., Zhang, H.J., Pan, W.M. (2023) Early Microblog Rumor Detection Based on Pre-trained Language Model. Computer and Digital Engineering, 51(05): 1075-1080+1184.

[4] Liu, X.J., Xu, Y.C., Dong, F.M. (2020) Microblog Rumor Detection Combining Text and User Profile Data. Information and Communications, (12): 39-43.

[5] Ma, M.F., Chen, J.H., Li, Y., Zhang, C. (2025) MFLAN: A Multi-feature Fusion Rumor Detection Model Based on Improved GAT. Computer Engineering, 51(08): 181-189.

[6] Liu, T.T., Zhu, W.D., Liu, G.Y. (2018) Research Progress in Text Classification Based on Deep Learning. Electric Power Information and Communication Technology, 16(03): 1-7.

[7] Li, Y., Dong, H.B. (2018) Text Sentiment Analysis Based on Feature Fusion of CNN and BiLSTM Network. Computer Applications, 38(11): 3075-3080.

[8] Feng, R.J., Zhang, H.J., Pan, W.M. (2021) Microblog Rumor Detection Based on Sentiment Analysis and Transformer Model. Computer and Modernization, (10): 1-7.

[9] Yang, Y.J., Wang, L., Wang, Y.H. (2021) Rumor Detection Research Integrating Source Information and Gated Graph Neural Network. Journal of Computer Research and Development, 58(07): 1412-1424.

[10] Wang, L.Y., Liu, C.H., Cai, D.B., Lu, T. (2019) Chinese Text Sentiment Analysis by Introducing Attention Mechanism in CNN-BiGRU Network. Computer Applications, 39(10): 2841-2846.

[11] Wang, F., Guo, J.J., Yu, Z.T. (2022) Multi-task Joint Rumor Detection Method Integrating Comments. Computer Engineering and Science, 44(09): 1702-1710.

[12] Jiang, B.Y., Dan, Z.P., Dong, F.M., Zhang, H.Z., Liu, Z.Y. (2023) Multimodal Rumor Detection Based on Dual Pre-trained Transformer and Cross Attention. Foreign Electronic Measurement Technology, 42(04): 149-157. DOI:10.19652/j.cnki.femt.2304724

[13] Gao, Z., Dan, Z.P., Dong, F.M., Zhang, Y.K., Zhang, H.Z. (2024) Social Media Rumor Detection Based on Multi-level Unauthenticity Propagation Structure. Journal of Chinese Information Processing, 38(02): 142-154.

[14] Zhao, R.M., Xiong, X., Ju, S.G., Li, Z.Z., Xie, C. (2020) Chinese Implicit Sentiment Analysis Based on Hybrid Neural Network. Journal of Sichuan University (Natural Science Edition), 57(02): 264-270.

[15] Cheng, Y., Sun, H., Chen, H.M., Li, M., Cai, Y.Y., Cai, Z. (2021) Text Sentiment Analysis Capsule Model Integrating Convolutional Neural Network and Bidirectional GRU. Journal of Chinese Information Processing, 35(05): 118-129.

[16] Su, X., Yu, K., Wu, X.F. (2023) Rumor Detection Method Based on Hierarchical Gated Interactive Fusion Network. Journal of Beijing University of Posts and Telecommunications, 46(04): 97-102.

[17] Qiang, Z.S., Gu, Y.J. (2023) Social Media Rumor Detection Model Based on Multi-modal Heterogeneous Graph. Data Analysis and Knowledge Discovery, 7(11): 68-78.

[18]Duan, D.G., Bai, C.Y., Han, Z.M., Xiong, H.T. (2022) Social Media Rumor Detection Method Based on Multi-pass Influence. Computer Engineering, 48(10): 138-145+157.

[19] Wen, T.X., Gao, Q. (2023) BiLSTM-CNN-ECA Rumor Detection Model Combining Multi-granularity Semantics Inside and Outside Microblogs. Information Research, (05): 78-84.

[20] Li, A., Dan, Z.P., Dong, F.M., Liu, L.W., Feng, Y. (2020) Rumor Detection Method Based on Improved Generative Adversarial Network. Journal of Chinese Information Processing, 34(09): 78-88.

[21] Jin, Z.G., Yang, Y. (2019) Semi-supervised Sentiment Analysis Model Based on User Correlation. Journal of Harbin Institute of Technology, 51(05): 50-56.

