*Artificial Intelligence Technologies and Applications*
*C. Chen (Ed.)*


*doi:10.3233/FAIA231401*

# Research on Rumor Detection Based on Sentiment Features

Ling YANG[1]
*Basic Research Institute, Engineering University of PAP Xi'an, Shaanxi, China*

**Abstract.** Due to the Internet and mobile terminal technology's rapid development, significant changes have taken place in culture, politics and social interaction on network. The rapid emergence of new social platforms like Weibo, Blogs and WeChat has steadily altered people's perceptions on information, the lower bar in informational distribution makes it easier for rumors to diffuse. At present, the majority methods employed in rumor detection focused on classifications and regional traits extraction, while ignored the emotional signals in publishers and receivers. So, we proposed an improved model based on dual sentiment features and commentators, then embedded in CNN, RNN and BERT network for rumor detection. These given results have demonstrated that the methods mentioned above are higher than 80% in terms of accuracy and F1. Besides, it also reflects a higher accuracy and better detection effect, compared with the single semantic detection model.

**Keywords-**component; rumor detection; false news; deep learning; feature extraction; dual sentiment features; pre-trained model

## 1. Introduction

Nowadays, network has become the main channel for people to impart information which informed a new type of interactive relationship [1]. On top of this, more netizens are used to capturing information, expressing opinions and participating in public topical conversations. WeChat, microblog and mobile client have become the originated and distribution center for mainstream media, which replaced the traditional media communication platforms. By taking advantages in characteristics of openness, concealment and wide spreading, some people with ulterior motives also make rhetoric in distorting facts, spreading false information, disturbing public views and misleading public opinions.

### *1.1 The Features in Text Sentiment*

Based on the new form of communication, the emergence of online media mentioned above has greatly shortened the dissemination cycle of information, and broke the single communication framework. As the detection model like BERT, GTP gains more maturity in use, it shows a marked improvement in model complexity and detection efficiency. While, these types of methods rely heavily on training model, it could cause the problem

[1] Corresponding author: Ling YANG, Basic Research Institute, Engineering University of PAP Xi'an, Shaanxi, China; email: 983275839@qq.com

in mismatch by pretraining and fin-tuning because of the missing MASK. Meanwhile, models often ignored the hidden emotional features contained in text when extracting eigenvalues.

*1.2 Trend in Rumor Detection*

Observe the impact of false news and achieve detection with high accuracy is also significant. These approaches can curb the generation of fake news, block the way of transmission, which contributes to the diminution of negative impact for society. The model like Deffuant, Krause-Hegselmann, and SIR lays the basis for rumor propagation research. 2009, Zanette. D. H [2], successfully brought in the concept of complex network, which verified the efficiency of modeling analysis. Pieri F [3] has enriched the detection content of Weibo and conducted rumor analysis, depending on the user relationship, comment content and forwarding rate. In order to excavate text features deeply, the scholars spend more effort on sentiment features. On the research of Twitter, Hamidian [4] mentioned that the features in reliability and emotion plays a significant role in analysis testing. 2018, Kim Jha [5] proposed the sentiment features in complex domains and built a semantic rule based on sentiment aware dictionary. Hannah Kim [6] raised an emotional classification based on CNN network and applied to rumor detection. The methods mentioned above have conducted in-depth study of blog information by establishing emotional dictionary and analyzing methods of extracting emotional features. Although it reflects the real structural characteristics of blog to some degree, but the fusion in text content and emotional signals was still passable.

## 2. The Improved Sentimental Feature Model

*2.1 The detection model based on sentiment feature*

Here, we proposed a detection model based on dual sentiment features which concentrated on the extraction analysis of Publisher and Social emotion. Firstly, validated the effectiveness of dual sentiment features by the comparison among baseline models, such as Emoratio [7], EmoCred [8], DTC [9] and SVM-TS [10]. Secondly, put the Text and Comment content into BIGRU, CNN and BERT to gain corresponding semantic features. Then, generated the feature vectors by splicing fusion of dual sentiment and semantic ones mentioned above. Finally, put them into MLP, made classifications by SoftMax fully connected layer, test evaluation metrics and validated effectiveness, as listed in Figure 1.

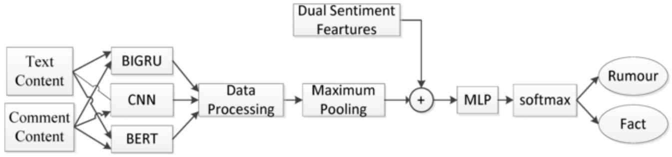


**Figure 1.** The network structure detection in dual sentiment detection

*2.2 The extraction of dual sentiment feature*

In this paper, we divided the content of Blog into two pieces: one is the Publisher Emotion that embedded emotions in text content, the other is Social Emotion by data preprocessing and feature extraction. Finally, we got the dual sentiment features by splicing two vectors mentioned above, as shown in Figure 2.

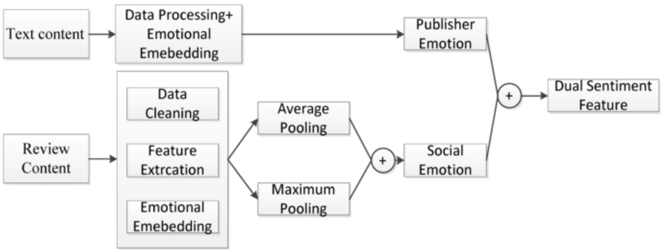


**Figure 2.** The structure in dual sentiment feature

*2.3 The Feature Extraction of Publisher Emotion*

Each blog can truly reflect the message and distinctive emotion conveyed by publishers. Here, we defined four metrics for fine-grained analysis in overall information and symbolic-level as follows: Emotion Class, make a sentiment classification on Weibo_16 dataset by using pre-trained model, then get the output feature $em_{PT}^{class}$, PT indicates the content of each blog, class indicates classifier; Emotion Degree, select the benchmark words in both positive and negative, then analyze the emotion degree by the closeness between emotion and reference words, Pwords and Nwords indicates the positive and negative emotion words respectively, for example

$$em_{PT}^{degree} = \frac{\sum_{pword_i \in Pwords, i=1}^{m} Sim(word, pword_i)}{m} - \frac{\sum_{nword_j \in Nwords, k=1}^{n} Sim(word, nword_j)}{n} \tag{1}$$

Emotion Score, here we used the sentiment dictionary of Dalian Technology University and the sentiment classification model in NLTK to analyze and calculate the positive and negative degree of the text, list the output feature as $em_{PT}^{score}$; Additional Features, emotions, punctuation frequency can assist in capturing non-lexical sentiment information, list the output feature as $emo_{PT}^{add}$. Above all, the feature of Publisher Emotion can be connected by the markers above, for example

$$emo_{PT} = emo_{PT}^{class} \oplus emo_{PT}^{degree} \oplus emo_{PT}^{score} \oplus emo_{PT}^{add} \tag{2}$$

*2.4 The Feature Extraction of Social Emotion*

Here, we made a comprehensive review as $C = \{C_1, C_2, C_3, \dots C_L\}$, $C$ stands for the toral review content, $C_i$ stands for the $i$-th message in the comment set. Firstly, as the measure of Publisher Emotion listed above, get the row vector $emo_{c_i}$ by calculating each emotion score of review, then obtain an approximate value $emo_C$ after transposed sentiment vectors for each review, for example

$$\widehat{emo_C} = \widehat{emo_{C_1}^T} \oplus \widehat{emo_{C_2}^T} \oplus \dots \oplus \widehat{emo_{C_L}^T} \tag{3}$$

In order to refine the sentiment feature vector, we imported the pooling operation to compress and filter the vectors: Average Pooling is used to extract the average sentiment information in comments, Maximum Pooling is used to extract the maximum value to extreme emotions in comments, for example

$$emo_C^{mean} = mean(\widehat{emo_C}) \tag{4}$$

$$emo_C^{max} = max(\widehat{emo_C}) \tag{5}$$

Overall, the feature in social emotion consists of the above two parts, for example

$$emo_C = emo_C^{mean} \oplus emo_C^{max} \tag{6}$$

*2.5 The Feature Extraction of Dual Sentiment*

The input parameters of the experimental model contain pre-processed semantic and dual sentiment feature vectors. Firstly, Combined the semantic features with reviewer vectors, then extracted the corresponding semantic features by the model in BIGRU, CNN and BERT. Thirdly, got the integrated emotion features after classification and extraction by Average Pooling and Maximum Pooling. Finally, set MLP and SoftMax layer to extract features and preform rumor classification, for example

$$R_{BIGRU} = SoftMax[MLP(BIGRU_C, emo_C)] \tag{7}$$

$$R_{CNN} = SoftMax[MLP(CNN_C, emo_C)] \tag{8}$$

$$R_{BERT} = SoftMax[MLP(BERT_C, emo_C)] \tag{9}$$

## 3. Experimental Results and Analysis

*3.1 Experimental environment*

The environmental configuration of this research listed in Table 1.

**Table 1** Experimental Environment

| Environment | Configuration |
|---|---|
| **CPU** | Inter Core i5-1135G7@2.40Ghz |
| **RAM** | 16GB |
| **OS** | Windows 11 |
| **Development Platform** | PyCharm 3.7 |
| **Deep Learning Framework** | Pytorch |

### *3.2 Data Prepossessing*

This research used the public Weibo-16 as experimental datasets with labels and contents. Due to the high repetition rate of the subset, we performed a filtering operation on the dataset to get Weibo_16_1, including 26285 contents and 1812351 reviews. Then divided the datasets into training and testing at the ratio of 6:4, as listed in Table 2.

**Table 2** weibo_16_1 dataset

| **Weibo_16_1** | **Training Set** | | **Testing Set** | |
|---|---|---|---|---|
| **Item** | **Content** | **Review** | **Content** | **Review** |
| **Fact** | 715 | 615537 | 474 | 415943 |
| **Rumor** | 15275 | 465129 | 13057 | 320487 |

The main processes include Text Cleaning, Word segmentation, and Spelling Correction. In order to understand the distribution in key words and high frequency words in dataset, we called the matplotlib to plot word clouds for analysis.

### *3.3 Sentiment Feature Categories Comparison Experiment*

In order to detect the experimental influence from different sentimental feature categories, we set Publisher Emotion, Social Emotion and dual sentiment into 4 layers MLP, then made a comparison with the base line rumor detection model in Emoratio, DTC, SVM-TS. The model mentioned above refers to earlier experimental research [11].

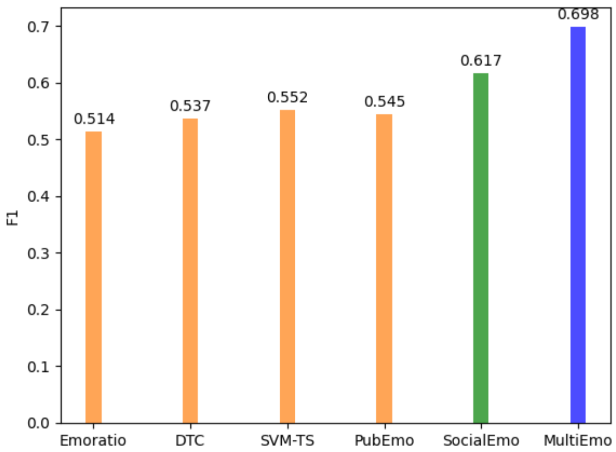


**Figure 3**. $F_1$ comparison of sentiment feature

Figure 3 shows the comparison results of the earlier different models as well as the baseline features in Weibo_16_1. The bars in yellow stand for the base line models of F1 scores, the green one stands for the model in Social Emotion, while the blue one is the Multi Emotion combined with Social and Publisher Emotion. As the trend shown in F1, it can be found that the social sentiment feature and dual sentiment ones proposed in this paper are better than Emoratio, DTC and SVM-TS. The detection effect of Publisher Emotion is slightly higher than Emoratio and DTC models; The F_1 Score of the Social Emotion is about 7% higher than the Publisher ones, which registered 6%-10% growth compared with the traditional model; The F_1 score of MultiEmo is significantly higher than the traditional baseline model, which increased by 15%. Meanwhile, the effect shown in dual sentiment feature is better than the Publishers and Social ones, also increased by 15.3% and 8.1%. Overall, we conclude than the dual sentiment features proposed in this paper is far more important in modeling. In the follow-up study, we should take MultiEmo as the main emotion feature, combined with BIGRU and CNN for comparative research.

### *3.4 Comparative Experiment Based on Dual Sentiment Features*

Due to the relatively single class of features in MLP, it can't take full advantage of contextual information in extracting hidden features, that caused the lack of representativeness. After verified the necessity of dual sentiment, the models in BIGRU, CNN and BERT were used to extract the semantic features from texts, also embed with Emoratio, DTC and MultiEmo to emotional signals above. Then, classified by MLP and SoftMax, as shown in Table 3, Table 4 and Table 5, Figure 4, Figure 5 and Figure 6. The item listed with BIGRU, CNN and BERT stand for the model without dual sentiment features which lies in Table 3, Table 4 and Table 5. From Tables mentioned above, we can see that the deep learning- model used in this paper performs an outstanding characterization in terms of Accuracy, $F_1$ and Recall compared with Emoration and DTC, also achieved the expected detection goals. This is because the earlier traditional models were based on Machin Learning, the selection and extraction of key features have been poorly targeted and insensitive. On the contrary, the algorithm based on neural network can analyze the text features effectively, which can be applied to classification tasks.

**Table 3** Multi-feature detection results in bigru

| Weibo_16_1 | Accuracy | F1 Score | Recall |
|---|---|---|---|
| **BIGRU** | 81.7% | 0.794 | 0.786 |
| **Emoratio** | 77.8% | 0.763 | 0.751 |
| **DTC** | 75.6% | 0.729 | 0.713 |
| **MultiEmo+BIGRU** | 82.5% | 0.831 | 0.844 |

**Table 4** Multi-feature detection results in cnn

| Weibo_16_1 | Accuracy | F1 Score | Recall |
|---|---|---|---|
| **CNN** | 83.4% | 0.812 | 0.786 |
| **Emoratio** | 83.7% | 0.811 | 0.801 |
| **DTC** | 82.9% | 0.846 | 0.851 |
| **MultiEmo+CNN** | 87.3% | 0.862 | 0.859 |

**Table 5** Multi-feature detection results in bert

| Weibo_16_1 | Accuracy | F1 Score | Recall |
|---|---|---|---|
| **BIGRU** | 90.3% | 0.895 | 0.871 |
| **Emoratio** | 82.5% | 0.817 | 0.819 |
| **DTC** | 80.3% | 0.791 | 0.785 |
| **MultiEmo+BERT** | 96.4% | 0.944 | 0.931 |

On the basis of Figure4, Figure5 and Figure6, we can conclude that the way mixed by dual sentiment features and BIGRU, CNN and BERT played an important role in testing index from Weibo_16_1, compared with the base-line model. In terms of accuracy, the model based on BIGRU, CNN and BERT show great outstanding performance in $F_1$ rate, which is 10.2%, 1.5% and 15.7% more than before. As the evaluation analysis introduced large amount of feedback in Weibo event, the whole process of the diffusion of information could be reflected more authentic, which provides a richer base for effective analysis of sentiment features. Therefore, the detection performance is much more efficient.

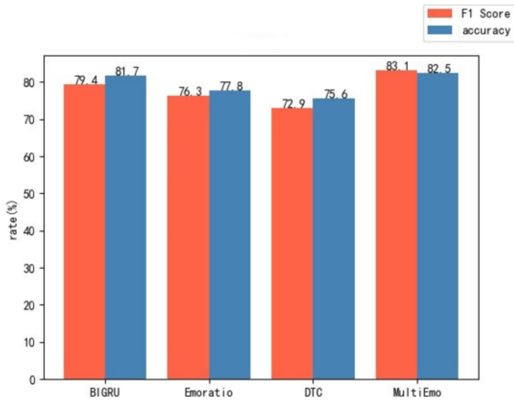


**Figure 4.** The comparison of BIGRU

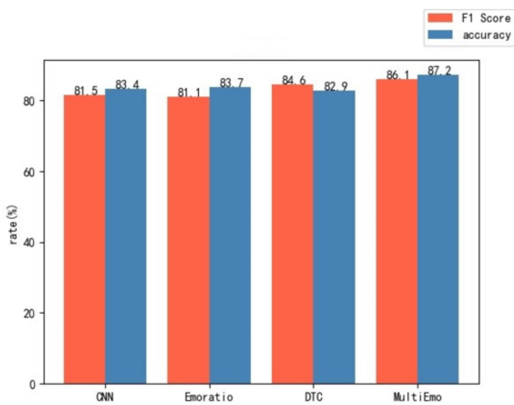


**Figure 5**. The comparison of CNN

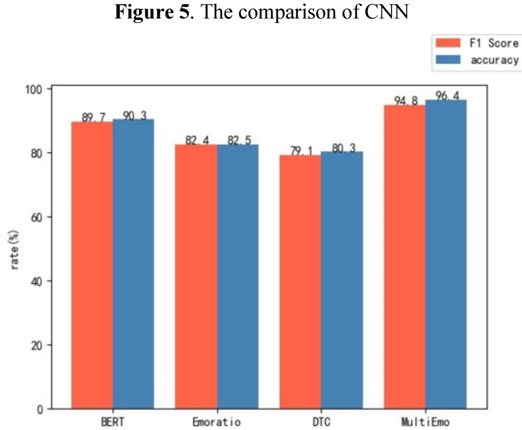


**Figure 6.** The comparison of BERT

Our experiments show that dual sentiment feature would be more significant in BIGRU, CNN and BERT than before. While, compared with BIGRU, CNN and BERT, we proposed the rumor detection model base on dual sentiment features increased by 3.7%, 4.6% and 5.1% on $F_1$ for data sets, this confirms the necessity of the introduction of dual sentiment features to the rumor detection on social platforms, especially for optimized BERT, which played the best performance on inspection.

## 4. Conclusion

This research concludes the network structure in analyzation and detection by dual sentiment features, examines the feature exaction methods in Blog content and Social Emotion. Then, analyzed the index of training model from Weibo_16_1 after filtered and reconstructed. By ablation study, we made a comparison among baseline, neural network and BERT model, the experiments have shown that Publisher and Review Emotion is also the key issue for rumor detection.

## References


[1] Guo Chen, How to build and guide public opinion in the area of mobile Internet[J].2021(2020-23):16-18.

[2] Zanette D H. Dynamics of rumor propagation on small-world networks [DB/PL]. [2009-03-01]. http://arxiv.org/PS_cache/cond-mat/pdf/0110/0110324v1.pdf.

[3] Pieri F, Ceri S. False news on social media: a data-driven survey[J]. ACM Sigmod Record, 2019, 48(2):18-27.

[4] Hamidian S, Diab M. Rumor identification and belief investigation on Twitter[C]. Proceedings of the 7th Workshop on Computational Approaches to Subjectivity, Sentiment and Social Media Analysis, San Diego, 2016, 3-8.

[5] J ha V, R S, Shenoy P D, et al. A novel sentiment aware dictionary for multi-domain sentiment classification[J]. Computers & Electrical Engineering, 2018,69:585-597.

[6] Hannah Kim, Young Seob Jeong. Sentiment Classification Using Convolutional Neural Networks[J]. Applied Sciences, 2019,9(11):2347.

[7] Ajao O, Bhowmik D, Zargari S. Sentiment aware fake news detection on online social networks[C]. ICASSP 2019-2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). IEEE, 2019: 2507-2511.

[8] Giachanou A, Rosso P, Crestani F. Leveraging emotional signals for credibility detection[C]. Proceedings of the 42nd International ACM SIGIR Conference on Research and Development in Information Retrieval. 2019: 877-880.

[9] Popat K. Assessing the credibility of claims on the web[C]// Proceedings of the 26th International Conference on World Wide Web Companion. 2017:1-7.

[10] Ma J, Gao W, Wei Z, et al. Detect rumors using time series of social context information on microblogging websites[C]// Proceedings of the 24th ACM international on conference on information and knowledge management, 2015:1751-1754.

[11] Yuan C, Ma Q, Zhou W, et al. Jointly embedding the local and global relations of heterogeneous graph for rumor detection[C]//2019 IEEE International Conference on Data Mining (ICDM). IEEE, 2019:796-805.