Multi-domain rumor detection method based on sentiment features and attention mechanism

Wen Jiang[1], *Mingshu* Zhang[2,*], *Bin* Wei[2], *Kelan* Ren[1], *Facheng* Yan[1], and *Xiong* Zhang[1]

[1]Armed Police Engineering University, Xi 'an, Shaanxi Province
[2]Key Laboratory of Network and Information Security of Armed Police Force, Xi 'an

Keywords: Rumor detection, Emotion feature, Domain embedding, Attention mechanism.

Abstract. The wide dissemination of rumor is increasingly threatening both individuals and society. In this paper, we propose a multi-domain rumor detection method based on sentiment features and attention mechanism. This method uses sentiment analysis technology to extract sentiment features from text, and uses attention mechanism to weight text features and domain features to detect rumor. Experimental results show that the proposed method has achieved good results on data sets from multiple fields and has good generalization performance.

1 Introduction

With the continuous development of social networks, rumor in different fields on social media is widely spreading. [1] Many areas of the real world, such as politics, are threatened by rumor. Traditional rumor detection methods are usually limited to a single domain or source, and it is difficult to train data from other domains.[2] Some researchers try to use cross-domain datasets to verify the applicability of the model in different domains, and multi-domain rumor detection has attracted more and more attention.

Sentiment feature is a very important feature in rumor detection.[3] rumor often tends to use extreme or emotional words to strike a chord with readers in order to achieve wide dissemination. By analyzing the emotional polarity of the text, we can capture this emotional language use, and then assist in judging the authenticity of the news. The same word may express different emotions in different fields of news, which mainly depends on the context of the word and the specific content described. For example, the word *change* may be used to express positive sentiments in news in the political domain, where a politician may say, Our government is making changes towards a more open and democratic way to meet the needs of the people. The *change* in this sentence conveys a positive sentiment, smelling of improvement and moving forward. In financial news, however, the same word *change* can take on different emotional overtones. The industry is undergoing tremendous change, with many companies facing transformation and layoffs,

* Corresponding author: ZMS2099@163.com

an economic commentator might write. *change* here implies market uncertainty and the risk of people losing their jobs, and therefore carries a more negative sentiment. Therefore, we need to be aware that the same word may carry different emotions in different fields of news.

Based on the above problems, this paper proposes a multi-domain rumor detection model based on sentiment features and attention mechanism. Firstly, the sentiment features and text features of multi-domain news texts are extracted as the input layer of the model together with domain features, and high-dimensional feature representations are obtained through convolutional neural networks. Then, the collaborative attention mechanism is combined to focus on capturing the emotional word weights and text feature weights of news in different fields. Finally, it is input into the fully connected layer to realize softmax classification, improving the rumor detection effect of the whole model.

2 Related work

2.1 Multi-domain rumor detection

At present, the research on multi-domain rumor detection is still in its initial stage. In order to fully capture the characteristics of text in different fields, researchers continue to explore more effective detection methods. Among them, Ran et al. [4] established a cross-domain learning model to learn embedded features from specific domains and cross-domains respectively. Then, the two features are combined for rumor recognition. Soft combination of multi-domain features is better than coarse decoupling of domain-shared features and domain-specific features. Nan et al. [5] proposed a multi-domain rumor detection model, which uses nine expert networks (TextCNN) to obtain different domain representations of each news, and then inputs them into a "domain gate" to aggregate multiple features for classification. However, news domain labels are not absolute, meaning that news articles in a specific domain also have characteristics of other domains, such as the use of words. Zhu proposed a memory-guided multi-view multi-domain rumor detection framework to model news segments from multiple perspectives, including semantics, sentiment, and style.[6]

2.2 Emotion feature

Emotional texts have different emotional tendencies, such as positive, negative, neutral, etc, especially for texts in different fields, different emotional words often represent different emotional characteristics. By pre-constructing a dictionary containing positive, negative and neutral words, and then according to the frequency or weight of sentiment words in the text, the text sentiment classification is performed. [7] However, this approach often ignores the differences of sentiment words in texts from different domains. To solve this problem, some research works focus on domain-adaptive sentiment lexicon construction. That is, when processing the sentiment text in a new domain, we first analyze the common sentiment words in the domain, and then construct the sentiment dictionary suitable for this domain according to these words. This method can better capture the sentiment characteristics of the new domain text and improve the accuracy of sentiment analysis.

2.3 Attention mechanism

Attention mechanism has become one of the most important concepts in the field of deep learning.[8] It is inspired by the human biological system, which tends to focus on unique parts when processing large amounts of information. With the development of deep neural networks, attention mechanism has been widely used in different fields.

Due to its unique functions, the field of natural language processing has begun to use this technology on a large scale to improve experimental performance. In 2014, the team at Google first used attention for image classification, thus opening the door to its popularity. Later, Bahadanau et al[9] also applied this method to machine translation, which promoted the development of attention mechanism in text processing, and the attention mechanism greatly improved the efficiency and accuracy of perceptual information processing.

3 Multi-domain rumor detection model based on Sentiment features and attention mechanism

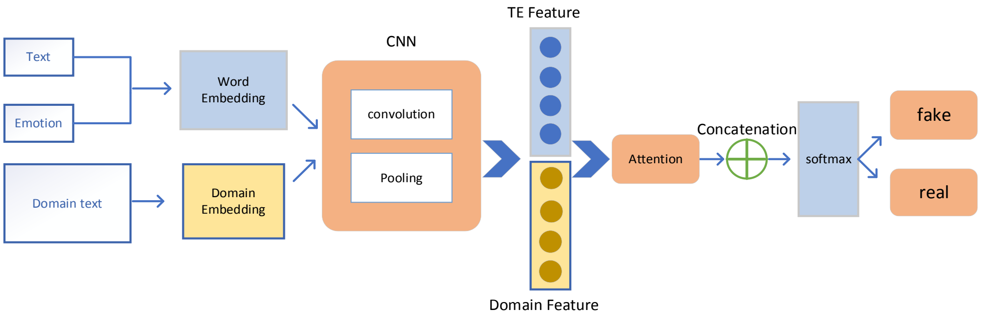

Fig. 1. TED-CNN-ATT Model.

3.1 Model introduction

Figure 1 is the structure of the TED-CNN-ATT model proposed in this paper, and the overall process of the model is shown in the following steps:

(1) The text information is transformed into word vector, and the word vector representation combining the text T and the sentiment feature E is used as the input layer, and the domain feature D is input into the CNN model.

(2) Give full play to the role of CNN model. Firstly, multiple convolutional layers are used to extract the feature information of the text, and then the feature dimension is reduced through the pooling layer to extract the key text feature information.

(3) The information obtained in the previous step is input into the attention layer ATT, and the results are weighted to distinguish the importance of different context information.

(4) The above results are input into the Softmax classifier to realize the true or false judgment of rumor detection.

3.2 Word embedding

The CBOW model of Word2vec[10] is selected for word embedding r , so that each word is mapped to obtain a vector with dimension 100. If the length of the rumor text l_r is more than the length of the rumor text, it is truncated, and the zero vector is added to obtain the vector matrix corresponding to the first rumor. The vectors j corresponding to the rumor text in the event are concatenated to form a word vector matrix $r_j = [r_1, r_2 \cdots r_{l_j}]$ of the rumor event. In the process of domain embedding, the news text is divided into 9 domains, which are science and technology, military, education, accident, politics, health, finance, entertainment, and society. The data set is divided into 9 domains, and the accuracy of false news detection can be further improved by combining text and emotional features.

3.3 Emotion feature extraction

Suppose a rumor consists of M sentences, where each sentence consists of m words and T^i is the first i sentence in the rumor, $T^i = [x_1^i, x_2^i, x_3^i, \cdots x_m^i]$

In order to fully express the sentiment, various sentiment features will be extracted from the rumor text, including sentiment category, sentiment vocabulary, sentiment intensity, sentiment score, and other auxiliary features. Among these five features, sentiment category, sentiment intensity, and sentiment score basically define the overall sentiment information of rumor, and the other two provide information at the lexical and symbol level.

(1) Emotion Category: Emotion classifier can effectively help it identify emotion types in a more detailed and accurate way. Here, Baidu AI model is used to directly output category features in rumor T .

(2) Emotion Lexicon[11]: high-frequency emotion words can express the specific emotion of rumor, which is realized by using the emotion lexicon annotated by experts. There is rumor T , t_i is the i word in T , and there are d_e emotional categories in the sentiment dictionary. $E = \left\{E_1, E_2, \cdots E_{d_e}\right\}$

Consider the frequency of occurrence of t_i in the sentiment lexicon and the semantic influence of hyponyms, such as degree words and negative words, which have the effect of changing the direction of sentiment (such as "unhappy", "very fond"). Secondly, the degree value of negative words and degree words in the sentiment dictionary and the probability of matching to the sentiment word are used to calculate:

$$s(t_i, e) = neg(t_i, w) * \deg(t_i, w) / L \tag{1}$$

Calculate the number $s(T, e)$ for all the words in a given sentiment e , adding up the scores for each word in the sentiment range:

$$s(t, e) = \sum_{i=1}^{L} s(t_{i,e}), \forall e \in E \tag{2}$$

W represents the window size, $neg(t_i)$ and $\deg(t_i)$ represent the negative word degree value and the negative word dimension value respectively:

$$neg(t_{i,w}) = \prod_{j=i-w}^{i-1} neg(t_j) \tag{3}$$

$$\deg(t_{i,w}) = \prod_{j=i-1}^{i-1} \deg(t_j) \tag{4}$$

Concatenate the scores for all the sentiment texts:

$$emo\frac{lex}{T} = s(T, e_1) \oplus s(T, e_2) \oplus \cdots s(T, e_{d_e}) \tag{5}$$

(3) Emotion Category: The process of extracting word intensity features is similar to the process of extracting word intensity features, which involves intensity scores. Ecstasy is higher than happiness; First, we compute the intensity-aware text-level score:

$$S'(T, e) = \sum_{i=1}^{L} s(t_{i,e}) = \sum_{i=1}^{L} \mathrm{int}(t_i) * s(t_{i,e}), \forall e \in E \tag{6}$$

$\mathrm{int}(t_i)$ is the sentiment strength score of the word t_i in the dictionary, or $\mathrm{int}(t_i)=0$ if t_i is not in the dictionary. The final emotion intensity feature is obtained by concatenating the intensity perception scores:

$$emo\frac{\mathrm{int}}{T}=s'(T,e_1)\oplus s'(T,e_2)\oplus\cdots s'(T,e_{d_e}) \tag{7}$$

(4) Sentiment Score: Typically, the sentiment score is a positive or negative value that represents the degree of positive and negative polarity of the entire text, and it can be calculated by using a sentiment lexicon or public toolkits.

(5) Other Auxiliary Features: Considering that the above features do not explicitly utilize information outside the sentiment lexicon, we introduce a set of auxiliary features to capture the sentiment signals of non-word elements, including emojis, punctuation, sentimental words, personal pronoun frequencies. Take emoticons and punctuation, which are universal symbols for emotional expressions around the world, such as ":)", "! Such punctuation can convey people's emotions. The feature results of the above five aspects are used to connect with each other to extract emotional features:

$$emo_T=emo_T^{cate}\oplus emo_T^{lex}\oplus emo_T^{\mathrm{int}}\oplus emo_T^{senti}\oplus emo_T^{aux} \tag{8}$$

3.4 Convolutional neural network

As one of the most commonly used detection models, Convolutional Neural Network (CNN) has been proposed to consider both local and global features[12]. The working process of the convolutional layer is to first select the input data with the same size as the convolution kernel, multiply the position elements checked by the convolution, and then sum and add the bias, using the resulting numerical value as a feature. Then, the convolution kernel slides over the input data according to the set step size until it covers all the data, and finally gets all the feature information of the input data. The convolutional layer uses a window that slides up and down to automatically extract local features to form a feature map. The general sliding window is 2 to 5 words, and the purpose of using multiple convolution kernels is to be able to extract richer feature information from multiple angles and further improve the effectiveness of the model. The specific process is calculated as follows:

$$f=g(W\oplus X+\mathrm{b}) \tag{9}$$

In this formula, g represents the Relu activation function[13], W represents the weight matrix, $\oplus$ represents the convolution operation, X represents the event vector matrix, b represents the bias term, and f represents the feature map obtained by convolution of the event vector matrix. In this paper, the two-dimensional text feature matrix is input into the multi-scale CNN, and two parallel convolution kernels are set to extract the features of two channels of different sentences of the event.

3.5 Attention mechanism combined with domain features

The working process of attention mechanism is mainly divided into three steps. The first step is to calculate the connection degree between the current word and the preceding and following text information to obtain the attention weight. The second step is to process the output of the previous layer and normalize it to obtain the text weight information. The third step performs the final weighted sum of the results and the Value obtained from the previous layer. Attention mechanism is essentially the process of giving different weights to each part of the text information, which can be understood as a mapping between a query (Q) sequence and a series of key-value (K-V) pairs.[14] It is calculated as follows:

$$Attention(K,Q,V) = soft\max(\frac{QK^T}{\sqrt{d}})V \tag{10}$$

Where, $Q \in R^{n\times d}$, $K \in R^{m\times d}$, $V \in R^{m\times d}$

The weight results calculated in the previous step are normalized by using the softmax function. The larger the output result is, the more important the information indicated by the current word is, which may directly affect the detection results of rumor.

4 Experimental results and analysis

4.1 Dataset

In this paper, we set two datasets, Twitter16 and Weibo-20. Among them, Twitter16 is a benchmark dataset for Chinese rumor detection, and each item is labeled as rumor or non-rumor [15]. It contains a total of 1355 rumors and 2351 non-rumors, which are divided into training set and test set. Weibo-20 is the data crawled by Q Nan et al from the microblog community management center from December 2014 to March 2021[16]. For each piece of data, the team collected multiple dimensions of information, including text content, caption, timestamp, comments, and refutation information. After deduplicating and cleaning, 4,488 rumor and 4,640 real news were finally obtained.

4.2 Evaluation index

This paper introduces confusion matrix and Accuracy, Precision, Recall and F1 are used as evaluation indicators.

Accuracy represents the proportion of the number of samples with correct classification in all samples.

$$Accuracy = \frac{TP+FN}{TP+FP+TN+FN} \tag{11}$$

Precision represents the proportion of the number of correctly classified positive samples to the number of all positive samples in the classifier.

$$Precision = \frac{TP}{TP+FP} \tag{12}$$

Recall represents the proportion of the number of correctly classified positive samples to the number of positive samples.

$$Recall = \frac{TP}{TP+FN} \tag{13}$$

F1-score represents the weighted harmonic average of accuracy rate and recall rate, where P represents accuracy rate and R represents recall rate.

$$F_1 = \frac{2*P*R}{P+R} \tag{14}$$

4.3 Experimental results and analysis

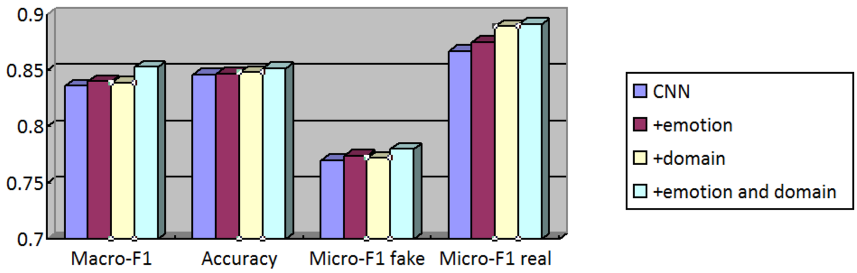

Fig. 2. A comparison of the results on Twitter16.

Figure 2 shows the comparison of the results of the four models. On the Twitter 16 dataset, the proposed model method of sentiment features and domain features has higher accuracy and higher F1 score than the other three methods. For a single CNN model, it has a certain detection effect, but adding sentiment features and domain features on the basis of CNN model can better analyze text information. In terms of macro F1, the proposed method has 2.5% improvement over a single CNN model, 0.7% improvement in accuracy, and 2.1% improvement in micro f1 value. Compared with the CNN model with emotional features, the proposed model has 1.9% improvement in macro F1 score, 0.6% improvement in accuracy, and 1.2% improvement in micro F1 score. Compared with the CNN model with domain features, the proposed method has 1.7% improvement in macro F1 score, 0.5% improvement in accuracy, and 1.1% improvement in micro F1 score. It can be seen that this method has a good role in detecting rumors.

Table 1. Results on Twitter16.

Model	Macro-F1	Accuracy	Micro-F1	
			fake	real
CNN	0.824	0.845	0.762	0.886
+emotion	0.830	0.846	0.771	0.888
+domain	0.832	0.847	0.772	0.890
+emotion and domain	0.849	0.852	0.783	0.893

Table 1 fully shows the experimental results, and the superiority of the proposed method can be better seen through clear data comparison.

Table 2. Results on Weibo-20.

Model	Macro-F1	Accuracy	Micro-F1	
			fake	real
CNN	0.836	0.846	0.770	0.867
+emotion	0.840	0.847	0.774	0.875
+domain	0.839	0.848	0.772	0.889
+emotion and domain	0.853	0.852	0.780	0.891

Table 2 and Figure 3 show the effect of the proposed method on the microblog dataset. Weibo 20 is a data set containing 9 domain labels, which highlights the domain coding and can further reflect the domain information of the text in the process of rumor detection. From the results, the detection effect is good. In terms of macro F1 value, the single cnn model detection value is 0.836, which is lower than the detection effect of CNN model with emotional features and domain features. In terms of accuracy, the proposed method is 0.6% higher than the single CNN model, and 0,5% higher than the CNN model with emotional features, indicating that emotional features play an important role in the process of rumor detection. On micro F1, the proposed method outperforms the CNN model with domain features by 0.8%, and domain features also play a good role in the rumor detection process.

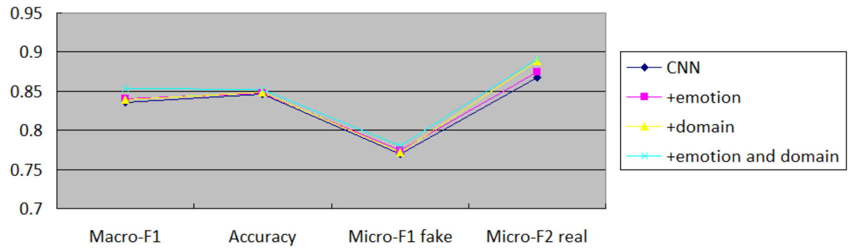

Fig. 3. A comparison of the results on Weibo-20.

5 Conclusion

The multi-domain false information detection model based on emotional features and attention mechanism proposed in this paper can play the role of emotional features and domain features, which plays an important role in improving the accuracy of rumor detection. In the next step, we should continue to mine multi-domain false information detection methods to improve the scientific nature of social information.

This research was funded by the National Social Science Foundation of China (20BXW101, 18XXW015). Innovation Research Project for the Cultivation of high-level Scientific and Technological Talents (Top-notch Talents of the Discipline) (ZZKY2022303)

References

1. Ma J , Li J , Gao W ,et al.Improving Rumor Detection by Promoting Information Campaigns with Transformer-based Generative Adversarial Learning[J].IEEE Transactions on Knowledge and Data Engineering,2021.
2. Zhou H , Ma T , Rong H ,et al.MDMN: Multi-task and Domain Adaptation based Multi-modal Network for early rumor detection[J].Expert Systems with Application, 2022.
3. Min Y, Qiang Q, Chen X J,et al. Feature-enhanced attention network for target-dependent sentiment classification[J] .Neurocomputing,2018.
4. Hongyan Ran and Caiyan Jia. Unsupervised Cross-Domain Rumor Detection with Contrastive Learning and Cross-Attention. [C]//AAAI Conference on Artificial Intelligence .2023
5. Nan Q, Cao J, Zhu Y, et al. MDFEND: Multi-domain fake news detection[C]//Proceedings of the 30th ACM International Conference on Information & Knowledge Management,2021.
6. Zhu Y C ,S Q and C J ,et al. Memory-Guided Multi-View Multi-Domain Fake News Detection [C]//IEEE Transactions on Knowledge and Data Engineering,2022.

7. Jia K L and Li Z N. Chinese Micro-Blog Sentiment Classification Based on Emotion Dictionary and Semantic Rules.[C]//International Conference on Computer Information and Big Data Applications,2020.
8. Niu Z Y, Zhong G. and Hui Y .A review on the attention mechanism of deep learning.[J]Neurocomputing ,2021.
9. Bahdanau D , Cho K , Bengio Y .Neural Machine Translation by Jointly Learning to Align and Translate[J].Computer Science, 2014.
10. Liu Y , Xu C , Kuai X ,et al.Analysis of the causes of inferiority feelings based on social media data with Word2Vec[J].Scientific Reports,2022.
11. Badaro G , Jundi H , Hajj H ,et al.ArSEL: A Large Scale Arabic Sentiment and Emotion Lexicon[C]//The 3rd Workshop on Open-Source Arabic Corpora and Processing Tools co-located with LREC,2018.
12. Na B, Wang Z X,Meng F R.A Stochastic Attention CNN Model for Rumor Stance Classification[J] IEEE Access,2020.
13. Bai Y. RELU-Function and Derived Function Review.[C]//SHS Web of Conferences,2022.
14. Ashish V , Noam M. Shazeer ,et al .Attention is All you Need.[J]Neural Information Processing Systems,2017.
15. Zhang X, Cao J, et al. Mining dual emotion for fake news detection.[C]//Proceedings of the Web Conference,2021.
16. Nan Q, Cao J,Zhu Y C, et al. MDFEND: Multi-domain Fake News Detection [C]//International Conference on Information and Knowledge Management ,2021.