DDCO model based false news detection research

Jing Zhao[1], *Bin* Wei[2,*] and *Yandong* Sun[2]

[1]Xi'an Key Laboratory of Human-Machine Integration and Control Technology for Intelligent Rehabilitation, Xijing University, Xi'an, China
[2]Engineering University of PAP, Xi'an, China

Keywords: Cooperative attention mechanism, DDCO, False news detection.

Abstract. With the rapid development of the information age, while the popularity of social media brings great convenience, it also brings some negative effects, such as the spread of false news. At present, the identification of fake news is still based on the personal screening ability, therefore, the intelligent and information-based automatic detection algorithm has become one of the hot issues of current research. Based on the characteristics of DCAN and DEFEND models, this paper proposes an novel model DDCO, which uses multi-layer collaborative attention mechanism to extract the most relevant information from the three dimensions of sentence level, word level and sentence-comment level respectively. Finally, the model designed in this paper is tested on Weibo and Twitter data sets, and the results show that the DDCO has a higher accuracy than the existing models, which provides an important reference for false news detection.

1 Introduction

With the rapid development of the information age, the network has become an important channel for information transmission. However, with the increasing speed and influence of information dissemination the network fake news [1] has become a major social problem.

At present, there are more and more studies on false news, including its causes, communication channels, detection and treatment, consequences and effects. Content-based detection requires that the document be long text in order to better learn the representation of words and sentences. However, social media is often short text, which leads to sparse data. In addition, some models require that each news item collect user reviews to understand the forwarder's comments. However, most users on social media tend to simply forward without commenting. Some models use cascading information paths to classify information, but for privacy reasons, many users hide or delete records of social

* Corresponding author: weibin82@126.com

interactions, limiting the application of the algorithm. Ma et al. use RNN to detect rumors, however, this network takes a long time to learn, which is not conducive to real-time detection. Lecun proposed the LeNet-5 model [2], which contains convolution, pooling, and full connection layers, that is CNN.

To sum up, the development of false news detection based on natural language processing can be roughly divided into three stages. (1)There is no uniform way to represent text. One-hot model can be used to express words [3], N-gram model [4] can predict text, but it cannot represent the similarity between words. (2)The launch of Word2vec is a milestone. It can be trained to get word vectors to represent the similarity between words. A large number of word vector methods, such as Glove or FastText, can then be used to derive vector features. (3)Stage 3: ELMo[5] is a two-way model that distinguishes polysemous words and generates word vectors based on the context of the text. GPT uses the Transformer [6] framework, which is a one-way language model for classification, similarity, and other operations. BERT is almost the same architecture as GPT.

This paper takes false news detection technology based on synergistic attention mechanism as the research object, puts forward a DDCO model, designs a three-tier structure of synergistic attention, expands sentence level to word level, makes the model finer granularity, and finally makes a synergistic attention analysis on sentence-comment. DDCO has performed experiments on two open datasets, Weibo-20 and Twitter15, and compared them with DCAN and DEFEND models. The results show that this method has higher accuracy than the existing models, which provides an important reference for false news detection.

2 False news detection algorithm based on three-layer cooperative attention mechanism

DEFEND uses RNN to model text information and commentary information, and then uses Co-attention to fuse the two information to determine whether news is false or not. The model improves detection by capturing the correlation between news content and user comments. DCAN is composed of five main components: news content encoder, user comment encoder, gating mechanism, content comment concentrating and word level concentrating.

By comparing the above two models, we can see that the attention mechanism plays an important role. This paper improves the attention level, puts forward three attention mechanisms: sentence-level synergistic attention, word-level synergistic attention and sentence-comment synergistic attention. At the same time, the gating mechanism is used to filter the input text for efficient detection.

Since sentences are the most intuitive reflection of news content information, this model first conducts sentence-level co-attention analysis, codes news sentences on this level, and obtains implicit information in sentences through sentence-level co-attention, which is the first interaction. Secondly, it is the second interaction to extend the sentence level to the word level, from the relevance of information conveyed by long sentences to the word level, and further focus on key words through the attention mechanism. Finally, the extracted information is input into the sentence-comment co-attention layer through the design of the first two layers, to study the degree of correlation between sentences-comments, to analyze the internal correlation between news sentences and user comments. Finally, the authenticity of the text is predicted. The model structure is shown in Figure 1.

Suppose for a news article $s=\{s^1,s^2,...,s^N\}$, including N sentences in the article, each sentence $S^i=\{w_1^i,w_2^i,...,w_n^i\}$ contains n words. $C=\{c^1,c^2,...,c^m\}$ is a comment set containing m related comments, in each comment $c^j = \{w_1^j, w_2^j, ..., w_m^j\}$ contain m words. Fake news detection can be viewed as a binary classification problem, where each news article can be true (y=1) or false (y=0).

2.1 Coding layer

2.1.1 Content coding

News content refers to the text information, pictures, videos and other modal information contained in the article. In this paper, GRU is used to model the word sequence. For one of the news content, the output formula as follows:

$$h_i = [\vec{h}_1 \oplus \vec{h}_2] \tag{1}$$

Because it contains contextual information, it can also be used to represent information throughout a sentence. To distinguish the importance of each word, this paper introduces the attention mechanism to learn the importance of words. Through the attention mechanism, sentence vectors $s \in p^{2d\times1}$ can be obtained as follows:

$$s^i = \sum\nolimits_{t=1}^{n} a_t^i h_t^i \tag{2}$$

Where, a_t^i stands for the importance of the word t^{th} to news content S. a_t^i is calculated as follows:

$$u_t^i = \tanh(W_w h_t^i + b_w) \tag{3}$$

$$a_t^i = \exp(u_t^i u_W^T) \Big/ \sum\nolimits_{k=1}^{n} \exp(U_k^i u_W^T) \tag{4}$$

Where, u_t^i is obtained from the hidden state h_t^i by fully embedding the layer, $W_w b_w$ is a trainable parameter.

2.1.2 Comment code

Most of the existing false news detection algorithms only focus on the news content, and few researchers take comments into account. However, these contents may provide a lot of information about their opinions on the news and help to improve the performance of the false news recognition algorithm. Therefore, this paper codes the comments to find out the important characteristics of comments for false news detection.

This paper uses two-way GRU to get forward and backward hidden states, and then connects those two states to get word annotations. Then, the attention mechanism has been used to get the vector. Set the comment vector as c^j , and the formula is:

$$c^j = \sum\nolimits_{t=1}^{m} \beta_t^j h_t^j \tag{5}$$

Where, β_t^j used to measures the importance of the word t^{th} to comment on c^j , for β_t^j :

$$\beta_t^j = \frac{\exp(u_t^j u_c^T)}{\sum_{k=1}^{m} \exp(u_k^j u_c^T)} \tag{6}$$

$$u_t^j = \tanh(W_c h_t^j + b_c) \tag{7}$$

In the formula, u_t^j represents the hidden representation after sending the hidden state h_t^j to the fully embedded layer, and u_c represents the weight.

2.2 Attention layer

The fusion optimization of the attention layer is a key part of the DDCO model. It uses a layer-by-layer filtering optimization method, which allows text information to be filtered step by step from the lower input layer to the higher input layer.

The first step is to input the news text into the sentence level to get the best relevant sentences. The second step is to input the most relevant sentences from the previous level into the next level, which expands the sentence level vector to the word level to extract the word level features. Step 3 sentence-comment co-attention.

2.2.1 Sentence-level co-attention

There are always words with greater meaning and less meaning in each text. Noise in user reviews can interfere with model judgment. In order to better play the role of sentence-level synergistic attention mechanism, this paper chooses meaningful comments through the relevant semantic affinity of content and comments.

Construct a feature matrix $S = [s^1, s^2, ..., s^V] \in R^{2d \times V}$ that reflects the content and a feature matrix $M = [\overline{m}^1, \overline{m}^2, ..., \overline{m}^T] \in R^{2d \times T}$ that represents user comments.

The affinity matrix $W \in R^{T \times v}$ is calculated as follows:

$$X = F(S)^T CF(M) \tag{8}$$

Where, C is an n-layer feed-forward neural network function. After obtaining the maximum values of rows and columns of the affinity matrix, the attention mechanism is introduced to weight the original content and comments. Thus, the original feature information of news content and user comments and the feature information of common concern can be retained.

2.2.2 Word-level co-attention

For more detailed information, this paper expands the sentence level to the word level, which is similar to the construction of sentence-level co-attention, and is calculated as follows:

$$u_{it} = \tanh(W_w h_{it} + b_w) \tag{9}$$

$$s_i = \sum_t a_{it} h_{it} \tag{10}$$

Where, u_{it} is the hidden representation of h_{it}, a_{it} is the normalized weight factor processed by the softmax function, and s_i is the vector representation of the expected *i-th* sentence.

2.2.3 Sentence-comment co-attention

Each sentence has a different importance in determining whether the news is false or not. Similarly, the same applies to user comment. Therefore, this paper uses Sentence-Comment Co-attention to capture the semantic affinity of news content and comments, decompose attention into two different feature codes, and learn the attention weights between sentences and comments.

The feature matrix $S=[s^1,s^2,...,s^V]\in R^{2d\times V}$ reflecting the content and the feature matrix $M=[\overline{m}^1,\overline{m}^2,...,\overline{m}^T]\in R^{2d\times T}$ representing the user's comments are constructed, and then the affinity matrix $F=\tanh(M^T W_l S)$, where W_l is the weight matrix, is calculated.

The attention weights for sentences and comments are calculated as follows:

$$a^s = soft\max(W_{hs}^T H^s) \tag{11}$$

$$a_m = soft\max(W_{hc}^T H^m) \tag{12}$$

Where, W_{hs} and W_{hc} are weight parameters.

Its weighted sum is:

$$\widehat{S} = \sum_{i=1}^{N} a_i^s s^i \tag{13}$$

$$\widehat{M} = \sum_{j=1}^{T} a_j^m m^j \tag{14}$$

3 Experiment and analysis

3.1 Experimental data

The Twitter15 dataset contains information about rumor content, user reviews, and the corresponding sequence of forwarded users. The Weibo-20 was built by Zhang [7]. The dataset contains three parts: rumor content, user reviews and tags.

3.2 Experimental results and analysis

First, to verify the validity of this model, we performed experiments on the model DDCO on two datasets, Weibo-20 and Twitter15, and compared the results with those of DCAN and DEFEND, which are shown in Table 1.

Table 1. Performance comparison of different models under the same data set.

Methods	Weibo-20				Twitter15			
	accuracy	precision	recall	F1	accuracy	precision	recall	F1
DCAN	0.669	0.671	0.622	0.645	0.772	0.763	0.763	0.761
DEFEND	0.703	0.710	0.632	0.671	0.771	0.820	0.661	0.730
Ours	0.793	0.732	0.930	0.812	0.824	0.821	0.832	0.820

From the table, we can see that this model achieves the best results on both datasets compared with DCAN and DEFEND. The experimental results show that the three-co-attention architecture can mine data correlation from a deeper level. In addition, the performance of the model on Chinese-English datasets is different, and the validation effect on English datasets is more obvious, which is determined by the complexity of the dataset, and the structure of Chinese components is more complex than that of English one.

Secondly, in order to further verify the applicability of this model in different scenarios, we have carried out experiments on the Weibo-16 dataset. Figure 2 shows the experimental results of this model on three datasets. On the Twitter15 dataset, its precision, recall, f1-score are all above 0.8, while the loss rate is below 0.4, which is nearly 15%-20% lower than that under Weibo-20 and Weibo-16 datasets. On the other two Chinese datasets, comparing the experimental results of Weibo-20 and Weibo-16 datasets, we can find that the model under Weibo-20 dataset has a higher accuracy rate of 80%, while the loss rate is relatively low. This shows that although both datasets are Chinese, their performance is slightly different due to their different data, capacity and text information. The above experimental results show that the model designed in this paper behaves slightly differently on different datasets, but on the whole, it can be applied to different application scenarios.

Thirdly, in order to analyze the effects of model parameters on the results, we experimented with the effects of two parameters, Batch Size and learning rate. As shown in Figure 3, when the learning rate is 0.0038, the change in the result is observed by changing the Batch Size value. When Batch Size was 32, the loss rate was 0.55 and the accuracy rate was 0.75. When Batch Size was 64, the loss rate was 0.57 and the accuracy rate was 0.74. When Batch Size was 128, the loss rate was 0.54 and the accuracy rate was 0.736. Thus, when Batch Size is 32, the DDCO model has the best performance, with a relatively high accuracy and a relatively low rate of test loss.

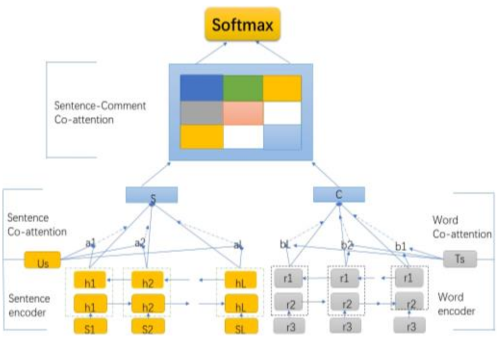

Fig. 1. DDCO model.

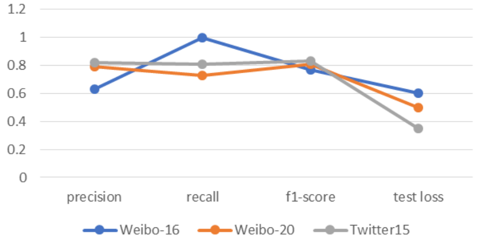

Fig. 2. Experimental results on three data sets.

As shown in Figure 4, when Batch Size is 32, the experiment results vary according to the learning rate. When the learning rate was 0.001, the loss rate was 0.58 and the accuracy rate was 0.71. When the learning rate was 0.01, the loss rate was 0.52 and the accuracy rate was 0.79. When the learning rate was 0.03, the loss rate was 0.57, and the accuracy rate was 0.73. Thus, the DDCO model works best when the learning rate is 0.01.

4 Conclusion

Fake news detection has become one of the research hotspots in the field of NLP. In this paper, to overcome the shortcomings of existing false news detection methods, a multi-layer detection model based on attention mechanism was proposed. In the process of building the model, the attention level is extended to three levels, which makes use of the three synergistic features of sentence level, word level, news content and comments to extract more accurate information, and improves the fineness and reliability of experimental data. Experiments on open datasets demonstrate that this model is advanced.

Fig. 3. Effect of learning rate on results.

Fig. 4. Effect of learning rate on results.

This study was supported by the Basic research plan of Natural Science in Shaanxi Province in 2021(Grant No. 2021JQ-878).

References

1. Deng Yijiao, Zhang Fengli, Chen Xueqin, etc Collaborative attention network model for cross-modal retrieval [J] Computer Science, 2020, 47 (04): 54-59
2. LeCun Y, Bottou L, Bengio Y, et al. Gradient-based learning applied to document recognition [J]. Proceedings of the IEEE, 1998, 86(11): 2278-2324.
3. Li J, Si Y, Xu T, et al. Deep Convolutional Neural Network Based ECG Classification System Using Information Fusion and One-Hot Encoding Techniques[J]. Mathematical Problems in Engineering, 2018: 155-218.
4. Yin Chen, Wu Min Overview of N-gram model [J] Computer System Application, 2018,27 (10): 33-38
5. Li Zheng, Chen Li, Zhang Shuang Emotional analysis of Chinese text based on ELMO and Bi-SAN [J] Computer Application Research, 2021, 38 (08): 2303-2307
6. Fan Wei, Liu Yong Social network information dissemination prediction based on space-time transformer [J] Computer Research and Development: 2022 (04): 1-12
7. ZHANG X, CAO J, LI X, et al. Mining dual emotion for fake news detection[C]//Proceedings of the Web Conference 2021. 2021: 3465-3476.