Rumor detection based on topic classification and multi-scale feature fusion

Tan Li¹, Ma Zihao¹, Cao Juan², Lv Xinyue¹

¹School of computer and information engineering, Beijing University of technology and industry, Beijing 100048
²Institute of computing technology, Chinese Academy of Sciences, Beijing 100190

Abstract. In recent years, with the rapid development of Internet technology, the spread of network rumors has become one of the important obstacles to maintain the stable development of social networks and ensure the public security. Most of the existing researches focus on the detection of rumors in general fields, ignoring the differences among different fields. According to the characteristics of rumor in the health field, this paper proposes a rumor detection method based on topic classification and multi-scale fusion. Different methods are used to extract features from different sub datasets of different scales, taking into account the overall, inter topic, and intra subject correlation and differences, and then judge after feature fusion. The experimental results show that this method is better than the general detection method in the data set of health field, and has some improvement compared with the algorithm in the same field.

1. Introduction
With the rapid development of social network, a large number of health information such as diet and health care burst out, seriously affecting people's daily life. According to the report on the governance of Internet rumors in 2019, in 2019, wechat platform released 17881 anti rumor articles, of which the health category is the disaster area of rumor propagation. For example, novel coronavirus outbreak in 2020, various rumors also broke out. Conniving at the spread of rumors in the field of health will lead to a series of social events, public panic, economic losses and so on.

At present, the existing social platform rumor detection is still mainly manual. With the increasing number of rumors, it is of great significance to automatically detect rumors in specific fields. To solve these problems, this paper constructs a health data set based on daily express. Through the multi-level analysis of data sets, it is found that there are great differences among the topics. Therefore, a rumor detection method based on topic classification and multi-scale feature fusion is proposed. Firstly, MLM and Bi LSTM + attention are used to extract features from different sub datasets, and then multi-scale algorithm is used to extract word level features. After normalization, feature fusion is carried out. Finally, neural network is used to output the prediction results.

2. Related work
Yang et al. [1] proposed to use SVM model to carry out automatic detection experiments based on traditional event content and user information, combined with event propagation characteristics, client type and event geographical location. Liang et al. [2] The rumor is detected by introducing user behavior features, and the traditional machine learning algorithm is applied to the comparative experiment. Shu et al. [3] According to the potential characteristics of users, this paper proposes a joint detection of false accounts and false news. Rosa et al [4] proposed a new rumor detection system based on the influence
potential and network characteristics of health rumor on twitter. In recent years, with the false news getting more and more attention from the public, many platforms have held relevant competitions. Dean Pomerleau and delip Rao held a fake news challenge in 2017[5]. In August 2019, The Institute of computing and data evaluation platform bienda of the Chinese Academy of Sciences, released the China multimodal false news data set and evaluation competition[6]. In February 2019, the "Challenge Cup" link of WSDM launched the challenge of anti rumor topic[7]. With the development of deep neural network, recurrent neural network is introduced to detect rumor events. Ma et al. [8] For the first time, recurrent neural network is introduced to detect rumor events. According to the data structure characteristics of event level rumor. Shu et al [9] proposed an interpretable detection model, which uses news content and user comments to capture interpretable Top-k to check valuable sentences and comments.

However, the above methods focus on rumor analysis in general fields. In the research texts, short texts account for a large proportion, while news in different fields are quite different.

3. DataSet

3.1. data set source

This paper collects all the news of the daily express from July 29, 2016 to November 30, 2018, with 79209 pieces of data. After manual screening and data annotation, 13754 pieces of effective data identified by authoritative organizations such as "doctor Dingxiang", "the whole people are serious" and "microblog refutation" were tested. 2511 rumor data and 5898 non rumor data are obtained by annotation, with an average length of 840, which belong to long text data. In the labeling process, the rumor paragraphs of rumor data are extracted respectively. 20% of them are test sets.

3.2. data preprocessing

Remove stop words and participles from the text. Because health data is mainly expressed in terms of terms, it has the characteristics of long text and wide content, and it is difficult to distinguish health rumors from the overall scale of the text. Therefore, we cluster the segmented text with LDA topic model. The selection range of topic number is [1,15], and the formula for calculating topic similarity is:

\[ D_{LS} (\tilde{\theta}_{\text{sd}}, \tilde{\theta}_{\text{ld}}) = \frac{1}{2} [D_{KL}(\tilde{\theta}_{\text{sd}} || \frac{1}{2} \tilde{\theta}_{\text{ld}} + \tilde{\theta}_{\text{ld}}) + D_{KL}(\tilde{\theta}_{\text{ld}} || \frac{1}{2} \tilde{\theta}_{\text{sd}} + \tilde{\theta}_{\text{sd}})] \]  

(1)

Taking the top 10 words with high weight, through comparison, we can find that the classification is the most obvious when choosing three topics, which are: (1) mother and infant gender, (2) food safety, (3) health care. And type mark the dataset.

4. feature extraction and fusion

4.1. model architecture

This paper proposes a rumor detection method based on topic classification and multi-scale feature fusion. The model is mainly divided into three parts: (1) extract multi-scale word level features of text under different topics and normalize them; (2) use MLM to pre train rumor data and generate local text features; (3) use bilstm + attention model to train global text features of all experimental data; (4) fuse three types of features and output final rumor detection results. The model architecture is shown in the Figure 1.
4.2. word level multiscale feature extraction

Word level features are different from other fields in the field of health, in which nominal statements account for a large proportion. The multi-level features are mined and fused. Three feature extraction methods are used. After cross screening results, the first 400 common feature words were selected for subsequent experiments. The specific process is shown in Figure 2:

TF-IDF is used to calculate the weight of feature words. The expected cross entropy reflects the distance between the probability distribution of the text topic class and the text topic class in the presence of a specific word. The calculation formula is as follows:

$$ECB(t) = p(t)\sum_{i}(Ci||t)\log\left(\frac{p(Ci||t)}{p(Ci)}\right)$$  \hspace{1cm} (2)

Where \(t\) is the characteristic word, \(Ci\) is the \(i\)th category, \(p(T)\) is the document frequency of the word, and \(p(Ci||t)\) is the document frequency of category \(Ci\) under the condition of the word \(t\).

The dominance rate is only suitable for binary classification, and only cares about the score of text features for the target class.

The first 2500 feature words were selected by ranking each word according to the dominance rate from large to small. The above three groups of characteristic words are normalized to \([0,1]\) interval, and then expressed by tanh activation function.

4.3. model pre training based on MLM

MLM (masked language model) uses context prediction to train cover words by covering a certain proportion of content in sentences. This method has been used in the Bert [10] model proposed by
Google, and has shown excellent results. In this paper, we simply adjust the rumor data and train the rumor paragraphs together with the complete rumor data. Finally output the transformed eigenvector \( v_2 \).

Fig 3: Structure of Bi LSTM + attention model

4.4. feature training based on BiLSTM + attention

In the bilstm model, attention mechanism is introduced to simulate the attention characteristics of human brain and pay attention to more important information. The model is applied to all experimental data to obtain complete context information. Refer to Huang et al. [11] The model architecture here shows the similarity between the x-th target output and the y-th input.

\[
S_{xy} = S(t_x, t_y) = \frac{W_a(t_x^* \cdot t_y^*)}{|t_x^*||t_y^*|} - W_a|t_x - t_y|^2 \| t_x - t_y \|
\]  

(3)

\( S_{xy} \) is defined as the difference between cosine distance and Euclidean distance. The closer the two are, the smaller the Euclidean distance is and the larger the cosine distance is.

4.5. feature fusion

The above three feature extraction vectors are fused by the method of vector splicing to get the vector \( V \), which is expressed as:

\[
V = v_1 \oplus v_2 \oplus v_3
\]  

(4)

In the part of convolutional neural network architecture, input \( V \) to the full connection layer to get vector \( V' \), and use softmax classification to get score \( scr \), among which the larger is the classification result. This model uses cross entropy loss function to compare the distance between the real value and the predicted value. The cost function formula is:

\[
L = \sum_{m=1}^{M} \left[ y_m \log(p_m) + (1-y_m) \log(1-p_m) \right]
\]  

(5)

5. Result analysis

The method in this paper is compared with the existing general method and domain method on the same data set, and the test results of each model are shown in Table 1 and table 2.
Table 1: Comparison of rumor data results

| Model         | Pre  | Rec  | F1   | Acc  |
|---------------|------|------|------|------|
| NB[12]        | .865 | .803 | .833 | .839 |
| SVM[2]        | .881 | .800 | .838 | .846 |
| BiLSTM[13]    | .861 | .828 | .845 | .851 |
| TextCNN[14]   | .861 | .843 | .852 | .853 |
| BiLSTM-DEC    | .877 | .833 | .855 | .858 |
| TCNN-DEC      | .890 | .860 | .875 | .877 |
| Ours          | .907 | .879 | .885 | .891 |

Table 2: Comparison of non rumor data results

| Model         | Pre  | Rec  | F1   | Acc  |
|---------------|------|------|------|------|
| NB[12]        | .816 | .875 | .845 | .839 |
| SVM[2]        | .817 | .892 | .853 | .846 |
| BiLSTM[13]    | .835 | .867 | .850 | .851 |
| TextCNN[14]   | .846 | .863 | .855 | .853 |
| BiLSTM-DEC    | .841 | .883 | .862 | .858 |
| TCNN-DEC      | .865 | .893 | .879 | .877 |
| Ours          | .888 | .890 | .894 | .891 |

It can be seen from the table that the accuracy of the topic based classification and multi-scale fusion rumor detection model proposed in this paper has reached 89.1% in the health data set, which is superior to the general model and 1.4% higher than the model in the field. Bilstm-dec and textcnn-dec fuse domain entity features, which shows that domain entities can improve the feature extraction accuracy of health data. In this paper, bilstm is also used in the model. The experimental accuracy is higher than that of domain entity features, which proves that multi-scale features play a role in rumor detection in the health field.

Table 3: Comparison of results of various topics

|        | Acc  | Pre  | Rec  | F1   |
|--------|------|------|------|------|
| topic1 | .883 | .901 | .867 | .874 |
| topic2 | .918 | .922 | .899 | .917 |
| topic3 | .872 | .869 | .855 | .848 |

6. Conclusion

In this paper, a novel rumor detection method based on topic classification and multi-scale fusion is proposed, which considers the long text characteristics of rumor information in the health field. After topic classification, the association and difference between and within topics are considered, and the features of specific sub data sets are extracted for multi-scale fusion and classification. The accuracy of rumor detection in the field of health is improved obviously by using this method, which proves that this method adapts to the characteristics of this field.

Acknowledgments

The work was supported by Beijing Municipal Natural Science Foundation-Haidian original innovation joint fund(L182007); The National Natural Science Foundation of China (61702020) and supporting projects (PXM2018_014213_000033).
Reference

[1] Yang F, Liu Y, Yu X, et al. Automatic detection of rumor on Sina Weibo[C]//Proceedings of the ACM SIGKDD Workshop on Mining Data Semantics. ACM, 2012: 13.
[2] Liang G, He W, Xu C, et al. Rumor identification in microblogging systems based on users’ behavior[J]. IEEE Transactions on Computational Social Systems, 2015, 2(3): 99-108.
[3] Shu K, Zhou X, Wang S, et al. The Role of User Profile for Fake News Detection[J]. 2019.
[4] Rosa Sicilia, Stella Lo Giudice, Yulong Pei, et al. Health-related rumour detection on Twitter[C]// IEEE International Conference on Bioinformatics & Biomedicine. IEEE Computer Society, 2017.
[5] http://www.fakenewschallenge.org/
[6] https://biendata.com/competition/falsenews/
[7] http://www.wsdm-conference.org/2019/wsdm-cup-2019.php
[8] Ma J, Gao W, Mitra P, et al. Detecting rumors from Microblogs with recurrent neural networks[C]// Proceedings of IJCAI2016.2016.
[9] Shu K, Cui L, Wang S, et al. dEFEND: Explainable Fake News Detection[C]// KDD. 2019.
[10] Devlin J, Chang M W, Lee K, et al. BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding[J]. 2018.
[11] Huang Dandan, Guo Yucui. Bi-lstm-crf Chinese word segmentation model integrated with attention mechanism [J]. Software, 2018,39 (10): 260-266.
[12] Castillo C, Mendoza M, Poblete B. Information credibility on Twitter[C]// Proceedings of the 20th International Conference on World Wide Web, WWW 2011, Hyderabad, India, March 28 - April 1, 2011. DBLP, 2011.
[13] Guo H, Cao J, Zhang Y, et al. Rumor detection with hierarchical social attention network[C]//Proceedings of the 27th ACM International Conference on Information and Knowledge Management. ACM, 2018: 943-951.
[14] Kim Y. Convolutional neural networks for sentence classification[J]. arXiv preprint arXiv:1408.5882, 2014.