Rumor detection in networks based on the BERT-CNN approach

Deming Huang^{1,4,*,†}, Sibo Yang^{2,5,†}, Shengwei Gao^{3,6,†}

¹Information Management and Information Systems, Tongji University, Shanghai, 201804, China

²School of Computer Science, China University of Geosciences, Wuhan, 430074, Chin

³Traffic and Transportation, Shandong University of Aeronautics, Shandong, 256600, China

⁴Corresponding author email: 2154355@tongji.edu.cn ⁵vangsibo23@outlook.com

⁶gaoshengwei707@163.com

[†]These authors contributed equally to this work and should be considered co-first authors.

Abstract. As of June 2023, China has witnessed a surge in internet users, reaching 1.079 billion, with an internet penetration rate of 76.4%. This rapid expansion of social media and online communication channels has not only enhanced the ease of information sharing but has also intensified the spread of rumors and misinformation. This trend poses substantial threats to social order and public safety. In response to this challenge, this work introduces an innovative approach to detect online rumors, utilizing the BERT-CNN methodology. By harnessing the contextual understanding capabilities of the BERT model and the efficient feature extraction prowess of CNN models, the study has crafted a framework that seamlessly integrates these technologies. This integrated approach proves effective in accurately identifying and categorizing rumors in online data. Through extensive experiments conducted on various realworld datasets, a comparative performance analysis had conducted against existing technologies. The results reveal significant enhancements in key metrics such as accuracy, recall, and F1 score when compared to traditional methods. This affirms the efficacy of the BERT-CNN approach in the domain of online rumor detection. This research not only introduces a fresh perspective to rumor detection technology but also furnishes a robust tool to tackle the challenges associated with the dissemination of rumors in the digital age.

Keywords: BERT, CNN, Rumor Detection, Online Data, Feature Extraction, Information Dissemination.

1. Introduction

Online rumors primarily rely on the Internet and utilize computers, mobile phones, and other mobile devices as means of dissemination. They are spread on the Internet and typically involve topics such as sudden events, public affairs, and prominent individuals [1]. As of June 2023, the number of Chinese

internet users has reached 1.079 billion, with an internet penetration rate of 76.4% [2]. In recent years, the exponential growth of social media platforms and online communication channels has exacerbated the problem of misinformation and rumors. These platforms provide fertile ground for the rapid dissemination of false information, often without proper verification or fact-checking. The consequences of these falsehoods can be profound, extending beyond individual reputations to impact societal dynamics, public sentiment, and even political landscapes. Instances of public panic, reputational damage, and social unrest have underscored the urgent need to develop effective techniques for detecting and combating online rumors. In the field of NLP, BERT and CNN methodologies have been successfully applied to various tasks. BERT has been utilized for text classification, sentiment analysis, and document summarization, among others, while CNNs have been widely employed for tasks such as image classification and speech recognition. The adoption of these techniques in rumor detection leverages their strengths and adapts them to the specific challenges posed by identifying rumors in networked environments. In this paper, a novel approach was proposed for rumor detection in networks based on the BERT-CNN methodology. It was aimd to exploit the contextual understanding of BERT and the feature extraction capabilities of CNNs to effectively identify and classify rumors in networked data. This work conducted experiments on real-world datasets to evaluate the performance of the approach and compare it with existing rumor detection methods. The results demonstrate the effectiveness of the BERT-CNN approach in detecting rumors and provide insights into its potential for addressing the challenges associated with rumor detection in networked environments.

2. Literature Review

In early research, researchers in the field of rumor detection relied on manual feature engineering based on the content of rumors, such as punctuation distribution, keyword distribution, and link features [3]. They then used traditional machine learning methods to automatically detect rumors. In recent studies, a method proposed by Zeng et al. combines BERT and topic modeling to optimize the effectiveness of rumor prediction, addressing the limitations of previous research in integrating contextual semantic information and topic semantic information [4]. In terms of detection methods, most researchers consider rumor detection as a binary classification problem. Early classification methods mainly relied on traditional machine learning methods, using text features such as bag-of-words, TF-IDF, etc., as inputs, and constructing classifiers like SVM, decision trees, etc., for rumor detection. Later, researchers introduced deep learning methods, utilizing deep neural networks such as convolutional neural networks (CNN), recurrent neural networks (RNN), etc., to learn text features, combined with other models like graph convolutional networks and attention mechanisms for rumor detection. For example, Liu et al. utilized CNN for rumor prediction [5] and achieved promising results. Feng et al. proposed a rumor detection method based on graph convolutional networks and attention mechanism, achieving accuracy rates of 0.86 and 0.87, and F1 scores of 0.86 and 0.87 on Twitter15 and Twitter16 datasets respectively. With the rise of pretrained language models like GPT and BERT, researchers started using these models for rumor prediction and obtained improved results [6,7]. In recent studies, researchers have begun to employ fusion models. For instance, the XLNet+BiGRU-MHA model proposed by Feng et al. demonstrated excellent performance in terms of F1 score, achieving an accuracy rate of 95.5% [8]. The BERT-RCNN model proposed by Li et al. achieved an accuracy rate of 95. 16% and an F1 score of 95. 14% on the Weibo rumor dataset [9], indicating the advantages of fusion models.

Regarding model improvement, Zhang et al. improved the generation loss by utilizing recurrent generative adversarial networks and Wasserstein distance, effectively enhancing the rumor detection capability of models under imbalanced data conditions [10]. In terms of model evaluation, currently, accuracy and F1 score are the main metrics used to compare the performance of models. Based on the analysis, traditional rumor detection methods mainly have the following issues: difficulties in handling imbalanced data due to a scarcity of rumor samples, resulting in the discriminator's limited ability to learn deeper features of rumor data [11]. These methods also suffer from inferior performance and long training times. Therefore, this paper proposes a rumor prediction model that

integrates BERT and CNN. It leverages BERT's contextual understanding and semantic representation capabilities, combined with CNN's local feature extraction abilities. The model is further enhanced using an adversarial generative network with Hinge Loss to mitigate the impact of imbalanced data on model accuracy and improve overall performance.

3. Datasets the work Used

This work utilized the Twitter15 and Twitter16 datasets in the field of rumor detection, which are widely used for this purpose [12]. Each dataset comprises 1490 and 818 labeled tweets from Twitter, respectively. All labeled tweets are annotated with four labels: non-rumors, false-rumors, true-rumors, and unverified-rumors. The specific information about the data is presented in Table 1. The textual data in this dataset exhibits the following characteristics: (1) Due to its origin on the web, the text is less formal and sentences are relatively casual. (2) The topics and emotions covered in the text are diverse. (3) There are significant differences in text length and content format among different texts. Therefore, in the work, the approach addresses the challenges posed by these three characteristics of textual data to a certain extent.

Statistic	Twitter15	Twitter16
# of source tweets	1490	818
# of non-rumors	374	205
# of false-rumors	370	205
# of true-rumors	372	207
# of unverified rumors	374	201

Table 1. Datasets Statistics.

4. Methodology

4.1. Why BERT?

BERT, short for Bidirectional Encoder Representation from Transformers, is a pre-trained language representation model. It breaks away from the traditional use of unidirectional language models or the simple concatenation of two unidirectional language models in the pre-training phase. Instead, BERT employs a novel approach known as the masked language model (MLM) to create comprehensive bidirectional language representations. The fundamental structure of the BERT model is depicted in Figure 1.



Figure 1. Structure of BERT.

4.2. Why CNN?

Convolutional Neural Networks (CNN) are a deep learning model that is widely used in image recognition and computer vision tasks. CNN can play an important role in rumor detection. Methods based on convolutional neural networks (CNN) can obtain the relevant features inside local neighbors, but cannot deal with the global structural relations of graphs or trees. However, rumor detection needs to mine the deep features of the text, and mining the deep features of the text through the learning and training of one-dimensional convolutional neural network can avoid the problem of feature construction, and can find those features that are not easy to be found, to produce better results.

4.3. Why does the work combine BERT and CNN?

However, the BERT model may not extract enough local information when dealing with long texts, and due to its powerful feature extraction ability, there may be overfitting. CNN can effectively extract local information through convolutional and pooling layers, and alleviate overfitting problems by using regularization techniques and data augmentation methods. The structure of the BERT-TF-IDF-CNN model is shown in Figure 2.



Figure 2. BERT-TF-IDF-CNN Model.

5. Results

5.1. Hyperparameter setting

Our hyperparameter settings are shown in Table 2.

Hyperparameters	Value
Learning Rate	0.0001
Maximum Text Length	40
Batch_size	32
Dense units	512
Dropout rate	0.6
Convolutional Kernel Size	5
Filters	256

 Table 2. Hyperparameter Setting.

5.2. Indicator selection

After feature extraction using the BERT model, TF-IDF transformation is performed, and the results are input into the convolutional layer of the CNN model for five fold cross validation. The model is evaluated using F1 score, Recall, and Precision metrics. Precision measures the proportion of real positive cases predicted by the model as positive cases. Recall measures the proportion of positive examples correctly identified by the model. F1 score is particularly effective in situations of imbalanced category distribution, and can comprehensively evaluate the accuracy and recall rate of

the model. The output of the confusion matrix is used to observe the corresponding relationship between the predicted label and the true label.

5.3. Model running results

The output of the model includes a confusion matrix, as shown in Figure 3, as well as the Average F1 Score, Recall, and Precision of the ten fold cross validation of the model. Output the indicators of the TF-IDF-CNN model, as shown in Figure 4, and compare them with the indicators of the BERT-CNN model, as shown in Figure 5.







Figure 4. TF-IDF-CNN Model Average F1 Score, Recall, and Precision.



Figure 5. BERT-CNN Model Average F1 Score, Recall, and Precision.

6. Conclusion

The purpose of this study is to explore the effectiveness of the BERT-CNN method, compare and analyze it with the TF-IDF-CNN method, and reveal the improvement of the BERT model over traditional models. Its unique advantages in identifying rumors in network environments are evident through key indicators. Rumor detection experiments were conducted on the widely used Twitter15 and Twitter16 datasets. By fine-tuning hyperparameters, including learning rate, maximum text length, batch

size, etc., it plays a crucial role in stabilizing the training process of the model and ensuring optimal performance. The limitation of this study is that the final model has limited exploration of different topic texts, which may lead to a decrease in model performance. In summary, the research validates the certain advantages of the BERT-CNN model in terms of precision in rumor detection. Through further exploration and improvement, this method will help solve the problem of rumor dissemination in society and cyberspace.

Acknowledgement

Deming Huang, Sibo Yang and Shengwei Gao contributed equally to this work and should be considered co-first authors.

References

- Xie, Y., Qiao, R., Shao, G., & Chen, H. (2017). Research on Chinese social media users' communication behaviors during public emergency events. Telematics and Informatics, 34(3), 740-754.
- [2] Yang, S. (2023, June). Current Situation and Digital Development Trend of E-Commerce Live Streaming in China. In Proceedings of the 6th International Conference on Economic Management and Green Development (pp. 381-389). Singapore: Springer Nature Singapore.
- [3] He, G., Lv, X., Li, Z., & Xu, L. (2013). Research on Weibo Rumor Recognition. Library and Information Service, 57(23), 114–120.
- [4] Zeng, J., Cheng, Z., Huang, Y., & Gao, P. (n.d.). Rumor Detection Method Combining BERT and Topic Model. Information Science, 1–27.
- [5] Chen, T., Li, X., Yin, H., & Zhang, J. (2018). Call attention to rumors: Deep attention based recurrent neural networks for early rumor detection. In Trends and Applications in Knowledge Discovery and Data Mining: PAKDD 2018 Workshops, BDASC, BDM, ML4Cyber, PAISI, DaMEMO, Melbourne, VIC, Australia, June 3, 2018, Revised Selected Papers 22 (pp. 40-52). Springer International Publishing.
- [6] Wang, X., & Lu, X. (2022). A Weibo Rumor Detection Method Based on Fine-tuning the General Language Model BERT. Computer Programming Skills and Maintenance, (5), 81–83. https://doi.org/10.16184/j.cnki.comprg.2022.05.012.
- [7] Liang, Z., Dan, Z., Luo, Y., et al. (2021). Rumor Detection Based on BERT Model and Enhanced Hybrid Neural Network. Computer Applications and Software, 38(03), 147–152+189.
- [8] Zhang, Q., Zhu, Y., Lan, C., Mao, Q., & Cui, Y. (2023, April). Weibo rumor detection based on GCN. In Third International Conference on Artificial Intelligence and Computer Engineering (ICAICE 2022) (Vol. 12610, pp. 839-844). SPIE.
- [9] Li, Y. C., Qian, L. F., & Ma, J. (2021). Early detection of micro blog rumors based on BERT-RCNN model. Information Studies: Theory & Application, 173-177.
- [10] Feng, L. Z., Liu, F., & Wang, Y. W. (2023). Rumor Detection Method Based on Graph Convolutional Network and Attention Mechanism. Data Analysis and Knowledge Discovery,1–15.
- [11] Zhang, H. Z., Dan, Z., Dong, F. M., Gao, Z., & Zhang, Y. K. (n.d.). Research on Rumor Detection Based on Recurrent Generative Adversarial Network and Wasserstein Loss. Data Analysis and Knowledge Discovery, 1–14.
- [12] Yuan, C., Ma, Q., Zhou, W., Han, J., & Hu, S. (2019, November). Jointly embedding the local and global relations of heterogeneous graph for rumor detection. In 2019 IEEE international conference on data mining (ICDM) (pp. 796-805). IEEE.