Frontiers in Computing and Intelligent Systems
ISSN: 2832-6024 | Vol. 4, No. 1, 2023

Fake News Detection based on Deep Learning

Rong Wang *

School of Information, Yunnan University of Finance and Economics, Kunming 650221, China
*** Corresponding author Email:** wr594998@163.com

Abstract: The rapid popularization of the Internet has broken the professional threshold of information dissemination, enabling more and more people to easily obtain information, share and express views through social media, which has greatly enriched people's daily life. However, due to the huge number of users of social media, false news fabricated for various purposes is emerging in endlessly. Moreover, with the progress of technology, false news is no longer simply spread in the form of text, but more spread through the combination of text, pictures and video, which greatly increases the confusion of false news. The experiment in this paper is based on tensorflow to detect false news. During the experiment, LR was used to obtain the fusion coefficients of CNN and LSTM models, that is the regression coefficient of LR, and then calculated the optimal threshold with the fused model on the verification set. In addition, in terms of model selection, lightgbm and xgboost were selected to train the model on the training set for false news, and predicted the news text on the testing set. The results of three experiments show that the effect of using xgboost model is the best, and the F1 score obtained in the experiment is the highest.

Keywords: False News Detection; Tensorflow; CNN and LSTM Fusion; Lightgbm; XGboost.

1. Introduction

1.1. Research Background and Significance

With the rapid popularity of mobile Internet, people's access to news has become more convenient. In particular, social media, mainly represented by Weibo and Twitter, have become an important channel for most people to obtain daily information. Nowadays, more and more people are gaining knowledge, sharing information, expressing opinions and exchanging experiences on these online platforms every day. However, to a certain extent, the rapid development of social media is a double-edged sword. On the one hand, it has facilitated the advent of the information age, where all kinds of information can be disseminated at a very low cost through social media, making it possible for "A genius to know everything about the world without leaving home". On the other hand, the Internet has become a breeding ground [1] for false information to grow and spread, and the vast amount of information is mixed with false news that is deliberately fabricated for various purposes. This kind of news is presented in various forms, often with text and pictures, which makes it very difficult for readers to get correct and effective information, and some of the widely disseminated false news has caused serious negative impact on individuals and society. Thus, it seems that there is a great need for us to study the fake news detection. Despite some recent scientific advances in false news detection, it is still a challenging problem due to its complex content, wide range of sources, diverse modalities, and the high costs involved in fact-checking.

Fake news spreads rapidly through social media platforms, which may affect social stability while negatively impacting social opinion. For example, false news such as "Temporarily exempting Wuhan Red Cross from managing disaster relief supplies" and "The private car was unclaimed for several months, passers-by asked and learned that the owner died after fighting the epidemic", etc. Although the official explained and refuted the rumors at the first time, such rumors still circulated widely on the Internet. It has attracted a lot of attention and caused serious negative impact on the parties concerned. At present, most of the news and information presented on social media are characterized by diverse modes and semantic diversity[2]. Along with the popularity of mobile devices, news is often published and disseminated rapidly on social media in the form of text, images, videos and other multimedia data. While this kind of news is convenient for readers to obtain and understand information, it also causes the traditional technique of relying solely on text analysis for false news detection to be no longer applicable. Another example is that various types of fake news about the location and scale of exercise in several Chinese military exercises often spread false information by stealing irrelevant pictures and videos.

1.2. Status of Domestic and International Research

As deep learning continues to make progress, more and more researchers are studying fake news detection based on neural networks. There are many branches in the field of fake news detection, such as unimodal-based fake news detection, multimodal-based fake news detection, etc. Unimodal-based fake news detection means that the text contains only one of text, picture or video, multimodal-based fake news detection refers to multiple types of fake news that contain text, pictures, videos, voices, etc.

(1) Unimodal-based Method for False News Detection

The fake news detection method[3] based on single modality mainly judges the authenticity of news by extracting text features from news text content or visual features from image information. For example, Castillo et al. [3]used decision trees to learn thematic features of news text content for classification. The model proposed by Yu et al. [4]obtained high-level interaction features and key features of relevant posts through convolutional neural networks.ma et al.[5] used recurrent neural networks to learn latent features of news text content. In MVNN[6], the authors used a multi-region visual neural network for fake news detection by targeting the rich visual information in different pixel regions.

(2) Multimodal-based Method for Fake News Detection

Fake news detection based on multimodality has attracted a lot of attention in recent years. Some of these methods concatenate the textual features in the post with the visual features of the images in the post[7]. However, this approach requires manual feature engineering on the one hand, and is not able to effectively obtain complex semantic representations in images on the other hand. At present, due to the excellent performance of deep neural network (DNN) in nonlinear representation learning[8], many multimodal representation learning methods use deep learning mechanism to learn feature representation, thereby improving the ability of fake news detection. Jin et al. [9]proposed a method based on deep learning, which can learn the multimodal content and social information of news posts, and then used the attention mechanism to fuse the multimodal features. In EANN[10], the authors learned the invariant features of events through an adversarial network containing a multimodal feature extractor to obtain multimodal features of each news item for fake news detection. In MVAE[11], Khattar et al. used a multimodal variational self-encoder for fake news identification, where the multimodal features of a post were fed into a bimodal variational self-encoder to obtain a multimodal feature representation of the news. Cui et al. [12]proposed an end-to-end deep embedding framework for fake news detection, in which the latent sentiment of the post publisher was used to distinguish fake news. SpotFake[13] used a pre-trained BERT [14]model to learn text features of news posts and a VGG-19 model pre-trained on ImageNet [15] to extract image features. SpotFake+ [16]was an improved version of SpotFake that used a modified version of BERT, the XLNet [17]model, to extract text features based on SpotFake. While learning news text features and visual information, the SAFE[18] model also learned the intrinsic connection between text content and vision to predict fake news. In M-GCN[19], the author focused on distinguishing different degrees of fake news according to the similarity between news, which used GCN modules of different depths to extract domain information of different scales, and fused these features through the attention mechanism.

2. Research Methods

2.1. CNN and LSTM Model Fusion

Before training the model, the text was processed, such as removing spaces and keeping only the text, then the text was divided into words and deactivated words, after the division of words, the training vocabulary was tracked, all the text was traversed and the words were counted, then the text was converted into a vector sequence, and filled the vector sequence according to the maximum length of the text, making the sequence of vectors to uniform length. Then word vector training was performed using word2vec_model. The attention mechanism was introduced in the model to capture the key points from the longer text without losing important information. After processing the text data, the training was performed with CNN and LSTM respectively, finally, LR was used to try to get the fusion coefficient of the two models of CNN and LSTM, here was the regression coefficient of LR, and then the best threshold was calculated on the validation set with the fused model.

2.2. Lightgbm and Xgboost Models

Segment the text data first, fill in the empty values after segmentation, and use the TF-IDF algorithm to extract the feature vector of the news text. The TF-IDF algorithm is one of the important algorithms for extracting feature word vectors, and it is also one of the main technologies for generating word vectors. The TF-IDF algorithm statistically evaluates the importance of a word to a document or other documents in the corpus to determine the feature words of the document. The basic idea: if a word appears frequently in a document, but appears infrequently in other documents in the corpus, it is determined that the word can be used as a feature word of the document to some extent. it has the ability to distinguish categories and can be used as the basis for classification. The TF-IDF algorithm is divided into TF (term frequency) algorithm and IDF (inverse document frequency) algorithm, where the TF algorithm represents the ratio of the count of a specific word to the total number of words in the document, representing the frequency of a specific word in the document. The IDF algorithm represents the logarithm of the ratio of the total number of documents in the corpus to the number of documents in which a particular word occurs in the corpus. Then set the model parameters respectively, and finally train the model.

3. Introduction of the Experimental Data Set

In the false news text detection task, the training set contains a total of 38,471 news items, including 19,186 real news items and 19,285 false news items. Each data consists of three elements [id, text, label], where id is the unique id of each data, which uniquely characterizes a news, text is the Chinese news text, label is represented by 0 and 1, 0 means real news, 1 means false news. The testing dataset needs to be submitted to the backend for determination. Analysis of the dataset yields the results shown in Table 1.

Table 1. Dataset Analysis

Maximum Sentence Length	Minimum Sentence Length	Average Data Length	Number of Words after the Participle
929	0	50.883678615060695	73915

4. Experiment and Results

4.1. Experimental Environment

(1) Experimental Hardware Environment

The laptop processor used in the experiment is i7-8550U, and the memory is 8.00GB. The experimental hardware environment is shown in Table 2.

Table 2. Experimental Environment

Equipment Name	DESKTOP-9C2CKR8
Processor	Intel(R) Core (TM) i7-8550U CPU
RAM	8.00 GB (7.9 GB available)

(2) Experimental Software Environment

The experimental code in this paper is completed in python language, and the python version used is 3.7.4. During the experiment, pandas, matplotlib, numpy, tqdm, jieba, scikit-learn, keras, tensorflow, xgboost, scipy, lightgbm packages were called, and their version numbers are pandas (1.3.5), matplotlib (3.1.1), numpy (1.21.5), tqdm (4.62.3), jieba (0.42.1), scikit-learn (0.21.3), keras (2.2.5), tensorflow (1.15.0), xgboost (1.6.1), scipy (1.3.1), lightgbm (3.3.2).

Table 3. Library Name and Version

Library Name	Version
Python	3.7.4
pandas	1.3.5
matplotlib	3.1.1
numPy	1.21.5
tqdm	4.62.3
jieba	0.42.1
scikit-learn	0.21.3
keras	2.2.5
tensorflow	1.15.0
xgboost	1.6.1
scipy	1.3.1
lightgbm	3.3.2

4.2. Experimental Procedure

4.2.1. Model Training

The training process of CNN and LSTM model fusion is shown in Figure 1 and Figure 2.

It can be seen from Figure 1 that it took 2.0396049999999377 Seconds to import word2vec during the training of the word vector model. The model was trained in two batches. And it can be seen from Figure 2 that the F1 scores on the verification set were 0.9429 and 0.9637 respectively.

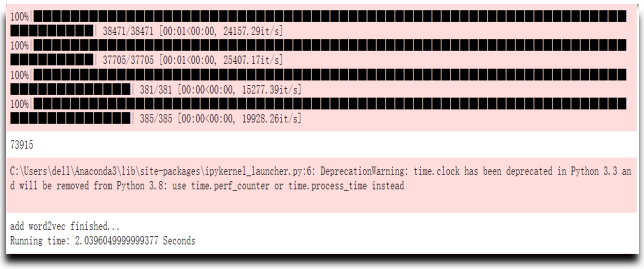

Figure 1. Model Training Process

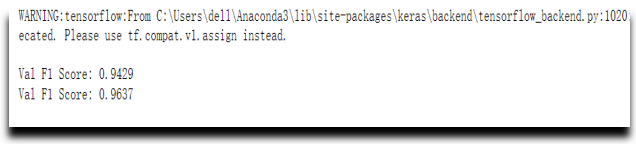

Figure 2. The F1 values in model training

It can be seen from Figure 3 that the number of positive samples used in the experiment is 15673, and the number of negative samples is 15487. During the model training process, 31160 pieces of data were used for training, and the number of features selected was 299. The minimum loss value obtained during training was 0.193351 and the highest accuracy value obtained on the validation set was 0.976451.

The xgboost model training process is shown in Figure 4, Figure 5 and Figure 6. It can be seen from Figure 4 that 31160 pieces of data were used for training during this training process. During the training process, it can be seen from Figure 5 that the highest training accuracy of the model during training could reach 0.99981. It can be seen from Figure 6 that through multiple iterations, the highest F1 score of xgboost on the validation set could reach 0.96589.

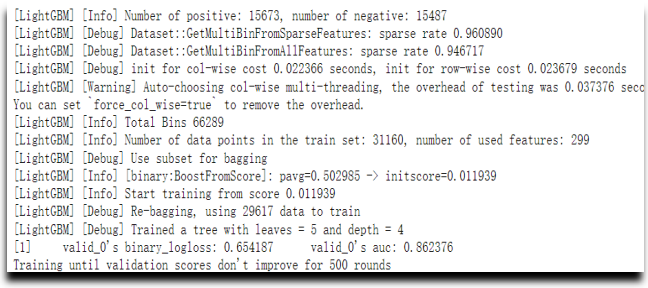

Figure 3. Lightgbm Model Training Process

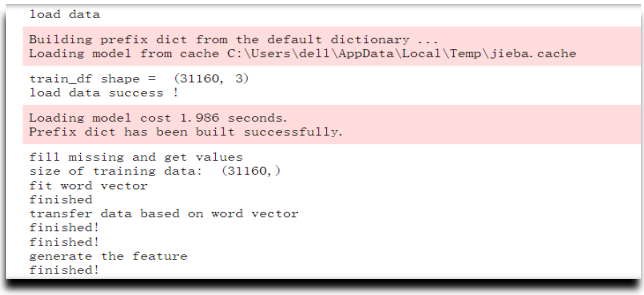

Figure 4. Xgboost Model Training Process

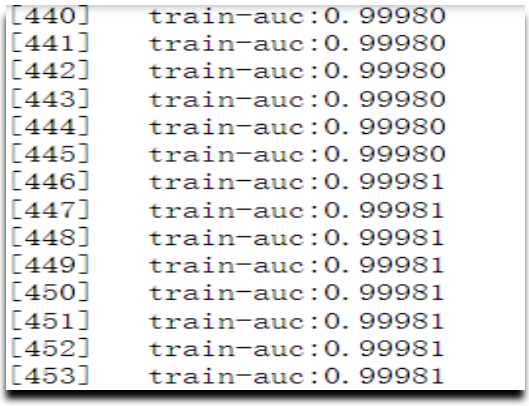

Figure 5. Xgboost Training Accuracy

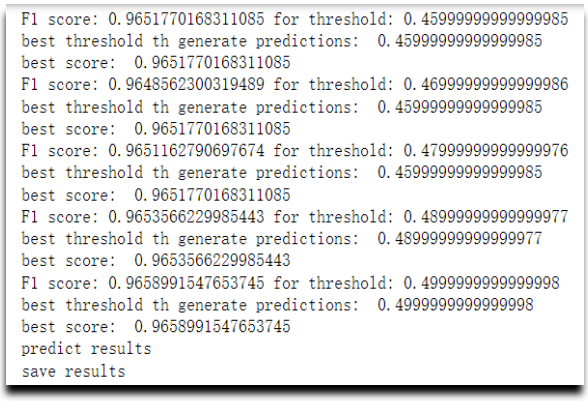

Figure 6. Xgboost Training F1 Value

4.2.2. Model Testing Process

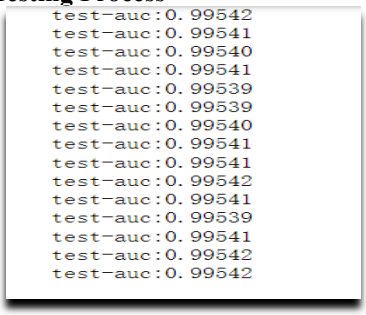

Figure 7. The test accuracy of lightgbm

The lightgbm test process is shown in Figure 7. During the test, 4000 pieces of data were selected for testing, and it can be seen from the figure7 that the test accuracy reached up to 0.99542.

4.2.3. Comparison of Experimental Results

The main performance indicators of the evaluation model are:

$$Precision = \frac{TP}{TP+FP} \times 100\% \quad (1)$$

Where TP denotes true cases, FP denotes pseudo-positive cases, and precision denotes the proportion of samples predicted to be true positive cases in the sample of positive cases.

$$recall = \frac{TP}{TP+FN} \times 100\% \quad (2)$$

where TP denotes true cases, FN denotes pseudo counter examples, and recall denotes the proportion of samples predicted as positive cases to all positive samples.

$$F1 = \frac{2 \times Precision \times recall}{Precision + recall} \quad (3)$$

The F1 values obtained by different models after training are shown in Table 4. After comparing different pre-trained language models under the same classification algorithm, the prediction result of the xgboost model is significantly higher than that of lightgbm, and it is also higher than the result of the fusion of CNN and LSTM models. The analysis shows that the xgboost model is more advantageous, and the F1 value reaches 0.972.

Table 4. Experimental Results

model	F1 score
CNN and LSTM model fusion	0.917
lightgbm	0.918
xgboost	0.972

5. Conclusion

This paper compares the detection results of CNN and LSTM model fusion, lightgbm, and xgboost models. The experimental results show that xgboost outperforms the other models in this study in terms of performance metrics, and its F1 values are all higher than the other models. Of course, there are still shortcomings in this paper for the study of fake news texts. Longitudinal analysis and comparison from multiple classification algorithms are still needed to find the best classification algorithm. In this paper, only single-mode false news was detected in the experiment. In the next step, the method proposed in this paper will be used to conduct experiments on multi-mode false news.

References

[1] Yujun Gao, Gang Liang, Fangting Jiang, Chun Xu, Jin Yang, Junren Chen,Hao Wang.A review of social network humor detection[J]. Journal of Electronics, 2020, 48(7): 1421-1435.

[2] Shengsheng Qian,Tianzhu Zhanng,Changsheng Xu.A review of multimedia social event analysis[J].Computer Science, 2021, 48(3): 97-112.

[3] CHEN L, LIANG J, XIE C, XIAO Y. Short Text Entity Linking with Fine-grained Topics[C]//Proceedings of the 27th ACM International Conference on Information and Knowledge Management. Torino, Italy: Association for Computing Machinery, 2018: 457-466.

[4] YU F, LIU Q, WU S, WANG L, TAN T. A Convolutional Approach for Misinformation Identification. [C]//IJCAI. 2017: 3901-3907.

[5] MA J, GAO W, WEI Z, LU Y, WONG K F. Detect Rumors Using Time Series of Social Context Information on Microblogging [M]//Social Media Content Analysis: Natural Language Processing and Beyond. World Scientific, 2018: 67-77.

[6] QI P, CAO J, YANG T, GUO J, LI J. Exploiting multi-domain visual information for fake news detection[C]//2019 IEEE international conference on data mining (ICDM). IEEE, 2019: 518-527.

[7] GUPTA M, ZHAO P, HAN J. Evaluating event credibility on twitter [C]//Proceedings of the 2012 SIAM international conference on data mining. SIAM, 2012: 153-164.

[8] ZHAO L, HU Q, WANG W. Heterogeneous feature selection with multi-modal deep neural networks and sparse group lasso [J]. IEEE Transactions on Multimedia, 2015, 17(11): 1936-19 48.

[9] JOULIN A, GRAVE E, BOJANOWSKI P, MIKOLOV T. Bag of tricks for efficient text classification[J]. arXiv preprint arXiv:1607.01759, 2016.

[10] WANG Y, MA F, JIN Z, YUAN Y, XUN G, JHA K, SU L, GAO J. Eann: Event adversarial neural networks for multi-modal fake news detection[C]//Proceedings of the 24th acm sigkdd international conference on knowledge discovery & data mining. 2018: 849-857.

[11] KHATTAR D, GOUD J S, GUPTA M, VARMA V. Mvae: Multimodal variational autoencoder for fake news detection [C] // The world wide web conference. 2019: 2915-2921.

[12] CHURCH K, HANKS P. Word association norms, mutual information, and lexicography[J]. Computational linguistics, 1990, 16(1): 22-29.

[13] SINGHAL S, SHAH R R, CHAKRABORTY T, KUMARAGURU P, SATOH S. Spotfake: A multi-modal framework for fake news detection[C]//2019 IEEE fifth international conference on multimedia big data (BigMM). IEEE, 2019: 39-47.

[14] DEVLIN J, CHANG M W, LEE K, TOUTANOVA K. Bert: Pre-training of deep bidirectional transformers for language understanding [J]. arXiv preprint arXiv:1810.04805, 2018.

[15] DENG J, DONG W, SOCHER R, LI L J, LI K, FEI-FEI L. Imagenet: A large-scale hierarchical image database[C]//2009 IEEE conference on computer vision and pattern recognition. Ieee, 2009: 248-255.

[16] SINGHAL S, KABRA A, SHARMA M, SHAH R R, CHAKRABORTY T, KUMARAGURU P. Spotfake+: A multimodal framework for fake news detection via transfer learning (student abstract) [C]//Proceedings of the AAAI conference on artificial intelligence: 34. 2020: 13915-13916.

[17] YANG Z, DAI Z, YANG Y, CARBONELL J, SALAKHUTDINOV R R, LE Q V. Xlnet: Generalized autoregressive pretraining for language understanding[J]. Advances in neural information processing systems, 2019, 32.

[18] ZHOU X, WU J, ZAFARANI R. Safe: similarity-aware multi-modal fake news detection (2020) [J]. Preprint. arXiv, 2020, 200304981: 2.

[19] HU G, DING Y, QI S, WANG X, LIAO Q. Multi-depth graph convolutional networks for fake news detection[C]//Natural Language Processing and Chinese Computing: 8th CCF International Conference, NLPCC 2019, Dunhuang, China, October 9–14, 2019, Proceedings, Part I 8. Springer, 2019: 698-710.