FNED: Integrated Learning-Based Fake News Inspection

Yuan Geng*

Economics and Management, Jiangsu University of Science and Technology, Zhenjiang, China, 212003

*Corresponding author: gengyuan0312@outlook.com

Abstract. This paper addresses the challenge of inadequate detection accuracy and generalization in false news detection amid the rapid spread of misinformation in the news environment, proposing FNED, an integrated learning-based false detection method. The method enhances detection performance by integrating multiple models; it employs a combination of three models—KNN, LSTM, and XGBoost—utilizing a stacking technique to leverage their predictions as features, followed by logistic regression as a meta-classifier for comprehensive evaluation. Augment the model's resilience across varying data distributions. Experiments utilize the public FakeNewsNet standard dataset, revealing that FNED enhances accuracy by approximately 1.3% relative to previous approaches, indicating superior detection capability and generalization performance.

Keywords: Ensemble Detection, FakeNews Detection, Logistic Regression, Word2vec.

1. Introduction

1.1. Background

Accurately predict the power load, based on BP neural network theory, combined with the advantages of Clementine in dealing with big data and preventing overfitting, a neural network prediction model for large data is constructed.

News can not only reflect the history and times, but also as an important force in the dissemination of social opinion, its role and impact on the world can not be ignored. However, with the development of intelligent network, new media gradually replaced the traditional paper media, the difficulty of false news detection and audit mechanism as well as the inability to achieve timely and effective false information detection. Social users are exposed to a large amount of new media information every day, and it is not credible and impractical to judge each piece of information by perceived awareness[1] , so there is an urgent need to find false information detection methods.

As the theory of news communication and technologies such as machine learning [2]continue to advance. Columns such as Guera[3] utilize deep learning models to generate audio and video that have never appeared in reality, which this time leads to more realistic disinformation, which is difficult to be identified by human beings. The task of disinformation detection is constantly improving the methods, and several literatures have been studied to explore the disinformation detection techniques in social media. The literature through machine learning with efficient analysis capabilities and the ability to process data, able to accurately identify news information security end with false information associated with multi-dimensional features. The current machine learning models for disinformation detection are mainly used by N. Conroy using[4] support vector machine, SVM), S. Helmstetter and H. Paulheim classifying each tweet/post as a binary classification problem before using random forest, RF, SVM, decision tree, artificial neural network, extreme gradient boosting tree algorithm, etc. The development of information technology, the length of news content and the number of information in the exponential increase in the early, how to quickly do a large number of judgments through the news headlines has become the focus. At present, short headlines often imply more complex misleading information due to the limitation of length, and the existing traditional single-model detection has the limitation of insufficient semantic capture. Although integrated learning shows the advantage of multi-dimensional feature fusion in long text detection, its application in short headline scenarios has not been explored. In this paper, we propose a novel weighted integration model based on the different degrees of difference of integration learning models to effectively enhance the feature complementarity of short text.

Based on the problems mentioned above, this paper proposes a model False news ensemble detection (FNED) based on integrated learning based on the shortcomings of the literature mentioned above, because integrated learning can fuse different machine learning by voting in order to avoid a single machine learning with a lower fit or a more fragile model for a certain type of problem. The model selection integrates three models by constructing three models (KNN, LSTM, XGBoost) after which the outputs of the three models are again used as inputs to the meta-classifier. Finally in this paper, by experimenting on the news content of FakeNewsNet dataset, the experimental results surface Accuracy, recall, Precision, F1-score are all improved, which can make up for the shortcomings of the appeal method.

To summarize, the innovations and contributions of this paper are as follows:

(1) A new model for detecting false information through integrated learning is proposed: the FNED model utilizes different base models selected in different situations, which greatly improves the model detection accuracy.

(2) In this paper, logistic regression is used as a meta-classifier, which has better accuracy compared to other meta-classifiers.

(3) In this paper, experiments have been conducted on publicly available datasets and compared with existing methods and models, and the experimental results show an average improvement in Accuracy and F1-score with reasonable interpretability.

The paper is structured as follows:

This paper describes in detail the work related to disinformation detection and the current direction of disinformation research in Section II, the model FNED is described in detail in Section III, describing the model principle and how it is constructed, and the dataset is described in detail in Section IV, along with the performance of the model under the dataset. Section V summarizes and evaluates this article.

2. Related work

2.1. Fake news detection

The following main approaches exist in existing research on methods of detecting false information on social networks:

Noteworthy classifiers based on traditional machine learning classification ,Evaluating ML models using various methods such as F1 scores, precision and accuracy is essential for evaluation, Sharma[5] produced a system which used applied ML and NLP to detect fake news using classifiers such as passive active classifier, random forest and logistic regression. Khanam[6] used XGBoost, while Pandey[7]achieved high accuracy and F1 scores using classifiers such as decision trees and logistic regression. To provide a comprehensive overview. In a social network news message such as microblogging, tweeting, which usually consists of a headline, a paragraph of text, and mechanism images, many articles have proposed salient features of rumor spreading in social media by using the features of words and phrases[8] . Kwon[9] . In addition Verónica[10] utilized SVM to implement a distinction between the differences in linguistic features that are directly present in true and false news to identify misinformation. Including many articles introducing features of FID, Potthast[11] utilizes differential writing segmentation to detect false information. Some of these texts utilize feature fusion Della Vedova[12] utilizes a TF-IDF word vector approach.M. S. Looijenga[13] Fake news detection has been found to be a predictive classification analysis application. Hybrid classification model that uses a combination of KNN and Random Forest. Integrated models, though, are more difficult to interpret due to the high complexity of including multiple model predictions and weight combinations, which may lead to the occurrence of overfitting. And it can be a significant drawback for application scenarios that require model interpretability. However, in this paper disinformation detection has a low need for interpretability and the integrated model has better prediction results.

2.2. Integrated learning

Regarding integrated learning, Ahmad[14] proposed to integrate machine learning models for automatic classification. To achieve this integration goal, Kaliyar[15] trained various machine learning models in a large and complete set and evaluated their performance using a comprehensive. The results of their novel approach demonstrate that ensemble training of models outperforms monolithic models.

Stacking is a hierarchical model integration algorithm that base models by using the outputs of the base model as feature inputs to secondary classifiers. The integrated learning model constructed by Stacking can obtain the advantages of different machine learning models, and usually gives better results than using one model. First, KNN, LSTM, and XGBoost models are used as the base models, and all the sample data are substituted into each model for the 50-50 cross validation training. The 50-50 cross validation prevents the model from falling into a local optimum due to the uneven selection of the samples, and increases the model's generalization ability. The new features are then combined with the original features, and the new feature set is rewritten and substituted into KNN, LSTM, and XGBoost to ultimately decide which category the sample belongs to. The metaclassifier reaggregates the KNN, LSTM, XGBoost model to improve the complexity of the model and greatly reduces the likelihood of model overfitting.

3. Methods

In this chapter, the model FNED was constructed to determine the output classification disinformation results.

3.1. Overall architecture of FNED

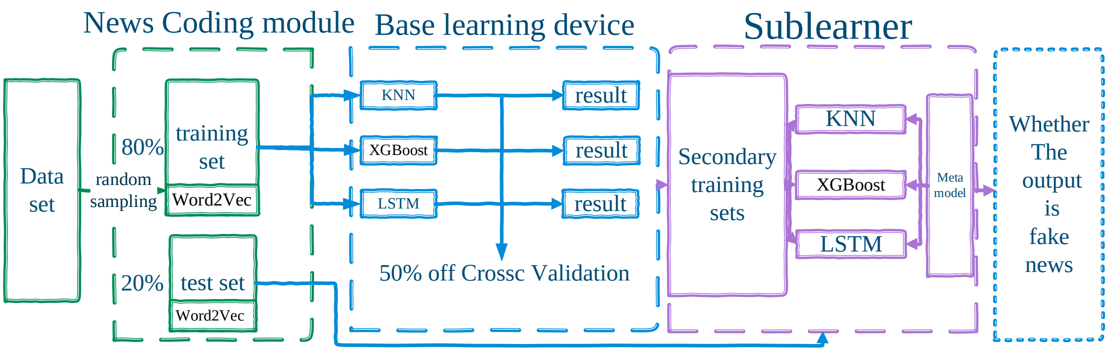

Figure 1 . Architecture diagram of FNED

The figure 1 above shows Stacking, followed by logistic regression used to improve the performance of a machine learning task by combining three machine learning models using the predictions of one or more layers of base models as inputs, using cross-validation to implement the stacking, and then combining these predictions with another model (often referred to as a meta-model or sub-model) to create the final output. We again train by Random Forest and Support Vector Machine to compare the accuracy. The algorithm parameters are also determined using 5-fold cross-validation. The performance of the model is assessed more representatively, which helps to reduce overfitting. More reliable assessment of the model's generalization ability: the model's generalization ability on unseen data can be more accurately assessed through multiple validations.

3.2. News coding module

This module focuses on feature extraction on the original input text,

$$C^{m} = \{(x_1, y_1)\ \ ,(x_2, y_2), \ldots , x_n, y_n)\} \tag{1}$$

where each x_i is a text (e.g., sentence or document), and all news headline text is converted to lowercase, corresponding to the label y_i. In this paper, pre-trained Word2Vec is utilized to convert words into vector form. These vectors are trained by neural network to capture the semantic features of the words. This step is the core of feature extraction. The semantic representation of the word vector V_{word}^{m} is obtained. If x_i is a text (e.g., a sentence) consisting of multiple words w_i, the sentence vector of the text is represented as V_{sen}^{m} by an aggregation operation (for each news text, the sentence vector is obtained by finding the mean of all the word vectors in the sentence. This sentence vector is the average of the word vectors of each word.

$$V_{word}^{m} = Word2Vec(C^{m}) \tag{2}$$

$$V_{sen}^{m} = Aggregate(V_{word1}^{m}, V_{word2}^{m}, ..., V_{wordn}^{m}) \tag{3}$$

Where $p(w_{t+j} \mid w_t)$ is usually calculated using the softmax function, which represents the probability of another word w_{t+j} occurring given the word w_t

3.3. Base model

The module contains several different base models, also known as first-level models. In this paper, three base models are used (KNN, LSTM, and XGBoost) Selecting diverse base models helps to improve the performance of the integrated model because different models may not have the same effect on different types of data, and cross-validation techniques are often used to obtain robust predictions when training base models. Modules are generated by dividing the training set into 5-fold sub-training sets.

Predicted output of KNN $H_1(x)$, predicted output of LSTM $H_2(x)$, predicted output of XGBoost $H_3(x)$.

The inputs to the metamodel are the predictions of these base models, denoted as meta-feature vectors:

where $f_i(x)$ is the feature vector of the sample C^m, of the three base model outputs

$$f_i(x) = [H_1(x), H_2(x), H_3(x)] \tag{4}$$

3.4. Meta modeling

The meta-modeling is done by Stacking, where each model is allowed to vote, and then the models are fused for output using a nonlinear approach. After training the base model, this paper generates a new dataset, the secondary training set. The features of this dataset are the predictions of the base model on the validation set, and the labels remain those of the original training set. These generated features are used to train the metamodel.

Construct a meta-training set to construct a new feature space.

matrix consisting of the feature vectors $f_i(x)$ of all the samples. The corresponding labels Y_i remain unchanged and are the same as the original true labels of the samples.

Selecting a meta-model: a simple and robust model (e.g., logistic regression) is selected as the meta-model. The logistic regression model is trained using the cross-entropy loss function as this handles the probabilistic output of the binary classification task.

Training meta-model: a logistic regression model is fitted using the constructed feature matrix F and the label Y_i. The optimization objective of the logistic regression model is to minimize the loss function, i.e:

$$Loss = -\frac{1}{N}\sum_{i=1}^{N}[y_i \log(h(f_i(x))) + (1-y_i)\log(1-h(f_i(x)))] \quad (5)$$

$$h(f_i(x)) = \frac{1}{1+e^{-(w \cdot f_i(x)+b)}} \quad (6)$$

Where $h(f_i(x))$ is the predicted probability calculated by logistic regression, the activation function of the model usually uses the sigmoida function, and finally the meta-model is evaluated by its performance on a test set, checking various metrics (e.g., accuracy, precision, recall, and F1-score) to make sure that it is improving its performance.

4. Experimental

4.1. Data sets

This experiment uses the details, experimental setup and evaluation criteria of the relevant publicly available FakeNewsNet standard dataset[16] ; the dataset comprises news articles classified as false and real, totaling 22,140 entries, with content lengths predominantly between 50 and 100 words. The data are specified in Figure 2 and Table 1.We randomly partition each dataset into training and testing sets in a ratio of 80 to 20. We will next describe the comparison techniques for the FNED model. The model's experimental results will next be summarised and discussed. A visualisation will be presented to demonstrate the effectiveness of the FNED model.

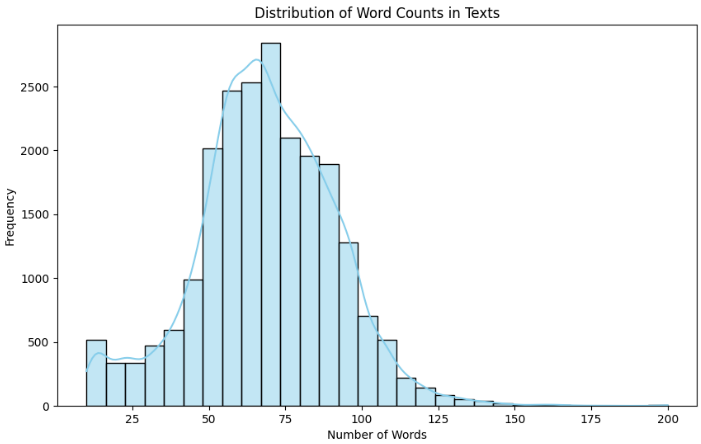

Figure 2. Distribution of word Counts in Texts

Table.1. Introduction to the dataset

Norm	Count	Fake_data	Real-data	Mean
Quantities	22140	5323	16817	68.70

4.2. Evaluation Indicators

The model's performance, whether positive or negative, requires specific standard indications for assessment. This research employs accuracy, precision, recall, and F1 score to assess the model. All aforementioned metrics are derived from the parameters within the confusion matrix. A confusion matrix, often known as an "error matrix," is a visualization tool commonly employed to juxtapose the outcomes of classifier training with the actual values. As shown in table 2.

Table.2. Confusion matrix

Parameters	Positive	Negative
Positive	TP	FN
Negative	FP	TN

Accuracy: the percentage of parts that the classifier classifies correctly, which in this experiment is a measure of the ability of the classifier to be able to accurately false information, which is calculated by the formula:

Precision rate: it is used to measure the percentage of actual positive examples that the classifier distinguishes between positive classes. It is calculated by the formula:

Recall: a measure of the ability of a classifier to correctly predict false information. Calculated by the formula:

Indicator F1: A composite measure of both precision and recall, which is calculated by the formula:

The larger the values of all the above indicators, the better the performance of the experimental model. Based on this, this paper uses precision, recall, and F1-score to evaluate the categorization performance of the data of each disinformation type and uses accuracy metrics to evaluate the overall performance. In this paper, we consider appropriately increasing the recall rate and then improving the precision, comparing the positive class in the proportion of actual positive cases, and the realistic application of the superiority when the ability to judge the false information is considered first.

4.3. Experimental details

KNN, LSTM and XGBoost were trained mainly by applying a 5-fold cross-validation surface. Specific model indicators are shown in Table 3.

Table.3. Parametric

parameter	value
KNN—nearest neighbors	7
XGBoost—learning_rate	0.1
XGBoost—max_depth	8
LSTM—Units	32

4.4. Comparison of Experimental results

In order to better evaluate the experimental performance of the FNED model, the following comparison experiments are conducted in this paper.

Table.4. Traditional machine learning models

Traditional models	Press data			
	Accuracy	Precision	recall	F1-score
KNN	0.778906	0.75	0.78	0.76
XGBoost	0.804652	0.79	0.80	0.78
LSTM	0.793134	0.78	0.79	075

Table.5. Comparison of various models

Base Classifier	Meta model	Press data			
		Accuracy	Precision	recall	F1-score
KNN+ XGBoost	Logistic regression	0.798780	0.78	0.80	0.77
KNN+ XGBoost+ LSTM	Logistic regression 2	0.810975	0.80	0.81	0.79
KNN+ XGBoost+ LSTM	Random forests	0.779358	0.77	0.79	0.77
KNN+ XGBoost+ LSTM	SVM	0.805555	0.80	0.81	0.78
KNN+ XGBoost+ LSTM	LightGBM	0.796973	0.78	0.80	0.77

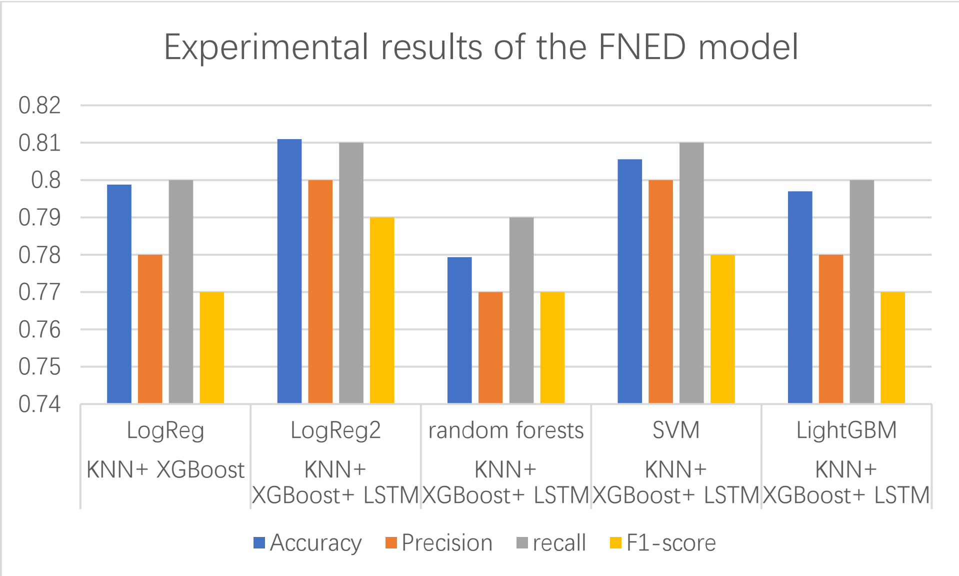

Figure 3. Comparison of visualization model parameters

To accurately assess the effectiveness of the FNFD model, various comparisons are made with all the base models and meta-models as shown in the table4, table5 and figure 3. Based on the experimental results shown in that table, this subsection provides an analysis and complete comparison.

From the above experimental results, it can be obtained that the FNED model developed in this paper achieves an ACC of 81.09%. the ACC increases by approximately 1.3%, but it is worth noting that the performance of meta-model-based is usually greater than that of traditional machine learning models. However, comparing the performance of the FNED model in terms of the performance of the meta-model on various modalities (Random Forest, SVM, LightGBm), the performance of the FNED model is still very impressive. Specifically, Random Forest performs the worst on all detections, probably because Random Forest requires a higher sample size to extract feature values. Compared to the base classifiers using only KNN and XGBOOST, this model has a better accuracy, the reason is mainly because using different algorithms or models can be able to capture different features under different types of data.

Obviously, the FNFD model outperforms all other methods. The advantage is that by training meta-classifiers with different base classifiers, can reduce the uncertainty of different feature recognition and sentiment in different social media short texts, which helps in disinformation recognition.

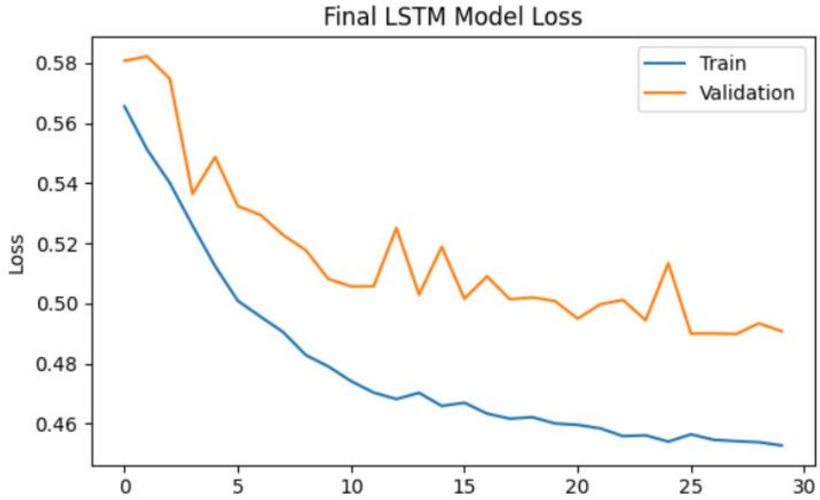

Figure 4. Comparison of visualization model parameters

According to the loss function curve of the model training Figure 4 shows that the LSTM model exhibits good convergence performance during the training process. The loss value (blue line) of the training set steadily decreases from the initial 0.56 to around 0.45, while the loss value (orange line) of the validation set also gradually decreases from 0.58 to around 0.49, keeping a moderate gap between the two curves, indicating that the model neither suffers from serious overfitting phenomenon nor underfitting. Especially in the late stage of training (20-30 epochs), the loss curve smoothes out, indicating that the model has reached a better convergence state. This performance of the training curve confirms the rationality of the model design and the effectiveness of the training process.

5. Conclusions

This paper presents an integrated learning-based Fake News Detection Model (FNFD), which improves the recognition of fake news. It mainly includes the proposal of integrated models, using a combination of three models, KNN, LSTM, and XGBoost, and the prediction results of these models are used as features through a stacking technique, and then logistic regression is used as a meta-classifier to make a comprehensive judgment. This design aims to combine the advantages of each model to improve the overall detection accuracy. The use of logistic regression for the meta-classifier is conducive to improving accuracy. Finally, experiments with the open-source FakeNewsNet dataset back up this paper. The results show that the FNFD model is much more accurate and has a higher F1 score, which means that the integrated learning approach works better at finding fake news. The FNFD model does better than the single model and other metaclassifier models in all classification metrics. This is especially true when using the integrated judgment of logistic regression to find fake news more accurately. This suggests that integrating different types of learning models can enhance the performance of detecting false information. In the future, this paper will use deep learning models with stronger performance for fake news detection.

Reference

[1] QIU X, OLIVEIRA D F M, SHIRAZI A S, et al. RETRACTION: Limited individual attention and online virality of low-quality information (Retraction of Vol 1, art no 0132, 2017) [J]. Nature Human Behaviour, 2019, 3(1): 102-.

[2] JORDAN M I, MITCHELL T M. Machine learning: Trends, perspectives, and prospects [J]. Science, 2015, 349(6245): 255-60.

[3] GUERA D, DELP E J. Deepfake Video Detection Using Recurrent Neural Networks [J]. 2018 15th IEEE International Conference on Advanced Video and Signal Based Surveillance (AVSS), 2018: 1-6.

[4] CONROY N, RUBIN V L, CHEN Y. Automatic deception detection: Methods for finding fake news [J]. Proceedings of the Association for Information Science and Technology, 2015, 52.

[5] SHARMA U, SARAN S, PATIL S M. Fake news detection using machine learning algorithms [J]. International Journal of creative research thoughts (IJCRT), 2020, 8(6): 509-18.

[6] KHANAM Z, ALWASEL B, SIRAFI H, et al. Fake news detection using machine learning approaches; proceedings of the IOP conference series: materials science and engineering, F, 2021 [C]. IOP Publishing.

[7] PANDEY S, PRABHAKARAN S, REDDY N S, et al. Fake news detection from online media using machine learning classifiers; proceedings of the Journal of Physics: Conference Series, F, 2022 [C]. IOP Publishing.

[8] RUBIN V L, LUKOIANOVA T. Truth and Deception at the Rhetorical Structure Level [J]. Journal of the Association for Information Science and Technology, 2015, 66(5): 905-17.

[9] KWON S, CHA M, JUNG K, et al. Prominent features of rumor propagation in online social media; proceedings of the 2013 IEEE 13th international conference on data mining, F, 2013 [C]. IEEE.

[10] PéREZ-ROSAS V, KLEINBERG B, LEFEVRE A, et al. Automatic detection of fake news [J]. arXiv preprint arXiv:170807104, 2017.

[11] POTTHAST M, KIESEL J, REINARTZ K, et al. A stylometric inquiry into hyperpartisan and fake news [J]. arXiv preprint arXiv:170205638, 2017.

[12] DELLA VEDOVA M L, TACCHINI E, MORET S, et al. Automatic online fake news detection combining content and social signals; proceedings of the 2018 22nd conference of open innovations association (FRUCT), F, 2018 [C]. IEEE.

[13] LOOIJENGA M S. The detection of fake messages using machine learning [D]; University of Twente, 2018.

[14] AHMAD I, YOUSAF M, YOUSAF S, et al. Fake news detection using machine learning ensemble methods [J]. Complexity, 2020, 2020(1): 8885861.

[15] KALIYAR R K, GOSWAMI A, NARANG P. FakeBERT: Fake news detection in social media with a BERT-based deep learning approach [J]. Multimedia tools and applications, 2021, 80(8): 11765-88.

[16] SHU K, MAHUDESWARAN D, WANG S, et al. FakeNewsNet: A Data Repository with News Content, Social Context, and Spatiotemporal Information for Studying Fake News on Social Media [J]. Big Data, 2020, 8(3): 171-88.