Research Article

Weibo Rumor Recognition Based on Communication and Stacking Ensemble Learning

Yu Wu ¹, Yan Zeng,² Jie Yang,² and Zhenni Zhao²

¹*School of Cyber Security and Information Law, Chongqing University of Posts and Telecommunications, Chongqing 400065, China*

²*School of Computer Science and Technology, Chongqing University of Posts and Telecommunications, Chongqing 400065, China*

Correspondence should be addressed to Yu Wu; wuyu@cqupt.edu.cn

Received 5 June 2020; Revised 23 August 2020; Accepted 2 September 2020; Published 18 September 2020

Academic Editor: Wei Wang

Copyright © 2020 Yu Wu et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Research on the identification of rumors in cyberspace helps to discover social issues that are of concern to the public and are not easily found, and it also can help to purify cyberspace and to maintain social stability. However, the real complexity of rumors makes it difficult for its recognition technology to bridge the semantic gap between qualitative description and quantitative calculation of rumors. Firstly, the existing rumor definitions are mostly qualitative descriptions, so we propose a technical definition of Internet rumors to facilitate quantitative calculations. Secondly, since the feature set used in rumor recognition research is not effective, by combining with communication, we construct a more suitable feature set for rumor recognition. Thirdly, aiming at the problem that traditional classification algorithms are not suitable for complex rumor information recognition, a rumor recognition method based on Stacking ensemble learning is proposed. Our experiment results show that the proposed method has higher accuracy, less algorithm execution time, and better practical application effect.

1. Introduction

Nowadays, the Internet is full of various information like Internet rumors, malware, and fake news, and it is difficult for people to tell the truth of that information. With the development of cyberspace, the researches on Internet misinformation have gradually mainly focused on Internet rumor recognition, Internet rumor propagation, malware and virus propagation, fake news detection, and water army detection.

The research on malware and virus propagation focused on modeling the spread of malicious malware to predict the spreading behaviour of it [1], while the research on water army is mainly about detecting water army among a large number of social media users in an online topic and preventing the negative effect on the development of the public opinion [2]. The spread of rumors may affect personal reputation, invade public privacy, or cause chaos in public order, lead to group events, and endanger the stability of the country. Therefore, modeling the propagation of Internet

rumors in social network to help control the spread of the misinformation is of great importance [3]. However, it is also vital to figure out what are Internet rumors and how to recognize them in a more effective way.

In paper [4], Peterson et al. proposed that rumor, in general, refers to an unverified account or explanation of events, circulating from person to person and pertaining to an object, event, or issue of public concern. Since then, rumors have been given some specific characteristics, such as ambiguity, transmissibility, and timeliness. With the development of the Internet, the spread of information has accelerated, and Internet rumors are derived. In paper [5], Chao et al. considered that Internet rumors were unconfirmed information transmitted by Internet users in specific ways. Most scholars believe that the spread of Internet rumors is carried out on the Internet, because the network connection is wide and arbitrary, which makes the spread of Internet rumors faster, the coverage of influence wider, and the extent of damage bigger [6]. However, the existing definitions cannot accurately describe the technical

characteristics of Internet rumors and other elements from the perspective of computability.

At present, on the one hand, the research on Internet rumors recognition focuses on extracting feature set that can be used to detect rumors. On the other hand, the classification modeling of Internet rumors, which does not need various features but a large amount of data, has become a research hotspot. However, feature selection of Internet rumors is more suitable for a small number of datasets.

The feature set used in current Internet rumor recognition is based on the features proposed by Castillo et al. [7] and Qazvinian et al. [8]. Generally, the feature set is divided into three types: content feature, user feature, and propagation feature; sometimes it is further subdivided into time feature, network feature, and combination of the two. These features are usually simple statistical characteristics, and the deep semantics of the text information are not mined. Therefore, the recognition accuracy is affected by the lack of key features. In paper [9], Kwon et al. applied RNN to learn the deep meaning between messages. Based on the protection mechanism, Chen et al. [10] proposed a new RNN model to recognize the Internet rumor by using time sequence to obtain the potential contextual change. Although the neural network model can overcome the problem of sparse features by using continuous vectors to represent text, it has too many parameters and slow convergence speed and needs a lot of corpus.

The classification algorithms often used in rumor recognition research include SVM, Decision Tree, Naive Bayes, and Neural Networks. For example, in paper [11], Duan et al. used SVM to detect fake information from the comment perspective of the source Weibo. In paper [12], Chen et al. used the regression method to recognize the online food safety rumors. However, this method is limited by the topic type of the rumor and can only determine whether the article is related to the rumor. In paper [13], Lu et al. proposed an improved method based on the Co-Forest algorithm, which improved the accuracy of prediction for unlabelled samples, to solve the problem of data imbalance.

Due to the fact that the dataset is difficult to obtain, the current research tends to extract the statistical characteristics of text information. Moreover, new experimental methods in which features are added on the basis of previous researches make the feature dimension continuously increase and lead to the inaccurate model parameter. Besides, classical algorithms such as SVM, Decision Tree, and Naive Bayes are no longer suitable for recognizing Internet information with complex content. In specific problems and scenarios, each model has its advantages and disadvantages. The results of recognition might be better by combining the advantages of multiple models [14]. For example, in paper [15], Xie et al. proposed a high-precision EEG emotion recognition model by integrating LightGBM, XGBoost, and Random Forest. In paper [16], Duan et al. classified the sentiment of Weibo text by using the Stacking ensemble learning method, and the accuracy rate was as high as 93%.

In this paper, we give the definition of Internet rumor based on communication's 5W Formula, and three characteristics of User Credibility, Emotional Consistency, and

Regional Correlation are constructed. Then, we verify the validity of the feature by using Chi-square, so as to better filter the feature set. Finally, analysing the existing classification algorithms, adopting the Stacking ensemble learning, we propose a rumor recognition method combined with different models and optimized using cross validation. Finally, the experiments are conducted among different methods and datasets.

The structure of this paper is as follows: the related work is introduced in the first part. The design of the Stacking model, which includes features construction and selection, and the construction of the model are described in the second part. In the third part, we conduct the experiments and analyse the results. The conclusion of this paper and the introduction of future work are given in the last part.

2. Related Work

2.1. Ensemble Learning and Stacking. For different problems, the speed, accuracy, and generalization ability of machine learning models are different. To obtain a model with strong generalization ability and high robustness, the ensemble learning method is proposed. The current mainstream methods are Boosting (typically AdaBoost, GBDT, and XGBoost), Bagging (typically Random Forest), and Stacking. The Stacking model applied in this paper has a layered framework. The original training dataset is inputted into multiple primary learners in the first layer, and the first layer's prediction results are used as the input training set for the next layer of learners. Finally, the prediction results obtained in the previous layer are inputted into the final metalearners to get the final prediction result (see Figure 1).

2.2. Other Methods. SVM [17] is a binary classification model that uses the margin maximization strategy. It minimizes the structured risk, the empirical risk, and confidence range to improve the generalization ability. Therefore, in a small statistical dataset, it can also get good statistical regularity.

Decision Tree J48 [8] is based on C4.5. It uses a divide-and-conquer strategy and has high credibility; and its results are easy to understand.

Random Forest [18] is one of the ensemble learning methods. By combining multiple weak classifiers, the final result is obtained by voting or calculating mean value. The model has high accuracy and generalization performance.

Logistic regression [19] is one of the generalized linear models, as well as a classical classification method used to solve the optimization problem with the likelihood function as the objective function.

3. Weibo Rumor Recognition

3.1. Rumor Definition. Based on the existing rumor definitions, this paper combines the 5W's of communication to divide the elements of rumor transmission into communicators, content, objects, effects, and channels. The communicator could be an individual or an organization. The content is the information that the communicator wants to

FIGURE 1: Stacking principle.

pass to the audience. The object is the receiver of information or the communicator of information processing. The effect is that the audience is affected by the message sent by the communicator, causing changes in their ideas, behaviours, and so forth. The channel is the mean by which communication is achieved. Sometimes the object of Internet rumors is also the communicator, so we combine them as the same element for analysis. Therefore, the qualitative definition of Internet rumor given in this paper is shown in Definition 1.

Definition 1. Internet rumors: internet rumors refer to the information published by Internet users through online media platforms, which is ambiguous in content, unconfirmed by the official, and to some extent harmful to the society. Its expressions include text, pictures, audio, and video.

The research object of this paper is Weibo. Weibo rumor, which can be divided into pure text, picture not matching the text, and fake images, is a type of the Internet rumors. Since most of Weibo contains text, the current recognition of Weibo rumors mainly focuses on text. In order to facilitate the recognition of Internet rumors, as shown in Definition 2, this paper gives its formal definition from computability.

Definition 2. Weibo rumors: the object of rumor recognition is a Weibo set $\mathbf{W}\{\mathbf{m}_1, \mathbf{m}_2, \mathbf{m}_3, \dots, \mathbf{m}_n\}$, and $\mathbf{m}_i = \langle \mathbf{X}_i, y_i, f \rangle$. $\mathbf{X}_i\{\mathbf{U}_i, \mathbf{T}_i, \mathbf{P}_i\}$ is the feature set of each Weibo. The user feature set $\mathbf{U}_i\{f_{u_1}, f_{u_2}, f_{u_3}, \dots, f_{u_n}\}$ represents the attributes of the i -th Weibo's publisher. The content feature set $\mathbf{T}_i\{f_{t_1}, f_{t_2}, f_{t_3}, \dots, f_{t_n}\}$ represents the attributes of the i -th Weibo's text. The propagation feature set $\mathbf{P}_i\{f_{p_1}, f_{p_2}, f_{p_3}, \dots, f_{p_n}\}$ represents the propagation attributes of the i -th microblog. y_i is the confidence value of whether $\mathbf{m}_i$ is a rumor, and $y_i = f(\mathbf{X}_i)$, $(0 \leq y_i \leq 1)$. When y_i is closer to 1, the probability of $\mathbf{m}_i$ being a rumor is higher and vice versa.

3.2. Rumor Recognition Process. At present, the research on rumor recognition mainly focuses on feature construction, and the method of adding new features based on previous

research work would make the feature dimension increase and estimation of model parameters inaccurate. Therefore, using the Chi-square test to test the validity of new features could obtain a feature set that is more suitable for Internet rumors recognition. Internet rumor recognition is regarded as a binary classification problem. In this paper, it is considered to use the idea of Stacking ensemble learning to build a new classification model because each algorithm has its advantages (see Figure 2).

3.3. Feature Construction. Combining the research studies of paper [8] and paper [20], we select 24 basic features, which are divided into content feature (CONT), user feature (USER), and propagation feature (TRAN). The content feature includes the length of text, the number of @, the number of #, the number of question/exclamation marks, whether there are pictures or URLs, and the number of positive/negative words. The user feature includes the length of username, gender, the number of friends, the number of followers, the number of mutual followers, the number of microblog posts, the number of favourite microblogs, certification information, personal description, and user's own influence. The propagation feature includes the number of reposts, the number of comments, the number of likes, the interval between user registration time and microblog post time, and microblog's attention.

The above features are mostly statistical features; in order to recognize Weibo rumor more effectively, this paper constructs new features in three aspects, user, content, and propagation features, to mine the hidden meaning behind text information.

Definition 3. User Credibility (UCRE): the user's credibility is determined by many factors. By integrating information such as the number of users' friends, followers, mutual followers, the number of microblogs posted, and certification information, the user's influence and activity are constructed to calculate the user's credibility. The more credible a user is, the more credible the information he/she posts. The calculation formula of the user's credibility is as follows:

FIGURE 2: Rumor recognition flowchart.

$$f_{\text{credibility}}(u_i) = f_{\text{verified}}(u_i) + f_{\text{influence}}(u_i) + f_{\text{InfoIntegrity}}(u_i), \quad (1)$$

where $f_{\text{influence}}(u_i)$ is the user's influence, f_{verified} is the user's certification information, and $f_{\text{InfoIntegrity}}$ is whether the user's information is complete; the user's information includes username, gender, personal description, registration place, and profile photo. The greater the user's influence, the greater the impact of the microblog they posted within a certain time and space. The user's influence is mainly determined by the number of users' followers and the number of mutual followers. The calculation formula is as follows:

$$\begin{aligned} f_{\text{influence}}(u_i) &= \frac{C_{\text{bifollowers}}(u_i)}{C_{\text{followers}}(u_i)}, \\ f_{\text{verified}}(u_i) &= \begin{cases} 0, & \text{unverified,} \\ 1, & \text{verified,} \end{cases} \\ f_{\text{InfoIntegrity}}(u_i) &= \begin{cases} 0, & \text{complete,} \\ 1, & \text{incomplete,} \end{cases} \end{aligned} \quad (2)$$

where u_i is user who posts i -th microblog, $C_{\text{bifollower}}$ is the number of u_i 's mutual followers, and C_{follower} is the number of u_i 's followers.

Definition 4. Emotion consistency (ECON): emotion consistency is whether the sentiment of the microblog is consistent with the sentiment of the microblog's comment. When the microblog shows a strong emotion, it may incite others' emotions; then the microblog is more likely to be a rumor. By segmenting the $\mathbf{m}_i$'s text and comments, we obtain the text's term vector set $\mathbf{V}_i\{v_1, v_2, v_3, \dots, v_n\}$, where v_i is the processed word, and the i -th microblog's j -th comment's term vector set $\mathbf{C}_j^i\{c_1, c_2, c_3, \dots, c_n\}$, where c_i is the processed word.

The number of positive/negative words is calculated by using Affective Lexical Ontology [21]. The specific formula is as follows:

$$S = C_{\text{pos}} - C_{\text{neg}}, \quad (3)$$

where S is the emotion of the term vector set, C_{pos} is the number of positive words, and C_{neg} is the number of

negative words. Then we can get the final emotion SO as follows:

$$SO = \begin{cases} 1, & S > 0, \\ 0, & S = 0, \\ -1, & S < 0, \end{cases} \quad (4)$$

where 1 represents positive, -1 represents negative, and 0 represents neutral. Calculating the emotion of each comment, the overall emotion of the comments is calculated as follows:

$$SO(c^i) = \begin{cases} 1, & \sum_{j=1}^n SO(c_j^i) > 0, \\ 0, & \sum_j SO(c_j^i) = 0, \\ -1, & \sum_j SO(c_j^i) < 0. \end{cases} \quad (5)$$

Comparing the emotions of $\mathbf{m}_i$ and c^i , the emotion consistency of $\mathbf{m}_i$ is calculated as follows:

$$SC = \begin{cases} 0, & SO(m_i) \neq SO(c^i), \\ 1, & SO(m_i) = SO(c^i). \end{cases} \quad (6)$$

Definition 5. Regional correlation (RCO): regional correlation refers to the distance between the place mentioned in the microblog and the user's registration place. The longer the distance is, the less credible the microblog is. This paper uses Euclidean distance to calculate the distance. The formula is as follows:

$$\text{dist}(x, y) = \sqrt{\sum_{i=1}^n (x_i - y_i)^2}, \quad (7)$$

where $\text{dist}(x, y)$ is the distance between city x and city y , the coordinate of city x is (x_1, x_2) , and the coordinate of city y is (y_1, y_2) . Calculating the distance between cities in China, the distance matrix is shown as follows:

$$\mathbf{M} = \begin{bmatrix} 0 & 96.06 & \cdots & 2433.9 & 2666.29 \\ 96.06 & 0 & \cdots & 2525.47 & 2759.9 \\ \cdots & \cdots & 0 & \cdots & \cdots \\ 2433.9 & 2525.47 & \cdots & 0 & 310.74 \\ 2666.29 & 2759.92 & \cdots & 310.74 & 0 \end{bmatrix}. \quad (8)$$

According to the difference between the user registration place and the place mentioned by microblog, it can be divided into 4 cases: ① Both the user registration place and the place mentioned in microblog are in China. ② The user registration place is in China, but the place mentioned in microblog is not. ③ The user registration place is not in China, but the place mentioned in microblog is. ④ Neither the user registration place nor the place mentioned in microblog is in China. Since most of Weibo rumors occur in China, the current research is mainly focused on case ①. In cases ②, ③, and ④, the distance would be set to 10000, which indicates the maximum threshold.

3.4. Feature Selection. In order to test the validity of the basic features and new features, we use Chi-square test to obtain the feature ranking results, as shown in Table 1.

As shown in Table 1, the Regional Correlation, Emotional Consistency, and User Credibility are ranked 3rd, 5th, and 9th, so the three new features we constructed are valid.

Two sets of control experiments are conducted on different models. One is that which is based on the new features and adds features one by one according to the features ranking results. In the other experiment, feature sets are added one by one according to the features ranking results. The experimental results are shown in Figures 3 and 4.

As illustrated in Figures 3 and 4, as the number of features in the feature set increases, the model's recognition accuracy gradually increases, but when the features exceed a certain number, the model's recognition accuracy tends to decrease.

In Figure 3, the accuracy of Naive Bayes is highest when the number of features is increased to 15; and when the number of features is increased from 3 to 12, the accuracy of SVM is significantly higher than that of Decision Tree. When the number of added features exceeds 12, the accuracy of Decision Tree starts to exceed SVM. As the number of features continues to increase, the accuracy of the Random Forest model continues to increase, but the accuracy decreases as the number of features exceeds 21. In general, the results of each model are the best when the number of features is around 13–14.

In Figure 4, the result is not much different from the result of Figure 3. The number of features in the feature set with better results is mostly around 16. In summary, we use the first 16 features shown in Table 1 as the feature set. The final feature set used in rumor recognition is shown in Figure 5.

3.5. Classification Algorithm. The Stacking method is adopted as a combination strategy of ensemble learning in

TABLE 1: Features ranking results.

Ranking	Description	Score
1	Verified	477.060
2	Time span	358.353
3	Regional correlation	305.972
4	The number of friends	268.123
5	Emotional Consistency	255.905
6	User Influence	159.063
7	The number of microblogs	147.373
8	The number of #	78.590
9	User Credibility	65.372
10	The number of mutual followers	47.535
11	The number of followers	36.855
12	The number of @	25.129
13	The number of reports	10.981
14	Picture	6.679
15	The length of text	6.468
16	The number of likes	5.991
17	The number of exclamation marks	5.841
18	The number of negative words	5.722
19	User information integrity	3.279
20	Personal description	3.279
21	The length of username	2.824
22	The number of positive words	2.058
23	Gender	1.604
24	The number of comments	1.546
25	The number of question marks	1.160
26	The number of favourite microblogs	0.652
27	User activity	0.309

FIGURE 3: Based on the new features.

this paper. We select SVM, Random Forest, and Naive Bayes as the primary learners and logistic regression as the meta-learner. SVM uses the hinge loss function to calculate surrogate loss, which makes it sparse. At the same time, it considers the minimization of the empirical risk and the structural risk to make it stable [22], so it has better generalization ability and a smaller computational cost when using kernel function [17]. Random Forest can estimate missing data and balance errors for imbalanced data [18]. When the correlation between attributes is small, the performance of the NB model is better. The model construction is shown in Figure 6.

FIGURE 4: According to the features ranking.

FIGURE 5: Final feature set.

The specific algorithm is described in Algorithm 1.

While time complexities of random forest and logistic regression model are $O(n \log n)$ and $O(n \cdot k + k)$ (where k is the number of features), that of SVM determined by the kernel function and Naive Bayes can reach $O(n^3)$. According to the strategy of the Stacking model, its time complexity equals the maximum value among the primary learners and metalearner. Therefore, the time complexity of Algorithm 1 is $O(n^3)$.

4. Experiment and Analysis

4.1. Dataset. We use the data from Ma et al. [23], which contain 2313 rumor events and 2351 nonrumor events, about 3.8 million pieces of microblog information, and 2.7 million pieces of user information. In the experiment, we split the dataset into training set and test set according to the proportion of 8:2.

At the same time, in order to verify the validity of the method in this paper on actual network data, we collected data on the Weibo platform and established an empirical database. The datasets used for empirical study in this paper are shown in Table 2.

4.2. Algorithm for Comparison. To verify the validity of the method proposed in this paper, we compare the following methods with our models: tanh-RNN [23], the method used in the paper where the data source is; SVM [20], the first method used in Weibo rumor recognition; Decision Tree J48 [8], the first method used in Twitter fake information recognition; AdaBoost and Random Forest, representative ensemble learning methods; and the method proposed in this paper. SVM and Decision Tree J48 are usually used as a benchmark for Internet rumor recognition in most research works.

4.3. Experimental Procedures

- Feature set comparison: based on the recognition model proposed in this paper, we conduct experiments on different feature sets to verify the validity of the new features proposed in this paper
- Algorithm comparison: in order to measure the accuracy and generalization ability of the recognition method proposed in this paper, we compared different algorithms through the accuracy, precision, recall, and F1-score

FIGURE 6: The model construction.

Input: a Weibo set $\mathbf{W}\{\mathbf{m}_1, \mathbf{m}_2, \mathbf{m}_3, \dots, \mathbf{m}_n\}$

Output: the confidence value y_i of $\mathbf{m}_i$

Step 1: extract features of $\mathbf{m}_i$ as shown in Figure 5

Step 2: calculate user credibility of $\mathbf{m}_i$

Step 3: segment $\mathbf{m}_i$ and its comments, and calculate emotion consistency of $\mathbf{m}_i$

Step 4: calculate regional correlation of $\mathbf{m}_i$

Step 5: standardize each feature of $\mathbf{m}_i$

Step 6: split the preprocessed data set into train set, test set and validation set, and input them into SVM, RF, and Naïve Bayes model

Step 7: input the new feature set in step 6 into logistic regression model

Step 8: calculate the accuracy, precision, recall and F1-score of the Stacking model

ALGORITHM 1: Weibo rumor recognition method based on stacking.

TABLE 2: Empirical study data.

Event	The number of Weibo	The number of users
Doctor accident	3,057	1,424
2019-nCoV	3,091	1,223
Community fire	2,447	896

- (c) Algorithm execution time comparison: we compare the training time and test time in different models to analyse model performance
- (d) Empirical analysis: the practicality of the recognition method proposed in this paper is verified by conducting experiments on the latest events

4.4. Feature Set Comparison. In this paper, we compare different feature sets by using the recognition model proposed in this article. The experimental results are shown in Table 3 ($\mathbf{F}_{\text{CONT}}$ is the Content Features set, $\mathbf{F}_{\text{USER}}$ is the User Features set, $\mathbf{F}_{\text{PROP}}$ is the Propagation Features set, UCRE is the User Credibility, ECON is the Emotional Consistency, ROC is the Regional Correlation, and $\mathbf{F}_{\text{SIFT}}$ is the Features Set shown in Figure 5).

Table 3 shows that the accuracy rate of using $\mathbf{F}_{\text{CONT}}$ for Weibo rumor recognition is as low as 70%, which indicates that it is difficult to detect rumors in the more complicated content. Compared with only relying on $\mathbf{F}_{\text{CONT}}$ for recognition, the recognition results by using $\mathbf{F}_{\text{USER}}$ and $\mathbf{F}_{\text{PROP}}$ are better; in particular, the accuracy is improved by 20%. In the experimental results of $\mathbf{F}_{\text{CONT+TRAN}}$, $\mathbf{F}_{\text{USER+TRAN}}$, and $\mathbf{F}_{\text{CONT+USER}}$, we can see that their accuracy is 0.2%, 1.9%, and 2.6% lower than that of $\mathbf{F}_{\text{SIFT}}$, respectively. The accuracy of $\mathbf{F}_{\text{CONF+SENT+LOC}}$, which is composed of only three new features constructed in this paper, is as high as 90.8%, which shows that the three new

features of User Credibility, Emotional Consistency, and Regional Correlation constructed in this paper have good effect on Weibo rumor recognition. However, the experimental results of $\mathbf{F}_{\text{SIFT-CONF-SENT-LOC}}$ without three new features are lower than those of $\mathbf{F}_{\text{SIFT}}$, and the accuracy of $\mathbf{F}_{\text{SIFT}}$ is 93%. The accuracy and recall rate of rumor and nonrumor recognition are all over 90%, and the F1-score is also stable at 93%. The feature set proposed in this paper has higher values than other feature sets, which indicates that our feature set is more effective to detect Weibo rumors.

In order to validate the effectiveness of each feature in $\mathbf{F}_{\text{SIFT}}$ selected in this paper, applying the rumor recognition method based on Stacking, we conduct 16 different experiments on $\mathbf{F}_{\text{SIFT}}$ with one feature removed every time. The results are shown in Table 4.

As is shown in Tables 3 and 4, the accuracy in each experiment is lower than that of $\mathbf{F}_{\text{SIFT}}$; and the rumor recognition accuracy based on the feature set without the regional correlation is the lowest, which indicates that the Regional Correlation has the biggest impact on the recognition results. In conclusion, the 16 features selected in this paper have a positive effect on rumor recognition.

4.5. Algorithm Comparison. We compare different algorithms with the rumor recognition model proposed in this paper to illustrate the accuracy and generalization ability of the model we proposed. The results are shown in Table 5.

As shown in Table 5, compared with tanh-RNN, SVM, and Decision Tree J48, the Stacking model has the highest accuracy rate of 93.5%. The Stacking model can recognize rumor with 96.5% recall rate and 91.4% precision, which shows that the model can recognize more rumors; and the Stacking model can recognize nonrumor events with 93%

TABLE 3: Feature set comparison results.

Feature set	Class	Accuracy	Precision	Recall	F1-score
F_{CONT}	N	0.704	0.662	0.783	0.717
	R		0.760	0.631	0.690
F_{USER}	N	0.913	0.938	0.877	0.906
	R		0.893	0.947	0.919
F_{PROP}	N	0.921	0.954	0.877	0.914
	R		0.895	0.961	0.927
$F_{\text{USER+PROP}}$	N	0.932	0.960	0.897	0.927
	R		0.911	0.933	0.932
$F_{\text{CONT+PROP}}$	N	0.916	0.945	0.877	0.910
	R		0.894	0.953	0.922
$F_{\text{CONT+USER}}$	N	0.909	0.929	0.877	0.902
	R		0.892	0.938	0.915
$F_{\text{SIFT-UCRE-ECON-RCO}}$	N	0.922	0.950	0.884	0.915
	R		0.899	0.957	0.927
$F_{\text{UCRE+ECONS+RCO}}$	N	0.908	0.909	0.897	0.903
	R		0.907	0.918	0.912
F_{SIFT}	N	0.935	0.960	0.902	0.930
	R		0.914	0.965	0.939

TABLE 4: The effectiveness of each feature in F_{SIFT} .

Ranking	Removed	Class	Accuracy	Precision	Recall	F1-score
1	The number of friends	N	0.9344	0.9573	0.9038	0.9298
		R		0.9159	0.9630	0.9388
2	The number of #	N	0.9344	0.9552	0.9060	0.9300
		R		0.9175	0.9609	0.9387
3	Emotional Consistency	N	0.9343	0.9531	0.9083	0.9301
		R		0.9191	0.9588	0.9386
4	The number of mutual followers	N	0.9335	0.9577	0.9028	0.9294
		R		0.9131	0.9630	0.9373
5	The number of followers	N	0.9334	0.9551	0.9038	0.9287
		R		0.9157	0.9609	0.9378
6	The number of @	N	0.9334	0.9487	0.9105	0.9292
		R		0.9206	0.9547	0.9374
7	Picture	N	0.9334	0.9529	0.9060	0.9289
		R		0.9173	0.9588	0.9376
8	Verified	N	0.9332	0.9508	0.9083	0.9291
		R		0.9190	0.9568	0.9375
9	User Credibility	N	0.9332	0.9508	0.9083	0.9291
		R		0.9190	0.9568	0.9375
10	The number of microblogs	N	0.9331	0.9541	0.9060	0.9294
		R		0.9180	0.9612	0.9391
11	The length of text	N	0.9325	0.9507	0.9060	0.9278
		R		0.9172	0.9568	0.9366
12	User Influence	N	0.9325	0.9550	0.9016	0.9275
		R		0.9139	0.9609	0.9368
13	The number of reports	N	0.9303	0.9505	0.9016	0.9254
		R		0.9136	0.9568	0.9347
14	The number of likes	N	0.9303	0.9401	0.9128	0.9262
		R		0.9218	0.9465	0.9340
15	Time span	N	0.9293	0.9441	0.9060	0.9247
		R		0.9167	0.9506	0.9333
16	Regional Correlation	N	0.9175	0.9405	0.8837	0.9112
		R		0.8986	0.9486	0.9229

TABLE 5: Algorithm comparison results.

Algorithm	Class	Accuracy	Precision	Recall	F1-score
SVM	N	0.917	0.917	0.911	0.914
	R		0.918	0.924	0.921
NB	N	0.813	0.819	0.815	0.817
	R		0.808	0.812	0.810
RF	N	0.935	0.945	0.917	0.931
	R		0.926	0.951	0.938
LR	N	0.905	0.918	0.873	0.898
	R		0.893	0.928	0.910
J48	N	0.923	0.916	0.931	0.920
	R		0.933	0.922	0.926
tanh-RNN	N	0.873	0.956	0.782	0.861
	R		0.816	0.964	0.884
DT-rank	N	0.732	0.726	0.749	0.737
	R		0.738	0.715	0.726
CAMI	N	0.933	0.945	0.921	0.932
	R		0.921	0.945	0.933
Stacking	N	0.935	0.960	0.902	0.930
	R		0.914	0.965	0.939

TABLE 6: Algorithm execution time results.

Time Algorithm	Training time (ms)	Testing time (ms)
SVM	2434.3073	24.2662
J48	32.0261	0.3138
NB	1.9568	0.9729
LR	15.342	0.2064
AdaBoost	2544.4393	31.1073
RF	532.5212	16.4673
Stacking	7954.252	0.0765

(a)

(b)

FIGURE 7: Continued.

FIGURE 7: Empirical analysis results-Stacking.

FIGURE 8: Empirical analysis results-feature set.

F1-score and recognize rumor with 93.9% F1-score, which are higher compared with other algorithms. The above results show that the Stacking model proposed in this paper has the best recognition effect.

4.6. Algorithm Execution Time. We calculate the training time and test time of each algorithm separately, and the results are shown in Table 6.

Table 6 shows that the Naive Bayes algorithm takes the shortest time in training, but the Stacking model proposed in

this paper takes the longest time in training because it ensembled multiple algorithms. In the testing phase, the Stacking model takes the shortest time, which only accounts for 7.8% of Naive Bayes, and the second shortest logistic regression model is 2.7 times longer than the Stacking model.

4.7. Empirical Analysis. In order to verify the practicability of the Weibo rumor recognition method proposed in this paper, we use three events for experimenting (see Table 2), and the results are shown in Figure 7.

Figure 7 shows that, in different events, the performance of each model is a little different. RF has 98% recall and 77% accuracy, which indicates that the RF model may recognize most nonrumor Weibo as rumor. Although the Stacking model has 93% recall and 80% accuracy, it is higher than other models. Therefore, the model proposed in this paper is relatively effective in actual recognition.

Besides, in order to verify the effectiveness of the new features proposed in this paper, the comparative experiments are conducted on the feature set with UCRE, ECON, and RCO, and the feature set excludes them. The results are shown in Figure 8.

As illustrated in Figure 8, evaluating on the accuracy, precision, recall, and F1-score, the Stacking model implemented on the feature set with UCRE, ECON, and RCO performs better. In conclusion, the new features proposed in this paper are suitable for actual rumor recognition.

5. Conclusion

In this paper, based on previous studies, three new features are constructed by combining the 5W's Formula in communication; and we obtain the optimal feature set of Weibo rumor recognition through the Chi-square test and other methods. Then, based on the idea of the Stacking method, we select SVM, Random Forest, and Naive Bayes as the primary learner and logistic regression as the metalearner to model and analyse the Weibo rumor recognition. The experimental results show that the features constructed in this paper and the proposed recognition method can effectively detect rumors in Weibo.

However, there are still some shortcomings in this paper. For example, in the case of 2019-nCov, while the content information is more complex and the amount of information is larger, the performance of the algorithm proposed in this paper is worse than that of the other two events. Therefore, it is necessary to carry out deeper semantic mining of microblog text content; for example, the content and emotion of the text using phonetic transcription abbreviations need to be specially processed. In addition, designing a more suitable and effective integration strategy of classification algorithms for Weibo rumor recognition is one of the future works. Besides, detecting rumors with Weibo text information is worthy of further study.

Data Availability

The experiment was conducted with open dataset, which can be downloaded at <http://alt.qcri.org/wgao/data/rumdect.zip>.

Conflicts of Interest

The authors declare that there are no conflicts of interest regarding the publication of this paper.

Acknowledgments

This work was supported by the Research on Cyber Group Events Management and Pre-Warning Mechanism, National Social Science Fund of China (17XFX013).

References

- [1] W. P. Liu and S. M. Zhong, "Web malware spread modelling and optimal control strategies," *Science Reports*, vol. 7, p. 42, 2017.
- [2] Z. H. Xie, Y. Zhang, and L. Zhang, "The recognition of public opinion viruses of micro-blog based on logistic regression," *Microcomputer & Its Applications*, vol. 36, no. 16, pp. 67–69, 2017.
- [3] W. Liu, X. Wu, W. Yang, X. Zhu, and S. Zhong, "Modeling cyber rumor spreading over mobile social networks: a compartment approach," *Applied Mathematics and Computation*, vol. 343, pp. 214–229, 2019.
- [4] W. A. Peterson and N. P. Gist, "Rumor and public opinion," *American Journal of Sociology*, vol. 57, no. 2, pp. 159–167, 1951.
- [5] N. P. Chao and X. Huang, "Research on "rumors" phenomenon in Internet communication," *Information Studies: Theory & Application*, vol. 27, no. 6, pp. 586–589, 2004.
- [6] G. He, X. Q. Lv, Z. Li, and L. P. Xu, "Automatic rumor identification on Microblog," *Library and Information Service*, vol. 57, no. 23, pp. 114–120, 2013.
- [7] C. Castillo, M. Mendoza, and B. Poblete, "Information credibility on twitter," in *Proceedings of the 20th International Conference on World Wide Web*, pp. 675–684, New York, NY, USA, 2011.
- [8] V. Qazvinian, E. Rosengren, D. R. Radev, and Q. Z. Mei, "Rumor has it: identifying misinformation in Microblogs," in *Proceedings of the 2011 Conference on Empirical Methods in Natural Language Processing*, pp. 1589–1599, New York, NY, USA, 2011.
- [9] S. Kwon, M. Cha, and K. Jung, "Rumor detection over varying time windows," *PLoS One*, vol. 12, no. 1, pp. 1–19, 2017.
- [10] T. Chen, X. Li, H. Yin, and J. Zhang, "Call attention to rumors: deep attention based recurrent neural networks for early rumor detection," *Lecture Notes in Computer Science*, vol. 41, pp. 40–52, 2018.
- [11] D. G. Duan, C. S. Wang, Z. M. Han, and B. Li, "A rumor detection model based on Weibo's reviews," *Computer Simulation*, vol. 33, no. 1, pp. 386–390, 2016.
- [12] Y. F. Chen, X. Y. Zhou, and L. Zhang, "Research on the methods to identify online food safety rumor-related documents based on semantic co-occurrence matching," *Information Studies: Theory & Application*, vol. 41, no. 6, pp. 130–136, 2018.
- [13] T. Q. Lu, B. Shi, and Z. M. Yan, "Semi-supervised learning algorithm applied to Microblog rumors detection," *Application Research of Computers*, vol. 33, no. 3, pp. 744–748, 2016.
- [14] W. Dong and X. H. Yue, "The weighted multiple meta-models stacking method for regression problem," *Proceedings of the 38th Chinese Control Conference*, vol. 33, pp. 7511–7516, 2019.
- [15] Q. Xie, Z. T. Liu, and X. W. Ding, "Electroencephalogram emotion recognition based on a stacking classification model," *Proceeding of the 37th Chinese Control Conference*, vol. 41, pp. 5544–5548, 2018.
- [16] J. D. Duan, S. R. Liu, K. Ma, and R. Y. Sun, "Text sentiment classification method based on ensemble learning," *Journal of Jinan University (Natural Science Edition)*, vol. 33, no. 6, pp. 43–488, 2019.
- [17] K. P. Kota and G. V. S. R. Anjaneyulu, "A comparative analysis of support vector machines & logistic regression for propensity based response modeling," *Social Science Electronic Publishing*, vol. 4, no. 3, pp. 7–16, 2015.

- [18] C. Su, S. Ju, Y. Liu, and Z. Yu, "Improving random forest and rotation forest for highly imbalanced datasets," *Intelligent Data Analysis*, vol. 19, no. 6, pp. 1409–1432, 2015.
- [19] H. Li, *Statistical Learning Methods*, Tsinghua University Press, Beijing, China, 2016.
- [20] F. Yang, X. H. Yu, Y. Liu, and M. Yang, "Automatic detection of rumor on sina Weibo," *Proceedings of the ACM SIGKDD Workshop on Mining Data Semantics*, vol. 19, pp. 13–20, 2012.
- [21] L. H. Xu, H. F. Lin, Y. Pan, and H. Ren, "Constructing the affective lexicon ontology," *Journal of the China Society for Scientific and Technical Information*, vol. 27, no. 2, pp. 180–185, 2008.
- [22] Z. H. Zhou, *Machine Learning*, Tsinghua University Press, Beijing, China, 2016.
- [23] J. Ma, W. Gao, P. Mitra, and S. Kwon, "Detecting rumors from Microblogs with recurrent neural networks," *Proceedings of the 25th International Joint Conference on Artificial Intelligence*, vol. 27, pp. 3818–3824, 2016.