Journal of New Media
DOI:10.32604/jnm.2021.019649
Article

Tech Science Press

Social Network Rumor Recognition Based on Enhanced Naive Bayes

Lei Guo*

School of Computer & Software, Nanjing University of Information Science and Technology, Nanjing, 210044, China
*Corresponding Author: Lei Guo. Email: guoleigray@hotmail.com

Received: 22 April 2021; Accepted: 29 June 2021

Abstract: In recent years, with the increasing popularity of social networks, rumors have become more common. At present, the solution to rumors in social networks is mainly through media censorship and manual reporting, but this method requires a lot of manpower and material resources, and the cost is relatively high. Therefore, research on the characteristics of rumors and automatic identification and classification of network message text is of great significance. This paper uses the Naive Bayes algorithm combined with Laplacian smoothing to identify rumors in social network texts. The first is to segment the text and remove the stop words after the word segmentation is completed. Because of the data-sensitive nature of Naive Bayes, this paper performs text preprocessing on the input data. Then a naive Bayes classifier is constructed, and the Laplacian smoothing method is introduced to solve the problem of using the naive Bayes model to estimate the zero probability in rumor recognition. Finally, experiments show that the Naive Bayes algorithm combined with Laplace smoothing can effectively improve the accuracy of rumor recognition.

Keywords: Rumor recognition; social network; machine learning; naive bayes; laplacian smoothing

1 Introduction

In recent years, various social network emerging technologies have developed extremely hotly, and social network media has performed prominently. Social networks enable people to obtain various news and information at a very low cost, and at the same time, they can instantly forward and comment on their views and opinions on the event. The information dissemination of social networks is simple and convenient, which is the main reason why social networks can spread rapidly [1], but it is also an important reason for the spread of rumors. Usually after a rumor-like social network is released, it can produce fission-level forwarding and dissemination effects, which can be spread on a large scale in a relatively short period of time, thereby causing greater social impact. Nowadays, a large number of rumors in the online network society have adversely affected users' social cognition, emotional orientation and daily life. It even reduces the effect of government affairs communication to some extent, reduces the government's public power, and causes great harm [2].

At present, the mainstream solution to identify social network rumors is to use manual review to identify the authenticity of various messages. However, the manual review is not only human energy and time, but also the economic cost of the enterprise is also huge and human Subjective assumptions will be biased in the judgment of the authenticity of the information, and it is impossible to conduct official research on the emergent information in real time to judge the authenticity. Therefore, it cannot realize the efficient and accurate judgment of rumors.

In recent years, machine learning technology has developed rapidly. The technical results produced by related research on machine learning can replace some of the mechanical and repetitive operations that were previously time-consuming and labor-intensive to improve efficiency. By studying the text characteristics, emotional expression and other factors of the rumors themselves, we find a practical method to use relevant algorithms in machine learning to analyze various text information spread on the network to identify rumors, thereby improving the recognition Efficiency and accuracy, so this technology is of great significance to cyberspace security.

From the early 1920s to the early 1960s, that is, during the Second World War, social psychologists carried out research on rumors; in the mid-1960s to the late 1980s, the society triggered by the intensification of social problems The decline in overall trust has promoted the upsurge of research credibility; in the 1990s, with the rapid development of the social network, online channels have become the main source of information for people, and researchers have paid more and more attention to the availability of the social network. Reliability research: From the beginning of the 21st century to the present, the development of Web2.0 technology has led to the emergence of a large number of social media platforms. The high interactivity of social media has also made social media platforms a breeding ground for rumors.

Rumor detection belongs to the category of social network information credibility research and is a new direction of social network information credibility research [3]. At present, most of the detections for rumors are carried out on larger platforms, such as Weibo and Twitter. Both are micro-blog social platforms and occupy a large market share in related fields. Therefore, research on rumors recognition technology on such platforms is pertinent and universal.

At present, the research topics in the field of social network rumor recognition involve the differentiation of users on social platforms, the identification of Weibo navy forces, the recognition of rumors based on the content of Weibo comments, the research on the evolution of public opinion in emergencies, and the classification based on machine learning algorithms. Most of the research is based on the credibility evaluation of social network texts and the recognition and detection of rumors. Usually, different social network text features are selected to construct the classifier feature set, or realized by improving machine learning algorithms [4,5].

The main research contents of this paper are as follows:

(1) Introduced the recognition methods of social network rumor text, including the current mainstream social network rumor recognition technology and related machine learning algorithms, such as Bayesian decision theory, naive Bayes classifier and the principle of Laplacian smoothing.

(2) The algorithm of this paper is verified through experiments. The first is to segment the text, remove the stop words after the word segmentation is processed, then build a naive Bayes classifier, and introduce Laplacian smoothing into the classifier, and finally train and test the model.

2 Related Work

2.1 Definition and Characteristics of Rumors

Rumors are a common phenomenon of public opinion [6], and different researchers have given different explanations about the meaning of rumors: in the dictionary, rumors are defined as "news without factual basis" [7]. The definition of a rumor is "an unclear source, unfounded transmission, unconfirmed content, and lack of factual information." This type of definition emphasizes the distinctive feature of "unfounded" rumors. Therefore, rumors and The biggest difference between other common social network texts is that the rumors are false. And Shou et al. [8] emphasized the characteristics of rumors: "Rumors lacking factual basis have the purpose of persuading others to believe certain information, and they often adopt exaggerated language styles and have strong emotional characteristics". Such texts are usually Contains a strong emotional color, many words express strong desires such as "urgent", "seeking", "important" and other emotions. Because the purpose of rumors is to make more people believe in the content of the rumors, and the exaggerated language style caters to the public's psychology and can capture the public's attention. Many users do not have the habit of verifying information, so when they see such information, they will not verify their authenticity and directly perform the following operations, such as comments and reposts, which also aggravates the further spread of rumors.

As a result of the development of the social network, online rumors are quite different from the common rumors previously recognized by the public. In the past, the spread of rumors was regional, transient, and limited in audiences. However, online rumors broke through these restrictions. First of all, there is no regional restriction. The development of the social network has made the flow of information convenient and fast, but at the same time it has also made the spread of rumors more harmful. In the past, the spread of rumors mainly relied on people's word of mouth, but now a rumor can be spread thousands of miles away by just moving your fingers. Then the spread of online rumors has broken the time limit. A sudden and widespread online rumor can be a reprint or modification of a certain online rumor many years ago. The existence of the social network prevents these rumors from disappearing with the public's memory. But it always exists on the social network, which gives people with the intention to take advantage of it. At the same time, because the past rumors were mainly spread by people through conversations, the amount of spread was limited. On social platforms, the spread of a hot information can get hundreds of millions of spreads in a very short period of time, resulting in the public The panic, such as the "salt grab" incident caused by the Japanese nuclear radiation incident in 2011.

As a social phenomenon, most of the themes of rumors are closely related to the public, and the content of the spread is often various types of information that the public pays more attention to. Zhang et al. [2] summarized the main themes of online rumors, including health, current affairs and society. Corresponding to its classification, this type of social network rumors have their special features in terms of topic characteristics, language styles, and "source" characteristics. Therefore, as a product of a new era, social network rumors have many differences compared with ordinary rumors in the past, which requires some targeted research and processing according to their own characteristics.

2.2 Characteristics of Social Network Text

Social network text has obvious characteristics. Take Weibo as an example. The main body of a Weibo is usually a short text of about 100 characters. Some Weibo will add other elements to enrich the main content while containing text content. Other materials, such as pictures and videos, are expensive to produce and have a relatively low weight in the spread of rumors. Therefore, the text becomes the main body of the spread of rumors. In addition to the characteristics of the text itself, social network text has some other characteristics: (1) It has high repetitiveness. Several current microblogs, after being forwarded and spread several times, can reach the majority of the content of all microblogs at the same time period in a short period of time. (2) It has high immediacy. For example, the hot search lists on Weibo reflect the most popular events of the moment, and they are changing almost every moment. (3) It has unique sparsity. Compared with long paragraphs and chapters, social network texts have fewer words, so it is more difficult to extract key information from their texts. (4) It has strong relevance. For example, when reposting and commenting on each Weibo, the user will unconsciously substitute some subjective factors such as the situation at the time and personal experience in it, which is also based on this correlation when interacting with others.

As a representative of social network text, Weibo also has a strong emotional tendency, which is also the biggest difference between Weibo and other social network texts. Weibo like rumors will use more emotional text to modify the text while transmitting information. Weibo content with strong emotions will more easily attract public attention to achieve a certain purpose. The key factor is the need for more people to receive such microblogs.

2.3 Related Principles of Social Network Rumor Recognition

At present, the mainstream methods of rumor recognition are based on emotional dictionaries and machine learning methods [8].

Rumor recognition technology based on the emotional dictionary method is mainly based on sentiment analysis to identify rumors, and through the emotional conflicts of words in the texts of different quality information sources, so as to identify social network rumors in a specific field. The core of this method is to construct a specific emotional dictionary. Because social network rumors have certain abnormal emotional characteristics that are different from ordinary social network texts, the use of sentiment analysis technology can be used to identify rumors.

The rumor recognition technology based on machine learning mainly adopts the classification method. It extracts word features from the text content, then selects a certain classification algorithm to train the classifier, and then uses the trained classifier to determine whether the information is a rumor [3]. Currently commonly used machine learning classification algorithms include naive Bayes, decision trees, random forests, logistic regression, support vector machines, adaptive enhancements, and K nearest neighbor algorithms. The core of this is to build a good classifier, and the effect of the classifier mainly depends on the choice of training set and features. Choosing different features will directly affect the recognition effect of rumors.

3 Method

3.1 Social Network Text Preprocessing

The first step before analyzing social network text is usually word segmentation. Because all social network texts are composed of words, and words form words, the essence of social network texts is actually a combination of words. By segmenting the sentence, the key information in the sentence can be better extracted and the features can be obtained. Therefore, word segmentation is generally the first step in text processing.

The current mainstream word segmentation methods are mainly based on three aspects: understanding, matching and statistics [9].

At the same time, there are more obvious differences between Chinese and English word segmentation. The composition of English sentences is mainly composed of words and non-letters, so it can be segmented by spaces, punctuation and other non-letters and non-digits. Chinese sentences do not have a regular sentence structure for simple and effective segmentation. Therefore, a third-party word segmentation component: jieba can be used to segment Chinese text.

After segmenting the text, you will find that the word collection formed by the training set segmentation results contains a lot of punctuation marks, and these punctuation marks cannot be used as classification features for rumor recognition. Therefore, the second step in the preprocessing of social network text classification is to remove the stop words. In addition, the number is removed and the number is not used as a classification feature. At the same time, delete some specific words, such as prepositions, conjunctions, pronouns, and auxiliary words that have no effect on the recognition of rumors. Removing these meaningless stop words can better extract the feature information of the text.

3.2 Naive Bayes Algorithm Combined with Laplace Smoothing

3.2.1 Bayesian Decision Theory

Bayesian Decision Theory is the basic method of decision-making under the framework of probability [10], that is, when there are multiple probability values, the decision with the highest probability is selected. The core idea is to assume that under ideal conditions, all relevant probabilities of the classification task are known in advance. By comparing the probabilities of the samples in different categories, the optimal category to which the sample belongs can be judged.

Assuming that there are N classification possibilities, λ_{ij} is the error caused by incorrectly classifying the sample c_j as c_i. Therefore, the expected error caused by classifying sample x as c_i based on the posterior probability $P(c_i|x)$ is the "conditional risk" on sample x:

$$R(c_i|x) = \sum_{j=1}^{N} \lambda_{ij} P\left(c_j \middle| x\right) \tag{1}$$

Then, in order to minimize the overall risk $R(h) = E_x[R(h(x)|x)]$, the Bayesian criterion $h: \mathcal{X} \mapsto$

$\mathcal{Y}$ is introduced, and the conditional risk minimum is selected for each sample $R(c|x)$, which minimizes $R(h)$, thereby minimizing the risk $R(h(x)|x)$, that is

$$h^*(x) = \underset{c \in \mathcal{Y}}{\operatorname{argmin}} R(c|x) \tag{2}$$

At this time, h^* is called Bayes optimal classifier, and the corresponding overall risk R(h^*) is called Bayes risk. At this time, the theoretical upper limit of the accuracy of the model established by the naive Bayes algorithm is $1 - R(h^*)$, which is the best performance that the classifier can achieve. If it is to minimize the classification error rate, then λ_ij can be written as

$$\lambda_{ij} = \begin{cases} 0, & if\ i = j\,; \\ 1, & otherwise. \end{cases} \tag{3}$$

The conditional risk at this time is

$$R(c|x) = 1 - P(c|x) \tag{4}$$

Therefore, the Bayesian optimal classifier changes from minimizing the classification error rate to selecting the classification that maximizes the posterior probability $P(c|x)$ for each sample x

$$h^*(x) = \underset{c \in \mathcal{Y}}{\operatorname{argmax}} R(c|x) \tag{5}$$

However, it is difficult to directly obtain the posterior probability $P(c|x)$ in actual tasks. Therefore, the role of machine learning is actually to estimate the posterior pro bability $P(c|x)$. Among them, one is to first give the sample x, and then predict c by directly modeling $P(c|x)$, so that the "discriminative model" (Discriminative Models) is obtained; so that the joint probability distribution $P(x, c)$ is modeled, and then $P(c|x)$ is obtained from this, which is a "generative model". For generative models, one must consider

$$P(c|x) = \frac{P(x,c)}{P(x)} \tag{6}$$

Based on Bayes' theorem, $P(c|x)$ can be written as

$$P(c|x) = \frac{P(c)P(X|C)}{P(x)} \tag{7}$$

$P(c)$ is called the prior probability, which represents the probability of event c occurring before event x occurs. $\frac{P(X|C)}{P(x)}$ is an adjustment factor, which is called the possibility function. Its role is to make the estimated probability closer to the true probability.

According to the law of large numbers, when the training set contains sufficient independent and identically distributed samples, $P(c)$ can be estimated by the frequency of various types of samples.

3.2.2 Naive Bayes Classifier

However, since the probability function $\frac{P(X|C)}{P(x)}$ involves the joint probability of all attributes of x, it is difficult to estimate directly from a limited training sample. This is based on the Bayesian formula (7) to estimate the main difficulty of the posterior probability $P(c|x)$. To avoid this problem, the naive Bayes classifier assumes that each probability distribution is conditionally independent, and that each attribute independently affects the classification. Based on this assumption, let n be the number of attributes, x_i is the value of x in the $i - th$ attribute, and formula (7) can be rewritten as

$$P(c|x) = \frac{P(c)P(X|C)}{P(x)} = \frac{P(c)}{P(x)} \prod_{i=1}^{n} P(x_i|c) \tag{8}$$

Because $P(x)$ is the same for each category, the expression of the naive Bayes classifier can be obtained based on Eq. (5):

$$h_{nb}(x) = \underset{c \in \mathcal{Y}}{\operatorname{argmax}} P(c) \prod_{i=1}^{d} P(x_i|c) \tag{9}$$

Therefore, the training process of the naive Bayes classifier is to estimate the class prior probability $P(c)$ based on the training set D, and estimate the conditional probability $P(x_i|c)$ for each attribute.

Assuming that there are enough independent and identically distributed samples, and assuming that D_c represents a set of samples of type c in the training set D, the prior probability is

$$P(c) = \frac{|D_c|}{|D|} \tag{10}$$

If it is a discrete attribute, assuming that D_{c,x_i} represents a set of samples in D_c, and its value is x_i on the $i-th$ attribute, then the conditional probability $P(x_i|c)$ is

$$P(x_i|c) = \frac{|D_{c,x_i}|}{|D|} \tag{11}$$

If it is a continuous attribute, the probability density function is used. Suppose $\mathrm{p}(x_i|c) \sim \mathcal{N}(\mu_{c,i}, {\sigma^2}_{c,i})$, then

$$p(x_i|c) = \frac{1}{\sqrt{2\pi}\sigma_{c,i}} \exp\left(-\frac{(x_i - \mu_{c,i})^2}{2\sigma_{c,i}^2}\right) \tag{12}$$

The Naive Bayes classifier pseudo algorithm is as follows:

1) Initialize the rumors datasets R;

2) For $i = 1$ to n do:

2.1) Randomly select a sample x_i from R;

2.2) Calculate the maximum value $\underset{c \in \mathcal{Y}}{\operatorname{argmax}} P(c) \prod_{i=1}^{d} P(x_i|c)$ and assign it to p_i;

2.3) $P_{jk} \leftarrow P(j|Y = C_k)x_{ji} + \left(1 - P(j|Y = C_k)\left(1 - x_{ji}\right)\right.$.

3) End.

3.2.3 Laplacian Smoothing

However, when performing probability estimation, if a certain attribute value does not appear at the same time as a certain class in the training set, no matter what the other attributes of this sample are, the probability value calculated by Eq. (10) will be zero, which is obviously not reasonable.

In order to avoid certain attributes of the samples in the training set from being "erased" by some non-existent attribute values, "smoothing" is usually performed when estimating the probability value, and "Laplace smoothing" is usually used.

Assuming that N represents the number of possible categories in the training set D, and N_i represents the number of possible values of the $i-th$ attribute, then formulas (10) and (11) are corrected as

$$\hat{P}(c) = \frac{|D_c| + 1}{|D| + N} \tag{13}$$

$$\hat{P}(x_i|c) = \frac{|D_{c,x_i}| + 1}{|D_c| + N_i} \tag{14}$$

The use of Laplace correction avoids the problem of zero-probability estimation due to insufficient training set samples, and as the training set becomes larger, the overriding influence introduced by the calibration process will become more and more insignificant. The valuation gradually tends to the actual probability value.

In practical problems, there are many ways to use the naive Bayes classifier. For example, you can perform a "look-up table" to predict and discriminate: When the task requires a high prediction speed, then for a given training set, you can store in advance all the probability estimates involved in the naive Bayes classifier. Value; the "lazy learning" method is suitable for the need to frequently replace task data: When a prediction request is received, the probability estimation is based on the current data set without pre-training; the incremental learning method is suitable for: The data is constantly increasing, based on the existing Estimation can only calculate and correct the probability estimate related to the attribute value of the newly added sample.

4 Experiment

The environment configuration of the test and experiment in this article is as follows:

Table 1: The environment configuration

Operating system	Win10
RAM	16 GB
CPU	2.5 GHz Intel Core i7
Development tool	Pycharm

4.1 Datasets

The main research object of this experiment is the social network text on the Weibo platform, and the experimental data is divided into two categories: rumors and non-rumors. The rumor data comes from the Chinese rumor data provided by the Weibo false information reporting platform. There are 27,151 rumors in total. The non-rumor data comes from Datatang, which is a Chinese Weibo content data randomly crawled from the Weibo platform, containing a total of 5,812 items.

4.2 Results

4.2.1 Analysis of the Impact of Text Data Preprocessing

Use jieba Chinese word segmentation component for word segmentation and class labeling for rumor and non-rumor text data. After the data is word-segmented, the next step is to obtain a vocabulary for the features of rumor recognition. For each recognition, 80% of the data is randomly selected as the training set data, and the remaining 20% is used as the test set data. Through word frequency statistics, all the words obtained by word segmentation in the training set are sorted in descending order. Finally, the returned list stores the vocabulary after word segmentation and sorted in descending order of word frequency.

The number of high-frequency words is determined by observing the relationship between the number of removed high-frequency words and the final accuracy rate. Remove 50, 100, 150... up to 1000 high-frequency words in turn, and draw an image of the relationship between the two. The final result is shown in the figure.

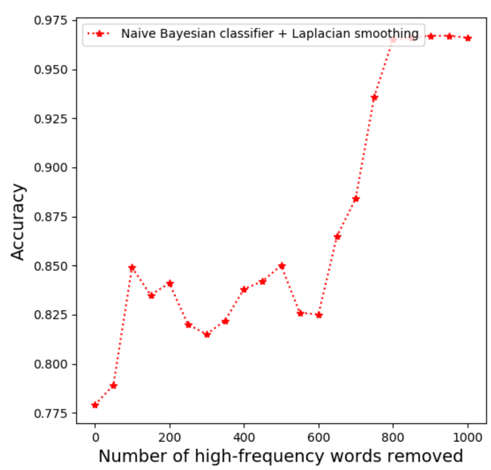

Figure 1: The impact of data preprocessing on accuracy

Through multiple tests, it can be seen from Fig. 1 that after the number of removed high-frequency words increases, the recognition accuracy rate increases to a certain extent. When the number of removed high-frequency words is about 800, the recognition accuracy starts to reach 90% or more, when the number of removed high-frequency words increases again, the accuracy rate does not change much, so the number of high-frequency words decided to be removed is 800. After determining the number of removed high-frequency words, a classifier for rumor recognition can be constructed.

4.2.2 Analysis of Rumor Recognition Effect

Rumors are classified into typical binary discrete values, and the final eigenvalues are only non-rumors and rumors. Through 5 experiments, observe the results of each experiment, the statistical results are shown in Tab. 2.

Table 2: Rumor recognition accuracy

	Group 1	Group 2	Group 3	Group 4	Group 5
Accuracy	0.964	0.965	0.962	0.965	0.961
Time Cost (s)	0.636	0.761	0.764	0.650	0.662

It can be seen from the experimental results in Tab. 2 that the final classification of the classifier is used to test the test set data, and the final recognition accuracy can reach 0.965. The final recognition accuracy rate is basically maintained at more than 90%, and the recognition result returned by the prediction is basically the same as the category corresponding to the test set data. Therefore, it can be seen that the naive Bayes algorithm can be used to construct a classifier with good performance to classify social network rumor texts, so as to achieve a certain effect of social network rumor recognition. Therefore, the research content of this article has certain practical value.

The main research object is the classification effect of the Naive Bayes algorithm on social network text. Through the experimental results of Tab. 2, we can see that the Naive Bayes algorithm is simple, fast, suitable for incremental training, thereby reducing time cost Increasing efficiency can better solve multi-classification problems. But it also has shortcomings. For example, it is very sensitive to the form of the input data and needs to pre-process the data in advance. The accuracy of Naive Bayes depends on the training data, and the data distribution needs to be independent of each other. The quality of the training data will directly affect the effect of the recognition results. In some cases, the calculation result of Naive Bayes may be 0, which makes it impossible to predict the result, but it can be corrected by Laplace smoothing, as shown in the Tab. 3.

Table 3: Compare results

	Group 1	Group 2	Group 3	Group 4	Group 5
Naive Bayesian classifier	0. 825	0.855	0.753	0.839	0.851
Naive Bayesian classifier+ Laplacian smoothing	0. 964	0. 965	0. 962	0. 965	0. 961

The comparison results show that by adding Laplace smoothing to the Naive Bayesian, the problem of zero probability estimation due to insufficient samples in the training set can be corrected, thus avoiding the “erasure” of some attributes of the samples in the training set by some non-existent attribute values, which improves the accuracy from the original Naive Bayesian classifier compared to the the accuracy of the final recognition.

4 Conclusion

This article is an exploration of the use of machine learning algorithms in natural language processing, and studies the application of text semantic recognition. This paper uses the Naive Bayes algorithm combined with Laplacian smoothing to identify social network rumor texts. First, the text is segmented. After the word segmentation is completed, the stop words are removed, and then the Naive Bayes are constructed. The classifier is finally used to train the data and test the data through the classifier. However, this article still has many shortcomings, mainly because it is too singular in the application of training algorithms. If you add more calculations for word weights, such as combining emotional dictionaries to highlight the weights of certain types of words with strong emotional colors in the text, and can apply them to machine learning algorithms, you can further improve rumor recognition The accuracy rate. These are expected to be further optimized and improved in future work.

Acknowledgement: This work was supported in part by NSFC (1736105).

Funding Statement: The author received no specific funding for this study.

Conflicts of Interest: The author declares that they have no conflicts of interest to report regarding the present study.

References

[1] Y. Han and J. W. Lu, "Summary of Weibo rumor spreading model," *Network Security Technology & Application*, vol. 11, pp. 41–43, 2018.

[2] Z. A. Zhang, K. R. Shu and L. N. He, "The theme and characteristics of WeChat rumors," *News and Writing*, vol. 1, pp. 60–64, 2016.

[3] S. Y. Jiang, D. Y. Chen and G. S. Pang, "A Summary of the research on the analysis of the reliability of Weibo information," *Library and Information Service*, vol. 12, pp. 136–142, 2013.

[4] A. Gupta and P. Kumaraguru, "Credibility ranking of tweets during high impact events," *ACM*, pp. 2–8, 2012.

[5] M. Gupta, P. Zhao and J. Han, "Evaluating Event Credibility on Twitter," *Society for Industrial and Applied Mathematics*, pp. 153–164, 2012.

[6] X. A. Guo, "*A Social Psychological Study of Internet Rumors in Contemporary China*," China Social Sciences Press, 2015.

[7] L. X. Sha, "*Social Psychology*," China Renmin University Press, 2006.

[8] H. R. Shou, S. Q. Deng and J. Xu, "Internet rumor recognition method based on sentiment analysis," *Data Analysis and Knowledge Discovery*, vol. 7, pp. 44–51, 2017.

[9] S. Zhou, "Social network text feature extraction," Nanjing University of Information Science and Technology, 2018.

[10] Z. H. Zhou, "*Machine Learning*," Tsinghua University Press, 2016.