Campus Network Public Opinion Monitoring System Based on Reptile Technology

Yingjie Tan, Qiang Lin*, Yuxin Luan, Tao Chen, Yifan Qiao, Yue Luan

School of mathematics and computer science, Northwest Minzu University, Lanzhou, Gansu, 730030

*Corresponding author e-mail: 1345817359@qq.com

Abstract. This paper conducts a lot of research on campus network public opinion analysis, and uses the hybrid crawling strategy commonly used in academic circles to focus on the key technologies of network public opinion analysis such as information preprocessing, Chinese word segmentation and topic recognition. Finally, the analytic hierarchy process and wavelet neural network are used to design the public opinion analysis system, and all the modules in the system are organically combined to develop an efficient and feasible network public opinion monitoring system, and the Sade event is taken as an example to express the public opinion. Analysis and trend prediction, the experimental results show that the system has good evaluation performance and estimation accuracy.

1. Introduction

The popularization and application of computer networks provide an unprecedented convenient way for the widespread dissemination of information. Internet public opinion is a hot spot and focuses issue in the real life that the public is concerned about on the Internet. It is the sum of different thoughts and attitudes expressed and disseminated by Internet users through the Internet. Internet sensation comes from real life, but it is a virtual reality different from real society. It extends the channels of people's communication and expression of sensation to the Internet. Compared with the traditional media, the development of a college campus network is likely to have a negative impact on the mental health of college students. Therefore, the management of colleges and universities should strengthen the mental health education of college students in a timely manner, explore the correct network public opinion guidance mechanism, effectively control the negative effects of the network, and control the development and spread of public opinion. How to timely resolve and realize the guidance and control of college network public opinion and maintain the harmony of college students' life is an urgent problem to be solved in building a harmonious socialist campus.

There are many mature public opinion monitoring systems at home and abroad. The internationally renowned monitoring systems include American TDT system, Pola emotional sentiment analysis software, Newsblaster system. [1] Domestic network Prison and jail according to system development status: China lyrics network palas network public opinion monitoring system, Mai information H-party network public opinion monitoring system, Goonie network public opinion monitoring system, eagle network sensation monitoring system, military dog network public opinion monitoring system. However, the more mature network monitoring system in China is mainly used for monitoring and
analyzing the public opinion information of various public events on the network by various government departments. For example, the military dog network public opinion monitoring system realizes network public opinion data collection, public opinion data analysis, public opinion response, network public opinion warning and public opinion analysis report generation, etc. through the automatic acquisition of public opinion information, public opinion classification, topic identification, and topic focusing. 16 main functions, mainly used in relevant security departments such as the military and public security. [2]

There are not many public opinion monitoring systems that are widely used in colleges and universities. Most of them are limited to a university to develop and use by themselves. However, in recent years, there have been many types of research on the monitoring of network public opinion in colleges and universities. Education departments and colleges have paid more and more attention to the network public opinion monitoring system. Construction. Chen Yan, from the Shaanxi University of Technology, discovered that the research on college public opinion began in 2006, and about 2,000 research papers related to college public opinion in the past decade. [3] Beijing Normal University Lu Xueliang proposed the combination of civil air defense and technical defense to construct a monitoring system through the characteristics of the network public opinion and public opinion monitoring system. [4] Xu Qi of Shaanxi Normal University took Shaanxi Normal University as an example to study the characteristics of school public opinion, and proposed the characteristics of network public opinion for his school, and proposed a public opinion monitoring strategy for his school. [5]

1 System Overall Architecture

The network public opinion monitoring system designed in this paper is mainly aimed at a large number of netizens directly involved in a specific field. The system is essentially a public opinion collection and analysis system. The system pays attention to the unified coordination of the acquisition module, processing module, and analysis module, and realizes the functions of public inquiry query, inclination analysis, and event trend analysis and corpus collection of the whole system. Among them, the main objects of corpus collection are news websites and blogs with a large number of netizens' comments and other media with strong interactivity. Because of the lyric corpus collection of these media, not only can the latest information be obtained through meta-search, but also the trend analysis technology and event tracking technology can be used to discover the emotional development direction among the topics, and the display module will also be based on graphs and tables. The form will visually display the results of the public opinion analysis.

As shown in Figure 1, the network public opinion monitoring system mainly includes three major functional modules, namely the public opinion acquisition module, the information processing module, and the public opinion analysis module.

The information collection module is mainly responsible for collecting information on events in the field of interest. The module also includes two sub-modules: Internet polls and focused crawlers. The former collects a large amount of data from the network through questionnaires and then summarizes and summarizes the information required for the project. The latter filters topic-independent links based on certain web analytics algorithms, retaining useful links and placing them in a queue of URLs waiting to be crawled. Then, it will select the URL of the web page to be crawled from the queue according to a certain search strategy, and repeat the above process until it reaches a certain condition of the system. The information processing module is mainly responsible for cleaning and stipulation of data. The information processing module further includes three sub-modules, namely the webpage de-emphasis module, the webpage cleaning module, and the text segmentation module; the public opinion analysis module is mainly responsible for the analysis of the public opinion information, and includes three sub-modules, namely the word segmentation recognition module, the word segmentation analysis module and the public opinion. Trend analysis module.
2. Lyric Information Processing

2.1. Webpage weight loss
The webpage is divided into two parts: the homologous webpage deduplication and the content deduplication. The reduplication of the homologous webpage is implemented by hash hashing of the webpage. This paper sets a hash function to hash the webpage. After the column, if the values are the same, the illustrated string is the same, which means that after being read and downloaded, the information of the page can be discarded, and the more appropriate parameters can be set, but the statistics need to be generated later. A number of related articles. Through the deduplication and denoising processing of the webpage, clean structured public opinion information can be obtained, and finally, the lyric webpage is stored in the database.

2.2. Web cleaning technology
The Web is a huge repository of information, and its content usually contains information about the subject matter and content information that is not related to the topic. This irrelevant information is called "noise." The noise of a webpage refers to the advertisements, copyright information, and page navigation, annotations, and scripts contained in the webpage, as well as some advertisements, useless images, useless links, etc. that do not match the theme. The existence of these noises directly affects the value of the crawling. Information, which causes the subject to deviating, affects the correct identification of the topic, and webpage purification is to denoise the webpage to obtain valuable information such as the title and body of the webpage. After the webpage is denoised, the subject content of the webpage becomes a direct processing object, and the text processing program improves
the accuracy in the process of processing. At the same time, the denoising of the webpage can also simplify the complexity of the label on the webpage, and therefore, the webpage denoising becomes information. An important part of the pre-processing process is also the main source of data for the public opinion analysis module.

After the webpage is denoised, the structure of the page is analyzed. The webpage is written in a hypertext markup language. The user can publish a web document with text, table, image, and other resources, and can click a hypertext link to browse. A document is a semi-structured page that uses tags to separate the various components of the text. The Document Object Model, as an interface and platform, allows programs or scripts to be dynamically compiled while storing and updating file content, structure, and style.

2.3. Text word segmentation technology
At present, there are three typical types of word segmentation. This paper mainly uses the string matching word segmentation method. It uses a prepared dictionary to formulate reasonable scanning rules, and then matches the Chinese word segmentation string with the dictionary. The recognition is successful and the word is divided.

3. Public Opinion Analysis

3.1. Word segmentation module
The implementation of word segmentation adopts the clustering algorithm. The multi-center representation can reflect the dynamics of the lyric topic. The use of double or multiple keywords to give higher weight method can more accurately identify the topic. The design idea of this module is: Using multi-center, using the title vector and the text vector to perform the double-vector comparison, the principle of the comparison process is to use double or multiple keywords to give greater weight and cluster the word segmentation into a word tree.

3.2. Word segmentation module
For the word segmentation analysis module, we must first establish a set of public opinion analysis and evaluation index system, which makes use of the system. Expert assessment, survey statistics and other methods and establish relevant data sheets. Second, to establish an evaluation model,

The model is a weight calculation module based on the analytic hierarchy process, that is, by constructing a judgment matrix and verifying the consistency thereof, the method can determine whether the matrix is reasonable and if reasonable, establish a hierarchical mechanism model and simultaneously use the summation the method calculates the weights of the indicators of each layer. The module for quantifying indicators is divided into qualitative index quantification and quantitative index quantification. Qualitative indicator quantification method is to comprehensively consider the hot topic of the past, corresponding to its index value, using expert evaluation method to quantify it, and quantitative index quantification, in the process of quantification, assigning a value to the indicator in stages, to process the quantified value of this indicator. Finally, the evaluation calculation module is established, and all the above index systems are comprehensively calculated. The weighted average method is used to obtain the network public opinion evaluation level, and the results are stored in the database.

3.3. Lyric analysis
Write the above total data into the analysis database, use MATLAB to analyze the overall trend, establish the relevant mathematical model, use the scatter plot to show the direction of the public opinion clearly and intuitively, and get a K value of the correlation degree, then write into the analysis database.
4. System Experiment

4.1. Data source
The sensational hotspots studied in this paper are the two events of the National People's Congress and the South Korean deployment of the Sade event in 2017. The keywords of the two lyric events, “Two Sessions” and “Sade”, are used to obtain their respective Baidu Index, Micro Index, and WeChat Index. The lyric time period is selected as From February 25, 2017 to March 26, 2017, a total of 31 days of data, of which the first 30 days of data for training data, the last day of data for expected forecast data, a total of 25 sets of data, this article uses 20 sets of data for training, 4 groups of data were verified, and 1 group of data was predicted, that is, the experimental prediction target value was the date of March 26, 2017. Baidu Index, Micro Index, and WeChat Index can reflect the degree of concern and hot agenda of public opinion events and are authoritative and comprehensive. Research papers on public opinion research, the data source is relatively single, so this article reflects the overall trend through multi-dimensional lyric sequence prediction.

4.2. Data preprocessing
For time series data, the sequence preprocessing process is first performed, that is, whether the sequence data is a random sequence and whether the data has stationarity. A purely random sequence, that is, a randomly generated sequence, without any regularity, is a white noise sequence. It is a stationary sequence with no information to extract, no sequence analysis is necessary, and it is not suitable for modeling prediction. By plotting the public opinion trend graph, as shown in Figure 2, it can be observed that the time series shows a certain trend, not a purely random sequence.

![Figure. 2 Lyric trend chart](image)

The time series of the public opinion event is not a white noise sequence, and then the sequence is analyzed for whether it is a stationary sequence. This paper will collect the data in the MATLAB tool for the Augmented Dickey-Fuller test, the unit root test, which was proposed by Dickey and Fuller in 1979. By calculating the unit root of each index, the test results show that the sequence data of the two public events is a non-stationary sequence.

4.3. Experiment analysis
In this section, we use the WeChat index of the Sade lyric event to analyze the example. The other data sequences use the same method for model training. The specific parameters may change due to the training optimization process.
According to the built model, the MATLAB development platform is used to write scripts to realize the model prediction process. Initialize the relevant parameters of the model, use the brand function for random initialization, use the wavelet base as the activation function, the neural network learning rate is 0.01 and 0.001 respectively, and the maximum number of iterations is 1000, find the model in the process of tuning, hidden layer nerve The number of elements is 4, and good prediction performance is obtained. In this paper, the number of training samples used in the model is 20, the number of test samples is 4, and the predicted sample is 1, and the predicted trend graph shown in Figure 3 is obtained.

![Figure 3](image)

**Figure 3** The preliminary Sade event WeChat index forecast for this system

The predicted trend graph of the Sade event WeChat index shows that the law of the public opinion trend is difficult to grasp, and the prediction results of the time series of the model construction often lag behind the existing values. When the public opinion trend changes greatly, the generalization error is large, and the fitting is large. Not effectively.

The neural network is then trained to obtain a predicted trend graph, as shown in the figure. As can be seen from Fig., the prediction model obtains the optimal value in the 12th iteration. For the error existing in the prediction process, the temporal correlation of the prediction error can be described by the error autocorrelation function. For a perfect predictive model, there should be only one non-zero autocorrelation function and the non-zero autocorrelation function value only appears at zero hystereses. In the actual modeling process, when the value of the autocorrelation function except the zero lag is about 95% of the confidence interval around zero, the model is valid. After the model training, the Sade event has multiple non-zero-value autocorrelation functions, and the value at zero hystereses conforms to the model result, that is, the error autocorrelation function value of the trend term of the Sade event is within the confidence range, as shown in Figure 4. Explain that the model is effectively available.
Due to the existence of the trend error, the error of the trend term has a certain law, but within the confidence range, the error is within the acceptable range. Thus, the final training result of the trending item is obtained, as shown in fig. 4.

The first graph in Figure 5 shows the relationship between the test output and the target value. The curve fits the training test value, the blue "+" represents the result of the training output, and the green "+" represents the result of the verification output. The red "+" represents the target prediction value of the model. The modeling results show that the training effect of the trending item is good, indicating that there is a certain regularity in the trend term obtained after the EMD decomposition, which is conducive to predicting the model for prediction. The second picture is the error value of the model training, and the error of the target value is large, but within the error confidence range.

Figure. 4 Trend item error autocorrelation graph

Figure. 5 The final prediction of the Sade event WeChat index based on this system
The predicted values of the various IMF components of the Sade event WeChat index and a trend term are reconstructed, and the predicted value of the Sade lyric event is finally obtained, as shown in Fig. 5. The fitting effect of Fig. 5 approximates the trend of public opinion, and some fluctuations have certain errors, indicating that the system can be applied to the time series prediction of the Sade event.

5. Conclusion

Campus network public opinion monitoring research belongs to the field of academic research and application, so the research in this paper has application innovation. At the same time, the integrated application of focused reptile technology and natural language-based guidance ensures the base of public opinion monitoring. In addition, it has the real-time and innovative nature of related applications. Focused reptile technology is more optimized than general web crawler technology, straightforward and saves time. In a word, the analytic hierarchy process and wavelet neural network are used to design the public opinion analysis system, and all the modules in the system are organically combined to develop an efficient and feasible network public opinion monitoring system.

Acknowledges

Project 201810742062 supported by National Training Program of Innovation and Entrepreneurship for Undergraduates

References

[1] Sun Nan. Design and implementation of network content and public opinion monitoring system [D]. Nanjing University of Posts and Telecommunications, 2016.
[2] Zou Yu. Design and implementation of network public opinion monitoring and analysis system [D]. Jilin University, 2015.
[3] Chen Yan. A Review of the Research on Public Opinion and Public Opinion in Colleges and Universities in Recent Ten Years [J]. Journal of Shaanxi University of Technology (Social Science Edition), 2017, 35(4): 85-90.
[4] Lu Xueliang, Xie Yifang, Zhang Xu, et al. Research on the Construction of Network Public Opinion Monitoring System in Colleges and Universities [J]. China education information, 2016(16):86-88.
[5] Xu Qi. Research on Publicity Monitoring and Guidance of Campus Network in Colleges Taking Shaanxi Normal University as an Example [J]. China Education Information Technology, 2013(19): 36-39.
[6] Zeng Runxi. Research on the Working Mechanism of Network Public Opinion Management [J]. Library and Information Service, 2009, 2009, 53(18): 79-82