Research and Implementation of Network Public Opinion Prediction System

Xuegang Chen, Shouping Gao*, Sheng Duan and Luda Wang

College of software and communication engineering, XiangNan University, Chenzhou, China
Corresponding Author: Shou ping Gao

Abstract. The Internet has become a distribution center of ideological and cultural information and an amplifier of public opinion. To dig, analyze and study hot public opinions on the Internet is an important means to fully understand what netizens are thinking and doing. This paper analyzes the function of public opinion system, and introduces the key technology to realize the prediction of network public opinion, which can well realize the prediction function of public opinion.

1. Introduction

In recent years, our country economy high speed development, into the deep waters of economic reform and crucial period of reform, but along with the social contradictions are accumulated and overlay, with economic and social transformation, the new problems and new contradictions appeared constantly, some hot issues of social public opinion, and even conflict starting or spread through the social network, Internet users use the Internet to fully express their beliefs, attitudes, opinions and emotions, in a very short time to build consensus, fermentation emotion, induce action, influence society, in the process of public opinion can push the government, enterprises and individuals to the forefront, until a fatal crisis, It poses a severe challenge to social harmony and stability. In the eighteenth big report, clearly put forward "to strengthen the network of social management, promote the network specification in accordance with the order running", in a big data environment, and early warning research hotspot in network society public opinion mining is social innovation in order to strengthen the network management, through public opinion massive amounts of data mining, statistics and analysis, prediction and guidance, improve the ability of guidance and control, for the network social governance provides an innovative theory and practice, to build a harmonious and make important contributions to the economic sustainable development. Therefore, in the face of an increasing number of hot events and mass emergencies, it is an urgent problem to develop a sound early warning system for online public opinion mining and risk situational awareness, so as to effectively dynamically detect, prevent and evaluate online public opinions when they appear, and effectively resolve and control risks when they develop.

Many researchers have done a lot of research on the prediction of network public opinion, and put forward many effective prediction methods. These methods can be roughly divided into data statistical prediction method, maximum probability prediction method and artificial neural network prediction method. The prediction method based on data statistics model is to express the law of data development by a simple function, that is, the development of data has a linear law. Usually, the function will solve the coefficients by Taylor polynomial and extend to N-level variables until the law of function development basically coincides with the law of data development. Kalloubi, F[4] et al. acquired a large number of words from the lexicon, took a certain time as the time node of sampling,
counted the emotional words, negative words, trained these words of different time periods by linear regression function, and judged whether there were major emergencies by analyzing the critical point of function fluctuation; Cao, D. L.[5] and others analyzed the existing grey system prediction (GM) method uses grey system prediction model for data model training, crawls public explosions in forums and news websites, selects the number of events forwarded and the degree of attention of events as variables, and compares with the actual results to verify the credibility of the prediction model. Experiments show that the model is suitable for network public opinion prediction. Data statistical prediction method can be used to calculate specific values, but it is difficult to achieve the desired results in the case of data analysis with multiple features. Maximum probabilistic prediction method[6] can estimate data with multiple features, assume that there are multiple states in the data, it can describe the state of the data at the next moment by using probability method. The commonly used probabilistic models include N-Bayes and Logistic regression models. Li, J. et al.[7] used the probability analysis model to predict the development of Weibo public opinion, in the performance test of Matlab, he regarded Weibo public opinion as hot news and hot information. According to the evaluation of these data by netizens, he divided the network public opinion into many states, and took the number of relevant articles and the dissemination of data as parameters of the model, based on the probability analysis model, he got the analysis network; Itoh, M. et al.[8] studied the stability of a country, he regarded the stability of a country as a potential influencing factor from the perspectives of democratic system, economy, population distribution and political composition, took historical data of national development as reference, put bayesian model into simulating the influence model of national stability, and predicted the possibility of future recession of a country. The forecasting methods of data statistics model and maximum probability model are suitable if the parameters of corresponding models are known, if the sample data or models to be analyzed have strong large-scale distribution and irregularity, the forecasting method of artificial neural network model is needed[9]. This method simulates the thinking mode of human brain, transforms the complex problems to be solved into non-linear problems with input and output. The forecasting methods include neural network and SVM. Some researchers take the grey prediction model as the research object, take the network text data as the data source, data mining the network hot spots, and train the network hot spots with the grey prediction model according to the hot spots obtained; Itoh, M.[10] and others take the Weibo as the data source of public opinion analysis, divide the data into training set and test set, and the ratio of data measurement is 4:1. Using BP neural network as data modeling method, the layer parameters and hidden layer variables of the model can be obtained by genetic algorithm.

2. Function Scheme Design of Public Opinion System

According to demand analysis, the public opinion management function is designed. Public opinion is the focus of this system. How to display the collected public opinions and make full use of the public opinion information obtained, including event title, event occurrence time, number of public opinion participants, propensity and its ratio, etc. The functional modules of the public opinion system designed now shall include the following six, as shown in figure 1.

1) Public opinion collection function. This function module provides a way to collect public opinion data. Relevant data can be collected through this function module.

2) Public opinion analysis function. Public opinion analysis is needed after the collected data are saved in the database. It includes "similarity analysis" and "affective tendency analysis".

3) Public opinion early warning parameter maintenance function. This function module provides an interface for parameter value setting, through which the value of early warning can be modified. If the corresponding value is not reached, the public opinion will be filtered out automatically.

4) Public opinion display. For example, table displays can be used, and pie charts can also be used. These various visual display methods can more intuitively reflect the percentage of public opinions, which is conducive to the public opinion managers' convenience in viewing the public opinion dynamics.

5) There is a public opinion level in the public opinion warning. If the public opinion level is important, relevant staff should take measures in time.

The function scheme design of the public opinion system is shown in figure 1.

In summary, the function module is mainly divided into five parts: public opinion collection function, public opinion analysis function, public opinion early warning parameter maintenance function, public opinion display function, and public opinion level warning function. The function module is mainly divided into five parts: public opinion collection function, public opinion analysis function, public opinion early warning parameter maintenance function, public opinion display function, and public opinion level warning function.
6) Public opinion decision-making, its main function is to provide effective intelligent decision-making for public opinion managers according to different public opinion levels after big data analysis.

![Diagram](image_url)

**Figure 1.** System functional design

### 3. The Key Technology of System Realization

#### 3.1. The Theme Crawler

News and news postings, BBS, weibo, blogging and social networking and other network platform for Internet users to provide a wealth of communication platform, has brought together a large number of public opinion information, in the era of big data, how to quickly for massive network data acquisition is the first step of network public opinion analysis, which includes a crawl, data storage and data cleaning technology, related to public opinion crawl is important one annulus, mainly USES the web crawler (also known as web spiders, network robot), divided into general web crawler, limited web crawler and three types of focused web crawler, focused web crawler is also called the topic crawler, is a kind of geared to the needs of a particular topic, A web crawler that automatically crawls (downloads) web page information in accordance with certain rules. The goal of a traditional web crawler is to crawl web pages into local storage as quickly and comprehensively as possible. The usual practice is to extract all URL links from an existing web page and crawl new pages based on those new URL links. Compared with traditional web crawlers, focused crawlers are more targeted and do not aimlessly crawl all the URL links contained in a web page. The goal of crawler focus is to effectively crawl the web pages related to the topic with limited resources, so as to minimize the crawling of irrelevant pages so as not to waste resources. In this way, we can get the public opinion information we want according to the demand, which can greatly reduce the cost of data mining.
To obtain the web pages, first of all, the web theme relevance assessment. Then, on the one hand, after the evaluation of the web page, combined with its relevance, from the web content to extract knowledge to expand the subject knowledge. On the other hand, based on the topic relevance of the web page, the extracted URL is evaluated, and the URL and the evaluation result are stored in the candidate queue. In the next round of crawling, the URL with the highest degree of correlation is extracted from the candidate queue for crawling.

As shown in figure 2, crawler is implemented by applying queue structure. It takes the URL of the content to be grabbed as the "URL to be grabbed." When the crawler works, it extracts a URL from the "URL to be grabbed" according to the traversal calculation method of depth or width for web page download and other operations. At the same time, it puts the URL into the "URL to be grabbed" queue and lists the URL from the "URL to be grabbed". In this way, the cyclic operation can theoretically crawl all the contents of the "URL to be grabbed". Save the target URL format you need to crawl into the seeds array as a string, such as "http://news.sina.com.cn/" for all target websites in sina news.

There are many links on the target web page, and there are many useless links such as advertisements, image urls. The crawler adds the seeds array to determine which urls are their targets, i.e., "URL to grab." Take sina news network as an example, the code is as follows.

```java
private void initCrawlerWithSeeds(String[] seeds) {
    for (int i = 0; i < seeds.length; i++) {
        LinkQueue.addUnVisitedUrl(seeds[i]);
    }
}
```

```java
ILinkFilter filter = new ILinkFilter() {
    public boolean accept(String url) {
        if (url.startsWith("http://news.sina.com.cn/"))
            return true;
        else
            return false;
    }
};
initCrawlerWithSeeds(seeds);
```

Figure 2. Principle of theme crawler
When the queue of "URL to be grabbed" is not empty and the number of columns in the queue of "URL to be grabbed" exceeds a certain value, the crawler is terminated. When you crawl the target URL and find that there are "formatted" links in the target URL, you add them to the "URL to be crawled" queue. The key code is as follows.

```
while (!LinkQueue.unVisitedUrlsEmpty() && LinkQueue.getVisitedUrlCount() <= 100) {
    String visitUrl = (String) LinkQueue.unVisitedUrlOutQueue();
    LinkQueue.addVisitedUrl(visitUrl);
    for (String link : links) {
        LinkQueue.addUnsiVitedUrl(link);}
}
```

3.2. Fuzzy Neural Network

Fuzzy neural network (FNN) applies the fuzzy theory to the prediction analysis of neural network, which effectively improves the ability of neural network to deal with non-linear data, enhances the expansibility of the network, and further improves the accuracy of data analysis. Fuzzy neural network is good at dealing with non-linear functions and relations. According to the previous analysis, people's emotional changes contain multiple emotional values. Emotional values and user attributes will have an important impact on user's emotional fluctuations, that is, emotional fluctuations have the characteristics of a non-linear function, so FNN model is used in public opinion prediction model. The structure of FNN model consists of multiple fuzzification layers and fuzzy reasoning layers. The general structure of FNN model is shown in Figure 3.

![Figure 3. Structure of fuzzy neural network](image-url)
In Figure 3, we can see that the fuzzy neural network model is roughly divided into five layers. The first layer is the input layer, in which each node represents the input of the sample, \( X_i \) represents the first input sample, and the number of internal storage nodes \( n \) represents the size of the input sample. The second layer is fuzzified, and the input sample is fuzzified, and each input sample is fuzzified by using the Gauss membership function \( U_{ij} \), in this process, the Gauss membership function needs to set reasonable center parameters and width range; the third level needs to be fuzzy reasoning, because each node contains certain fuzzy rules, and multiple nodes represent multiple fuzzy rules. Therefore, it is necessary to calculate the fuzzy rule strength \( K_i \) of each node, where nodes contain multiple samples, and the fuzzy rule strength of nodes adopts membership function. The product sum is expressed; different intensity of fuzzy rules as a whole "1", need to be partially-totally proportional processing, that is, the proportion of different intensity of fuzzy rules is calculated, the number of instances of the intensity of fuzzy rules is the same as the number of nodes; the fifth level, output, the intensity of fuzzy rules and the corresponding weight coefficient \( P_j \) as the variable of calculation, output the number of nodes.

4. System Implementation

4.1. The Realization of Public Opinion Display

WEB customer interface: based on the cloud computing model, customers can use the WEB browser to log in the system at any time and have a comprehensive understanding of the public opinion on the customer interface. The customer interface includes such main contents as public opinion trend, detailed information of public opinion, latest microblog information, carrier coverage and so on. It is also presented in a visual way such as list and pie chart display to facilitate customers to view. The navigation bar is clear and clear, convenient for our center to view various categories; Real-time search function, you can customize the search conditions to view the latest monitoring information captured by the system.

4.2. Public Opinion Predicting

It is difficult to judge the level of online public opinion warning and predict the trend of public opinion. After many studies, the predecessors concluded that some judgment algorithms only said that "most of the early warning situations are consistent with the expert's judgment". Therefore, the value synthesized by relevant algorithms such as neural network is used as the public opinion prediction.

4.3. Public Opinion Decision

Its main function is to provide effective trend prediction and intelligent decision making for public opinion managers according to different public opinion levels after big data analysis. In order to adapt to the needs of adaptive growth of multiple decisions, a new decision can be created, or a row can be selected to modify the decision and delete the decision. For example, I now create a new decision "negligible" with the lowest value and the highest value of 0.1. Then it can be seen that the emotional tendency between 0 and 0.1 can be ignored.

5. Conclusions

The implementation and application of the system provides a systematic, integrated and macroscopic basic platform, forming an efficient and smooth online public opinion discovery, analysis, supervision, early warning, disposal and feedback mechanism.

6. Acknowledgments

The authors would like to thank for financial support by youth fund project of the humanities and social sciences of Education Ministry (No.15YJC870004), science and technology innovation team of XiangNan University (Recognition and analysis based on big data public opinion), Social Sciences fund project of Hunan Province (No.13YBA302) and education Science in Hunan province in 12th Five-Year planning project(XJK014CGD081, XJK011BXJ004).
7. References

[1] Cao, D., Ji, R., Lin, D., & Li, S. (2016). A cross-media public sentiment analysis system for microblog. *Multimedia Systems, 22*(4), 479-486.

[2] De Maio, C., Fenza, G., Loia, V., & Parente, M. (2015). Time aware knowledge extraction for microblog summarization on twitter. *Information Fusion, S156625351500055X*.

[3] Liu, M., Liu, S., Zhu, X., Liao, Q., & Pan, S. (2015). An uncertainty-aware approach for exploratory microblog retrieval. *IEEE Transactions on Visualization and Computer Graphics, 22*(1), 1-1.

[4] Kalloubi, F., Nfaooui, E. H., & El Beqqali, O. (2016). Microblog semantic context retrieval system based on linked open data and graph-based theory. *Expert Systems with Applications, S0957417416000300*.

[5] Cao, D. L., Ji, R. R., Lin, D., & Li, S. Z. (2014). Visual sentiment topic model based microblog image sentiment analysis. *Multimedia Tools and Applications*.

[6] Zhao, F., Zhu, Y., Jin, H., & Yang, L. T. (2016). A personalized hashtag recommendation approach using lda-based topic model in microblog environment. *Future Generation Computer Systems, 65*(C), 196-206.

[7] Li, J., Wen, J., Tai, Z., Zhang, R., & Yu, W. (2016). Bursty event detection from microblog: a distributed and incremental approach. *Concurrency & Computation Practice & Experience, 28*(11).

[8] Itoh, M., Yoshinaga, N., & Toyoda, M. (2016). Word-Clouds in the Sky: Multi-layer Spatio-Temporal Event Visualization from a Geo-Parsed Microblog Stream. *2016 20th International Conference Information Visualisation (IV) IEEE*.

[9] Zhao, Y., Qin, B., Liu, T., & Tang, D. (2016). Social sentiment sensor: a visualization system for topic detection and topic sentiment analysis on microblog. *Multimedia Tools and Applications, 75*(15), 8843-8860.

[10] Itoh, M., Yoshinaga, N., & Toyoda, M. (2016). Spatio-temporal Event Visualization from a Geo-parsed Microblog Stream. *Companion Publication of the International Conference on Intelligent User Interfaces. ACM*. 

