Research on Network Public opinion Analysis Platform Architecture Based on Big Data

Research on Network Public opinion Analysis Platform Architecture Based on Big Data

Fusheng Yuan, Jingxiong Yang, Que Zheng*

BeiJing Branch, State Grid Information & Telecommunication Group Beijing, China

*Corresponding author e-mail: zhengque001@sina.cn

Abstract. With the development of Internet technologies and applications, network public opinion has undergone tremendous changes in terms of data volume, complexity, and speed of production. Existing public opinion analysis techniques and monitoring systems are difficult to adapt to changes in this scenario, and must be innovative under the thinking of big data analysis. Therefore, this paper designs a network public opinion analysis platform through the use of big data technology, including information collection, information processing, information analysis, public opinion display and information security. And from the three aspects of government decision-making, hotspot tracking, and public opinion guidance, the application effect of the platform is proposed, and the innovation of thinking, mode and technology of network public opinion management is realized. It has important theoretical and practical value for network public opinion management.

1. Introduction

With the rapid development of society, the rapid spread of information technology has greatly increased the number of people making speeches on the Internet. Events such as cyber violence have occurred from time to time. The impact of online public opinion on the society is getting higher and higher, and the management of online public opinion is becoming more and more difficult [1-3]. How to implement effective management and control of the network of public opinion, public opinion analysis by accurately grasp the evolution of its spread and reduce the negative impact on public opinion, is the Internet age urgent problem. In the 21st century, with the exponential growth of network data volume, the concept of big data came into being [4, 5]. In the context of big data, how to use big data technology to realize the innovation of thinking, mode and technology in network public opinion management, improve the accuracy of public opinion analysis results, and provide decision support for government management and control. Formulating a public opinion guidance strategy and strengthening green network construction are of great significance for promoting economic development and social progress.

2. Summary of research at home and abroad

2.1. Status of foreign research

Network public opinion analysis, that is, through the integration of data and information resources such as speech and communication generated by an event on the network, using scientific methods, big data and other technical means to sort out and classify the collected data resources. At the same time, it

predicts the future development trend of the event, and then proposes rationalization and control of the event [6, 7]. The core purpose of network public opinion analysis is to provide decision makers with decision support for public opinion management by predicting the development law of public opinion, so that the network can run and develop healthily. From the existing research results and development status of public opinion analysis at home and abroad, the main analysis methods include network survey method, pattern recognition method based on statistical rules, and topic monitoring method based on content mining.

Foreign research on public opinion analysis started earlier, and Stat Pac survey software developed the survey software solutions system. The company designed a questionnaire for a topic and threw it on the network. The system then automatically collected, collated and analyzed the results of the questionnaire, which led to the analysis of the company's decision-making support. In 2006, Google launched Google Trends application, collected web search logs and analyzed logs, and found out real-time search hotspots and keywords on the web to help analyze the hot topic of the network. With the development of big data, foreign universities and research institutions have carried out research on big data applications. In 2013, the University of California conducted research, analysis, and management of community software platforms with the help of big data [8]. In 2014, Optimal Solutions adopted big data technology, analyzed the industrial supply chain, and proposed a supply chain optimization solution, which provided effective help for the company's development [9].

2.2. Status of domestic research

Domestic research on public opinion has started late. In recent years, some network public opinion monitoring and analysis platforms have also been launched in the market. For example, Founder Zhisi Internet public opinion monitoring system, Baidu public opinion and so on[10]. Lian Wei believes that the initial information of microblog public sentiment has a great influence and role in public opinion communication. After comprehensively analyzing the characteristics of the first-time information, the micro-blog initial information heat prediction equation model is constructed, and the model can obtain more accurate prediction results [11]. Gao Xinqing and others based on the perspective of stakeholders, from the perspective of the content of public opinion events, reveals the evolution of public opinion [12]. Sun Rong analyzed the difficulty of the government's control over public opinion and proposed a suggestive solution [13].

2.3. Summary of analytical studies abroad

Judging from the current domestic and international research on the development and development of big data and public opinion, the research results of the current public opinion analysis are still immature, and most of them are only based on a certain angle to analyze the public opinion, and the results are more one-sided. Moreover, the existing research results use more traditional lyric analysis methods, and the traditional methods can’t meet the efficiency and accuracy requirements of current public opinion processing.

Combine big data with public opinion analysis, transform all metadata into valuable information through quantitative associations, and achieve multiple uses. In addition, through the multi-dimensional interpretation of public opinion and the revealing of new insights, the comprehensiveness and objectivity of the results of public opinion analysis greatly surpass the traditional network public opinion management, and provide more accurate forecasting trends for government decision-making. Big data has become a source of strength for quantitative management of online public opinion. Therefore, the author relies on big data technology to design a network public opinion analysis platform architecture. Realize the collection, filtering, storage and in-depth analysis of massive network data. Excavate the law of public opinion communication, predict the development direction of public opinion, and provide data support for the development of public opinion control strategy.

3. Platform architecture design

The network public opinion analysis platform designed by the author consists of five modules: information collection, information processing, data storage, public opinion analysis and information security. As shown in Figure 1. Through the above functional modules, to achieve the status of public opinion information analysis, forecasting future trends, the analysis results are presented as well as public sentiment warning alarm. Provide data support for the government to monitor the development trend of public opinion and develop a public opinion control strategy.

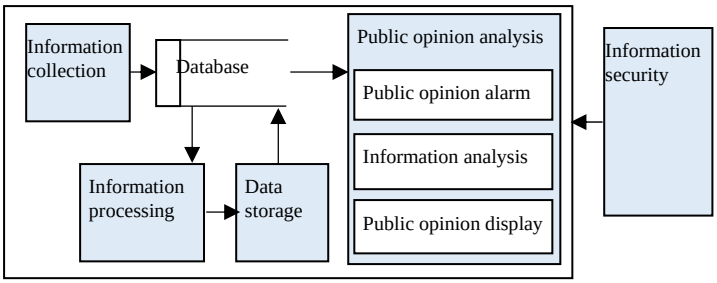

Figure 1. Network public opinion analysis platform.

3.1. Information Collection

With the use of web crawler tools, distributed crawler technology is used to crawl the information of the network. In order to efficiently crawling data, using parallel crawling mode. The crawler system consists of a service controller, a task allocator, and multiple crawler clients. The service controller is responsible for connecting to the client and controlling the crawl time of each client. The task allocator is responsible for assigning a crawl task to each client. In the process of crawling data, a crawling strategy based on depth-first traversal and breadth-first traversal strategy is used. While improving capacity and coverage, focus on the related information of the event of interest, reduce the crawling of useless information, and improve the efficiency of information update. The crawling process is shown in Figure 2.

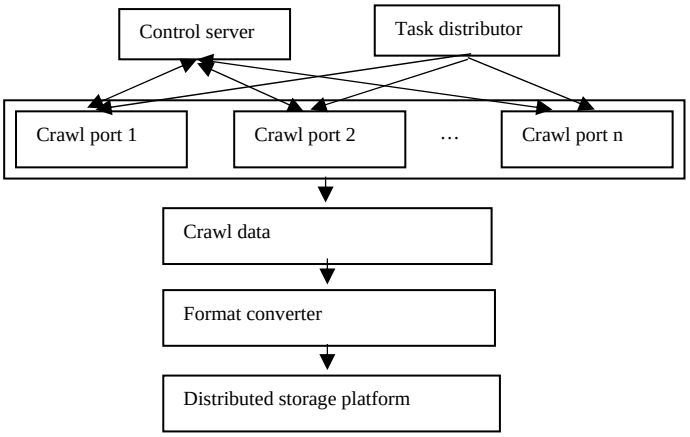

Figure 2. Crawling process.

3.2. Information Processing

The main function of the information processing module is to clean, filter, and extract the collected information. Then convert the valuable information content into formatted information text, which facilitates the subsequent operations such as storage and analysis. First, the information is deduplicated, noise is removed, and the like, and the web page information of no value is stripped. The word segmentation technique is used to cut the text string into the term information, and these terms are

defined as the feature items of the text. Finally, using vector space model, probability model and other methods, the mathematical model is used to extract feature records from feature items to form a text vector set. It is convenient for sorting, comparing, calculating and other operations in the later stage of public opinion analysis. Because the amount of information collected in the network is too large, the stand-alone system can no longer meet the requirements for such a large amount of data processing. Therefore, the information processing module also adopts a distributed processing method to improve processing efficiency through parallel computing, and to mine public opinion information in a short time.

3.3. Data Storage

Regardless of any platform or system, you need to consider how to store large amounts of data. With the rapid development of information technology, the more advanced and advocated storage method is cloud storage. The public opinion analysis platform is deployed on the cloud server, and the collected data for filtering processing is directly stored on the cloud. Not only realizes storage management automation and intelligence, but also improves storage efficiency. Moreover, virtualization technology can realize automatic redistribution of data, improve storage resource utilization, and reduce operating costs. Therefore, the network public opinion analysis platform is arranged on the cloud computing server, and the processed information is automatically classified into the cloud database according to the extracted feature items, and the user can remotely access and retrieve the data resources at any time.

3.4. Public Opinion Analysis

Information analysis. The public opinion analysis module is the core module of the platform. From the database, the hidden information containing potential value is refined and aggregated to realize the functions of sensation topic identification, topic tracking, text orientation analysis, hotspot mining, and hotspot prediction. Topic recognition is machine learning of a set of text vectors, and a large number of documents on the same event are clustered to identify the main topic of the document. Topic tracking calculates the similarity of each subsequent updated vectorized text to determine whether the text is related to an existing topic. If relevant, the text is classified into this topic, and irrelevant is counted as a new topic. Text orientation analysis can obtain text semantics by computer mining non-content or non-fact information such as various viewpoints, preferences, attitudes, and emotions contained in the text content of the network. Help relevant departments to detect negative grievances in a timely manner. Hotspot mining is to monitor the excavated topics and find that the amount of propagation within the set standard time period exceeds the critical value. It is determined that the topic has jumped to a hot topic and needs to be increased. Hotspot prediction is based on the results of hotspot mining, by comparing with similar events in the database, combining topic categories, topic content, public attitudes and so on. Forecasting the traffic volume curve of hot topics in a future period of time, laying the foundation for the formulation of public opinion management strategies.

Public opinion alarm. The sudden outbreak of public opinion information will have a huge adverse impact on the network. The public opinion analysis platform has the responsibility to assist in predicting and controlling the outbreak of public opinion. By setting warning and alarm warning lines, if the information spread range or influence level exceeds the warning line, an alarm will be issued in time. When the warning line is exceeded, all the relevant content of the topic is marked as yellow, and an early warning alarm is issued. When the alarm warning line is exceeded, the relevant content of the topic is marked in red, and an alarm is issued. In addition, the alarm is divided into two situations. One is a monitoring alert for real-time information, and the other is an alert for predicting trends. The predicted trend of a hot topic is marked with an alarm color corresponding to a node when it finds that it exceeds the warning line.

Lyric display. Statistical topic source information such as authority, number of comments, number of reprints, time of speech, intensity, etc., graphically display the data visually. And label the development trend of information dissemination. Excavate the hot topics in the stage time, and sort

according to the heat, vividly express the status quo, development and change of the public opinion, and improve the decision-making efficiency of decision makers.

3.5. Information Security

In the era of big data, there is a large amount of data generated every day, which contains a large amount of user privacy, business data and information. Along with the development of the era of big data, a large amount of data generated greatly increased the risk of confidence leakage. Through the establishment of a sound big data security information system, promote the development of related security technologies, and increase the way information security analysis. Detailed analysis and analysis of massive data information, timely identification of existing insecure factors and network anomalies, and accurate grasp of the risk problems in massive data.

4. Platform architecture design

4.1. Provide support for government decision-making and improve the accuracy of decision-making

In the context of globalization and post-industrialization, the information dissemination of virtual society has leapt to the mainstream, and the scope of influence is broader, which has adversely affected the stability of society. n addition, in order to improve the government's governance level and improve the accuracy of policy decisions, big data analysis has become an effective method. he public opinion analysis platform constructed by constructing the data analysis model with data analysis as the core can explore the correlation between social ecological elements. he use of public opinion analysis platform can timely grasp the transient and synchronicity of the public opinion outbreak, and effectively improve the enthusiasm and flexibility of the grassroots administrative work in dealing with sudden problems. t the same time, the public opinion analysis platform can predict the development trend of public opinion and provide decision makers with basic data for emergency plan formulation. Therefore, with the help of the public opinion analysis platform, the use of big data technology to form a more complete decision support system. ot only can it face the challenges of virtual social governance, but it is also the fundamental requirement for building a national governance system and realizing the modernization of governance capabilities.

4.2. Realize real-time tracking, assessment and trend forecasting of hotspots, and carry out precautionary activities for public opinion supervision

Network hotspots are usually the first stage of the public opinion outbreak, and hotspots will cause large changes to the network structure. Monitoring hotspots can greatly reduce the outbreak and impact of network public opinion. Therefore, the process of establishing social evolution of public opinion in the crowd behavior modeling system to achieve real-time tracking of the hot issues. By monitoring people's communication behaviors and comparing with historical data, mining the law of hotspots and predicting the development trend of events. At the same time, it evaluates the impact of hot events on the network and improves the level of network risk management and control.

4.3. Improve the scientific and rationality of public opinion guidance strategies

Take full advantage of big data to improve the ability to guide public opinion work, including the following three aspects:

Use big data to improve the predictability and purpose of network public opinion guidance: Through data capture and correlation analysis, a netizen opinion inclination analysis model is constructed. Understand the preferences and characteristics of netizens, build and improve government websites, official microblog nd use opinion leaders to achieve "good speech, dialogue, grounding, and practical matters."

Through the value conversion of data, realize the value guidance of network public opinion: Through data capture and correlation analysis, a netizen opinion inclination analysis model is constructed. Understand the preferences and characteristics of netizens, build and improve government

websites, official microblog nd use opinion leaders to achieve "good speech, dialogue, grounding, and practical matters."

Improve the credibility of public opinion guidance: On the one hand, we will strengthen the interaction between the old and new media, give full play to their respective advantages and communicate with the public, crack down on rumors and rumors, and achieve the dual guarantee of timeliness and authority. On the other hand, it is necessary to prevent public opinion analysts from being influenced by experience preferences in the process of processing data, and to prevent big data from being a means for some institutions and individuals to manipulate public opinion more conveniently.

5. Conclusion

Big data is an opportunity for the transformation of online public opinion management. Establish a public opinion analysis platform based on big data to accurately grasp the internal characteristics of the public opinion and its evolution. Optimize network public opinion management and control, Strengthen the construction of green networks,. Important for social stability and development.

References

[1] Chen Bikun, Wang Rifen, and Liao Haihan. “Current situation and development trend of social public opinion analysis and decision making based on the background of big data era,” Information Science, vol. 10, pp. 8-14, 2016.

[2] Chen Yonghua and Wang Rifen. “Social public opinion analysis and decision making support with big data,” New Technology of Library and Information Service, vol. Z1, pp. 3-11, 2016.

[3] Ma Mei, Liu Dongsu and Li Hui. “Research on network public opinion analysis system model based on big data,” Information Science, vol. 34, pp. 25-28+33, March 2016.

[4] Huang Wei, Li Rui and Meng Jialin. “The study on information literacy of university library based on xMOOC curriculum mode,” Library and Information Service, vol. 21, pp. 38-44+62, 2015.

[5] Chend Hongbing, Rong Chunming, Wang Kai, et al. “Secure Big Data Storage and Sharing Scheme for Cloud Tenants,” China Communications, vol. 06, pp. 106-115, December 2015.

[6] Xia Huosong and Zhen Huachun. “Public opinion analysis and decision support study under big data surroundings,” Journal of Intelligence, vol. 34, pp. 1-6+21, February 2015.

[7] Tao Xu, Wang Dongsheng Wang, Liu Guodong Liu. “Banian: A Cross-Platform Interactive Query System for Structured Big Data,” Tsinghua Science and Technology, vol. 20, pp. 62-71, January 2015.

[8] Wu Xia, Xu Lele and Yao Li. “Big data analysis of the human brain's functional interactions based on fMRI,” Chinese Science Bulletin, vol. 35, pp. 5059-5065, 2014.

[9] Jiang Changjun, Ding Zhijun, Wang Junli,et al. “Big data resource service platform for the internet financial industry,” Chinese Science Bulletin, vol. 35, pp. 5051-5058, 2014.

[10] Zhang Huaping, Zhang Ruiqi, Zhao Yanping, et al. “Big data modeling and analysis of microblog ecosystem,” International Journal of Automation and Computing, vol. 11, pp. 119-127, February 2014.

[11] Lian Zhixuan, Lan Yuexin, Xia Yixue, et al. “The research of Micro-blog public opinion popularity forecasting model based on the firstly published information,” Information Science, vol. 36, pp. 107-114, September 2018.

[12] Gao Xinqing, Yang Fan, Shang Chao and Huang Dexian. “A review of control loop monitoring and diagnosis:Prospects of controller maintenance in big data era,” Chinese Journal of Chemical Engineering, vol. 24, pp. 952-962, Aguest 2016.

[13] Sun Rong. “The Difficulties and Solutions of the Grassroots Government's Response to Internet Public Opinion,” People’s Tribune, vol. 24, pp. 53-55, 2018.