

A Survey of Key Technology of Network Public Opinion Analysis

Su Ying Li^a

Department of Literature and History at Party school of the C.P.C. Shandong Province Committee,
Shandong, China

Abstract: The internet has become an important base for internet users to make comments because of its interactivity and fast dissemination. The outbreak of internet public opinion has become a major risk for network information security. Domestic and foreign researchers had carried out extensive and in-depth study on public opinion. Fruitful results have achieved in the basic theory research and emergency handling and other aspects of public opinion. But research on the public opinion in China is still in the initial stage, the key technology of the public opinion analysis is still as a starting point for in-depth study and discussion.

Keywords: network public opinion; information collection; information extraction; public opinion analysis

Introduction

After newspapers, radio, television, the internet has become the fourth media. It has become the main medium of information dissemination and main carrier of social public opinion reflected ^[1]. Internet public opinion refers to the internet users' subjective reflection of the various phenomena and problems in society, and the public's comments and views with a certain tendency propagated through the network ^[2]. Because of the internet's virtuality, arbitrariness, and rapid propagation etc., false information, reactionary remarks, malicious speculation and other negative things are developing rapidly when the normal public dissemination of information. It disrupts the social order and harms the network information security. If we let this phenomenon go unchecked, negative public opinion will form a larger threat to public safety. In order to maintain social stability and prevent danger, monitoring online public opinion and early warning have become increasingly important.

With the emphasis on the network public opinion, the analysis about public opinion has become a hot research. Internationally, the United States' TDT ^[3] (Topic Detection and Tracking) system is the most well-known network public opinion analysis system.

TDT had developed a number of related algorithms about unknown topics to identify and existing topics to track. In 2005, the British software company developed a public opinion analysis software about emotion. The software is able to analyze online news stories and judge their emotion is positive, negative or neutral. Domestic public opinion research network units and agencies have Chinese Academy of Sciences, Academy of Social Sciences and some colleges and universities. Liu Yi wrote "Introduction to network public opinion research," this literature conducted in-depth discussions about the network of public opinion from theory and practice ^[4]. Pan xin etc.put forward new public opinion propagation model based on social network analysis^[5]. Zhang Yu carried out fruitful exploration about actual forum, blog, website with news comment function and proposed a comprehensive variety of algorithms for network public opinion^[6]. Jiang Fan ^[7] etc. conducted research about the forum network, and set up a theme discovery system. Zhou Yadong ^[8] analyzed the demand for network public opinion and delimited the hotspot of public opinion. Li hongtao ^[9] proposed gray evaluation method for network public opinion. Panchong Xia ^[10] completed the network public opinion evolutionary trend forecast with gray theory.

This paper analyzes the main technology in online public opinion. To collect the information rapidly and

^a Corresponding author: lisuyign@sina.com

accurately, and achieve the extraction about acquired information and denoising, and ultimately finish accurate judgment the trends of public opinion and forecast analysis of public opinion is the core of the work in the network public opinion analysis and early warning.

A public opinion information collection technology

Because of diversity and complexity of public opinion, public opinion information collection network has a certain degree of difficulty. The traditional method of information gathering public opinion is finished manually, it has large limitations and it is inefficient. Collecting information on the internet depends on web crawler technology. It mainly collects web pages through making use of the relationship between the pages. Research directions about crawler technology are introduced as follows.

(1)Information collection based on the entire web. It can expand from some seeds URL to the entire web information collection and finish the entire web collection. It is dominant in practical applications currently.

(2)Incremental web information collection. It only collects the pages which has changed or generated newly when the page is refreshed, but it can not collect pages with no change.

(3)Web information collection based on the topics. Collection is finished through selective search for those pages associated with the theme predefined.

(4)Personalized information collection. It can collect information meeting the different needs of users by means of user interaction.

(5)Information collection based on agent. Intelligent agent system refers to a computer system, It can be flexible activities independently in the environment like humans and it can perceive changes in the user's interest and adjust acquisition strategy independently .

The results of different information collection methods are different in dealing with the same query asked to return, and the quality is uneven. This is mainly due to its own advantages and disadvantages. It was found in actual collection process that a method

alone is not very satisfactory. It will miss a lot of qualified page. So we usually use two or more kinds of ways to complete the acquisition of information..

Public opinion information extraction technology

Web content will generally contain navigation, title, text, related links and advertising information. Web information extraction is to extract the web page content from the source file and identify the information related with the themes ^[11]. Extraction techniques commonly used are the following:

(1) the extraction technology based on the feature of web pages

Web page layout uses HTML tags generally. It has a hierarchical structure between the data with each other and is arranged in chronological order. To use structure characteristics of the html page layout, or to use a regular expression technology can get relatively pure text messages. The technology can remove additional noise information of web pages, such as page ads, links to other pages, descriptive information pages and so. It can effectively make the information extracted more accurate and pure.

(2) information extraction based on natural language processing

The technology completes processing by taking the entire web page as a text document. Therefore it is more suitable when the relatively large of information needs extraction. This method is the application of the traditional language processing technology, including parsing, syntactic analysis, semantic analysis and identification. The extraction process also includes the use of some extraction rules.

(3) information extraction technology based on ontology

Ontology is a formal concept. And it is used to describe the relationship of the related fields in order to provide for a common understanding of knowledge and realize shared knowledge ultimately. The approach based on ontology utilizes information to describe the data itself and it can establish an ontology library and use the specified extraction rule to complete information extraction. The advantage of this method is that does not depend on the pages' structure. As long as

the body is large enough, it can extract various information in this field.

(4) information extraction technology based on Hidden Markov Model

Hidden Markov model is a finite state automaton and it is an important means of signal processing. Currently it has been quite successful applications in speech recognition, behavior recognition and other fields. In recent years, this model is widely applied to information extraction, and there is now a very significant effect. It is possible to deal with new data robustly.

Public opinion analysis and prediction

Public opinion analysis is the core of the public opinion analysis system. At present clustering and classification methods in data mining are used to solve this kind of problem. The ultimate goal is to carry out public opinion analysis to predict the future trend of public opinion. Internet public opinion prediction is to study how to use the massive alarm data to predict future network of public opinion, that is, to predict the future by analyzing the past. Public opinion analysis and forecasting technology mainly consists of the following tasks.

(1) identify hot topics and sensitive topics

Mass information from the internet contains a large number of topics. According to topic areas they can be divided into: military, sports, entertainment, technology, society and so on. According to the importance of the topics, they can be divided into general topics and hot, sensitive topic. Currently, mainly found in the web-based text mining on the topic, and then measure the relevant reports on the topic of the total amount, the growth rate, fluctuation level, there are users of the topic replies intensive, emotional and other dimensions tendency to identify certain period of hot topics. At the same time, we determine whether it belongs to the sensitive topic through keyword matching as well as further semantic analysis.

(2) the topic tendentious analysis

Topics propensity analysis mainly refers to the emotional orientation analysis. Further extension may be subdivided into political orientation analysis,

product preference analysis AND interest preference analysis. Its main purpose is to analysis published articles, blog, reply and microblog, semantic analysis on the network. It can determine its emotional tendencies, such as commendatory, derogatory or neutral.

(3) the trend forecasting network public opinion analysis

Currently, the main focus of analysis to predict, there are two aspects, one is affected by the degree of concern of public opinion, mainly to predict trends reports its numbers and number of replies, the other is the degree of complexity of contradictory public opinion, mainly on trends to predict their emotional tendency. There are now widely used algorithm BP neural network model based on the integration of autoregressive moving average model (ARIMA), decision tree model and gray theory.

Conclusion

This article focuses on the key technology of internet public opinion analysis and introduces technical points and the corresponding research status. It can lay the foundation for further study of public opinion. When network provides a convenient way to send message, virtualization and concealment of the network also become a major social stability risk on the political and cultural security. Therefore, strengthening the monitoring of public opinion information and timely grasping public opinion trends are important.

Acknowledgments

This work is supported by the National Natural Science Foundation of China(Grant No.61402271), and Shandong Social Science Planning Project(15CXWJ08 and 14CZXJ14), and the Natural Science Foundation of Shandong Province (Grant No.ZR2015JL023 and Grant No.ZR2015FL025),and Key R & D program in Shandong Province(Grant No. 2015GGX101012).

References

- [1]Dai Yuan, Cheng learned Flag practical overview of the key technologies for the Internet public opinion analysis [J]. Computer Research and Development, 2008: 62-65.

-
- [2]Liu Zeguang network public opinion analysis[D].
Northeastern University Master Thesis, 2013.
- [3]Leek T, Schwartz R, Sista S. Probabilistic approaches
to topic detection and tracking [M].New York:.
Springer US, 2002.
- [4]Liu Yi Public Opinion Research Network
Introduction [M]. Tianjin: Tianjin People's
Publishing House, 2007.
- [5]Pan new Denggui Shi Tong Bin public opinion
propagation model based on social network
Construction and analysis [J]. Operations Research
and Management, 2011: 176-179.
- [6]Wu Yu. entertainment network public opinion
analysis[D].Chengdu: University of Electronic
Science and Technology, 2011.
- [7]Jiang Fan, high Junbo, Zhang Min and other .BBS
themes found in the design and realization of the
prototype system [J]. Computer Engineering and
Applications. 2005: 155-157.
- [8]Zhou Yadong, Sun Qin East, Guan et Traffic
content words related to the extraction of the hot
topics of the network [J].2007 Xi'an Jiaotong
University: 1142-1145.
- [9]Lee Yun Tao, Liu Yan, Liu Yi Grey alarm
assessment of network public opinion [J].
Information] .2011: 24-27.
- [10]Pan Chongxia stage evolution of network public
opinion analysis [J].Computer and Modernization,
2011 (10): 179-182.
- [11]parties pay to build a network of public opinion
research to quantify Situation Analysis Application
[J]. Journal of Information, 2014 (33): 47-51.