Impact of Network Public Opinion Disposal Model on Public Value Factors under the Background of Big Data

Peng Bo, Xu Xiao-long[*]

School of Business Administration,Shanghai Lixin University of Accounting and Finance, Shanghai, China

*Corresponding Author.

Abstract

It is the key for the government to control the degree of information alienation to study the mechanism and control model of network public opinion information alienation for big data. This provides a theoretical basis for the government to deal with and manage the network public opinion. This paper uses qualitative analysis of the information alienation mechanism of network public opinion under the big data environment, and expands the evolution mechanism model of network public opinion to the information alienation control model. On this basis, the classification of government control information alienation is studied by numerical simulation. This paper takes the actual forum, blog, website with news comment function as the research object, and proposes a prediction platform construction scheme based on Java, which integrates a variety of prediction models. This provides useful exploration and ideas for quantitative research on the complex social phenomenon of network public opinion.

Keywords: *Big Data, Network Public Opinion, Public Opinion Evolution Mechanism Model, Public Value.*

I. Introduction

Public opinion is also called public opinion, which is expressed by "public opinion" in the West. The narrow concept of social public opinion refers to "in a certain social space, around the occurrence, development and change of intermediary social matters, the social and political attitude of the people as the subject to the state manager as the object" [1-2]. The narrow sense of social public opinion is the component and core content of the broad sense of social public opinion. In China, social public opinion refers to the new situation and changes of public opinion which are not included in the traditional vision of the party and the government, and which are of great significance for building a socialist harmonious society and maintaining the sustainable and stable development of society and economy [3].

In the network environment, the form of social public opinion is network public opinion, which has the characteristics of fast expression, multiple information and interactive way [4-5]. It fundamentally changes the relationship between the disseminator and the receiver, and has the incomparable advantages of traditional media, which is the deconstruction and subversion of the traditional media communication mode. The main carriers of network public opinion are BBS, blog and website with news comment function, which constitute the main information source of network public opinion. According to the statistical report on China's Internet Development released by China Internet Network Information Center in January 2008, as of December 2007, the number of Internet users in China had increased to 210 million, slightly lower than 215 million in the United States, ranking second in the world [6].

Such a huge audience and such a rich application make a lot of progressive, healthy and beneficial advocacy information to be shared, but at the same time, it also provides a communication channel for many reactionary, superstitious and low-level misleading remarks, which makes it an important hidden danger to trigger mass

incidents and social emergencies. Therefore, it is particularly important to build the network public opinion prediction platform and prevent the negative network public opinion from harming the society.

II. Development environment

2.1 Introduction to eclipse

At present, there are a large number of development tools based on integrated development environment (IDE) in Java development field, such as IBM's Visual Age for Java, WebSphere Studio, Borland's JBuilder, NetBeans and Eclipse, etc.

Eclipse project is the next generation of extended development environments (xdes) or collaborative development environments (cdes), which is researched and promoted by many software tool development companies such as IBM, Borland, RedHat, SuSE, TogetherSofl and QrCX. Other companies have joined, including Hewlett Packard, Fujitsu and Sybase. The goal of Eclipse project is to provide a collaborative development environment that can use any language for software development with plug-in architecture. At present, Eclipse has been able to develop various Java-based applications and provide functional plug-ins developed in C, C++, C+and other languages [7-9].

It can be seen from figure 1 that the core components of Eclipse platform: Workbench, plug-in development environment PDE, Java development toolset JDT, Eclipse platform microkernel Platform Runtime, and its bottom layer is supported by standard Java9 virtual machine. The basic function of Eclipse platform is to provide software tool developers with the use mechanism and rules of seamless integration tools. These mechanisms and rules are expressed through well-defined API interfaces, classes and methods. Eclipse platform also provides some useful frameworks for developing new tools. On Eclipse platform, users can independently develop various functional components in the form of plug-ins and integrate them into the platform.
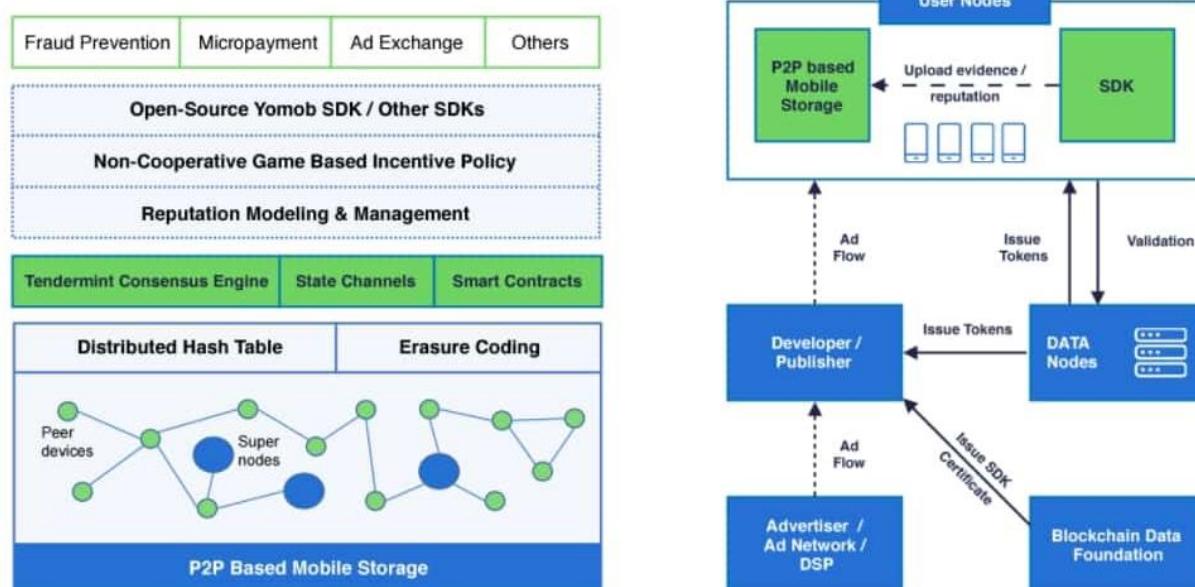

Fig 1: Eclipse platform architecture

2.2 Introduction to Java

In the white paper of Sun company, the definition of Java is "Java：A Simple, object—oriented, distributed, robust,

secure, architecture—neutral, portable, high—performance, multi．threaded, and dynamic language". That is, Java is a simple, object-oriented, distributed, strong, secure, architecture neutral, portable, high-performance and multi-threaded dynamic language.

 This definition fully illustrates the characteristics of Java language.

(1) Simplicity
Java is a simple language. Language designers try to reduce the size of the language. Java cancels many complicated and difficult contents in many languages, such as C + + pointer, operator overloading, multi class inheritance, etc., and greatly reduces the memory management of programmers by implementing the automatic garbage collection function. Java is familiar to most programmers in appearance and easy to learn. At the same time, java compiler is also very small, easy to implement on various models.

(2) Object oriented
Java is an object-oriented language. The object here refers to the program entity that encapsulates the data and its operation method. Java program design focuses on object and its interface. Java provides a simple class mechanism and dynamic interface model to realize modularization and information encapsulation. Java classes can provide a prototype of a class of objects, and then reuse the code through inheritance mechanism.

(3) Distribution
Java is a distributed language. It has a complete set of communication and related function library, which can handle tcpsp protocol and other protocols. Users can easily access other objects on the network with URL address. Using java to develop distributed network program is one of the main applications of Java.

(4) Robustness
The program written in Java can be executed stably in many situations, mainly because it needs to check the possible problems during compilation and runtime. Java has a special pointer model, its function is to exclude the possibility of data in memory being covered or damaged. Java also integrates object-oriented exception handling mechanism to prompt possible but unhandled exceptions at compile time to prevent system crash.

(5) Security
Java is a secure network programming language, which does not support pointer type. All memory access must be realized through object instances. This can prevent others from using deceptive means to access the private members of the object, and also can avoid the error in pointer operation. In addition, Java has multiple levels of interlocking protection measures, which can effectively prevent virus intrusion and deliberate destruction.

2.3 Introduction to MySQL

MySQL is a real, fast and stable, powerful, low-cost multi-user, multi-threaded relational database server. It adopts client and server architecture and has rich application interfaces. Its existing security mechanism is based on the user's autonomous access control mechanism. It has the following advantages:

(1) Compared with the complex setting and management of Oracle, SQL server, DB2 and other large databases, MySQL is flexible and easy to use and manage. It supports a variety of common operating systems, such as windows, UNIX, Linux, OS / 2 and so on. MySQL supports network applications, and users can access MySQL database through the network.

(2) In addition to using structured query language (SQL), MySQL also supports open database connectivity (ODBC).

(3) MySQL is the most popular open source database in the world. Users can add content to MySQL, fix its vulnerabilities, improve MySQL or make optimization suggestions.

2.4 JDOM software introduction

JDOM (Java document object model) can be used to generate and process XML documents and quickly develop XML applications. It is an API (Application Programming Interface) based on Java2. First, get and install JDOM. You can download the latest version of JDOM, JDOM. 1.1, from http://www.jdom.org. After downloading, unzip and add JDOM. Jar to classpath. JDOM makes up for the shortcomings of DOM and Sax; Sax does not have the functions of document modification, random access and output), which effectively combines the advantages of the two.

JDOM models the whole XML data file as a Document object, which can also exist independently. A complete JDOM Document contains a list of nodes, usually including Comment, ProcessingInstruction, a DocType and an Element object of the root element. The rest of the structure of the file is mainly built from the Element object, each containing a content list, as well as the attribute list and namespace list of this element. To input data into XML file, go through org. jdom. input package, and output requires org. jdom. output

III. Design and implementation of prediction platform

3.1 The overall process of forecasting platform

Figure 2 is the overall flow diagram of the network public opinion prediction platform. First, input the original data (the time sequence of the total number of topic replies about a hot topic, including the number of replies in BBS, the number of replies on the blog message board and the number of online news comments). Then the exponential smoothing module, ARIMA module and BP neural network module are used to predict and store the number of network topics in short time. Then the average absolute percentage error is used to evaluate the prediction effect, and the best prediction model of the number of online topic replies is selected to predict the development trend of the topic in the next period of time, and judge the topic type (continuous type and peak stop type). It is of great theoretical and practical significance to provide basis for early warning and guidance of network public opinion in the critical stage of the transformation of network speech from potential public opinion to dominant public opinion, to play a forward-looking role for decision makers or public opinion supervision departments, to grasp the initiative of public opinion guidance, to improve the level of network supervision, to timely respond to unexpected public events and to comprehensively grasp social and public opinions.
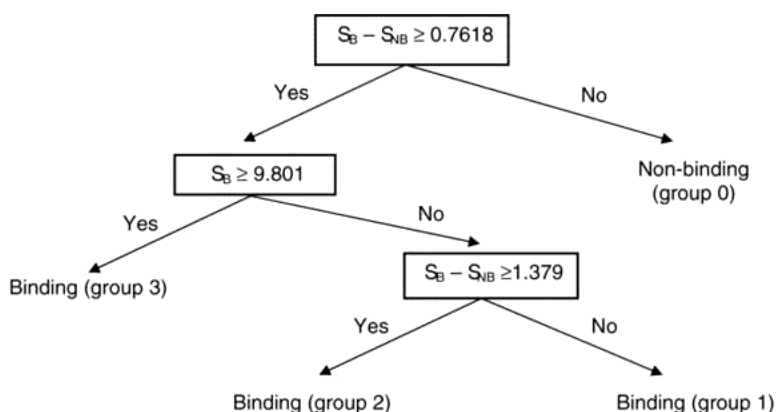

Fig 2: Overall flow diagram of prediction platform

3.2 Network crawler

Network crawler provides basic data for the prediction of the development trend of network public opinion. At present, there are 1.5 million websites and 8.47 billion web pages in China. In view of the efficiency of network crawler and the accuracy of information, this module mainly collects some large-scale and influential websites. The data capture process includes the following steps:

1) Give the initial URL and put it in the URL queue.

2) The grab management process starts the grab thread.

3) The capture thread obtains the target URL from the URL queue. It connects to the server through the h1tp protocol to obtain the page data pointed to by the target URL.

4) The captured web page is sent to the page parsing process to remove the useless information such as advertising links, redundant format tags, automatic recognition of paragraphs, etc., and parse out the URL and useful data contained in the page.

5) The URL obtained from page parsing is put into URL queue for grabbing thread to use; The data obtained from page parsing is stored in MySQL database.

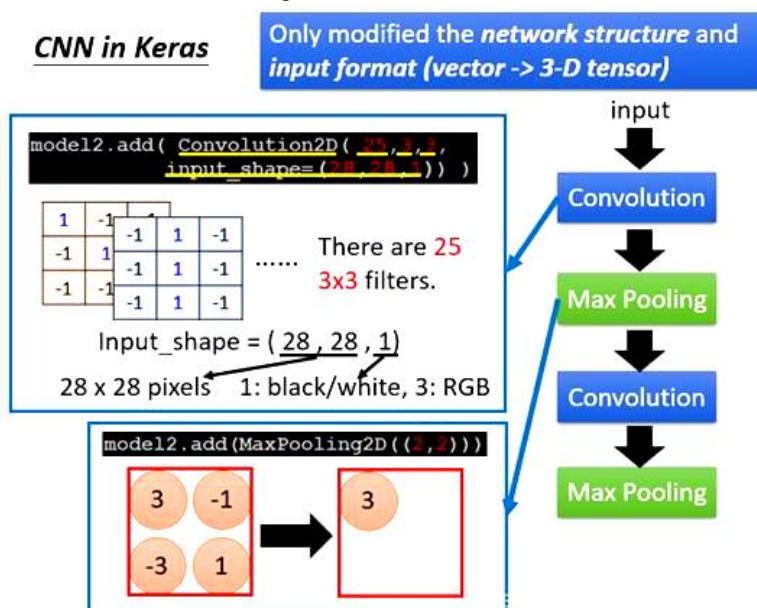The structure of network crawler is shown in Figure 3.

Fig 3: Network crawler structure

3.3 ARIMA dynamic model

In 1970s, American and British scholars bocks and Jenkins put forward the modeling method of auto. regressive integrated moving average models (ARIMA(p, d, q) model for short), which had a great influence on the time series analysis method and its application.

Basic types of ARIMA model [10]:

Let the time series x_1, x_2, ..., x_n be real stationary random sequences, which satisfy the following random equations:

Autoregressive model AR(p): $x_t = \phi_1 x_{t-1} + \phi_2 x_{t-2} + \cdots + \phi_p x_{t-p} + \delta + \varepsilon_t$ (1)

Moving average model MA(q): $x_t = \mu + \varepsilon_t + \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2} + \cdots + \theta_q \varepsilon_{t-q}$ (2)

In formulas (1) and (2), t: 1, 2, 3,..., n, p, q are the autoregressive order and moving average order of the model, θ, ϕ are the autoregressive coefficient and moving average coefficient, εt is the white noise residual sequence, μ is the mean value of x_t, δ is the constant term (related to the mean value of x_t).

Differential autoregressive moving average model ARIMA(p,d, Q): The above model is only applicable to stationary time series. In practice, time series are often non-stationary, so it is necessary to use finite difference to transform non-stationary time series into stationary series. The definition of difference operator ∇ is as follows:

First order difference: $\nabla x_t = x_t - x_{t-1}$; Second order difference: $\nabla^2 x_t = \nabla x_t - \nabla x_{t-1}$;...;D-order difference:

$\nabla^d x_t = \nabla^{d-1} x_t - \nabla^{d-1} x_{t-1}$,If the sequence x_1, x_2,..., x_n is a stationary time series after d-order difference (in practical application, d generally does not exceed 2), it is called "differential autoregressive moving average model", which is denoted as ARIMA (p, d, q). The formula (3) is as follows:

$$\phi(B)\nabla^d x_t = \delta + \theta(B)\varepsilon_t \text{ (3)}$$

3.4 BP neural network model

BP neural network model is completely proposed by Rumelhart, Hinton and Williams. It is a multi-layer feedforward network including input layer, hidden layer and output layer. Its structure is shown in Figure 4. It does not need to establish a complex mathematical model, after learning, it can establish the implicit complex relationship between samples; It has strong adaptability and fault tolerance, and allows data input with certain noise; Distributed parallel storage mode; Self organizing and adaptive data processing.

Considering that the network public opinion itself is a complex dynamic nonlinear system and the time series is generally continuous, the transfer functions of hidden layer and output layer are sigmoid (hyperbolic tangent function) type differentiable functions. Sigmoid function is a strict increasing function, showing a good balance between linear and nonlinear, so it can realize any nonlinear mapping between input and output, and is suitable for medium and long-term time series prediction.
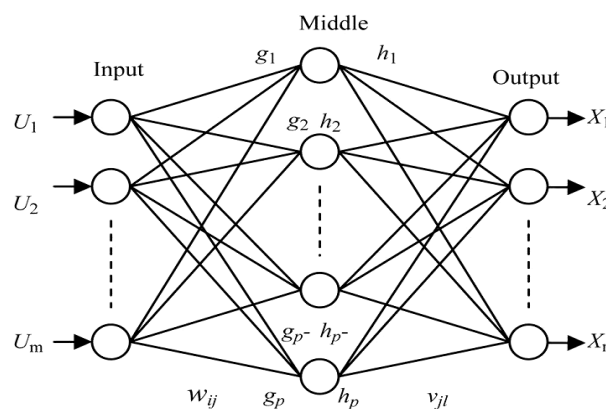

Fig 4: Structure of BP neural network

According to the network structure, the prediction model can only make one-step prediction, that is to predict the event value at t + 1 time according to the observation value at t time and the previous time. In order to predict the event value at t + 2 time, all inputs of the network model are moved forward one sampling period at t time. However, the event value of t + 1 in the input is still unknown, so it can only be replaced by the pre sliding value of the previous step. By analogy, the multi-step prediction value can be obtained by one-step prediction network model.

IV. Conclusion

Taking Eclipse, MySQL and JDOM as the main development tools and Java as the main programming language, this paper designs a prediction platform for network public opinion. It can be used as an important module of Internet public opinion comprehensive processing system to form a complete system.

The prediction model of network public opinion based on time series is based on the inertia principle of human society and the persistence of network public opinion. Although the public opinion orientation is time-varying and random, it is difficult to accurately predict. The platform can reflect the general trend of network public opinion development to a certain extent, and has high feasibility and effectiveness. Just for specific topics in a certain point in time, the prediction platform can not predict well, which is the next step to focus on solving the problem.

Acknowledgements

References

[1] Li, D. , Zhang, Y. , & Li, C. . (2019). Mining public opinion on transportation systems based on social media data. Sustainability, 11.

[2] Zhang, Y. X. , Feng, Y. X. , & Yang, R. Q. . (2019). Network public opinion propagation model based on the influence of media and interpersonal communication. International Journal of Modern Physics B, 1950393.

[3] Feng, Z. . (2018). Hot news mining and public opinion guidance analysis based on sentiment computing in network social media. Personal and Ubiquitous Computing, 23(3-4), 373-381.

[4] Alcala-Fdez, J. , Bonarini, A. , Flugy-Pape, N. , & Herrera, F. . (2010). Analysis of the effectiveness of the genetic algorithms based on extraction of association rules. Fundamenta Informaticae, 98(1), 1-14.

[5] Degroat, R. D. , Vis, M. , & Bliss, W. G. . (2001). Experimental characterization of media noise based on equalized, synchronized drive data. IEEE Transactions on Magnetics, 37(2), 633-638.

[6] Xia, P. , & Zhou, H. . (2017). A novel popular tourist attraction discovering approach based on geo-tagged social media big data. ISPRS International Journal of Geo-Information, 6(7), 216.

[7] Neville, M. C. , Keller, R. , Seacat, J. , Lutes, V. , Neifert, M. , & Casey, C. , et al. (1988). Studies in human lactation: milk volumes in lactating women during the onset of lactation and full lactation. American Journal of Clinical Nutrition, 48(6), 1375-1386.

[8] Wong, A. K. Y. , Ray, P. , Parameswaran, N. , & Strassner, J. . (2006). Ontology mapping for the interoperability problem in network management. IEEE Journal on Selected Areas in Communications, 23(10), 2058-2068.

[9] Tianyi Qin, Drivers drowsiness detection in embedded system, IEEE International Conference on Vehicular Electronics and Safety, 2007.ICVES

[10] Yan Chen, Shunqing Zhang, Shugong Xu, G.Y. Li, "Fundamental tradeoffs on green wireless networks," Communications Magazine, IEEE , vol.49, no.6, pp.30,37, June2011.