

<http://ansinet.com/itj>

ITJ

ISSN 1812-5638

INFORMATION TECHNOLOGY JOURNAL

ANSI*net*

Asian Network for Scientific Information
308 Lasani Town, Sargodha Road, Faisalabad - Pakistan

Research on Network Public Opinion Warning Index System Based on Feature Analysis of the Public Opinion

Peiguang Lin, Lei Liu, Peiyao Nie and Xiaoguang Zhu
School of Computer Science And Technology Shandong University of Finance,
Economics, Jinan, 250014, China

Abstract: Based on the analysis of characteristics of the network public opinion in different periods, this study presents three key elements to build the network public opinion early warning index system. We analyzed secondary events which had not been considered in other index systems and modified the definition of some index. Then, we added a part of new indicators and validated it in combination with the qualitative analysis, quantitative analysis and concrete examples. Finally, we built a scientific, rational and comprehensive warning index system of public opinion. The index system can provide theoretical support for monitoring public opinion by the government and relevant departments.

Key words: Network public opinion, characteristics of public opinion, early warning, index system

INTRODUCTION

In a narrow sense, public opinion refers to the social and political attitudes generated and held by the public against the social managers. The attitudes primarily surround the appearance, development and changes of intermediary social events in certain social space (Wang, 2003). In a broad sense, the network public opinion information refers to the combination of emotions, attitudes, beliefs, consciousness, thoughts, opinions, requirements and behavior modes expressed by the social people through the medium of the Internet. The information can reflect four civilization activities of material, political, spiritual and social aspects in the modern society (Dai and Yao, 2008).

Currently, the network public opinion has become a key problem of social science research, while the early warning of it is regarded as a crucial direction. Due to the complexity and variability of the Internet, the network public opinion may also be accompanied by the generation of secondary events. However, all the existing warning index systems of the network public opinion have ignored the influence of secondary events, making the performance of early warning not comprehensive and accurate. Based on the analysis of characteristics of the network public opinion, this paper made innovation of index and built a scientific and comprehensive warning index system of public opinion.

NETWORK PUBLIC OPINION EARLY WARNING INDEX SYSTEM

Characteristics of network public opinion: From the generation to fade-away process of the network public opinion, different periods of it has different characteristics and manifestations. In general, the network public opinion will probably go through four periods: the incubation period, the development period, the outbreak period and the fade-away period (Lan, 2011).

Network public opinion is very covert in the incubation period and whether it will be future hot topics in the network mainly depends on the sensitivity of the information content of public opinion (Gao *et al.*, 2011). In addition, the intuitional nature of the theme contents of public opinion also has certain influence on whether the information will be broadly paid attention to and whether public opinion will have further development; it's obvious that highly intuitive information content is more likely to be accepted and understood by the public.

In the development period, the network public opinion has become a hotspot and focus issue in the network. The information of the public opinion involves broad participation of people which is mainly reflected in the large amount of clicks, replies and reprinting; more active participants even express their views and opinions by means of blogs. Public opinion information is spread and reproduced within a certain range (such as a QQ group and a forum) in the period and receives certain

amount of clicks and replies and the number is still rising. With the development of the issues and increased concerns, exchange or opinions collision in small groups may be formed in the network. However, the scope of the impact of public opinion information is limited in this period and the number of participants is not too large (Wang *et al.*, 2011).

A concrete manifestation of the outbreak period of the development of the public opinion is the large-scale dissemination and reprinting of the information of network public opinion, along with a high rate of clicks and response as well as the formation of intense discussion in groups both in the online and offline world. Participants express their emotions and opinions through different channels. It may even turn into extreme forms like demonstrations and petitions which may lead to social unrest (Zeng and Xu, 2009). Therefore, the network public opinion in this period is similar to the one in the development period. The characteristic of public opinion is reflected in the active participation of people, while what is different is that people participating in the outbreak period are much more than that in the development period (Gold, 2007).

After the outbreak period and the successful settlement of incentive events of the network public opinion, the public's interest in participation is gradually reduced and the network public opinion becomes stabilized and gradually disappears. But the incentive events of the network public opinion leave their marks in the minds of people and once they are stimulated again by similar incentive events, it may generate secondary events with even greater impact and even deeper harm.

Establishment of the early warning index system of network public opinion: According to the analysis on

characteristics of public opinion in different periods of the network public opinion, it can be observed that the level of harm of the network public opinion can be judged and its trend of development can be predicted according to the presented characteristics of the current public opinion; therefore, the early warning index system of the network public opinion constructed and based on analysis of the characteristics of it is scientific and feasible. Specific indexes are shown in Table 1.

Final-level indexes' meaning and quantifying: The final-level indexes can be divided into two main categories of quantitative and qualitative indexes.

The first category is quantitative index which is obtained by means of the Web crawler through Nutch in major portals, forums and blog sites. Its quantitative ways include:

- The total number of words is that the words contained in the information content of the network public opinion in a certain statistical period
- The total number of pictures is that the pictures contained in the information content of the network public opinion in a certain statistical period
- The total time length of videos refers to the videos contained in the information content of the network public opinion in a certain statistical period
- The total number of clicks refers to the clicks on the information content of the network public opinion in a certain statistical period; it is derived by adding the number of clicks on news sites, blog sites and community forums
- The total amount of reprinting refers to the reprinting of relevant major posts of the information content of the network public opinion in a certain statistical period

Table 1: Early warning index system of network public opinion

1st-level index	2nd-level index	3rd-level index	Nature of index	Method of calculation
me attraction (A)	Sensitivity of me (A ₁)	Sensitivity of me (A ₁₁)	Qualitative index	Assessment by expert
		Total No. of words (A ₂₁)	Quantitative index	Data mining
	Intuitional nature of contents (A ₂)	Total No. of pictures (A ₂₂)	Quantitative index	Data mining
Total time of video (A ₂₃)		Quantitative index	Data mining	
Participation of public (B)	Amount of public opinion (B ₁)	Total No. of clicks (B ₁₁)	Quantitative index	Data mining
		Total No. of reprinting (B ₁₂)	Quantitative index	Data mining
		Total No. of comments (B ₁₃)	Quantitative index	Data mining
		Total No. of blog-writing (B ₁₄)	Quantitative index	Data mining
		Total No. of search (B ₁₅)	Quantitative index	Data mining
		Change rate of amount of public opinion (B ₂)	Change rate of clicks (B ₂₁)	Quantitative index
	Active level of public (B ₃)	Change rate of amount of reprinting (B ₂₂)	Quantitative index	Data mining
		Change rate of amount of comments (B ₂₃)	Quantitative index	Data mining
		Change rate of amount of blog-writing (B ₂₄)	Quantitative index	Data mining
		Change rate of amount of search (B ₂₅)	Quantitative index	Data mining
		Active level of public (B ₃₁)	Quantitative index	Data mining
Influence of public opinion (C)	Influence of event (C ₁)	Direct loss (C ₁₁)	Quantitative index	Official data
		Indirect loss (C ₁₂)	Qualitative index	Assessment by expert
	Influence of 2nd event (C ₂)	Influence of secondary event (C ₂₁)	Qualitative index	Assessment by expert

- The total number of comments is that the comments on the information content of the network public opinion in a certain statistical period
- The total amount of blog-writing refers to the released texts in blogs and micro blogs about the information content of the network public opinion in a certain statistical period
- The total number of searches refers to the searches for the information content of the network public opinion in a certain statistical period, the main source of which includes two parts: the first part is the search amount of main professional search tools (e.g., Google, Baidu); the second part is the search amount of search tools by various online media (e.g., blog forums, news sites)
- The change rate of public opinion refers to the change amount of public opinion within a certain period of time; it is calculated as follows: the change amount of public opinion over a period of time relative to the previous period is divided by that of public opinion in the previous period. The index can well reflect the development trend of public opinion.
- Direct loss refer to economic and property loss which can be obtained through the official data

The second category is qualitative index which is obtained mainly by scoring of public opinion expert team:

- As for theme sensitivity, Internet generates thousands or even tens of thousands of public opinion themes every day and different themes have different attractive ability to people. For example, Internet users are very concerned about the drug event as well as the problem of corruption and house prices, but the former will only affects the image of a star, while the latter may bring social harm and need the government's attention. The higher the sensitivity of the theme is, the greater harm it may bring and the greater crisis it may cause; this index is obtained through the scoring of the panel of experts of public opinion with the score set in [0,1]. 0, 0.3, 0.7 and 1 respectively means not sensitive, sensitive, very sensitive and most sensitive
- The index of indirect loss refers to negative impact of an event on the society which includes the reduction of the credibility of the government, the increase of social instability factors and endangering the government's public administration order. This index is obtained through the scoring of the expert panel of public opinion with the score set in [0, 1]. 0, 0.3, 0.7 and 1 respectively means no loss, loss and harm existing, serious harm and loss and very serious harm and loss

- The index of secondary events' influence refers to the degree of harm on the society caused by secondary events which are generated by public opinion including direct and indirect loss. This index is obtained through the scoring of the experts panel of public opinion according to the previous public opinion events with the score set in [0, 1]. 0, 0.3, 0.7 and 1 respectively means no loss, loss existing, serious harm and loss and very serious harm and loss

CALCULATION OF THE WEIGHT OF PUBLIC OPINION INDEXES

The study calculates the weight of public opinion early warning indexes by using the analytic hierarchy process; the characteristic of the analytic hierarchy process is to use less quantitative information to make the thought process of decision-making mathematical on the basis of in-depth analysis of nature, influencing factors and its intrinsic relationship of the complex decision-making problem, so as to provide a simple method for complex decision-making problems of multi-objective, multi-criteria or non-structural properties, which is especially suitable for occasions when it's difficult to directly and accurately measure the decision-making results.

Construct the pairwise comparison judgment matrix: The judgment matrix can be built based on the proportional scale table provided by the analytic hierarchy process, as shown in Table 2.

According to the proportional scale table, the public opinion expert group score and assess the pairwise comparison judgment matrix.

The first-level index judgment matrix:

$$\begin{pmatrix} 1 & 1/5 & 1/3 \\ 5 & 1 & 4 \\ 3 & 1/4 & 1 \end{pmatrix} \quad (1)$$

The second-level index has three judgment matrixes and is three first-level indexes respectively aimed at the theme attraction, the public participation and the influence of public opinion:

Table 2: Proportional scale table

1	It represents that two comparing elements are of same importance
3	It shows that as for two comparing elements, one element is slightly more important than anothe relement
5	It shows that as for two comparing elements, one element is obviously more important than anothe relement
7	It shows that as for two comparing elements, one element is strongly more important than anothe relement
9	It shows that as for two comparing elements, one element is extremely more important than anothe relement
2, 4, 6, 8	Is adjacent values for above judgment

$$\begin{pmatrix} 1 & 4 \\ 4 & 1 \end{pmatrix} \quad (2)$$

$$\begin{pmatrix} 1 & 2 & 3 \\ 1/2 & 1 & 2 \\ 1/3 & 1/2 & 1 \end{pmatrix} \quad (3)$$

$$\begin{pmatrix} 1 & 2 \\ 1/2 & 1 \end{pmatrix} \quad (4)$$

Third-level index has a total of four judgment matrixes which are four second-level indexes respectively aimed at the intuitional nature of contents, the public opinion amount, the change rate of public opinion and the influence of the event:

$$\begin{pmatrix} 1 & 6 & 5 & 4 \\ 1/6 & 1 & 1/2 & 1/4 \\ 1/5 & 2 & 1 & 1/2 \\ 1/4 & 4 & 2 & 1 \end{pmatrix} \quad (5)$$

$$\begin{pmatrix} 1 & 1/3 & 1/4 & 1/3 & 1/2 \\ 3 & 1 & 1/2 & 1/2 & 2 \\ 4 & 2 & 1 & 2 & 3 \\ 3 & 2 & 1/2 & 1 & 2 \\ 2 & 1/2 & 1/3 & 1/2 & 1 \end{pmatrix} \quad (6)$$

$$\begin{pmatrix} 1 & 1/3 & 1/4 & 1/3 & 1/2 \\ 3 & 1 & 1/2 & 1/2 & 2 \\ 4 & 2 & 1 & 2 & 3 \\ 3 & 2 & 1/2 & 1 & 2 \\ 2 & 1/2 & 1/3 & 1/2 & 1 \end{pmatrix} \quad (7)$$

$$\begin{pmatrix} 1 & 3 & 2 \\ 1/3 & 1 & 1/2 \\ 1/2 & 2 & 1 \end{pmatrix} \quad (8)$$

Calculate the weight vector and make consistency test:

This study adopts the eigenvalue method of calculating the sorted vector and its normalized eigenvector is corresponding weight vector. It can be observed from the analytic hierarchy process that conformance test can be made by judging whether $CR = CI/RI$ is less than 0.1, where, in $CI = (\lambda_{max}-n)/(n-1)$ and n is the matrix order. If CR is less than 0.1, the weight vector is acceptable; otherwise, it's not acceptable.

By calculating the matrix (1), the largest eigenvalue of the normalized feature vector $W = (W_1, W_2, W_3)^T = (0.1007, 0.6738, 0.2255)^T$ is: $\lambda_{max} = 3.0858, CI = 0.0429, CR = 0.0825 < 0.1$. it passes the consistency test.

Table 3: Early warning index weight of network public opinion

First-level	Second-level	Third-level	Weight of index
F1-0.1007	F11-0.8000	F111-1.0000	0.0806
		F121-0.1634	0.0033
	F12-0.2000	F122-0.2970	0.0060
		F123-0.5396	0.0109
F2-0.6738	F21-0.5396	F211-0.0728	0.0265
		F212-0.1875	0.0682
		F213-0.3718	0.1352
		F214-0.2487	0.0904
		F215-0.1192	0.0433
	F22-0.2970	F221-0.0728	0.0146
		F222-0.1875	0.0375
		F223-0.3718	0.0744
		F224-0.2487	0.0498
		F225-0.1192	0.0239
F3-0.2255	F32-0.1634	F321-1.0000	0.1101
	F31-0.6667	F311-0.7500	0.1128
		F322-0.2500	0.0376
		F321-1.0000	0.0752
F1-0.1007	F11-0.8000	F111-1.0000	0.0806
		F121-0.1634	0.0033
		F122-0.2970	0.0060
		F123-0.5396	0.0109
F2-0.6738	F21-0.5396	F211-0.0728	0.0265
		F212-0.1875	0.0682
		F213-0.3718	0.1352
		F214-0.2487	0.0904
		F215-0.1192	0.0433
	F22-0.2970	F221-0.0728	0.0146
		F222-0.1875	0.0375
		F223-0.3718	0.0744
		F224-0.2487	0.0498
		F225-0.1192	0.0239
F3-0.2255	F32-0.1634	F321-1.0000	0.1101
	F31-0.6667	F311-0.7500	0.1128
		F322-0.2500	0.0376
		F321-1.0000	0.0752

By calculating the matrix (2), the matrix of normalized feature vector $W = (W_1, W_2)^T = (0.8000, 0.2000)^T$ is a two-order matrix and the consistency test is not needed.

By calculating the matrix (3), the largest eigenvalue of the normalized feature vector $W = (W_1, W_2, W_3)^T = (.5396, .2970, 0.1634)^T$ is: $\lambda_{max} = 3.0092, CI = 0, 0046$ and $CR = 0.0088 < 0.1$. It passes the consistency test.

By calculating the matrix (4), the matrix of normalized feature vector $W = (W_1, W_2)^T = (0.6667, 0.3333)^T$ is a two-order matrix and the consistency test is not needed.

By calculating the matrix (5), the largest eigenvalue of the normalized feature vector $W = (W_1, W_2, W_3 \text{ and } W_4)^T = (0.6000, 0.0687, 0.1178, 0.2135)^T$ is $\lambda_{max} = 4.0916, CI = 0.0305, CR = 0.0343 < 0.1$. It passes the consistency test.

By calculating the matrix (6), the largest eigenvalue of the normalized feature vector $W = (W_1, W_2, W_3 \text{ and } W_4, W_5)^T = (0.0728, 0.1875, 0.3718, .2487, 0.1192)^T$ is: $\lambda_{max} = 5.0916, CI = 0.0229, CR = 0.0200 < 0.1$. It passes the consistency test and matrix (7) is the same as the matrix (6).

By calculating the matrix (8), the largest eigenvalue of the normalized feature vector $W = (W_1, W_2,$

$W_3)^T = (.5396, 0.1634, 0.2970)^T$ is: $\lambda_{\max} = 3.0092$, $CI = 0.0046$, $CR = 0.0088 < 0.1$. It passes the consistency test.

The weight of third-level index can be drawn by the weighted total rank (Table 3), for example, the weight of the index $A_{111} = 0.1007 * 0.8000 * 1.000 = 0.0806$.

CONCLUSION

On the basis of analysis on characteristics of the network public opinion, the paper established the network public opinion early warning index system, made clear the meaning and calculation method of each index, validated it in combination with the analysis on concrete examples, proved the feasibility of the index system not only theoretically, but also practically and provided theoretical support for monitoring public opinion by the government and relevant departments.

ACKNOWLEDGMENT

This study was supported by MOE Project of Humanities and Social Sciences (Project No.10YJC880076, 12JD710120) and Natural Science Foundation of Shandong Province (Project No. ZR2010FL008).

REFERENCES

- Dai, Y. and F. Yao, 2008. Research into Information mining and evaluation index system based on the security of public opinion on the internet. *Inform. Stud. Theory Appl.*
- Gao, C.S., X. Rong and Y. Chen, 2011. Research on micro blog public opinion monitoring index system. *J. Inform.*, Vol. 30.
- Gold, J.B., 2007. Survey methods and strategies of the network public opinion. *Henan Soc. Sci.*
- Lan, Y.X., 2011. Establishment of security assessment indexes system of the network public opinion. *J. Inform.*, Vol. 39.
- Wang, L.H., 2003. Introduction to the Public Opinion Research: Theory, Method and Reality Hotspot. Tianjin Academy of Social Sciences Press, China.
- Wang, Q., Y. Cheng and N.P. Cao, 2011. Research overview of public opinion monitoring and early warning index system. *Inform. Sci.*, Vol. 29.
- Zeng, R.X. and X.L. Xu, 2009. Early warning system, indexes and mechanisms of network emergency event. *J. Inform.*, Vol. 28.