TELKOMNIKA, Vol.15, No.4, December 2017, pp. 1808~1816
ISSN: 1693-6930, accredited **A** by DIKTI, Decree No: 58/DIKTI/Kep/2013
DOI:10.12928/TELKOMNIKA.v15i4.6832

How to Calculate the Public Psychological Pressure in the Social Networks

Rui Jin, Hong-Li Zhang, Xing Wang, Xiao-Meng Wang
School of Computer Science and Technology, Harbin Institute of Technology, Harbin, 150001,
telp: +86-0451-86418273, China
*Corresponding author, e-mail jinrui@hit.edu.cn[1], zhanghongli@hit.edu.cn[2], yeahwx@gmail.com[3], wangxiaomeng1123@163.com[4]

Abstract

With the worldwide application of social networks, new mathematical approaches have been developed that quantitatively address this online trend, including the concept of social computing. The analysis of data generated by social networks has become a new field of research; social conflicts on social networks occur frequently on the internet, and data regarding social behavior on social networks must be analyzed objectively. In this paper a type of social compouting method based on the principle of maximum entropyis proposed, and this type of social computing method can solve a series of complex social computing problems including the calculation of public psychological pressure. The quantitative calculation of public psychological pressure is so important to the public opinion analysis that it can be widely applied in a lot of public information analysis fields.

Keywords*: public psychological pressure, social computing, information entropy, principle of maximum entropy, public opinion analysis*

Copyright © 2017 Universitas Ahmad Dahlan. All rights reserved.

1. Introduction

In contemporary society the social networks has become the key infrastructure services. A lot of social information wasdistributed, transmitted, oracquired through social networking services. These sercices have a profound influence on the nation's politics, economy, culture and the organization of social activities. The data from SNS (Social Network Service) commonly have various media such as text, location, time, audio, and video and so on [1] [2]. At the same time, computer experts and data analysts have begun to show great interest in the data that are generated by these networks, prompting a new computing approach called "social computing" [3].

A seminar on social computing was held at Harvard University in 2007, and the U.S. military held a seminar at Arizona State University in April 2008 on behavior modeling and behavior forecasting. In 2009, the thinking of social computing began to become the formal academic concept developed by David Lazer et al. in the journal Science. They considered that a great deal of information contained in social networks such as blogs, BBS, chats, records of consumption, and E-mails can be analyzed by data analyst, which involved a lot valuable information [4].

The people expressed their personal views in the social networks, in which involve a lot of public mood. How to calculate the public psychological pressure in the social networks is an extremely essential subject in the public opinion analysis [5] [6] [7]. During the Arab Spring [7], a series of national governments collapsed. The social networks such as Twitter and Facebook play a vital role in it. When social networks make it easier or more convenient for the public to access information, countries face inevitable social and political consequences [8].

This article proposes a type of general-purpose social computing method that could be widely applied in complex social computing situations that have the characteristic of uncertainty. The proposed method can be used to apply quantitative calculations to complex social information systems. In this paper we try to calculate the public psychological pressure quantitatively for social networks, and this is a very meaning work.

Received August 28, 2017; Revised November 10, 2017; Accepted November 25, 2017

2. Related Works

2.1. The Related Theories of Social Computing

Social computing is a kind of calculation thinking which is used to quantitatively calculate the social properties or the social tendency.

1) The concepts behind Computational Social Science

In 2009, an epoch-making calculation was proposed in the field of social networking research when David Lazer and others proposed the concept of Computational Social Science. This concept expounded on the idea that much of the information found on networks such as blogs, BBSs, chats, log files, and E-mails are mappings of individual and organizational behavior in society [3]. This description is considered a complete representation for social computing theory.

2) The small world theory

In 1998, Duncan J. Watts and Steven H. Strogatz published a famous paper in the journal *Nature*, and in this paper they described the small world theory for social networks [9].

2.2. The Mathematical Modeling of Social Networks

Since the seminal works of Watts and Strogatz (1998) and Barabási and Albert (1999), network modeling has undergone rapid development. Researchers in the field of network modeling discovered that large-scale networks across different domains follow similar natural laws such as scale-free distributions, the small-world effect and strong community structures [9] as shown in Figure 1.

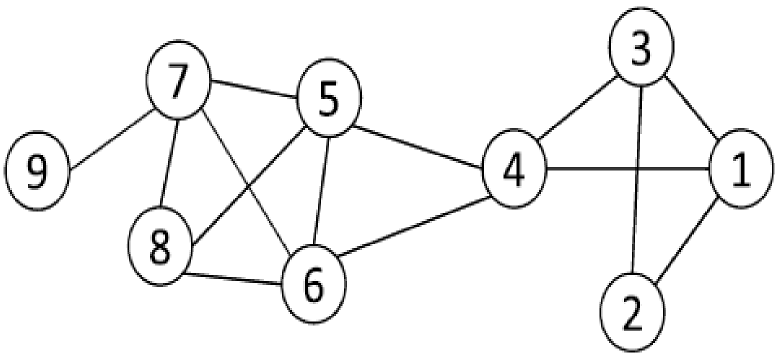

Figure 1. A social network

3. A New Type of Social Computing Method

3.1. The Mathematical Model of Social Event

3.1.1. The Structure of a Social Event

Based on the concept of ontologya social event is described in Figure 2 [10]. The mathematical description: Suppose that the social event information is the complete set U ,which is composed of the subsets Child content 1, Child content 2, …, Child content *n*-1, and Child content $n, U_1, U_2, ..., U_n$, satisfying the condition that $U_n = \overline{U_1 \cup U_2 ... \cup U_{n-1}}$,where U_1 represents Child event 1, U_2 represents Child event 2,…, and U_n represents Child event *n*. According to set properties we know that $U = \overline{U_1 \cup U_2 ... \cup U_n}$.

As shown in Figure 2, "Child content 1" is an attribute of the "Social event", and "Child content *i*" is an attribute of "Child content 2".These Child contents form the subsets of the "Social event," for which the set U is the complete set. Based on general philosophical principles, parts do not exist in isolation within any system; instead, they mutually influence each other.

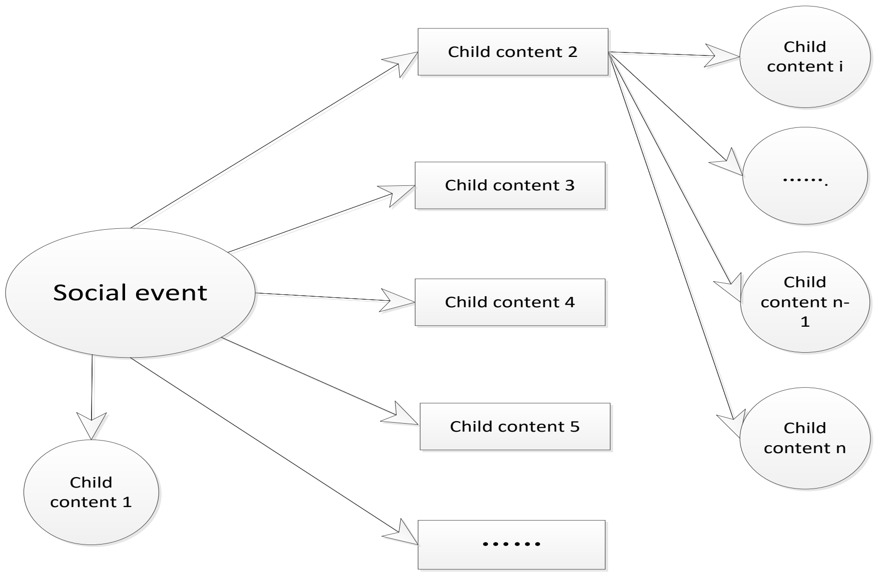

Figure 2. The social event structure

3.1.2. The Modeling of Multidimensional Random Variables

Suppose that the discrete random variable X represents the "social event", and the event itself is a random information system in which X_1 represents "child content 1", X_2 represents "child content 2",…, and X_n represents "child content *n*". Therefore, the discrete random variable X isequivalent to $\left(X_1, X_2,..., X_i,..., X_n\right)$, or can be represented as $X \square \left(X_1, X_2,..., X_n\right)$, where $\left(X_1, X_2,..., X_n\right)$ is the multidimensional random variables for which the function relation between every $X_i (0 \le i \le n)$ is either unknown or is complex and cannot be described with a specific mathematical formalization.

3.1.3. The Vector Space of Multidimensional Random Variables

Element-event: $\left(X_1, X_2,...X_i..., X_n\right)$ contains the basic values in its domain of values,forming an element-event in the vector space that can be represented by $\left(x_1, x_2, \dots, x_n\right)$, and all element-events constitute the vector space A.

Suppose that the number of elements in set U_i is q_i. In U_i, each $X_i (1 \le i \le n)$ obtains its values, and then are all $\prod_{i=1}^{n} q_i$ element-events, which for the vector space A can be represented as the matrix A, $\prod_{i=1}^{n} q_i$ rows, n columns.

$$A = \begin{bmatrix} x_{1,1} & x_{1,2} & \cdots & x_{1,n-1} & x_{1,n} \\ x_{2,1} & x_{2,2} & \cdots & x_{2,n-1} & x_{2,n-1} \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ x_{q_{1,1}} & x_{q_{2,2}} & \cdots & x_{q_{n-1,n-1}} & x_{q_{n,n}} \end{bmatrix}$$

3.2. A Method to Calculate Social Event Information Entropy

3.2.1. Shannon Entropy

In the 1940s, Claude Shannon, the father of information theory developed the information theory [11]. The definition: Suppose that X is a discrete random variable (i.e., one whose range $R=\{x_1,x_2,...,x_n\}$ is both finite and countable),let $p(x_i)=P\{X=x_i\}$. The information entropy $H(X)$ of X is defined by

$$H(X)=-\sum_{x} p(x)\log p(x) \quad (1)$$

The entropy $H(X)$ of the random variable X is a function of the random variable probability distribution $p(x)$. Because $X \square (X_1,X_2,...,X_n)$, it follows that

$$H(X)=-\sum_{x} p(x_1,x_2,\ldots,x_n)\log p(x_1,x_2,\ldots,x_n) \quad (2)$$

3.2.2. The Calculation Method Based on the Principle of Maximum Entropy

1) A description of maximum entropy theory

In 1950, E. T. Jayness proposed the principle of maximum entropy. The maximum entropy statistical model is a type of selection methodused to select the optimal distribution among eligible probability distributions [12].

$$p'=argmaxH(p) \quad (3)$$

2) The maximum entropy distribution of the discrete random variables

Suppose that X is a discrete random variable whose range $R=\{x_1,x_2,...,x_n\}$ is finite and countable. Let $p(x_i)=P\{X=x_i\}$; the corresponding probability value is then $p_1,p_2,...,p_i,...,p_n$ and the necessary and sufficient condition is $p_1=p_2=\ldots=p_n=\frac{1}{n}$ [13]. Proof, Because $\sum_{i=1}^{n} p_i=1$, based on the Lagrange multiplier method we try to solve the probability distribution under the maximum entropy constraint condition $\sum_{i=1}^{n} p_i=1$

$$F(p_1,p_2,\ldots,p_n)=-\sum_{i=1}^{n} p_i \ln p_i+\lambda\left(\sum_{i=1}^{n} p_i-1\right).$$

According to the necessary condition for obtaining the maximum, we take the partial derivative of p_i to get the equation set. Thus, we would have:

$$\partial F/\partial p_i=-\ln p_i-1+\lambda=0, \quad i=1,2,\ldots,n.$$

To solve this, $p_i=exp(\lambda-1)$, and p_i is a constant. According to the constraint condition $\sum_{i=1}^{n} p_i=1$,we know that $np_i=1$—that is, $p_i=1/n$. Therefore, the entropy function would be

$$H(x)=-\sum_{i=1}^{n}(1/n)\ln(1/n)=\ln(n).$$

The conclusion is that, monotonic behavior of the entropy function $H(x)=\ln(n)$ as shown in Figure 3.

$$H(x) = \ln(n). \quad (4)$$

The entropy calculation formula for multidimensional random variables is similar to that for a one-dimensional random variable. Calculating the information content of a social event is a maximum entropy issue for which the constraint condition is $\sum p(x_1, x_2, \ldots, x_n) = 1$, the form of the entropy function is similar to the form of a one-dimensional random variable, and the information entropy value can be any positive number.

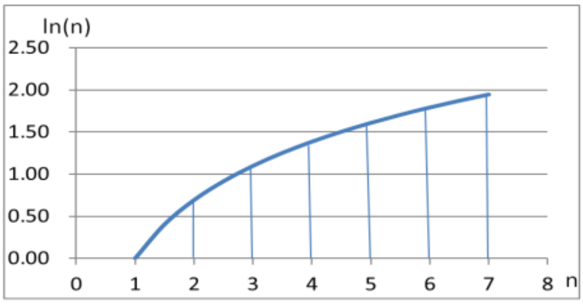

Figure 3. The monotonic behavior of the entropy function $H(x) = \ln(n)$

When the entropy function is maximized, the joint probability distribution is the uniform probability distribution [13].

$$\begin{aligned} & H(X_1, X_2, \ldots, X_n) \\ & = -\sum_x p(x_1, x_2, \ldots, x_n) \log p(x_1, x_2, \ldots, x_n) \\ & = -\log p(x_1, x_2, \ldots, x_n) \end{aligned}.$$

Suppose that q_i represents the amount of X_i after obtaining the values as shown in Table 1. While X_i obtains its value in its basic set of values once, $q_i = 1$.

Table 1. The Value of Sub-Variables

X_i	X_1	X_2	......	X_n
q_i	q_1	q_2	......	q_n

Then, the discrete multidimensional random variables $(X_1, X_2, \ldots, X_n)$ have $\prod_{i=1}^{n} q_i$ distribution items. According to the constraint condition

$$\sum p(x_1, x_2, \ldots, x_n) = 1,$$

We would then have $p(x_1, x_2, \ldots, x_n) = \frac{1}{q_1 q_2 \cdots q_n}$, and the entropy function is

$$\begin{aligned} H\left(X_1, X_2, \ldots, X_n\right) &= -\log p\left(x_1, x_2, \ldots, x_n\right) \\ &= -\log \frac{1}{q_1 q_2 \ldots q_n} \\ &= \log\left(q_1 q_2 \ldots q_n\right) \end{aligned}.$$

That is, the entropy function of social event multidimensional random variables is described by Formula (5).

$$H\left(X_1, X_2, \ldots, X_n\right) = \log\left(\prod_{i=1}^{n} q_i\right) \quad (5)$$

3) The monotonicity proof of the social event multidimensional random variables entropy function. The entropy function is

$$\begin{aligned} H\left(X_1, X_2, \ldots, X_n\right) &= -\log p\left(x_1, x_2, \ldots, x_n\right) \\ &= \log\left(q_1 q_2 \ldots q_n\right) \end{aligned}.$$

Proof, when $\left(X_1, X_2, ..., X_n\right)$ obtains a set of values $(q_1, ..., q_i^{'}, ..., q_n)$, the entropy is $H^{'}$; however, when $(X_1, X_2, ..., X_n)$ obtains another set of values $(q_1, ..., q_i^{"}, ..., q_n)$, the entropy is $H^{"}$. When $q_i^{"} > q_i^{'}$ because $q_i > 0$, then $q_1 ... q_i^{"} ... q_n > q_1 ... q_i^{'} ... q_n$; thus, we would have

$$\log(q_1 ... q_i^{"} ... q_n) > \log(q_1 ... q_i^{'} ... q_n).$$

Therefore, $H^{"} > H^{'}$ and the entropy function has strict monotonicity,Q.E.D. Based on a theoretical analysis, the calculation method is shown to be appropriate for almost all social computing situations.

4. A Calculaiton Method for the Entropy of Public Psychological Pressurein the Social Networks

4.1. The Analysis for Public Psychological Pressure

4.1.1. The Structure of Public Psychological Pressure

There are multiple categories of public moods involved in the public psychological pressure, and the structure of public psychological pressure is shown in Figure 2. In recent years the related researchers analyzed the text content of daily Twitter feeds by two mood tracking tools, namely OpinionFinder thatmeasures positive vs. negative mood and Google-Profile of Mood States (GPOMS) that measures moodin terms of 6 dimensions [14] [15].

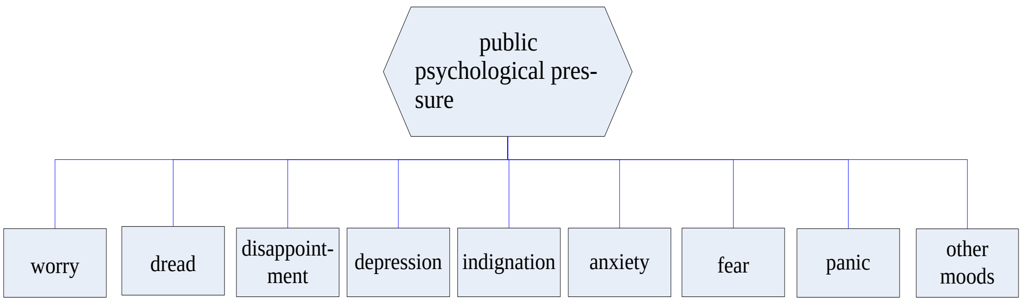

Figure 4. The structure of publicpsychological pressure

4.1.2. The Mathematical Model for Public Psychological Pressure

Based on Figure 4, the mathematical model is described as follows. Assume that the discrete random variable$X$ represents the“public psychological pressure” such that X_1 represents “**worry**”, X_2 represents “**dread**”, X_3represents “**disappointment**”, X_4 represents “**depression**”, X_5 represents “**indignation**”, X_6 represents “**anxiety**”, X_7 represents “**fear**”, X_8 represents “**panic**”, X_9 represents “**other moods**”, and $X \sim (X_1, X_2 \ldots\ldots X_9)$.

Furthermore, assume that the domain of X is U, and the probabilitydistribution is $p(x)$; the domain of X_1, X_2…, and X_{10} are U_1, U_2…, and U_9, respectively; and the probabilitydistributions are $p_1(x)$, $p_2(x)$, …, and $p_9(x)$, respectively.

4.2. The Calculation Method

According to the mathematical model, we can obtain the entropy function H of X as follows:

$$H = \log\left(q_1 q_2 \ldots q_9\right) \quad (6)$$

5. Experiments

5.1. The Experiments Data

In July 2012 a number of important public events occurred that garnered extensive attention in Chinese social networks. We calculated the corresponding public psychological pressurefor each event as shown in Table 2.

Table 2. Public Events

Sequence number	Event	Time
1	The 7/21 rainstorm in Beijing.	2012.7.21
2	A Japanese company discharged carcinogenic sewage into Nantong city, Jiangsu Province.	2012.7.25
3	A 3.3- magnitude earthquake occurred in Dali, Yunan Province.	2012.7.27

We collected corresponding social media comments concerning these events from social networks over a 30-day period. These comments tended to be highly integrated and traditionalwith abundant emotional coloring. For example, the number of comments collected over 30 days for Event 1 is shown in Table 3.

Table 3. Event 1

Time(day)	No.1	No.2	No.3	No.4	No.5	No.6	No.7	No.8
Number	1356	1795	1837	1936	1846	2305	2541	2252
Time	No.9	No.10	No.11	No.12	No.13	No.14	No.15	No.16
Number	2071	1803	1595	1502	1396	1203	1181	1063
Time	No.17	No.18	No.19	No.20	No.21	No.22	No.23	No.24
Number	936	982	782	657	662	608	573	562
Time	No.25	No.26	No.27	No.28	No.29	No.30		
Number	441	382	321	309	331	341		

Subsequently, the data for these comments were processed, and the statistical totals of the pessimistic public sentiments on Day 1 are shown in Table 4.

Table 4. Pessimistic Public Moods on Day 1 for Event 1

X	X_1	X_2	X_3	X_4	X_5	X_6	X_7	X_8	X_9
q_i	521	637	726	581	672	832	693	825	1

The public psychological pressure $H_1$of the public Event 1 on the firstday can be calculated as follows.

$$H_1 = \ln\left(q_1, q_2, \ldots, q_9\right)$$
$$= \ln(521*637*726*581*672*$$
$$832*693*825*1)$$
$$= 52.16$$

For example, the number of comments collected over 30 days for Event 1, Event 2, Event 3 is shown in Table 5.

Table 5. Public Events

Time(day)	No.1	No.2	No.3	No.4	No.5	No.6	No.7	No.8
H_1	52.16	54.71	54.08	54.92	54.83	55.52	56.30	55.81
H_2	47.81	48.72	49.18	46.82	46.93	44.78	42.93	43.52
H_3	33.29	33.92	28.93	24.21	16.83	13.92	12.81	10.38
Time(day)	No.9	No.10	No.11	No.12	No.13	No.14	No.15	No.16
H_1	56.47	55.92	54.71	53.82	52.06	52.27	51.83	52.01
H_2	42.68	42.39	42.01	43.29	41.48	42.82	43.92	42.52
H_3	12.28	12.74	10.82	10.77	11.69	10.01	11.92	9.88
Time(day)	No.17	No.18	No.19	No.20	No.21	No.22	No.23	No.24
H_1	51.93	52.63	52.73	51.39	48.29	45.92	45.83	46.72
H_2	42.28	38.58	39.73	38.96	39.72	37.39	36.30	35.93
H_3	9.92	8.39	8.38	9.92	10.39	8.82	8.30	7.20
Time(day)	No.25	No.26	No.27	No.28	No.29	No.30		
H_1	42.21	42.93	43.76	42.92	42.02	41.83		
H_2	36.43	35.72	35.83	36.92	33.62	33.39		
H_3	6.39	7.30	8.20	6.93	8.36	7.72		

Using those data, we can describe the tendency of public psychological pressure over 30 days, as shown in Figures 5.

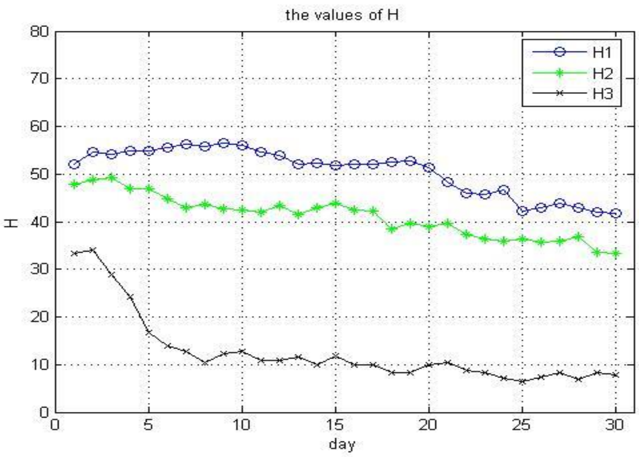

Figure 5. The H values for Event 1, Event 2, and Event 3

6. Conclusion

This paper presents a new type of social computing method based on maximum entropy theory. The method has a solid theoretical foundation and can be used to quantitatively analyze uncertain issues using data gathered through social networks. Applying the proposed method the public psychological pressure can be quantitativelycalculated for public information analysis.

7. Acknowledgments

Supported by the Major State Basic Research Development Program of China (973 Program) under Grant 2013CB329602; the National Natural Science Foundation of China under Grant No.61202457，No.61472108，No.61402149.

References

[1] Setiawan Assegaff, H Hendri, Akwan Sunoto, Herti Yani, Desy Kisbiyanti. Social Media Success Model for Knowledge Sharing (Scale Development and Validation). *TELKOMNIKA. Telecommunication Computing Electronics and Control.* 2017. DOI: http://dx.doi.org/10.12928/telkomnika.v15i3.5569
[2] Louis Y, Asur S, Huberman BA. What Trends in Chinese Social Media. 5th SNA-KDD Workshop'11 (SNA-KDD'11), San Diego, CA, USA. 2011.
[3] Social Computing.2016. [URL:http://en.wikipedia.org/wiki/Social_computing].
[4] Lazer D, Pentland A, Adamic L, Aral S, Barabasi AL, Brewer D, Christakis N, Contractor N, Fowler J, Gutmann M, Jebara T, King G, Macy M, Roy D, Van Alstyne M. Social science. *Computational social science. Science*. 2009; 323(5915): 721-723. [doi:10.1126/science.1167742].
[5] Russell G. Brooker and Todd Schaefer. "Methods of Measuring Public Opinion". PUBLICOPINION IN THE 21ST CENTURY. 2015.
[6] JA Balazs, JD Velásquez, Opinion mining and information fusion: a survey, Inf. Fusion. 2016; (27): 95–110.
[7] "Arab Spring". Available at https://en.wikipedia.org/wiki/Arab_Spring.
[8] Khondker, Habibul Haque. *Role of the New Media in the Arab Spring*. Special Forum on the Arab Revolutions. 2011: 675-679. http://dx.doi.org/10.1080/14747731.2011.621287
[9] DJ Watts, SH Strogatz. "Collective dynamics of 'small-world', networks". *Nature*. 1998; 393(6684): 440-442. doi: 10.1038/30918.
[10] Raimond Y, Abdallah S, Sandler M. The Music Ontology.International Conference on Music Information Retrieval. *Austrian Computer Society*. 2007: 417-422.
[11] Doran RS, M Ismail, TY Lam, E Lutwak R. *The Theory of Information and Coding*. Cambridge Books Online © Cambridge University Press. 2010.
[12] Jaynes ET. Information and statistical mechanics. *Physical Review*. 1957; 32(1): 3-7.
[13] L, XD. *The Method Study About Probability Distribution Based On The Principle Of Maximum Entropy*. Master Degree Dissertation, North China Electric Power University, Beijing, China. 2008.
[14] Aisah Rini Susanti, Taufik Djatna, Wisnu Ananta Kusuma. Twitter's Sentiment Analysis on Gsm Services using Multinomial Naïve Bayes. *TELKOMNIKA Telecommunication Computing Electronics and Control*. 2017. DOI: http://dx.doi.org/10.12928/telkomnika.v15i3.4284
[15] Johan Bollen, Huina Mao, Xiaojun Zeng. Twitter mood predicts the stock market. *Journal of Computational Science*. 2011.