HK.NCCP
International Journal of Intelligent Information and Management Science
ISSN: 2307-0692 *Volume 2, Issue 8, December 2013*

Research on the Indicator System and Method of Internet Public Opinions Evaluation

Gensheng Wang
Electronic Business department
Jiangxi University of Finance and Economics
Nanchang, China

Abstract: Internet public opinions evaluation is always an important research contents in network and its application fields and existing methods about Internet public opinions evaluation has no satisfactory solution in particular for its defects in calculation speed. So the paper present a new algorithm for internet public opinion evaluation based on improved ant colony algorithm. First, the paper designs a new evaluation indicator system based on the characteristics analysis of internet public opinions. Second, the ant colony algorithm is improved and a new internet public opinion evaluation model is presented to speed up the convergence and simplify the model structure of the model. Finally the experimental results verify that the effectiveness and validity of the model can be guaranteed when used for evaluating internet public opinions practically.

Keywords: Internet public opinions evaluation, Ant colony algorithm, Evaluation indicator system

1. Introduction

Open internet, with its marked features like convenient communication and fast spreading, becomes the most frequently-used window to reflect social conditions and public opinions; internet public opinions are the real-time barometer of social public opinions. However, the liberalization and diversification of internet public opinions is a double-edged sword. If effective management methods and means fail to be adopted, negative impact will be exerted on the safety of social public opinions, national stability and harmony. In recent years, there are dramatic increases in major public opinion events broken out through internet; making the situation of safety supervision and warning of internet public opinions in all countries become very serious, causing the high attention of countries and researchers. Safety evaluation of internet public opinions provides a systematic and overall basic method for the realization and application of internet public opinions analysis and warning techniques, helping government and national public opinions supervision departments promptly find public opinions information which may bring disastrous consequence for the safety of national public opinion and make warning, preventing the negative internet public opinions from harming the society. Therefore, evaluation on internet public opinions safety (including establishing evaluation indicator system and evaluation method study of internet public opinion safety) has become one of the research hotspots and difficulties in the industry.

2. Literature Review

Up to now, mathematical models adopted by evaluation of internet public opinions mainly include the following categories. (1) Analytic hierarchy process is a good method for quantitative evaluation via quantitative method, having the functions of establishing the ideal weight structure of evaluated object value and analyzing the weight structure of actually-built value by evaluated object; however, the method has strong limitations and subjectivity, with large personal error, not suitable for complicated system with lots of evaluation indicators[3,4,5]. (2) BP neural network evaluation method makes use of its strong capability in processing nonlinear problems to carry out evaluation of Music education performance; the method has advantages like self-learning, strong fault tolerance and adaptability; however, the algorithm is easy to be trapped into defects like local minimum, over-learning, strong operation specialization[6,7]; (3)Fuzzy comprehensive evaluation is a method carrying out comprehensive evaluation and decision on system through fuzzy set theory, the greatest advantage of which is that it works well on system evaluation of multi-factor and multi-level complicated problems. However, the membership of fuzzy evaluation method as well as the definition and calculation of membership function are too absolute, difficult to reflect the dynamics and intermediate transitivity of evaluation indicators of internet public opinions [8,9].

The paper uses cnt colony evaluation algorithm to correct and modify standard to overcome the question of slow convergence speed of BP neural network algorithm. In so doing, not only the problem of convergence speed of BP neural network algorithm has been solved, but also the simplicity of the model structure and the accuracy of system evaluation are ensured, and then a new BP neural network algorithm is advanced which is used toevaluat internet public opinions.

3. Establishment of Evaluation Indicator System

Based on the deep analysis of the characteristics of internet public opinions, referring to the studied literature, in consideration of the measurability principle, reliability principle, orientation principle, continuity principle and minimum indicators principle, this paper has designed a set of evaluation indicator system of public opinions warning. The system includes 4 hierarchies, 4 first-grade indicators, that are theme popularity, theme strength, tendency of net citizens, and theme time effect respectively. Theme popularity includes four second-grade indicator, that are explosive Power of theme(including one third-grade indicator, ie category of information theme); Media conditions(including three third-grade indicator, ie website contents distribution, website contents distribution, website popularity), Media influence(including two third-grade indicator, ie website credibility, attribution); Publisher influence(including three third-grade indicator, ie occupation attribute, religious faith, education); Theme hazard(including three third-grade indicator, ie theme spreading, theme sensitivity, theme importance); Theme spreading degree(including four third-grade indicator, ie browse number, times of report, reply ratio, republish ratio); Tendency distribution of net citizens(including three third-grade indicator, ie distribution proportion of different tendency, forward tendency proportion, reversed tendency proportion); Theme timeliness(including three third-grade indicator, ie theme status, theme inflection point, reversed theme duration).

4. Derivation of Algorithm

4.1. Ant Colony Evaluation Algorithm

Ant colony evaluation algorithm is a kind of swarm intelligence algorithm generated through human's studies on the formation principle of ant heap. Main idea of ant colony evaluation algorithm is to randomly place the data objects to be clustered in a two-dimensional grid environment, each data object occupying a grid, artificial ants randomly moving in the two-dimensional grid; each ant, according to the level of similarity of data objects and local environment, decides the probability of picking up or putting down the data objects. The higher of level of similarity is, the smaller the probability of picking up is, the smaller the probability of putting down is; the movement of the ant makes it interplay and interact with its neighbor nearby. After certain iterations, ant colony acts jointly, causing the objects of the same category to be clustered together in the same spatial region, realizing the evaluation process of self-organization.

4.2. Neighborhood Similarity Function

Neighborhood similarity function $f(i)$ is the average similarity of ants picking up or putting down object i and the neighborhood object within its observation radius; $f(i)$ can be calculated by equation 1.

$$f(i)=\max(0,\frac{1}{\sigma^2}\sum_{j\in I}(1-\frac{\sigma(i,j)}{\alpha(1+(v-1)/v_{\max})})) \quad (1)$$

In equation 1, $\sigma(i,j)\in[0,1]$ represents the distance between object i and object j, normally adopting Euclidean distance or cosine distance; $\alpha\in[0,1]$ is similarity adjustment factor; σ^2 is the size of local neighborhood L; the ant lies in the center of the region, and the neighborhood visible radius r of the ant is given by $(\sigma-1)/2$; v means the velocity of movement of the ant, and $v_{\max}$ means the maximum velocity of movement of the ant.

Neighborhood Similarity Function.The ant keeps moving in the two-dimensional grid and repeatedly picks up or puts down the objects; the influences of similarity function $f(i)$ on the probabilities of the ant picking up and putting down the objects while moving are expressed as p_{pickup} and p_{drop}, see equation 1 and equation 2.

$$p_{pickup}(i)=(\frac{k_p}{k_p+f(i)})^2 \quad (2)$$

$$p_{drop}(i)=\begin{cases}2f(i), & f(i)\prec k_d\\ 1, & f(i)\geq k_d\end{cases} \quad (3)$$

Improved Ant Colony Evaluation Algorithm

As there are some problems in ant colony evaluation algorithm, some improvements have been done on basic ant colony algorithm (BACA).

In the ant colony evaluation algorithm, it always takes the ants a lot of time to look for the appropriate data objects to pick up, resulting in the non-loaded idling of the ants in the iterations, thus leading to low time efficiency of the algorithm and taking a long time to form high-quality cluster. In the improved algorithm, apply rule 1 to improve the strategy for the ants picking up data objects.

Rule 1: After putting down objects, the ants will immediately allocate an appropriate idle data object in the iteration and try to pick up; if failed, they will re-allocate, until picking up the appropriate data object.

While the ant puts down the data object, if it finds the appropriate location s, but s has been occupied by other data, the strategy of original algorithm is to randomly move the ant to other position, which is a little blind, increasing the time cost of the algorithm but decreasing the evaluation effect of the algorithm. The treatment strategy for such situation in the improved algorithm is to look for other empty position within the visible radius scope of the ant around s; specific method is to improve the strategy for the ant putting down data objects by adopting rule 2.

Rule 2: first check whether there is empty position where the distance from s is 1. If empty, put down the object in the empty position; otherwise, check whether there is empty position where the distance from s is 2. Do the above repeatedly, until the visible radius r. If there is no empty position in the end, randomly move the ant to other position.

Determination of parameter α is depended on the experience of users, making evaluation algorithm lack of universality, affecting the effect of clustering. α determines the similarity of data objects in local neighborhood. If α is too big, it is difficult for the ant to pick up objects, but it is easy for the ant to put down the objects, leading to a cluster combined by different clusters; in turn, if α is too small, it will prevent the objects of one cluster to amalgamate one cluster. In the improved algorithm, through rule 3, make use of adaptive strategy to determine the value of parameter α, thus improving evaluation quality. Parameter α always keeps the adaptive change in the entire evaluation process, which makes the evaluation process more strong.

Rule 3: The initial value of parameter α is set to be constant 0.1. In the process of algorithm operation, for every N times of iteration, investigate the ratio r_f of the failed times N_f the ant putting down the objects and N; if $r_f \succ 0.99$, add 0.01 to α, otherwise, reduce 0.01 from α.

While BACA is finished, there are some data objects loitering not being designated to certain heap, appearing as certain isolated points and small heaps, including the data objects borne by the ant while finishing and some data objects forced to be put down in the process of iteration of certain algorithm. As for these two situations, there are two circumstances when the ant puts down the data objects; one is that the ant has no “short-term memory”; at this time, the ant places the objects in its position; the other is that the ant has the “short-term memory”; at this time, the ant places the objects in the best position in its memory. All these two circumstances cause the data to be put in wrong cluster or become isolated points, especially in the first circumstance. Through the experiment, we find that the number of the cluster with intensive objects obtained through ant colony evaluation is the same as that of the true cluster. Through setting threshold, we can screen out the wanted cluster. Treatment strategy for objects loitering is as follows.

Rule 4: First mark the cluster number of the results of ant colony algorithm, compare the number of objects of all clusters with threshold, obtain each cluster and respective data object according to comparison results, and then obtain the center of each cluster (average value of the cluster in data space). Move the objects loitering in the two-dimensional grid to the cluster with the minimum distance from the evaluation center; finally, re-mark the cluster number of the results of ant colony algorithm.

5. Experiment Confirmation

This paper chooses cousin event in Shaanxi (i.e. Director of Shaanxi Administration of Work Safety, Yang Dacai, smiling, wore luxuries and savings event) as study object. Data are chosen in the order to time. Setting the disclosure of smiling event dated August 26, 2012 as starting point, and the dismissal of Yang Dacai dated September 20, 2012 as terminating point, the event lasted 26 days. Here omits the specific calculation.

This paper realizes the improved algorithm and calculation results can be see in table 1, In view of the limited space, here only list the evaluation results of several time points of first-class indicator evaluation. While setting warning grades, this paper adopts traditional method, i.e. evaluation result lower than 0.5 is safe; value of evaluation result lying in $[0.5, 0.65)$ is slight warning; value of evaluation result lying in $[0.65, 0.8)$ is medium warning; value of evaluation result lying in $[0.8,1]$ is severe warning.

Table 1. Final Evaluation Results of Different Time Point

	Time 1	Time 3	Time 5	Time 7	Time 9
Final Evaluation Results	0.297	0.633	0.761	0.896	0.196
Warning Level	Safe	slight warning	medium warning	severe warning	Safe
Warning Signal	Green	Blue	Orange	Red	Green

As for the time consuming, calculation time needed of the improved model presented in the paper is 21 seconds and calculation time for the original BP neural and network is 521 seconds with the calculation platform as follows: hardware is Dell Poweredge R710, in which processor is E5506, memory 2G, hard disk 160G; software platform is Windows XP operating system, C programming language environment, so the model can be used in internet public opinions evaluation practically.

6. Conclusion

The evaluation indicator system and evaluation method of internet public opinions put forward in this paper can help the country or supervision departments of social public opinions better keep dynamic tracking of real-time information, able to focus on the public opinions of key themes and key regions, finding clues needing to be tracked from sources of information, as well as conductive to keep the continuity of dynamic tracking. As to the public opinions situation of specific theme, the overall condition of the event and the context of development change can be mastered through the evaluation method in this paper. Qualitative analysis and quantitative analysis have their own strengths; while designing the weight of evaluation indicator of internet public opinions, appropriate qualitative study shall carve out a way before the quantitative analysis; quantitative analysis can provide forceful argument and support for qualitative statement; reaching qualitative cognition based on quantitative analysis of information; sometimes, qualitative study is also used for explaining results obtained by quantitative analysis. Hence, the combination of qualitative analysis with quantitative analysis in this paper to set the weight of each evaluation indicator has favorable scientificity. Gathering and supervision of internet public opinions is a long-term and arduous study task, for which the study in this paper provides some beneficial idea. But such aspects as the automatic discovery of themes of internet public opinions and automatic evaluation public opinions information study shall be strengthened.

References

[1] Shouhua Zhang, Zhenpeng Liu, "Hotspot Detection Method of Internet Public Opinion based on Key Words Extraction", International Journal of Digital Content Technology and its Applications, Vol. 6, No. 16, pp. 340 ~ 347, 2012

[2] Zhao Hongyan, He Mingsheng, "Formation Path of Network Public Opinion Based on Collaborative Interaction Using Social Network Analysis", Advances in Information Sciences and Service Sciences, Vol. 4, No. 22, pp. 247 ~ 255, 2012

[3] Honglu Liu, Zhihong Tian, Dongpo Xiao, "Research on Driving Force Mechanism of the Network Public Opinion Waveform based on Wave Mechanics", Advances in Information Sciences and Service Sciences, Vol. 3, No. 6, pp. 278 ~ 284, 2011

[4] Cheng Xian-Yi, Zhu Ling-ling,Zhu Qian,Wang Jin, "The Framework of Network Public Opinion Monitoring and Analyzing System Based on Semantic Content Identification", Journal of Convergence Information Technology, Vol. 5, No. 10, pp. 48 ~ 55, 2010

[5] Sheng Li, Lingling Liu, Zenggang Xiong, "Ontology-Based Sentiment Analysis of Network Public Opinions", International Journal of Digital Content Technology and its Applications, Vol. 6, No. 23, pp. 371 ~ 380, 2012

[6] Smithras S T., "A Curve matching method based on Ant Colony ", Journal of Computer, Vol. 11,No.2, pp,134-141,2009.

[7] John S. K., Brooks K. H., "Research on the Methods of System", Journal of Computer Application, Vol.21,No.2,pp.197-204, 2011.

[8] David L.Post, Christopher S. Calhoun, "An evaluation of methods for producing CRT monitors", Color Research and Application, Vol. 14,No.4,pp.172-186,2007.