TELKOMNIKA, Vol. 11, No. 7, July 2013, pp. 3990 ~ 3996
e-ISSN: 2087-278X

A Novel Algorithm of Internet Public Opinion Evaluation

Gensheng Wang
Electronic Business department, Jiangxi University of Finance and Economics, Nanchang, 330013, China
e-mail: 397382377@qq.com

Abstract

With the speedy interaction and spread of the network information, it is great significance to collect emerging massive internet information and discover hotspots of network public opinion. The paper presents a new model for evaluating internet public opinions based on improved BP neural network. First, a new evaluation indicator system containing 4 hierarchies and 26 third-grade indicators is designed based on the characteristics analysis of internet public opinions. Second, a new internet public opinion evaluation model is presented using genetic algorithm to speed up the convergence and simplify the model structure of the model, and the calculation steps of the model are designed. 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, BP neural network, genetic algorithm, evaluation indicator system*

1. Introduction

Internet public opinions mean the opinions and viewpoints of net citizens on people and things in the society published on BBS, blog, SNS (virtual community) and all kinds of network media. Some of the net citizens make use of the anonymity of network to publish and spread inappropriate contents, causing negative effects to people's cultural life, even the economic construction of the country. Therefore, how to realize the correct guidance to internet public opinions and regulation control has become a focus of attention and research hotspot of all countries. At present, studies on internet public opinions at home and abroad are mainly starting from the perspectives of communication, sociology, education and politics, mainly involving internet public opinions and ideological and political education, guidance to the supervision and management of internet public opinions, analysis technology study on internet public opinions, communication influence of internet public opinions as well as emergency management strategies, and etc. However, there is lack of the indicator system of the gathering and analyzing mechanism of internet public opinions, and evaluation methods have different application disadvantages; hence, study on the safety evaluation system and evaluation method of internet public opinions is more of a hot issue for internet public opinions study [1].

2. Literature Review

Currently, models or methods used for crisis warning of internet public opinions at home and abroad are mainly quantitative analysis method and qualitative analysis method. Qualitative analysis methods mainly include viewpoints classification method, public opinions warning grades method and case analysis method. Viewpoints classification method, on the basis of statistics of net citizens' emotions, classifies net citizens' emotions into support, objection and neutrality through the digging and paraphrasing of key words, thus obtaining the warning grades of internet public opinions [2]. Public opinions warning grades method, according to the hotspots grades of internet public opinions, classifies internet public opinions into such four grades as slight, medium, severe and urgent crisis, with such four different corresponding colors as blue, yellow, orange and red, adopting different countermeasures in the case of different warning grades [3]. Case analysis method, through the analysis and summary of some typical cases, finds out the evolvement rule of internet public opinions, warning according to the rule. This kind of method is able to reflect the evaluators' subjective

Received February 9, 2013; Revised April 25, 2013; Accepted May 10, 2013

judgments, also easy in operation, but strong in subjectivity as to judgment basis, difficult to master in actual operation [4].

Qualitative analysis methods mainly include mathematical statistics method, analytic hierarchy process and data mining method based on clustering. The former two methods are simple in application theory, clear in algorithm structure and fast in computation speed, but they are strong in judgment subjectivity and low in judgment precision; evaluation accuracy is even difficult to guarantee for the warning evaluation of internet public opinions with lots of evaluation indicators. Data mining method based on clustering has such two models as fuzzy clustering method and neural network currently. Fuzzy evaluation method is the most frequently-used fuzzy clustering method, which can better reflect the dynamic nature, fuzziness and transitivity of evaluation indicators, and can better deal with quantitative and qualitative classification indicators. However, this method is sensitive to the choosing of initial clustering center of the cluster. In fact, once choosing the wrong clustering center, it is difficult to guarantee the clustering accuracy [5]. Neural network evaluation method provides a relatively effective and simple method to solve complicate problems. Neural network algorithm can easily solve the evaluation problem with hundreds of parameters. This method has such advantages as self-learning, strong fault tolerance and adaptivity, high evaluation accuracy, and etc. However, this method is easy to fall into the defects such as local minimum and over-learning, causing that calculation convergence problem of the algorithm must be solved in the specific application [6].

Internet public opinions evaluation models based on neural network have high accuracy but leaves behind the question of slow convergence speed of its algorithm. Therefore, it is hard to put it into effect in internet public opinions evaluation. Genetic algorithm is used to improve BP neural network in this paper. In so doing, not only the problem of convergence speed has been solved, but also the simplicity of the model structure and the accuracy of the transformation are ensured.

3. Indicator System Construction for Evaluating Internet Public Opinions

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, 8 second-grade indicators, 22 third-grade indicators; see Table 1 for details [2,3].

4. Evaluation Algorithm Design

4.1. BP Neural Network Structure

BP neural network is generally comprised of input layer, hidden layer and output layer, each layer connecting to the other, the node of each layer not connecting. The number of nodes of input layer generally adopts the dimension of input vector, and that of output layer generally adopts the dimension of output vector; there has no certain standard to obtain the number of nodes of hidden layer which shall be obtained through repeated cut-and-try methods. According to Kolmogorv law, three-layer BP neural network with one hidden layer (sufficient nodes of hidden layer) is able to approximate any nonlinear continuous function in any accuracy on a closed set. Therefore, this paper adopts BP neural network with single hidden layer to illustrate, topological structure as shown in Picture 1 [7].

Suppose that the input vector of X is $x \in R^n$, in which $x = (x_0, x_1, x_2 ..., x_{n-1})^T$; there are n_1 neurons in the hidden layer, the output of which is $x^{'} \in R^{n1}$, $x^{'} = (x^{'}_0, x^{'}_1, x_{^{'}2} ..., x^{'}_{n-1})^T$; there are m neurons in the output layer, output $y \in R^m$, $y = (y_0, y_1, y_2 ..., y_{n-1})^T$, the weight from input layer to hidden layer is $w_{i,j}$, threshold is θ_j; the weight from hidden layer to output layer is $w_{j,k}$, threshold is $\theta^{'}_k$; hence, output of neurons in each layer is as shown in Formula 1 [7].

Table 1. Indicator System of Internet Public Opinions Evaluation

Target Hierarchy	First-class Indicator	Second-class Indicator	Third-class Indicator
Evaluation of internet public opinions	theme popularity	explosive power of theme	category of information theme
		media conditions	website contents distribution
			regional distribution of net citizens
		media influence	website popularity
			website credibility
		publisher influence	attribution
			occupation attribute
			religious faith
	theme strength	publisher influence	education
			theme spreading
			theme sensitivity
			theme importance
		theme spreading degree	browse number
			times of report
			reply ratio
			republish ratio
	tendency of net citizens	tendency distribution of net citizens	distribution proportion of different tendency
			forward tendency proportion
			reversed tendency proportion
	theme time effect	theme timeliness	theme duration
			theme inflection point
			theme status

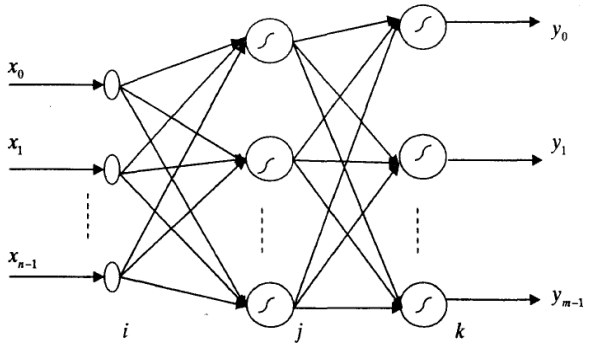

Figure 1. Topological structure of BP neural network used in the paper

$$\begin{cases} x_j^{'} = f(\sum_{i=0}^{n-1} w_{ij} x_i - \theta_j), & j = 0,1,2...,n_1 - 1 \\ y_k^{'} = f(\sum_{i=0}^{n1-1} w^{'}_{jk} x^{'}_j - \theta^{'}_k), & k = 0,1,2...,m-1 \end{cases} \quad (1)$$

Obviously, it will complete the mapping from n dimensional space vector to m dimension, in which activation function $f(x)$ is unipolar. Sigmoid function is as shown in Formula 2. $f(x)$ is continuous differentiable and meets Formula 3 [8].

$$f(x) = \frac{1}{1+e^{-x}} \quad (2)$$

$$f'(x) = f(x)(1 - f(x)) \quad (3)$$

4.2. Improving BP Algorithm with Genetic Algorithm

BP algorithm adopts gradient descent, poor in global searching ability, strong in local searching ability; while genetic algorithm is global optimization searching algorithm, strong in global searching ability, poor in local searching ability. In order to conquer the shortcomings of such two algorithms and develop the strength of such two algorithms, this paper combines such two algorithms to make their respective advantages complementary to each other. The basic concept is that first train the network with genetic algorithm to find a relative optimal solution; then take this network parameter as the initial parameter of network in BP algorithm to carry out the training, so as to enhance the classification ability of network.

Basic problems needing to be solved for genetic algorithm to improve BP algorithm are encoding of chromosome, formation of fitness function, design of selection operator, crossover operator and mutation operator, as well as the combination of best individual and BP algorithm.

4.2.1. Encoding of Chromosome

Learning process of BP neural network is the optimization learning process on such two continuous parameters as weight and threshold of network. If choosing wrong initial parameters, BP algorithm is easy to fall into local optimization. This paper adopts genetic algorithm to determine the initial parameter of BP network so as to avoid the defect of BP algorithm, which is easy to fall into local optimization. In the encoding process of chromosome, if binary encoding is adopted on parameters, the encoded string will be too long and shall be reverted to real numbers while decoding, thus influencing the learning accuracy of network and the running time of algorithm. Therefore, this paper adopts real number decoding, i.e. code string form is shown as formula 4 [9].

$$X = (w_{n1,1},...w_{n1,2},\theta_1,...,\theta_{n1},w_{m1}...w_{mn1},\theta_1^{'},...,\theta_m^{'}) \quad (4)$$

4.2.2. Formation of Fitness Function

Genetic algorithm basically makes no use of external information in the evolution search, only taking fitness function as reference, making use of the fitness value of each individual in the group to search and judging the excellence of individual with fitness value. Therefore, it is critical to choose fitness function, directly influencing the rate of convergence of genetic algorithm and whether able to find optimal solution. Generally, fitness function is transformed from objective function. This paper defines the network error as Formula 5.Error function is also the objective function in this paper. As the small the objective function value is, the larger the fitness value is, and the larger the objective function value is, the smaller the fitness value is, fitness function shall take the reciprocal of objective function, i.e. fitness function as shown in Formula 6.

$$E_A = \sum_{p=1}^{P} E^{(P)} = \frac{1}{2}\sum_{p=1}^{P}\sum_{k=0}^{m-1}(d_k^{(p)} - y_k^{(p)})^2 \quad (5)$$

$$F(E_A) = 1/E_A \quad (6)$$

4.2.3. Design of Selection Operator

Selection strategy adopts commonly-used proportion in genetic algorithm to choose operator, suppose that the group size is M, the fitness of individual i is F_i, then the probability P_i for individual i to be selected is Formula 7 [9].

$$P_i = F_i / \sum_{i=1}^{M} F_i \quad (i = 1,2,3,...,M) \tag{7}$$

4.2.4. Design of Crossover Operator

As real number encoding is adopted in this paper, crossover operator adopts arithmetic crossover strategy, suppose that there are two individuals X_A^t and X_B^t, carry out arithmetic crossover between them, and the generated two new individuals after crossover operation are expressed with Formula 8 and Formula 9 in which α is a parameter which can be both a constant and a variable decided by evolution generation [10].

$$X_A^{t+1} = \alpha X_B^t + (1-\alpha) X_A^t \tag{8}$$

$$X_B^{t+1} = \alpha X_A^t + (1-\alpha) X_B^t \tag{9}$$

4.2.5. Design of Mutation Operator

Mutation operator adopts uniform mutation strategy, suppose that there is an individual $X = x_1 x_2 ... x_k ... x_t$, if x_k is mutation point, the value range of which is $[U_{\min}^k, U_{\max}^k]$, after the point carries out mutation operation on the individual X, a new individual $X = x_1 x_2 ... x_k^{'} ... x_t$ can be obtained, in which the new gene value of mutation point is as shown in Formula 10, in which r is a random number meeting uniform probability distribution within the range of $[0,1]$.

$$x_k^{'} = U_{\min}^k + r(U_{\max}^k - U_{\min}^k) \tag{10}$$

4.2.6. Combination of Optimal Individual and BP Algorithm

After completing genetic algorithm training, find out the individual with the largest fitness value, decoding each component of the individual into corresponding parameter values, then train with BP algorithm until the algorithm meeting the termination condition [11].

4.2.7. Improved Algorithm Operation Process

Improved operation process can be listed as follows.

① Reduce dimension of samples with factor analysis, establish sample set;

② Calculate the fitness value of each individual in the group, save the optimal fitness value;

③ Turn to the 4th step if reaching the set evaluation generation or current optimal individual meeting conditions; otherwise, turn to the 2nd step after such genetic operations as selection, crossover and mutation;

④ Decode the optimal individual in the 3rd step into network parameter to serve as the initial parameter of BP neural network algorithm;

⑤ Modify current network parameter with BP neural network algorithm;

⑥ Terminate if reaching the condition for terminating BP algorithm; otherwise, turn to the 5th step.

5. Results and Analysis

5.1. Events and Data Choosing

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. Data sampling takes even numbers as time points, totaling into 13 sampling points of sequential sample, taking September 24 as the 14th sampling point. In the specific calculation, the value range of each indicator is among [0,1]; quantitative indicators shall be directly measured in specific value assignment of indicators; indicator weights of qualitative indicators shall be determined by questionnaire, expert consultation, literature reference, and etc. Here omits the specific calculation.

5.2. Empirical Results of Samples and Analysis

In view of the limited space, here only list the evaluation results of several time points of first-class indicator evaluation. Specific evaluation process sees Table 2 and Table 3. 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 2. Evaluation Results of Different Time Point of the First-grade Indicators

	Time 2	Time 5	Time 8	Time 11	Time 14
Theme Popularity	1.12	1.47	1.89	1.99	0.92
Theme Strength	1.32	1.68	1.98	2.21	1.13
Tendency of Net Citizens	1.19	1.66	1.87	2.18	1.08
Theme Time Effect	0.65	0.89	1.11	1.54	0.32

Table 3. Final Evaluation Results of Different Time Point

	Time 2	Time 5	Time 8	Time 11	Time 14
Final Evaluation Results	0.32	0.64	0.78	0.89	0.21
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 by the model presented in the paper is 19 seconds and calculation time for the original BP neural and network is 542 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.

6. Conclusion

This paper applies BP neural network algorithm to carry out deep study on internet public opinions warning evaluation; test result shows that the algorithm put forward in this paper not only gives play to the existing advantages of BP neural network method but also uses genetic algorithm to conquer its defects in the aspect of BP neural network, having favorable practicability. In the next step of study, first, scientificity of indicator choosing shall be further considered in the establishment of indicator system as well as the operability of each indicator while measuring. The calculation of each qualitative indicator weight shall possess more controllable regulation; second, automatic choosing of the hot theme of internet public opinions shall be also involved, instead of choosing theme according to evaluators' subjective judgment. Only through this can scientific methods be used on the premise of scientific choosing of internet warning sign indicator to carry out comprehensive evaluation on these indicators, for the sake of building scientific internet public opinions warning mechanism, so as to realize the correct guidance to internet public opinions.

Acknowledgements

This work is supported by the National Social Science Foundation of China under the grant No. 12CTQ042.

References

[1] Zhang SH, Liu P. Hotspot Detection Method of Internet Public Opinion based on Key Words Extraction. *Journal of Digital Content Technology and its Applications*. 2012; 6(16): 340-347.
[2] Zhao HY, He MS. Formation Path of Network Public Opinion Based on Collaborative Interaction Using Social Network Analysis. *Advances in Information Sciences and Service Sciences*. 2012; 4(22): 247-255.
[3] Lamsid H, Zhu LL. Research on Driving Force Mechanism of the Network Public Opinion Waveform based on Wave Mechanics. *Advances in Information Sciences and Service Sciences*. 2011; 3(6): 278-284.
[4] Chenis XL, Zenis HH. The Framework of Network Public Opinion Monitoring and Analyzing System Based on Semantic Content Identification. *Journal of Convergence Information Technology*. 2010; 5(10): 48-55.
[5] Sheng L, Tomas H. Fuzzy Analysis of Network Public Opinions. *International Journal of Digital Content Technology and its Applications*. 2012; 6(23): 371-380.
[6] Qian GL, Shu WH. A New Method for Monitoring Internet Opinion Based on Machine Learning. *Journal of Software*. 2008; 19(1): 845-850.
[7] Roy SK, Mark E. CRT Colorimetry Based on Improved BP Neural Network. *International Review on Computers and Software*. 2009; 38(4): 299-313.
[8] Guo MZ, Wang YD. Research of Curve Matching Method Based on BP Network. *Chinese Journal of Computers*. 2010; 33(8): 819-827.
[9] Post LP, Chiristopher C. An Evaluation of Methods for Producing CRT Monitors. *Color Research and Its Application*. 2007; 14(4): 172-186.
[10] Li JJ, Rui L.Construction Equipment Control Research Based on Predictive Technology. *TELKOMNIKA Indonesian Journal of Electrical Engineering*. 2012; 10(5): 960-967.
[11] Chen Xiaohui, A Rough Neural Network Algorithm for Multisensor Information Fusion. *TELKOMNIKA Indonesian Journal of Electrical Engineering*. 2013; 11(2): 747-753.