The Evaluation of Public Opinion in the Internet Using Optimized Neural Network and Genetic Algorithm

Fei Wang

Information Engineering College of Gansu University of Political Science and Law, China
wangfei6128@126.com

Abstract. Over the past era of time, with the bursting interaction and propagation of the Internet information, collecting emerging plenty of internet information and fining out the hot topic of network public opinion becomes a hotspot research branch. In the paper, we propose a novel model and prototype aiming at evaluating internet public opinions based on neural network (NN). Firstly, a novel evaluation indicator system is proposed and designed based on the characteristics analysis of internet public opinions. Later, we propose a novel evaluation model for the public opinion analysis, the detailed steps are discussed. Finally, we conduct experiment to test the robustness and effectiveness of our proposed methodology with detailed explanation.

Keywords: Internet Public Opinion; Neural Network; Genetic Algorithm; Evaluation Methodology

1 Introduction

Internet public opinions mean the opinions and viewpoints of net citizens on people and things in the society published on BBS, blog, 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].
The state-of-the-art Internet public opinion evaluation models based on NN (neural network) gain acceptable accuracy and correctness, however, most of the algorithms are time-consuming. Under this condition, we decide to optimize the traditional back-propagation neural network in this paper. 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 Figure1 illustrate the system.

2 Traditional Back-Propagation Neural Network

BP (back-propagation) neural network is generally composed 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 Kolmogorov 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 Figure2. Moreover, with the development of machine learning, deep neural network could also be applied to simulate for our proposed method [2].

Fig. 1. Indicator System

<table>
<thead>
<tr>
<th>Target Hierarchy</th>
<th>First-class Indicator</th>
<th>Second-class Indicator</th>
<th>Third-class Indicator</th>
</tr>
</thead>
<tbody>
<tr>
<td>topic power of theme</td>
<td>category of information theme</td>
<td>website contents distribution</td>
<td></td>
</tr>
<tr>
<td>theme popularity</td>
<td>media conditions</td>
<td>regional distribution of net citizens</td>
<td>website popularity</td>
</tr>
<tr>
<td>media influence</td>
<td>website popularity</td>
<td>website credibility</td>
<td>occupation attitude</td>
</tr>
<tr>
<td>publisher influence</td>
<td>religious faith</td>
<td>education</td>
<td>religious faith</td>
</tr>
<tr>
<td>theme strength</td>
<td>theme spreading</td>
<td>theme sensitivity</td>
<td>theme spreading</td>
</tr>
<tr>
<td>theme spreading degree</td>
<td>theme importance</td>
<td>theme sensitivity</td>
<td>theme importance</td>
</tr>
<tr>
<td>times of report</td>
<td>browse number</td>
<td>times of report</td>
<td>browse number</td>
</tr>
<tr>
<td>reputation ratio</td>
<td>distribution proportion of different</td>
<td>reputation ratio</td>
<td>distribution proportion of different</td>
</tr>
<tr>
<td>tendency of net citizens</td>
<td>tendency distribution of net citizens</td>
<td>tendency</td>
<td>tendency distribution of net citizens</td>
</tr>
<tr>
<td>theme time effect</td>
<td>theme timeliness</td>
<td>theme timeliness</td>
<td>theme timeliness</td>
</tr>
<tr>
<td>theme timeliness</td>
<td>theme states</td>
<td>theme states</td>
<td>theme states</td>
</tr>
</tbody>
</table>

Copyright © 2016 SERSC
Fig. 2. Explanation of the Structure

\[
\begin{align*}
  x_j &= f \left( \sum_{i=0}^{n} w_{ij} x_i - \theta_j \right), \quad j = 0, 1, 2, \ldots, n - 1 \\
  y_i &= f \left( \sum_{j=0}^{m} w_{ij} x_j - \theta_i \right), \quad k = 0, 1, 2, \ldots, m - 1
\end{align*}
\]

The propagation function is shown above, 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.

\[
f(x) = \frac{1}{e^{-x} + 1}
\]

\[
f'(x) = f(x)(1 - f(x))
\]

3 Proposed Genetic Based BP Neural Network Algorithm

**Chromosome Encoding Process.** 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. Therefore, this paper adopts real number decoding, i.e. code string form is shown as formula 4[3].

\[
X = \left( w_{n1,1}, \ldots, w_{n1,2}, \theta_1, \ldots, \theta_{n1}, w_{m1,1}, \ldots, w_{m1,2}, \theta_1', \ldots, \theta_m' \right)
\]

**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.

$$E_A = \sum_{p=1}^{P} E^{(p)} = \frac{1}{2} \sum_{p=1}^{P} \sum_{k=1}^{M} (d_k^{(p)} - y_k^{(p)})^2$$  \hspace{1cm} (5)

Selection Operator. By using the state-of-the-art proportion in genetic algorithm(GA), we settle down our selection strategy. Suppose 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 the following Formula [4].

$$P_i = \frac{F_i}{\sum_{i=1}^{M} F_i} \quad (i = 1, 2, 3, ..., M)$$ \hspace{1cm} (6)

Mutation 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 the following Formulas.

$$X_{A_i}^{t+1} = \alpha X_A^t + (1- \alpha) X_A^t$$  \hspace{1cm} (7)

$$X_{B_i}^{t+1} = \alpha X_A^t + (1- \alpha) X_B^t$$

Crossover Operator. Mutation operator adopts uniform mutation strategy, suppose that there is an individual $X = x_1 x_2 ... x_i ... x_n$, if $x_k$ is mutation point, the value range of which is $[U_{min}, U_{max}]$ after the point carries out mutation operation on the individual, a novel individual $X' = x_1' x_2' ... x_k' ... x_n$ could be captured. The following Formula shows this property.

$$x_k = U_{min} + r(U_{max} - U_{min})$$ \hspace{1cm} (8)

Detailed Steps for Proposed 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 [5]. The steps are: (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.

4 Experimental Result and Analysis

Data and Event Selection. This paper chooses cousin event in Shaanxi as study object. Data are chosen in the order to time. Setting the disclosure of smiling event dated August 26, 2014 as starting point, and the dismissal of Yang Dacai dated September 20, 2014 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.

Result 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 Figure 3 and Figure 4.

Fig. 3. Evaluation Results of Different Time

Fig. 4. Final Evaluation

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.
5 Conclusions and Summary

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 was supported by the Humanities and Social Science Research Project of the Ministry of Education (No. 14XJC840002), the Scientific Research Project of Colleges and Universities in Gansu Province (No. 2014B-064), Education Science Planning Projects of Gansu Province (No. GHBZ019) and Project Teaching Reform of Gansu Institute of Political Science and Law (No. GJG2014B01).

References