Journal of Applied Sciences

ISSN 1812-5654

science alert

ANSInet
an open access publisher
http://ansinet.com

Journal of Applied Sciences 13 (19): 4062-4066, 2013
ISSN 1812-5654 / DOI: 10.3923/jas.2013.4062.4066

Improvement of BP Neural Network and its Application in Internet Public Opinion Evaluation

Gen-Sheng Wang
Department of Electronic Business, Jiangxi University of Finance and Economics,
330013, Nanchang, Jiangxi, China

Abstract: BP neural network is a hot research field in various disciplines for its powerful simulation calculation ability in recent years, but the algorithm has some shortages such as low convergence which limited the usage of the algorithm. Based on the genetic algorithm, the study improves the BP neural network to speed up its calculation. First, structure and shortages of BP neural network are analyzed for subsequent improvement; Second, the BP neural network algorithm is improved through parameter setting, initialization and evaluation, selection operator design, crossover and mutation operator design and algorithm calculation process design. Third, the improved algorithm is realized and the experimental results show that the improved BP algorithm has better performance in iterations and expectation fitting and can be used in internet public opinion evaluation practically.

Key words: BP neural network, genetic algorithm, internet public opinion evaluation, expectation fitting

INTRODUCTION

Up to now, people have put forward hundreds of neural network models, not to mention the numerous learning algorithms; however, from the perspective of the application of neural network, there are only dozens studied the most, among which the network with the most extensive and representative study is BP (Back propagation) neural network (Chen *et al*., 2011).

Studies on the improvement of the theory of BP neural network are comprised of the following five aspects: (1) Improve the gradient of excitation function, (2) Improve error curved surface, (3) Selection of initial weight value of network, (4) Improving and optimizing algorithm, (5) Optimizing network structure; This study, starting from the problem that BP neural network is complicated in network structure and poor in classification capacity for the classification of complex samples, improves BP algorithm with genetic algorithm, so as to enhance the convergence speed of BP neural network towards, forming a kind of BP neural network algorithm with strong system evaluation capacity.

MATERIALS AND METHODS

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 (Qian and Chen, 2010; Zhu and Jiang, 2012).

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_l 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 θ'_k; hence, output of neurons in each layer is as shown in Eq. 1 (El Kadhi *et al*., 2012):

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

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 Eq. 2. F (x) is continuous differentiable and meets Eq. 3 (Wang *et al*., 2012; Shih and Hu, 2008):

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

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

Improvement of BP neural network algorithm: From the perspective of the characteristics of neural network BP and genetic algorithm, the training of BP algorithm is based on the weight modification principle of error gradient descending, inevitably having the problem of falling into local minimum point; genetic algorithm is good at global search, while insufficient in local accurate search. Therefore, the combination of making use of genetic algorithm to optimize the initial weight and threshold of neural network with using neural network algorithm to finally complete network training realizes the complement of advantages, beneficial for better solving practical problems.

An Ann (Artificial Neural Networks, ANN) model can be described by the connecting method of finite parameters such as neuron, network layers, neuron number of each layer and neuron, weight of each connection and transfer function. So, we can encode an ANN model and realize the learning process of neural network with genetic algorithm.

- **Parameter setting:** Input population size P, network layers (not including input layer), neuron number of each layer. Genetic algorithm has excellent robustness towards the setting of these parameters; changing these parameters won't exert great impact on obtained results
- **initialization and evaluation:** ① Randomly generate initial population $P = (x_1, x_2,..., x_n)$, any $x_i \in P$ being a neural network weight, which is comprised of a weight vector and a threshold vector, weight vector being n-dimensional real vector, n being the number of all the connection weights, threshold vector also being n-dimensional real vector (not including neuron of input layer). Each network weight x_i is equal to a chromosome; there are N such chromosomes, i.e., population size. The neurons are numbered from the bottom to the top, from the left to the right (including input neuron) (Wang *et al*., 2012; Gao *et al*., 2010).

  ②According to corresponding neural network of randomly generated weight vector and threshold vector, as for the given input set and output set, calculate the global error of each neural network, as genetic algorithm can only evolve towards the direction of increasing fitness. So, the fitness function can be formed according to Eq. 4 and 5, among which f_i is the adaptive value of the ith individual, i = 1, 2, ..., N being the number of chromosome, k = 1, 2, ..., n being the number of nodes of output layer, p = 1, 2, ,..., m being the number of learning samples, v_{pk} being the output value of the kth node while inputting the pth training sample, T_{pk} being the anticipated output value:

$$f_i = \frac{1}{E_i} \quad (4)$$

$$E_i = \sum_{p=1}^{m}\sum_{k=1}^{n}(v_{pk} - T_{pk}) \quad (5)$$

- **selection operator:** This study adopts the mass selection operator combining spinning roulette wheel strategy with optimal retention strategy. Selecting process takes the spinning roulette wheel as basis, which is a kind of playback random sampling method. All the selections are to select good individual according to individual fitness from current population in the light of certain criterion to enter the next generation population, the basic ideal of which is that the selective probability of each individual is equal to the ratio of its fitness to the individual fitness among the entire population. The higher the individual fitness is, the greater the possibility to be selected is and the greater the probability to enter next generation is. However, due to random operation, the selection error of this method is relatively big, sometimes even making the individual with high fitness be selected. In order to improve the convergence of genetic algorithm, this study adopts optimal retention strategy, selecting individual with the largest fitness as seeded player, directly retaining to the next generation. Substitute the worst individual in the population with the optimal individual recorded by the preceding generation while forming new population every time, so as to preventing the individual with optimal fitness in current population from being destroyed
- **crossover and mutation operator:** ① Improved adaptive crossover probability and mutation probability

  In the parameters of genetic algorithm, the selection of crossover probability P_c and mutation probability P_m is the key to influence the behaviour and performance of genetic algorithm, exerting a direct impact on the convergence of algorithm. In the simple genetic algorithm, as the values of P_c and P_m are constant, it is not efficient enough to solve multivariable complication optimization problems, having the problems of prematurity or misconvergence. Srinivas and etc., put forward adaptive genetic algorithm, AGA, the basic idea of

which is that the individual with fitness higher than average fitness in the population adopts the smaller crossover probability P_c and mutation probability P_m, aiming at retaining individual with favorable structure so as not to be destroyed and to enter the next generation; as for individual with fitness lower than average fitness, using higher crossover probability and mutation probability to facilitate the elimination of such individual. Although, this method is improved compared with simple genetic algorithm, there are still some problems. For example, while the fitness is close to the largest fitness, the crossover probability and mutation probability are; while equal to the largest fitness, the crossover probability and mutation probability are zero, which makes AGA undesirable in the early stage of evolution. As in the population of early stage of evolution, more optimal individuals are in an unchangeable state and the favorable individual at this time is not always the globally optimal solution, which is easy to make the evolution tend to be locally converged . Hence, this thesis, based on this, adopts improved adaptive algorithm, making the individual crossover probability and mutation probability of largest fitness in the population be not zero, as shown in Eq. 6 and 7, in which f_{avg} represents the average fitness of population of each population; f_{max} represents the largest fitness in the population; f' represents larger fitness of two individuals to be crossed over; f represents the fitness of individual to be mutated in the population. P_{c1}, p_{c2} p_{m1} and p_{m2} are design parameters, which are 0.9, 0.6, 0.1, 0.001, respectively:

$$P_c = \begin{cases} P_{c1} - \dfrac{(p_{c1} - p_{c2})(f' - f_{max})}{f_{max} - f_{avg}}, & f' \geq f_{avg} \\ p_{c1} & f \leq f_{avg} \end{cases} \quad (6)$$

$$P_m = \begin{cases} P_{m1} - \dfrac{(p_{m1} - p_{m2})(f' - f_{max})}{f_{max} - f_{avg}}, & f' \geq f_{avg} \\ p_{m1} & f \leq f_{avg} \end{cases} \quad (7)$$

Improved AGA not only keeps the adaptive advantage of AGA but also conquers the shortage of slow evolution of population in the early stage, having favorable optimization function

② Crossover operator. First, in the population, according to the crossover probability P_c obtained in ①, randomly select certain quantity of chromosomes as parents and randomly select a breakpoint, exchanging the gene strand on the right (or top) of the breakpoints of parents, generating new filial generation; finally, substitute the paternal chromosome with filial generation chromosome, generating new population

③Mutation, similar to the selection of paternal generation in crossover process, as for each selected chromosome to be mutated, in order to get better mutation, multiple mutation is permitted. While mutating, first randomly generate a vector with the same dimension as each weight and threshold of chromosome and add to the selected vector to be mutated. As to the result of each mutation, restore neural network and carry out performance evaluation. If the descendant is better than paternal generation, the mutation of paternal generation shall be ended; otherwise, carry out next mutation on paternal generation, until finding out descendant better than paternal generation

- **Immigration operator:** It is found through the test that in the search process of genetic algorithm, the individual with highest fitness in the population at present is possible to participate in crossover and mutation calculation, just with small probability; on the contrary, the lower the fitness of the individual is, the larger the probability to be selected to participate in crossover and mutation is, but the generated individual fitness is very low and the global search performance on algorithm is not obviously increased. Therefore, this thesis introduces immigration operator which is a good method to avoid prematurity. In the immigration process, it can only accelerate the elimination of bad individual, but also increase the diversity of solution, further meeting the evolutionary mechanism of creatures. Immigration operator eliminates the worst individual with certain elimination rate (generally 15~20%) in the evolutionary process of each generation and generates part of excellent immigrants to supplement the population. Excellent immigrants here are mutated and generated through the multiple crossovers on those individuals to be eliminated. Thus, not only fully retain the good gene genetic pattern of paternal generation but also guarantee the diversity of population, improving the optimization searching performance of GA
- **End of operation:** If the network error meets the requirement or reaches certain evolution generations, the evolution shall be stopped and the evolution result shall be outputted; otherwise, turn to Step (3)

Process of improved algorithm: The process of the improved algorithm can be listed as follows. ① Initial population, including the population size and the initialization of each weight (generate according to the method for neural network to generate initial weight) and encode it, ② Calculate the selection probability of each

individual and sort them, ③ Select good individual to enter next generation population according to spinning roulette wheel selection strategy, ④ In the new generation population, select adaptive individual to carry out crossover and mutation operation according to adaptive crossover probability and mutation probability to generate new individual, ⑤ Insert the new individual into the population and calculate the fitness of new individual, ⑥ Immigration operator operation. Judge whether there is "prematurity phenomenon", if there is, immigration strategy shall be adopted and turn to step ②, ⑦ If the satisfactory individual is found, it shall be ended; otherwise, turn to step ② (Cui and Xu, 2012; Lv, 2012).

RESULTS AND DISCUSSION

Indicator system design for internet public opinions evaluation: Based on the deep analysis of the characteristics of internet public opinions, referring to the studied literature, this study 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. 4 first-grade indicators are theme popularity, theme strength, tendency of net citizens, theme time effect. Popularity for example includes for second-grade indicators, that are explosive power of theme, media conditions (which contains three third-grade indicators, ie website contents distribution, regional distribution of net citizens, website popularity), media influence (which contains 2 third-grade indicators, ie website credibility, attribution), publisher influence (which contains three third-grade indicators, ie occupation attribute, religious faith, education). Limited the space of the study, he analysis of the other three first-grade indicators are omited here.

Experiment results and analysis: This study 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, 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.

Table 1: Evaluation results of different time point of the first-grade indicators

Parameters	Time 1	Time 3	Time 5	Time 7	Time 13
Theme popularity	1.061	1.403	1.901	2.001	0.919
Theme strength	1.261	1.704	1.997	2.312	1.138
Tendency of net citizens	1.172	1.702	1.904	2.254	1.106
Theme time effect	0.638	0.904	1.210	1.617	0.347

Table 2: Final evaluation results of different time point

Parameters	Time 1	Time 3	Time 5	Time 7	Time 13
Final evaluation results	0.309	0.656	0.789	0.921	0.241
Warning level	Safe	Slight warning	Medium warning	Severe warning	Safe
Warning signal	Green	Blue	Orange	Red	Green

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 1 and 2. While setting warning grades, this study 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.

As for the time consuming, calculation time needed by the model presented in the study is 16 sec and calculation time for the original BP neural and network is 542 sec 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.

CONCLUSION

Both of the analysis of simulation performance in section 5 of the study and experimental results in section 6.2 in the study can show that the improved algorithm in the study has more advantages in iterations and expectation fitting curve than original algorithm and can be used in practice.

ACKNOWLEDGMENT

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

REFENENCES

Chen, Q., K. Huang, X. Yang, M. Luo and H. Zhu, 2011. A BP neural network realization in the measurement of material permittivity. J. Software, 6: 1089-1095.

Cui, S.Y. and A.F. Xu, 2012. A study on performance evaluation model of public oopinion. J. Conv. Inform. Technol., 7: 238-245.

El Kadhi, N., K. Hadjar and N. El Zant, 2012. A mobile agents and artificial neural networks for intrusion detection. J. Software, 7: 156-160.

Gao, S., Z. Zhang and C. Cao, 2010. A novel ant colony genetic hybrid algorithm. J. Software, 5: 1179-1186.

Lv, H.M., 2012. Research on the evaluation methods on internet public opinion with intuitionistic fuzzy information. J. Digital Content. Technol. Appl., 6: 60-68.

Qian, G.L. and B. Chen, 2010. A color matching method based on machine learning. Chinese J. Inst., 9: 529-551.

Shih, Y.Y. and J.S. Hu, 2008. Fuzzy quality attributes for evaluating internet marketing system performance. Total Qual. Mana. Bus. Excellence, 19: 1219-1234.

Wang, J., Y. Gu and X. Li, 2012. Multi-robot task allocation based on ant colony algorithm. J. Comput., 7: 2160-2167.

Zhu, C.J. and E.H. Jiang, 2012. A study on roughness coefficient based on improved BP neural network. J. Math. Appl., 5: 166-174.