6. MeinshausenN.,BuhlmannP.(2006) High-dimensional Graphs and Variable Selection with the Lasso.*The Annals of Statistics*, 27(3), p.p.169-185.
7. TibshiraniR.(1996)Regression Shrinkage and Selection via the Lasso.*Journal of the RoyalStatistical Society*, 58(1), p.p.267-288.
8. WangH.,LiR.,Tsai,C.(2007)Tuning Parameter Selectors for the Smoothly Clipped Absolute Deviation Method, *Biometrika*, 94(6), p.p. 553-568.

A Method of Optimizing the Feedforward BP Neural Network Based on Elitist Ant Colony Optimization

Huikang Tang

Information Engineer Department, Henan Vocational and Technical Institute, Zhengzhou 450046, Henan, China

Abstract
Artificial neural network and the intelligent optimization algorithm are hotpots and the cutting-edge ones of the current information science technology, which are of great theoretical and application significance for the fields of pattern classification and identification as well as prediction, etc. This paper put forward a learning method, which is EACONET (Elitist Ant Colony Optimization NET), to optimize the feedforward neural network based on the elitist strategy ant colony optimization, targeting at the weight optimization of the feedforward neural network, in order to solve problems like prematurity and slow rate of convergence of the ant colony optimization in training the neural network. This method combines the global parallel search with the local certainty of BP network, to search for optimal points. At the same time, it can enhance the rate of convergence, avoid the local extremum, can be applied to the functional approximation and the nonlinear system identification, etc. The results of simulation experiment show the effectiveness of the proposed algorithm.
Key words: BP NEURAL NETWORK, ANT COLONY OPTIMIZATION, ELITIST STRATEGY

1. Introduction

Artificial neural network (ANN) is a calculation model that simulates how human brain works based on basic principles of the neural networks in biology. It is a complex network system formed by the broad interconnections among a large amount of neuron units, which reflects a number of basic features of the functions of the human brain, also it is a highly complicated non-linear dynamic system. ANN has a strong non-linear mapping ability and functions like self-organizing, self-learning and strong fault tolerance, etc[1].

Through the mutually corresponding input and output data provided beforehand, ANN is able to learn the underlying rules between them, learn, analyze and infer the output results by using the new input data based on these rules[2]. W. S. Mcculloch, a psychologist and W.Pitts, a mathematical logician, proposed an M-P Model, which was the first model using mathematical language to describe the process of information processing of the brain. Then, D.O.Hebb, a psychologist, prosoped the hypothesis of variable synaptic connections. Rosenblatt, a computer scientist, prosoped the notorious perceptron model, which was the first intact artificial neural network[3]. John. J. Hopfield, an American physicist, summarized various structures and network algorithms, built a Hopfield network. Rumelhart and Cun and other scholars prosoped the back-propagation algorithm of the multi-layer perception. BP network, which is short for back-propagation, is a multi-layer feed-forward network trained by error back-propagation algorithm. The traditional BP training algorithm has deficiencies like slow rate of convergence and easiness to involve in the local minimum, which have hindered the application of neural networks in various practical fields, at present there are many researchers have applied the intelligent search algorithm into the training of neural networks[4]. The ant colony algorithm is a heuristic bionic evolutionary system based on the population, which has adopted the distributed parallel computing mechanism with the positive feedback. It is easy to combine with other methods, and with such strong robustness it can be used to train the parameters and structures of neural networks[5].

Firstly this paper analyze the working principles of models of artificial neurons and the BP neural network, then it illustrate the rationale of the ant colony optimization and study the implementation method of the elitist strategy ant colony optimization, in addition on the basis of the above-mentioned research, this paper put forwards a learning method to optimize the feed-forward neural network based on the elitist strategy ant colony optimization. The last part of this paper is the experiment simulation and results analysis.

2. The Back Propagation (BP) Neural Network

2.1. The Model of Artificial Neurons

There is more than one input and generally multiple inputs for a neuron. The model of artificial neuron has simulated the three most basic and significant functions of biological neurons, which are weighing, summation and transferring. For each artificial neuron, it can accept a set of input signals from other neurons of the system, and each input corresponds to one weight and the weighing of all the input decides the activated state of this neuron. Here each weight means the synaptic connection strength $P_i(i=1,2,\ldots,n)$, the n amount of input components of i, the artificial neuron, connects with $W_i(i=1,2,\ldots,n)$,which is the weight component that multiplies it, to form another input to activate the function W_i, the weight, and the matrix form of the input component can be explained through the row vector of W, which is $W=[W_1W_2 \ldots W_n]$, and the column vector of P, which is $P=[p_1p_2\ldots p_n]$. ε is the neuron deviation[6]. The activation function is as shown in formula (1) and the universal model is as shown in Fig. 1.

$$f(net_i)=\sum_{i=1}^{n} w_i p_i+\varepsilon \tag{1}$$

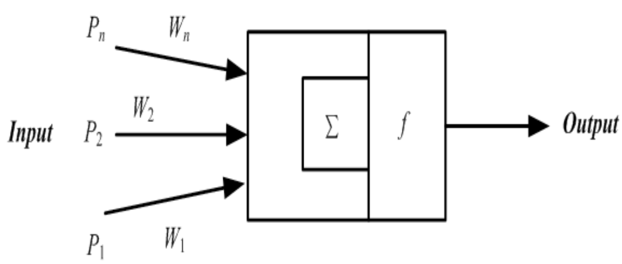


Figure 1. The universal model of artificial neuron

2.2. The BP Network

The Back Propagation (BP) neural network is an artificial neural network with supervised feed-forward operation, which can learn and save a large amount of mapping relations between input and output without the need of revealing the mathematical formula beforehand to describe this mapping relation. It is consisted of the input layer, the hidden layer, the output layer and the connection weights on the nodes of various layers, and the algorithm of this training process is consisted of the forward propagation of information and the back propagation of errors. The neuron of each layer of the network only contains the output signals from the previous input layer, and the output signals of the neuron of the output layer of the network has consisted all the response that generates the motivation model at the originating modes of the input layer, which means the signal is input at the input layer, transmitted to the output layer through the hidden layer and the output layer gets the output signal. Its' training rule is to use the steepest descent method to minimize the error sum of squares by constantly adjusting the network weights and threshold values through the back propagation[7]. The structure diagram of BP neural network is as shown in Fig. 2.

The BP network can learn and save a large amount of mapping relations between the input and output modes without the need of revealing the mathematical formula beforehand to describe this mapping relation.

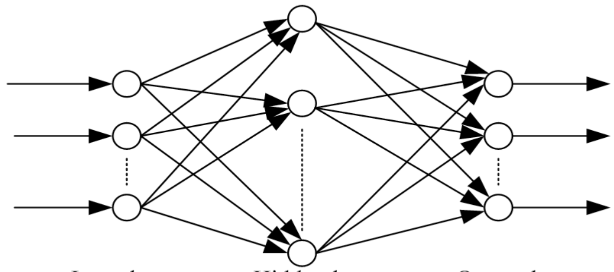


Figure 2. Structure diagram of the BP network

The BP network is trained according to this training method, which is that the input signal is transmitted to the output layer from the input layer through the hidden layer and generates the output signal at the output end, which is the forward propagation of the working signal[8]. The input information is processed at one layer by another one layer from the input layer through the hidden layer and is transmitted to the output layer, of which the neuron of each layer only influences the output of the neuron of the next layer. If there is no anticipated output at the output layer, it turns to the back propagation, which starts from the output layer through the hidden layer and gets back to the input layer with revising individual connection weight at various layers based on the principle of reducing the error between the anticipated output and the actual output. During the back propagation process of the error signal, the network weight is adjusted by the error feedback, of which the process is conducted from the output layer to the hidden layer and to the input layer. With the continuous training on the back propagation of errors, the response accuracy of the network to the input model will be enhanced continually[9].

3. Mathematical Model of TSP and Ant Colony Algorithm

3.1. Mathematical Model of TSP

Firstly the artificial ant colony algorithm was presented to solve the traveling salesman problems, which is short for TSP. TSP is one of the most typical problems in the mathematics field. It can be described as that there is a traveling salesman who needs to visit n cities, and it is restricted that he has to visit each of the cities and gets back to the city where he starts, in addition it requires that the path length acquired eventually must be the minimum value among all the length of paths. According to the ant colony algorithm model, the TSP of the n cities is to find the shortest path visiting n cities and getting back to the starting city. Assuming there are n cities, m ants with $d_{ij}(i,j=1,2,\ldots,n)$ as the length between city i and city j and $\tau_{ij}(t)$ as the pheromone intensity on the path between city j and city j at the moment of t. While moving, Ant K decides its next path based on the pheromone intensity left on each path, and $p_{ij}^{k}(t)$ means the probability of Ant K moving from city i and city j at the moment of t, and there is:

$$p_{ij}^{k}(t)=\begin{cases}\dfrac{\tau_{ij}^{\alpha}(t)\eta_{ij}^{\beta}(t)}{\sum_{F\in allowed_k}\tau_{ij}^{\alpha}(t)\eta_{ij}^{\beta}(t)}, & j\in allowed_k\\ 0, & otherwise\end{cases} \tag{2}$$

When the circulation is completed, it releases pheromone with corresponding intensity according to the length of the whole path and updates the pheromone intensity it has reached.

$Allowed_k$ is the set of next cities allowed for Ant K to visit and it changes dynamically with the moving process of Ant K.

$j_k(i)(i=1,2,\ldots,n)$ means the set of cities allowed for Ant K to visit, namely the list of is the $tabu$ list of Ant K. When all the cities are listed in $tabu$, Ant K has visited all the cities and when it gets back to the starting city, it finishes its journey, and all the path it has been through is a feasible solution to this TSP. As for α, it means the significance of the accumulated pheromone intensity during the traveling on the path choice made by Ant. η_{ij} means the expectation level for Ant K to move from city i and city j and it can be decided by certain heuristic algorithm, for example it can be set as $\eta_{ij}=1/d_{ij}$. When all ants have finished their visits, pheromone on various paths will be updated entirely based on Formula 3.

$$\tau_{ij}(t+1)=(1-\rho)*\tau_{ij}(t)+\Delta\tau_{ij}$$

$$\Delta\tau_{ij}=\sum_{k=1}^{m}\Delta\tau_{ij}^{k} \tag{3}$$

Among which, $\rho(0.5\le\rho\le0.99)$ means the volatilization coefficient, means the persistent coefficient, $\Delta\tau_{ij}$ means the variation of the pheromone intensity on the paths, and $\Delta\tau_{ij}^{k}$ means the pheromone intensity Ant K left between city i and city j in this circulation, of which the computing method is decided by the calculation model[10,11].

3.2. Ant Colony Algorithm of Elitist Strategies

Although there are some improvement measures like enhancing the population size and adjusting coefficients dynamically to improve the optimal performance, yet some of the inherent problems of the ant colony algorithm are not solved. The idea of the elitist strategy is to preserve the most adaptable individual from the last generation, intensify the influence of the ant that finds the path that is better than the current optimal path on the pheromone intensity, and lower the influence of the ant that finds the path that is worse than the current optimal path on the pheromone intensity. After each circulation, all optimal solutions will be given extra intensified pheromone intensity in order to make the current optimal solutions to be more attractive for the ants in the next circulation. After each iteration, to enhance the employment on the local optimal solutions when updating the pheromone. And the pheromone will be updated according to the following formula.

$$\tau_{ij}(t+1) = \rho \cdot \tau_{ij}(t) + \Delta\tau_{ij} + \mu_{ij} \tag{4}$$

In which,

$$\mu_{ij} = \begin{cases} \omega \dfrac{Q}{L} & \text{if line } (i, j) \text{ is a part of founded optimal solution} \\ 0 & \text{Otherwise} \end{cases}$$

μ_{ij} means the increase on the pheromone intensity on the path between (i, j) caused by the elitist ants, ω means the amount of elitist ants and L means the length of path of the optimal solution[12].

4. The method to optimize the BP network weight based on the elitist ant colony optimization

The following is to illustrate the implementation of the algorithm in this paper.

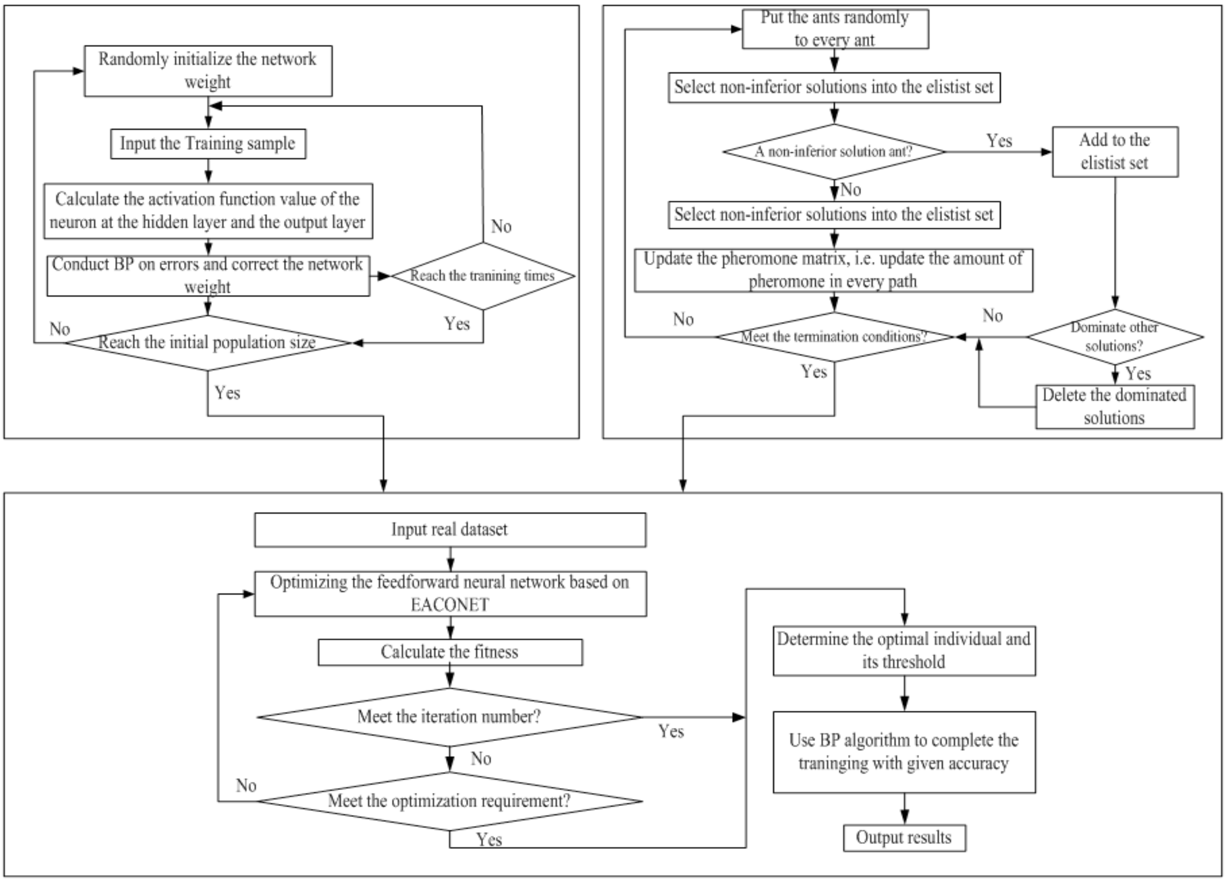


Figure 3. Structure flowchart of EACONET

(1) Variables and parameters setting

$X_k = [x_{k1}, x_{k2}, \cdots x_{kM}], (k = 1, 2, \cdots, N)$ is the input vector or the training sample, and N is the amount of the training samples. First to initialize and settle the structure parameters of the network and then randomly initialize parameters with relatively small weight.

(2) To conduct X_k on the input sample and to conduct feed-forward calculation on the input signal u and output signal v of the neuron of each network layer. Calculate on unit j of the l layer. The activation function of the neuron adopts the double Sigmoid function, and

$$y_j^l(n) = 1 \Big/ 1 + \exp(\sum w_{ji}^l(n) y_i^{k-1}(n)) \tag{5}$$

(3) To calculate the output unit, to correct various weights

$$w_{ij}^l(n+1) = w_{ij}^l(n) + \tau_{ij}(n) y_i^{l-1}(n) \tag{6}$$

(4) For the input sample, conduct reverse calculation on the local gradient of the neuron at each layer

until expectation E meets the requirement, O_j is j unit output.

$$E = \sum_{j=1}^{N} (x_j - O_j)^2 / 2N \tag{7}$$

(5) To input the actual testing data and set various parameters of the algorithm. To uniformly and randomly generate the original ant colony, with the population size as N, and calculate the objective function value $f_i(x), i = 1, 2, \ldots N$, of each ant.

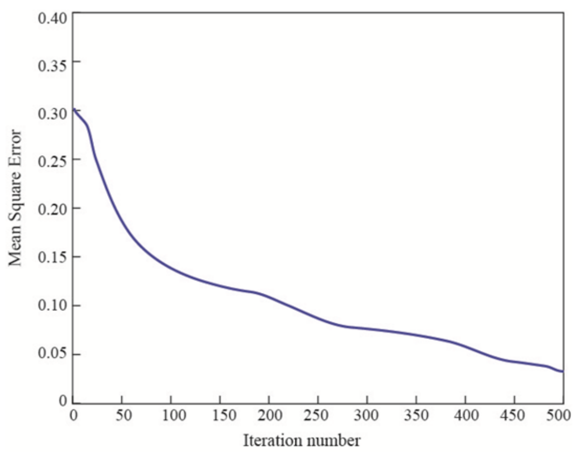
(6) To move Ant i within in the range of ants activities and to reevaluate Ant i and by adopting Formula (2), (3) and (4) to calculate its objective function value and threshold. To update the elitist set and if Ant i is a feasible solution and a non-dominated solution for the BP, then add Ant i into the elitist set .

The structure flowchart of EACONET algorithm as shown in Figure 3.

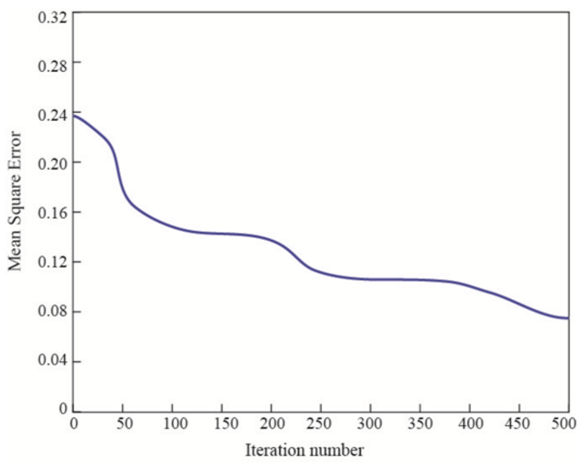
5. Experiment Simulation Test and Analysis

Respectively curve simulation and identification have been tested in order to evaluate the performance of training algorithm EACONET, on optimizing the weight of the feedforward neural network and all the data used comes from the database for studying the neural network. The algorithm tests are conducted on the MATLAB and the parameters of the ant colony $N = 30, \rho = 0.7, \alpha = 2, \beta = 3, Q = 200$.

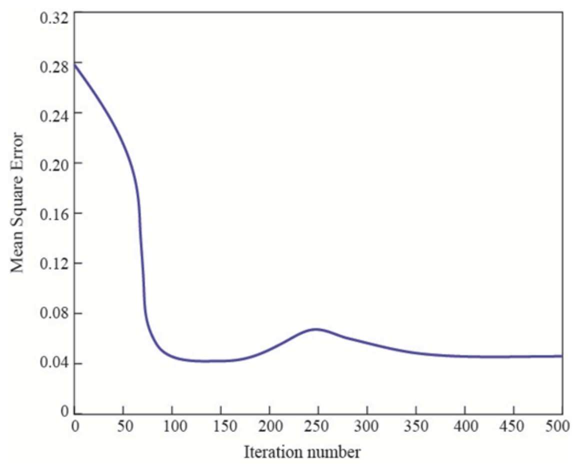
The first data set has 816 groups of data samples, which are consisted of 8 properties and 2 classifications. It is an issue of binary classification, thus to explain the generalization ability of the trained optimal network. The second data set has 553 groups of data samples, which are consisted of 7 properties and 2 classifications. The third data set has 340 groups of data samples, which are consisted of 11 properties and 2 classifications. Fig.4 is the convergence curves of various issues in the training of the algorithm of EACONET.



(a) Test 1-convergence curve



(b) Test 2-convergence curve



(c) Test 3-convergence curve

Figure 4. Training convergence curves

The above curves have reflected the convergent tendency and efficiency of training various issues with the algorithm of EACONET. Table 1 has provided the minimum value, the maximum value and the standard deviation to test the accuracy of the classification and identification of the network as follows.

Table 1. Test results of the network training

Data set	Standard deviation	Average accuracy
Set 1	2.74e-2	80.15%
Set 2	2.86e-2	71.53%
Set 3	9.12e-3	93.76%

From the above, it can be observed that the optimal performance of EACONET is better. In terms of the three training datasets, the training results with EACONET have relatively low mean value of error of mean square, higher accuracy on the average identification of the three testing datasets, and the trained network has better propagation ability.

6. Conclusion

On the basis of in-depth researches on the BP neural network and principles of the ant colony optimization, this paper has analyzed the merits and demerits of both of them, and has adopted the EACONET to optimize the weight of the feedforward neural network to promote the network weight, reach the best performance through the EACONET, thus to enhance the stability of the model of the neural network.

References

1. Mona Mahrous Mohammed, Amr Badr, M.B. Abdelhalim (2015) Image Classification and Retrieval Using Optimized Pulse-Coupled Neural Network. *Expert Systems with Applications*, 42(11), p.p.4927-4936.
2. Roberto Vega, Gildardo Sanchez-Ante, Luis E. Falcon-Morales, Humberto Sossa, Elizabeth Guevara. (2015) Retinal Vessel Extraction Using Lattice Neural Networks with Dendritic Processing. *Computers in Biology and Medicine*, 58(1), p.p.20-30.
3. Adam P. Piotrowski (2014) Differential Evolution Algorithms Applied to Neural Network Training Suffer from Stagnation. *Applied Soft Computing*, 21(8), p.p.382-406.
4. D. Jude Hemanth, C.Kezi Selva Vijila, A.Immanuel Selvakumar, J. Anitha (2014) Performance Improved Iteration-Free Artificial Neural Networks for Abnormal Magnetic Resonance Brain Image Classification. *Neurocomputing*, 130(23), p.p.98-107.
5. Mina Jafari, Hassan Khotanlou (2013) A Routing Algorithm Based an Ant Colony Local Search and Fuzzy Inference to Improve Energy Consumption in Wireless Sensor Networks. *International Journal of Electrical and Computer Engineering (IJECE)*, 3(5), p.p.640-650.
6. Bruno J.T. Fernandes, George D.C. Cavalcanti, Tsang I. Ren (2013) Auto Associative Pyramidal Neural Network for One Class Pattern Classification with Implicit Feature Extraction. *Expert Systems with Applications*, 40(18), p.p.7258-7266.
7. Anna Aprile, Giovanna Castellano, Giacomo Eramo (2014) Combining Image Analysis and Modular Neural Networks for Classification of Mineral Inclusions and Pores in Archaeological Potsherds. *Journal of Archaeological Science*, 50(10), p.p.262-272.
8. M. Saritha, K. Paul Joseph, Abraham T. Mathew (2013) Classification of MRI Brain Images Using Combined Wavelet Entropy Based Spider Web Plots and Probabilistic Neural Network. *Pattern Recognition Letters*, 34(16), p.p.2151-2156.
9. Saeed Soltanali, Rouein Halladj, Shokoufe Tayyebi, Alimorad Rashidi (2014) Neural Network and Genetic Algorithm for Modeling and Optimization of Effective Parameters on Synthesized ZSM-5 Particle Size. *Materials Letters*, 136(1), p.p.138-140.
10. Dorigo M, Gambardella LM. (1997) *Ant Biosystem*, 43(2), p.p.73-81.
11. Naimihm T. (2009) New Robust and Efficient Ant Colony Algorithms: Using New Interpretation of Local Updating Process. *Expert Systems with Applications*, 36(8), p.p.481-488.
12. Feng Xiao-rong, Feng Xing-jie, Liu Dong (2014) A Dynamic Multinest Ant Colony Algorithm for Aircraft Landing Problem. *TELKOMNIKA Indonesian Journal of Electrical Engineering*, 12(3), p.p.2196-2202.