Research on Self-regulation of Hidden Neuron Number of BP Network Based on Simulated Annealing Algorithms

Qi Anzhi

Liaoning Jianzhu Vocational College Liaoning Liaoyang 111000, China

Keywords: BP Network; Simulated Annealing Algorithms; Number of Hidden Layer Nodes

Abstract: BP is based on the steepest descent method and its application range is made. The algorithm is one of the most widely used neural network learning algorithms, which has the disadvantages of slow convergence speed and easy to fall into local minimum, which limits its application scope. When the scale reaches a certain level, the solution becomes impossible in terms of time, and the neural network is a good solution for approximate solution. In this paper, a hidden layer node estimation algorithm based on simulated annealing algorithm for single hidden layer BP neural network is proposed. The lower bound of the number of hidden nodes is determined based on experience. The number of hidden nodes is increased by simulated annealing until the end of the algorithm, and the optimal solution is obtained. The network was tested by test samples and compared with the predictions of nonlinear regression. Experiments show that the accuracy of the hidden layer nodes in the BP network hidden layer is higher and the speed is faster.

1. Introduction

BP network; simulated annealing algorithm; hidden layer node number simulated annealing algorithm is derived from the principle of solid annealing, which simulates the process of solid rising to high temperature and then slowly cooling. Because of its strong self-learning ability, BP neural network can approximate any complex non-linear function, and can also solve the problem that traditional parameter methods are difficult to find appropriate rules [1]. It is a theoretical mathematical model of human brain neural network and a non-algorithmic information processing system based on imitating the structure and function of brain neural network. The basic elements of neural networks are neurons. Neurons come from the study of the biological nervous system. They are models obtained by mathematically calculating biological neurons [2]. The input information is processed in a manner similar to a biological neural network, expressing problems that cannot be accurately described by the mechanism model, but with a certainty between the input and the output. Increasing the number of hidden layers can effectively reduce the network training error and improve the accuracy of network recognition [3]. However, the BP algorithm converges slowly and is prone to fall into the local minimum point of the objective function. Therefore, it is of theoretical and practical value to study learning algorithms with strong global search ability.

In this paper, a simulated annealing algorithm with adaptive cooling schedule is proposed, which adaptively adjusts the parameters of the tempering temperature and the attenuation function of the control parameters according to the network model error, thereby further improving the global simulation annealing algorithm. Search ability. Compared with the multiple regression method, the results show that the prediction accuracy of the neural network is much better than the multiple regression method, and it has good nonlinear fitting ability and prediction accuracy for the BP network hidden layer neuron self-adjustment. It takes full account of the accuracy of the algorithm and the global optimization characteristics of genetic algorithm, so that the corpse algorithm can get rid of the local minimal problem, and the network trained monthly can achieve the required accuracy [4]. Each neuron consists of a weighted value and a threshold, which are connected together to form a neural network. To verify whether it can escape from the local minimum, converge to the global minimum and the most popular algorithm to escape from the local minimum. Simulated annealing algorithm is compared to see whether its convergence rate reaches or exceeds the convergence rate of simulated annealing algorithm.

  DOI: 10.25236/isaicn.2019.001

2. Methodology

The learning process of BP network consists of two processes: the forward propagation of signals and the reverse propagation of errors. The purpose of error back propagation is to correct the weights. It distributes the errors among all units in each layer and obtains the error signals of each unit in each layer. Compared with the selection of excitation function of learning algorithm, the research on determining the number of hidden layer nodes is less. For the problems of pattern recognition and data set classification, if the number of hidden layer nodes is too small, the recognition accuracy of the network will be lower [5]. There is no connection between neurons in the same layer, and there is no reverse connection between the posterior layer and the anterior layer. The dimension of input and output vectors varies with the task, and generally corresponds to the number of neurons in the input and output layers of the network. With the continuous decline of temperature parameters, combined with the probability jump feature, the global optimal solution of the objective function is randomly found in the solution space, that is, the local optimal solution jumps probabilistically and eventually becomes global optimal. The traditional BP network uses the gradient descent algorithm to calculate the connection weight, which is easy to fall into the local minimum. If there are too many hidden layer nodes, the training time will become longer and cause training. Therefore, as long as we submit an instance of the reactive network behavior to the network, the network can automatically learn from the initialized weights and offset values based on the instance.

Since the input data in this paper are all positive numbers, in order to ensure the convergence speed, the data needs to be normalized to a certain range by the linear normalization method. The formula is:

$$S(\mathrm{t}) = \frac{y(t)}{l} \tag{1}$$

Among them, S (t) is the minimum value of input data and Y (t) is the maximum value of input data.

Since the input data is normalized to the [0,1] interval, the layer 1 activation function selects the sigmoid function, and the layer 2 activation function selects the linear activation function, where the sigmoid function formula is:

$$S_{\mathrm{j}} = \frac{1}{\sum_{i=1}^{n} (S_r)}, (0 < (S_{\mathrm{j}}) \le 1) \tag{2}$$

Although neurons are simple, a large number of neurons can be combined to express complex things and phenomena in the physical world, just like the various memory and thinking computing powers stored in the human brain. In the forward transfer, the input signal passes through the hidden layer from the input layer to the output layer, that is, the state of each layer of neurons can only affect the state of the next layer of neurons. Simulated annealing seeks a global optimal solution of a large search space in a fixed time by a probability algorithm. The simulated annealing algorithm is an algorithm that simulates the annealing process of a metal material. Usually, the larger one of the input node and the output node is selected as an initial value according to experience, and then the network is trained a certain number of times. When input acts on the network, the actual output of the network is compared with the target output, and then the weights and biases of the network are adjusted by learning rules, so that the actual output of the network is closer to the target output. In a non-feedback network, if the neurons are arranged hierarchically and each layer of neurons only connects the neurons of the upper layer, the network is called a feed forward network. Because of its unsupervised characteristics, the dynamic clustering method based on the spindle kernel function may have poor results when the feature space of various clouds is close and their share is similar.

In order to obtain the optimal weight Rj of the BP neural network, the square error function of the actual output and the predicted output of the BP neural network is used as the optimization function.

$$R_j = \frac{f_{ij}}{Tj} \times S_j \tag{3}$$

The solution defining this optimization problem is the connection weight R_j. Assume that the generating function of k_j solution Dj in the jth search step is a random perturbation model:

$$k_j = \frac{R_j}{Dj} \times Uj \times \mathrm{Re}_j \tag{4}$$

In the algorithm, the weight and threshold of the individual are coded. The reciprocal of the error norm between the predicted output and the expected output is taken as the fitness function Dj, and the calculation formula is as follows:

$$D_{\mathrm{j}} = \sum_{i=1}^{n} (H_P \times V_p) \tag{5}$$

Among them, Dj is the fitness function, Hp is the error norm, and Vp is the expected output of the neural network.

Neural network consists of input layer, hidden layer, output layer and node connection weight between layers. The learning process of network consists of forward propagation of information and back propagation of error. If the output layer can't get the expected output, the back-propagation error is transferred to adjust the weights and thresholds of the network according to the prediction error, so that the prediction of BP neural network keeps approaching the expected output. For some networks with a large number of nodes, the computational efficiency is low; empirical formula method lacks relevant theoretical basis, because its formula comes from the experience of projects and experiments, it can only be effective for specific data sets, and can not be used as a general method to determine hidden layer nodes. In the process of layer-by-layer transmission, each layer of neurons only affects its next layer of neurons. Then, use the BP network to find the best advantage. If the termination condition is satisfied, the current solution is output as the optimal solution, and the program is terminated. The termination condition is to terminate the algorithm when several consecutive new solutions are not accepted. When the temperature is lowered and the above process is repeated, the SA is declared to have a global optimal solution, which is the optimal connection weight of the BP neural network.

3. Result Analysis and Discussion

The simulated annealing algorithm is an algorithm based on statistical physics. By introducing a natural mechanism such as the crystal annealing process in the physical system, the optimal problem solution is accepted by the Metropolis criterion. The input information is processed layer by layer from the input layer input layer to the output layer output. If the output of the output layer does not match the expected output, the error is reversed. Small-scale neural networks often fail to achieve the accuracy of solving problems, and it is difficult to complete learning tasks, resulting in neural networks not converging, or slow convergence. However, the inherent problem of the BP method is that the network can only converge to a local minimum, and cannot be sure to converge to the global minimum. Therefore, the initial threshold weights are different, and the obtained errors are different. Three-point search may miss the global optimal solution. In addition, these algorithms need to continually simulate the whole network training process, which runs for a long time and costs a lot. Using the chain rule of derivative, the derivative of connection weight about error function is returned along the original connection path, and the error function is reduced by modifying the weights of each layer. If the parameters are not selected properly, BP neural network is easily trapped in local extremum, which can not achieve the desired results. In addition, long time-consuming is also a common problem.

Under the number of different hidden layer nodes, the error of the network after training several times is shown in Figure 1. When the number of hidden layer nodes is 10, the simulated annealing algorithm judges the end of the operation, and finally determines the number of network hidden layer nodes of the recognition model to be 18.

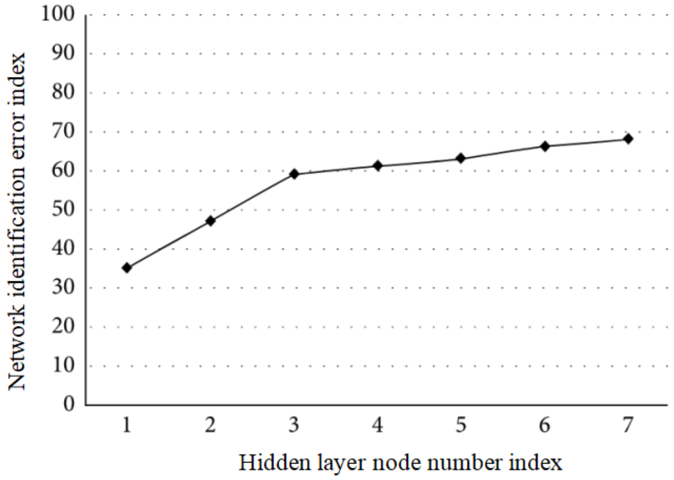

Fig.1. Network training error rate

The generalization ability of BP neural network refers to the ability of the learned network to approximate and predict unlearned samples. It can be seen that the prediction accuracy of the BP network is determined by its generalization ability. The most critical point of the algorithm is to use a certain probability for the local minimum solution of the problem to judge whether to accept or not, in order to jump out of the local extremum. In the training, we use the learning method to make the error between the actual output and the expected output to be minimized when we satisfy the given input. Since the error is positive or negative, we take the sum of the squares of the errors. The decision tree is used to determine the structure of the neural network, including the number of hidden layers and the number of hidden cells. The selected qualified input and output measurements are arranged into sample vectors suitable for BP network training format, and the input data are normalized to meet the requirements of input layer neuron non-linear function. Finally, the category corresponding to one of the larger values is selected as the result of network recognition to verify the recognition accuracy of the network. The convergence speed of the algorithm is determined by the minimum eigenvalue and the learning speed. The larger the difference between the eigenvalues, the slower the convergence of the steepest descent algorithm.

In order to verify the accuracy of the experimental results, the number of hidden layer nodes in the experiment is 15, and the BP neural network is established. The experimental data sets are trained. The training results are shown in Figure 2.

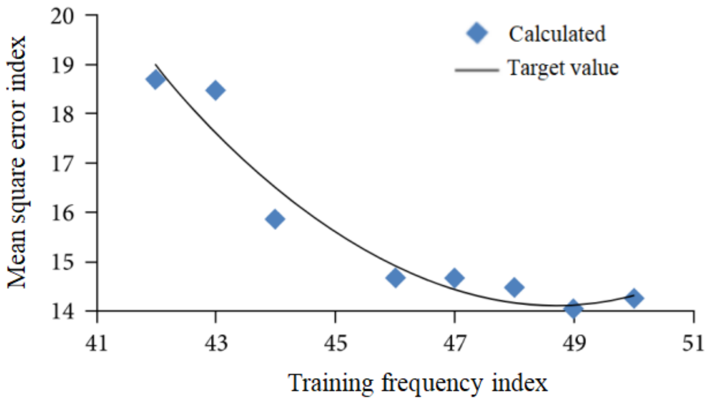

Fig.2. Network training result information

After determining the structure of the BP neural network, the network is trained through the input and output sample sets, that is, the threshold and weight of the neural network are learned and corrected, so that the network achieves a given input-output mapping relationship, which is Learning of BP neural networks. When the neuron activation function is selected, considering the learning phase, the threshold in the local gradient of the output unit is relatively large, which may cause uneven learning speed of each neuron. The weights and thresholds of the BP network are optimized, and the optimal weights and thresholds are obtained to realize the prediction of BP neural network. Since the BP network is associated by similarity, the prediction accuracy of the state with high similarity to the sample is high. For this reason, when selecting the training sample vector, a possible extreme value combination of three input color values should be included. For input data, it will be processed after reading into the neural network; according to the different types of flowers, the corresponding output node value is 1, the other value is 0, which is the output data of training data. BP neural network is based on the steepest descent method, and the convergence speed of the steepest descent method depends on the characteristics of the surface. It does not distinguish the minimum point, the maximum point and the saddle point. The steepest descent method may converge to the saddle point. The advantage of this improved algorithm is that it can get more stable weights and thresholds, but if the convergence speed is slow, the network will easily fall into local extremum and the learning process will oscillate.

4. Conclusion

In order to enhance the generalization ability of BP neural network, this paper studies the characteristics of BP neural network based on simulated annealing algorithm, and improves it from the aspects of sample space, initial weights and network structure parameters. The accuracy of the prediction is further improved from two aspects: the selection of training samples and the optimization of hidden layer nodes. Although the number of sample data is extremely limited for the training and prediction of neural networks, the prediction results show that the method has high accuracy and fast convergence speed. It can adjust the parameters of the tempering temperature of each stage and the attenuation function of the control parameters adaptively according to the network model. It is related not only to the output of the output layer, but also to the output of the hidden layer, so that the modification of the weight of the network during the iteration is as synchronous as possible with the difference between the actual output and the expected output of the output layer. Experiments show that the algorithm can find the appropriate number of hidden layer nodes effectively and quickly, and the network works normally and the recognition rate is high. Since this precision is very close to the original function, there is no need to continue learning, of course, this is the reason for time. The higher the accuracy, the slower the convergence speed.

References

[1] Lihua Y, Jingjing W U, Yao W, et al. Optimized BP Neural Network Based on Simulated Annealing Algorithm for pH Value Prediction[J]. Chinese Journal of Sensors and Actuators, 2014, 27(12):1643-1648.

[2] Manoochehri M, Kolahan F. Integration of artificial neural network and simulated annealing algorithm to optimize deep drawing process[J]. The International Journal of Advanced Manufacturing Technology, 2014, 73(1-4):241-249.

[3] Lin Y J, Chen X J. Simulated Annealing Algorithm Improved BP Learning Algorithm[J]. Applied Mechanics and Materials, 2014, 513-517:734-737.

[4] Ul Islam B, Baharudin Z, Nallagownden P. Short Term Electric Load Forecasting with Back Propagation Neural Network and Simulated Annealing Genetic Algorithm[J]. Applied Mechanics and Materials, 2015, 785:14-18.

[5] Zuo H J, Zhang J Y, Ying Y L, et al. Rigorous microlens design using vector electromagnetic method combined with simulated annealing optimization [J]. Optics Express, 2014, 22(10):12653.