Optimal Structure and Parameters of BP Neural Network for Curve Fitting Problem

Hongjie Yi[1, 2, a *], Guangrong Ji[1, b], Jinghua Liu[2, c] and Lin Jia[1, d]

[1]Ocean University of China, Qingdao, China

[2]Qingdao Technological University, Qingdao, China

[a]yihongjie@sina.com, [b]grji@ouc.edu.cn, [c]l_jinghua@163.com, [d]jialin@ ouc.edu.cn

Keywords: BP neural network; Optimal structure and parameters; Curve fitting; Learning rate; Momentum

Abstract. BP neural network is wildly used because of its strong nonlinear processing ability, self-learning capability, fault tolerance capability. Therefore, the structure and parameters of artificial neural network determine the performance of neural networks. The performance of a BP neural network is not only affected by the network structure, but also affected by its parameters. In this article we will discuss the learning rate and momentum parameters matching relationship and its impact on network performance. The experimental results show that for the curve fitting problem there will be an optimal structure and parameters for the BP neural network.

Artificial Neural Network and Its Structural Characteristics

BP neural network is wildly used because of its strong nonlinear processing ability, self-learning capability, fault tolerance capability. Curve fitting especially for higher-order function is a complex process. The processing of BP neural network curve fitting, in fact is adjusting the connection weights of all artificial neurons in the BP neural networks. It is a process of matching optimal state of the network and curve function. The training convergence process of the BP neural network is the matching process of parameters adjustment. Generally with the size and complexity of the problem is growing, the number of neurons in the neural network will be more and more, the network structure will be more complex. If a neural network contains very little neurons, there are not enough weights to be adjusted, so it cannot remember the appropriate mode, it cannot be expected to complete the mappings. But too complex network structure will lead to neural network training convergence slows down, sometimes even lead to the training process cannot converge, making it impossible to complete the mapping of expected. Therefore, the structure and parameters of artificial neural network determine the performance of neural networks. In the curve fitting problems, fitting precisely and quickly is an important performance indicator of BP neural network. This includes samples that have been learned and not been learned, that is called the generalization ability. In addition, in practical applications it should also consider the cost of time and space.

In theory, the number of the hidden layers of BP neural network may be 0, 1, 2…n. It has been proved that only one input layer and output layer BP neural network cannot solve nonlinear problems [1]. It also has been proved that a BP neural network with a single hidden layer structure can approximate any derivable function and its derivatives at any precision [2, 3]. Therefore, the BP neural networks were selected for the single hidden layer structure in most practical applications. The number of the neurons of the input layer and the output layer depends on the dimensions of the problem to be solved. In general, these two parameters are determined according to the actual problems. Therefore, to determine the BP neural network topological structure can be attributed to the selection of the number of the hidden layer units.

Some scholar [4] has proposed the number of the hidden layer units and the input layer units should meet the following equation as Ep.1.

$$N_{hid} = 2 * N_{in} + 1 \tag{1}$$

Somebody else has considered the impact of training samples [5]. When the number of inputs

N_{in} and outputs N_{out} are given, based on the number of training sample set N_{tr}, the upper bound of the number of the hidden layer units should meet the following formula.

$$N_{hid} \leq N_{tr} /[R+(N_{in}+N_{out})] \tag{2}$$

Some scholars continue to consider the impact of the total connection weights of the BP neural network and give a boundary formula as follow.

$$\begin{cases}(N_w / N_{out}) \leq N_{tr} \leq (N_w / N_{out}) \bullet \log_2^{(N_w / N_{out})} \\ N_w = N_{tr} \bullet \log_2^{N_{tr}}\end{cases} \tag{3}$$

In which N_{tr} is the number of training samples, N_{out} is the output layer nodes and N_w is total weights of the BP network. Some scholars give a suggestion for the number of the hidden layer units based on the particular problem [6]. For example, according to the given character recognition problem raise the empirical equation as Ep.4.

$$N_{hid} = \sqrt{N_{in} \bullet (N_{out}+1)+1} \tag{4}$$

In addition, we also can set a random number of the hidden layer units at first, and then adjust it by some methods. The methods include Trial-error method [7], Cross-validation method [8], increased method, decreased method [9], GA method [10] or even PSO method [11] etc. The methods above try to build a best network from a different viewpoint. But the performance of a BP neural network is not only affected by the network structure, but also affected by its arguments. In this article we will discuss the learning rate and momentum parameters matching relationship and its impact on network performance.

BP Neural Networks Construction and Parameter Selection for Curve Fitting Problem

Curve to Be Fitted. In this experiment, we use curve fitting problem as an example. The curve to be fitted is a higher-order function as Ep.5.

$$y = 2x^6 + 3x^3 - 2x^2 \quad -2 \leq x \leq 1 \tag{5}$$

It is sampled once every 0.01 in the function definition, a total of 301 sample points sequence is get. One point is chosen in every 15 sampling points, a training set consists of 20 sampling point is build. Then we randomly selected 15 points from 301 sample points to build a test set. Because the neural network convergence curve can reflect the convergence speed and the ultimate convergence error, so we can compare the performance of the network by comparing different network convergence curve and fitting results.

Determine the Number of Hidden Layer Units. First, we should determine the topology of the BP neural network. According to the previous analysis, we use single hidden layer network architecture. The number of hidden units can be from 10 to 100 according to the formula above. We set the network parameters for the learning rate at 0.1 and the momentum at 0.5 to compare how the different number of hidden layer units impact on network performance. After 1000 times training the network convergence curves and time-consumings are as follows.

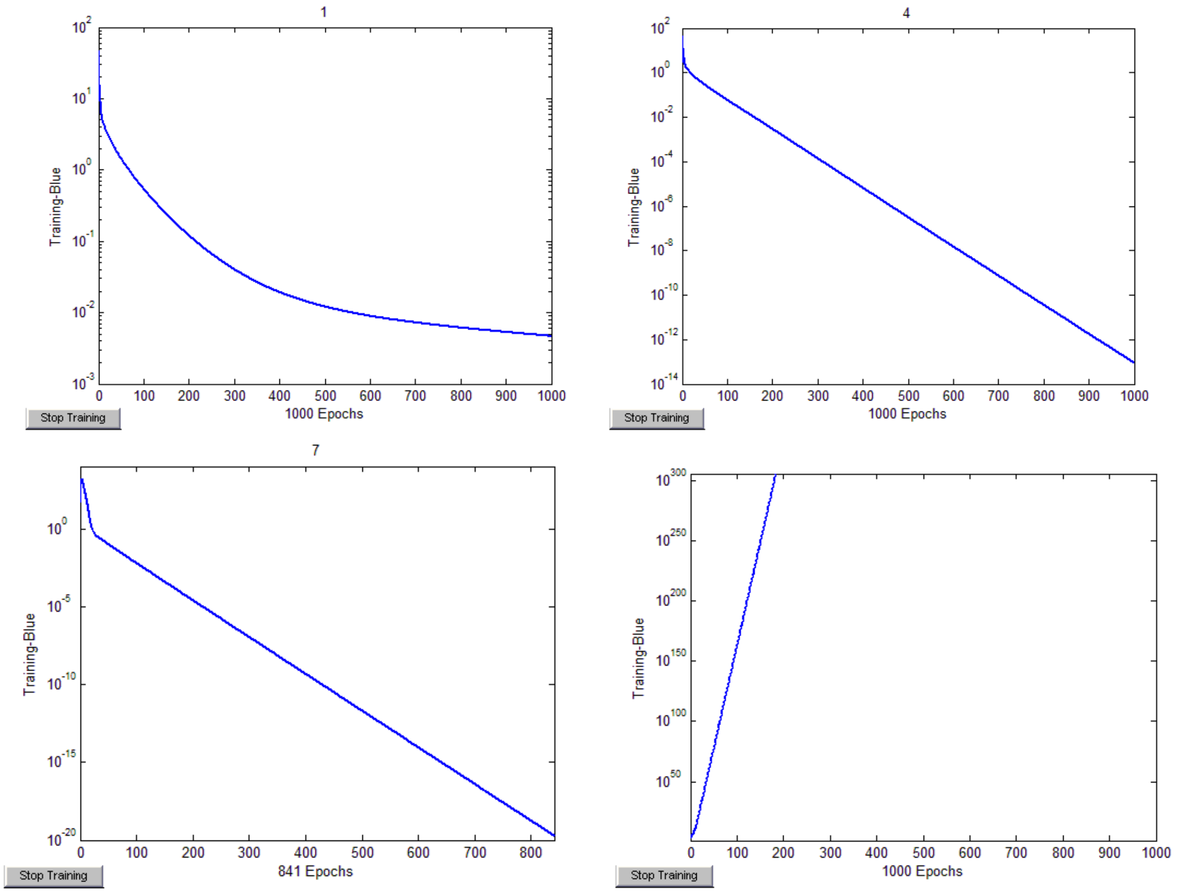

Figure 1. Convergence curve of the different number of hidden layer units (10, 40, 70, 100)

Table 1 Convergence error, time-consuming and fitting accuracy comparison

number of network hidden layer	Convergence error	fitting accuracy(mse)	time-consuming(s)
10	0.0046633	0.097929	0.954
20	00158171	0.033216	1.078
30	1.46393e-009	3.0742e-008	1.203
40	9.09086e-014	1.9091e-012	1.391
50	9.88607e-019	2.0761e-017	1.438
60	2.20182e-020	4.6238e-019	1.547
70	1.9073e-020	4.0053e-019	1.585
80	NaN	NaN	7.563
90	NaN	NaN	8.625
100	NaN	NaN	8.656

From the Fig. 1 and Table 1 we can find that when a smaller number of hidden units, the network topology are fewer complexes and less time-consuming, but the error is larger. As the increase of the number of hidden units, network topology complexity increases, time-consuming increases too, but in the while the convergence error decreased. However, when the number of hidden layer units exceeds a certain threshold the network will not converge, and time-consuming will be significantly increased.

Select the Optimal Parameters of the Learning Rate and Momentum. Then we set the number of hidden layer unit at 60 and analysis how the different learning rate and momentum impact on the performance of network. We make 81 different BP neural networks with different learning rate and momentum and then train and test them with same set. A few representative convergence curves, fitting accuracy and time-consuming are selected and showed as follows.

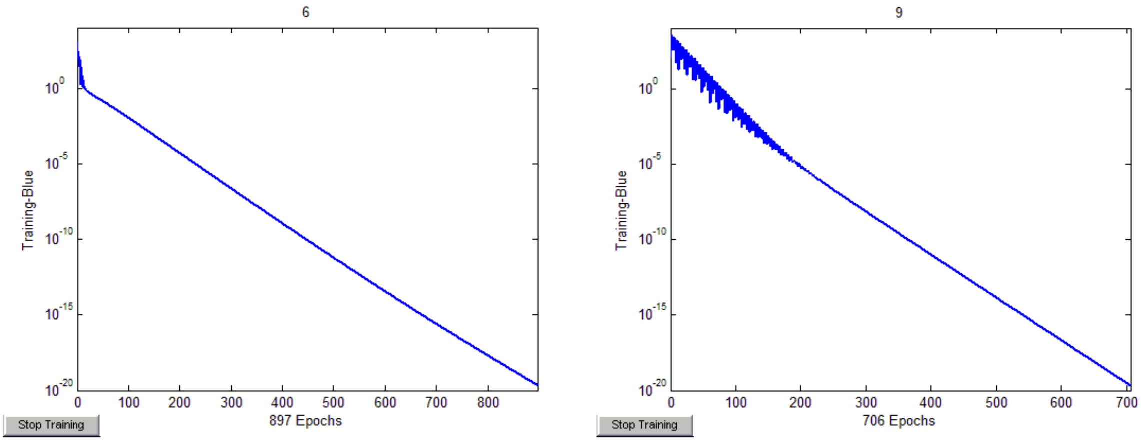

Figure 2. The convergence curves with same learning rate at 0.1 but different momentum(0.6 and 0.9)

As can be seen at the same learning rate, enhanced momentum parameters can speed up network convergence. But in the same time it increased the dither which may make it easier fall into the local minimum in the complex problems.

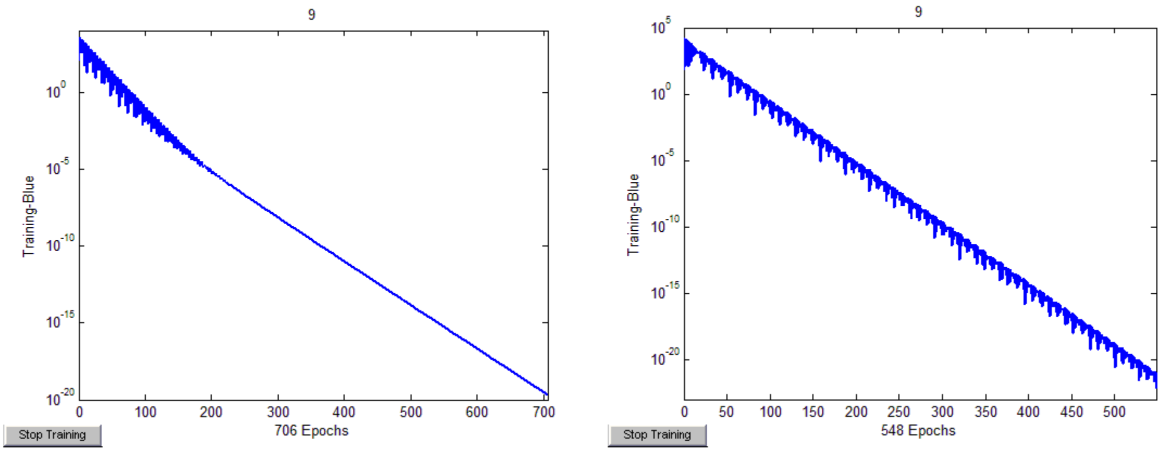

Figure 3. The convergence curves with same momentum at 0.9 but different learning rate (0.1 and 0.4)

As can be seen at the same momentum, enhanced learning rate can speed up network convergence. But in the same time it increased the dither or even makes the network divergence.

Table 2 Fitting accuracy with different learning rate and momentum

Lr\mc	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
0.1	NaN	NaN	NaN	NaN	2.2734e-009	2.1852e-009	3.7892e-009	2.0125e-009	1.9592e-009
0.2	NaN	NaN	NaN	NaN	NaN	NaN	1.5424e-009	1.8604e-009	4.2461e-010
0.3	NaN	NaN	NaN	NaN	NaN	NaN	NaN	1.2598e-010	1.3446e-010
0.4	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	1.5019e-010
0.5	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	1.9501e-008
0.6	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
0.7	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
0.8	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
0.9	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN

Table 3 Time-consuming with different learning rate and momentum

Lr\mc	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
0.1	6.338	6.452	6.593	6.499	1.616	1.536	1.848	1.6	1.211
0.2	6.736	8.593	7.05	8.086	6.968	7.151	1.124	0.75	0.891
0.3	6.757	6.806	6.773	6.951	7.054	6.552	6.775	0.545	1.026
0.4	7.022	6.845	7.043	6.916	6.896	6.719	6.614	6.394	0.954
0.5	7.024	6.86	7.014	7.327	7.228	6.871	6.848	6.671	1.673
0.6	6.782	6.8	6.715	6.766	6.7	6.649	6.6	6.395	5.742
0.7	6.783	6.802	6.709	6.714	6.682	6.689	6.579	6.467	6.069
0.8	6.703	6.746	6.74	6.715	6.696	6.725	6.85	6.6	6.797
0.9	6.953	6.849	6.825	7.001	7.104	7.584	8.497	8.929	9.728

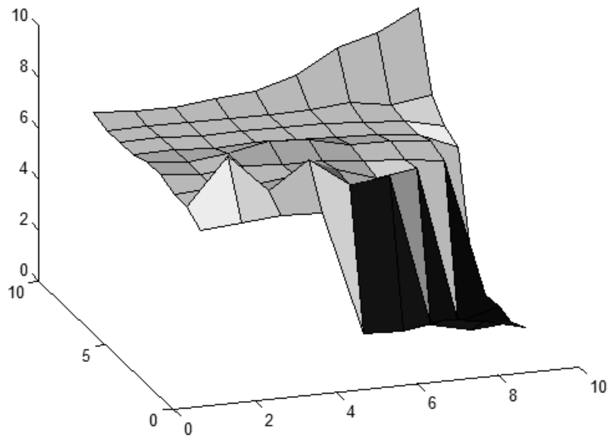

Figure 4. Three-dimensional map of the performance of network with different learning rate and momentum

From the Fig. 4 and Table 2, 3 we can find that the BP neural network with different learning rate and momentum has different performance. For curve fitting problem a bigger momentum with a smaller learning rate has a better performance on time-consuming and fitting accuracy.

References

[1] 1M. Minsky, S. Papert, Perceptrons, An Introduction to Computational Geometry, MIT Press, 1969: 12-15

[2] Hornk K, Stinchcombe M., White H, Mulilayer feedforward networks are universal approximates[J], Neural Networks, 1989, (2): 359-366

[3] Hornk K, Stinchcombe M, White H. Universal approximation of an unknown mapping and its derivatives using mulilayer feedforward networks [J], Neural Networks, 1990, (3): 551-560

[4] Eberhart R. C., Dobbins R. W., Neural network performance metrics for biomedical applications, Computer-Based Medical Systems, 1990.282 - 289

[5] Hamid Beigy, Mohamad Reza Meybodi, A learning automata-based algorithm for determination of the number of hidden units for three-layer neural networks, International Journal of Systems Science, 2009, Vol.40 (1): 101-118

[6] Hongjie Yi, Guangrong Ji, Haiyong Zheng, Optimal Parameters of Bp Network for Character Recognition, Industrial Control and Electronics Engineering (ICICEE), 2012 1757-1760

[7] BP Lveczky, TM Otchy, JH Goldberg, A Dmitriy, MS Fee, Changes in the neural control of a complex motor sequence during learning, Journal of Neurophysiology, 2011, 106(1):386-397

[8] Y Zhang, Y Yin, D Guo, X Yu, L Xiao, Cross-validation based weights and structure determination of Chebyshev-polynomial neural networks for pattern classification, Pattern Recognition, 2014, 47(10):3414-3428

[9] Fahlman S E,Lebiere C., The Cascade-Correlation Learning Architecture[A]. Touretzky D S.Advances in Neural Information Processing Systems[C].San Mateo, CA:Morgan Kaufmann 1990,2:524-532

[10] L Zhuo, J Zhang, P Dong, Y Zhao, B Peng, An SAfmann 1990,2:524-532Information Processing Systems[C].San Mateo,omial neural, Neurocomputing, 2014, 134(9):111-116

[11] JY Li, Bp Neural Network Optimized by PSO and its Application in Function Approximation, Advanced Materials Research, 2014, 945-949:2413-2416