Algorithm of BP neural network based on parameterization of excitation function

Dongbo Zhao [1, a], Hui Li [2]

[1]School of Electronic Engineering, Xi'an Aeronautical University, Xi'an, 710077, China.

[2]School of Electronic Information, Northwestern Polytechnical University, Xi'an, 710129, China.

[a]alien_dffy@163.com

Abstract

In view of the shortcomings of slow convergence rate and low convergence precision of BP neural network algorithm, a BP network algorithm with parametrization of excitation function is proposed, that is to introduce adjustable parameters(θ_0 and θ_1)in the S transfer function (excitation function), and the parameter θ_0 is used to change the slope of the excitation function, θ_1 is used to change the displacement of the excitation function. The experimental results show that the convergence speed and convergence accuracy of the improved algorithm are greatly improved compared with the standard BP algorithm, and the feasibility of the improved function is verified.

Keywords

BP neural network, excitation function, adjustable parameter.

1. Introduction

The BP neural network (Back Propagation NN) is a multi-layer forward neural network, which can realize any nonlinear mapping from input to output, and its weights are adjusted by the back propagation artificial neural network. In the standard BP neural network algorithm, the neuron excitation function usually adopts single and bipolar S functions. But the standard sigmoid function is single, its mapping range and location are fixed, lack of flexibility, which is incompatible with the generalization ability of neural network. It can't improve the convergence speed of neural network, low convergence precision and easy to fall into local extreme small solution. In view of the shortcomings of the standard BP algorithm, many scholars have studied and proposed an improved scheme. For example, literature [5] has proposed a neuron activation function containing 4 undetermined parameters. However, in practical applications, because of the more parameters of the method, a lot of time has been consumed. Based on the study of literature [5], this paper proposes an efficient excitation function with adjustable parameters. The simulation experiment shows that the function can satisfy the fast convergence requirement of the BP algorithm.

2. Algorithm of BP Neural Network Based on Incentive Function Parameterization

2.1 BP Neural Network Algorithm BP

BP neural network main idea: For Q input learning samples P1, P2, Pq, known its corresponding output samples are: T1, T2,. Tq. The purpose of learning is to modify the weight of the network by using the error between the actual output Ai of the network and the target vector Ti, so that Ai can be as close as expected to the Ti. That is, minimizing the minimum mean square error of the output layer of the network.

In the standard BP algorithm, the learning algorithm converges slowly near the target point, mainly because the error space is a very complex surface in the N dimension space, the "height" of each point on the surface corresponds to an error value, and the coordinate vector of each point corresponds to the N weight value. There exists a flat area in the network error surface, causing the error to slow down and affect the convergence speed. The feature of all the smallest points including the global best advantage is the error gradient is zero, which leads to the BP algorithm based on the gradient as the weight adjustment can't distinguish the smallest point, so the training is often trapped in a local point. To achieve the global minimum solution, we must try to jump out of the local minimum point. To solve this problem, this paper proposes another algorithm to solve the "local minimum" problem.

2.2 The Effect of Excitation Functions on Network Performance

The three node of network is represented as input node x_j, hidden node y_i and output node O_i respectively. The weights of the network between the input node and the hidden node are ω_{ij}, the network weights of the hidden nodes and the output nodes are ω_{ij}. The calculation formula of the BP model is as follows:

(1) the output of the hidden node is :

$$y_i = f(\sum_j \omega_{ij} x_j - \theta_i) = f(net_i) \tag{1}$$

Where θ_i is the threshold of hidden node neurons, and net_i is the input sum of hidden neurons.

(2) the output of the output node is :

$$O_i = f(\sum \omega_{ij} y_i - \theta_i) = f(net_i) \tag{2}$$

Where θ_i is the threshold of output node neurons, and net_i is the input sum of output node neurons.

(3) the error of the output node is:

$$E = \frac{1}{2}\sum_t (t_l(t_l - O_l))a^2 \tag{3}$$

Where t_l is the desired output of the output node.

(4) the weight value of the hidden layer between the output layer is adjusted to:

$$\frac{\partial E}{\partial \omega_{li}} = \sum_{k=1}^{N} \frac{\partial E}{\partial O_k} \cdot \frac{\partial O_k}{\partial \omega_{li}} = \frac{\partial E}{\partial O_l} \frac{\partial O_l}{\partial \omega_{li}} \tag{4}$$

Because E is a function of multiple O_k, and only one O_k is related to E, and each O_k is independent of each other. Thus, we can get the following:

$$\frac{\partial E}{\partial O_l} = \frac{1}{2}\sum_k -2(t_k - O_k) \cdot \frac{\partial O_k}{\partial O_l} = -(t_l - O_l) \tag{5}$$

$$\frac{\partial O_l}{\partial \omega_{li}} = \frac{\partial O_l}{\partial net_l} \cdot \frac{\partial net_l}{\partial \omega_{li}} = f'(net_l) \cdot y_i \tag{6}$$

Formula (4) can be simplified by using formula (5) and formula (6):

$$\frac{\partial E}{\partial \omega_{li}} = -(t_l - O_l) \cdot f'(net_l) \cdot y_i \tag{7}$$

(5) the adjustment of the weight of the input layer between the hidden layer

$$\frac{\partial E}{\partial \omega_{ij}} = \sum_l \sum_i \frac{\partial E}{\partial E} \frac{\partial O_l}{\partial y_i} \frac{\partial y_i}{\partial \omega_{ij}} \tag{8}$$

Similarly, because E is a function of multiple O_l, and because it only corresponds to one y_i for a ω_{ij}, and is related to all O_l ((the upper form is only L summation). Thus, we can get the following:

$$\frac{\partial E}{\partial O_l} = \frac{1}{2}\sum_k -2(t_k - O_k) \cdot \frac{\partial O_k}{\partial O_l} = -(t_l - O_l) \tag{9}$$

$$\frac{\partial O_l}{\partial y_i} = \frac{\partial O_l}{\partial net_l} \cdot \frac{\partial net_l}{\partial y_i} = f'(net_l) \cdot \frac{\partial net_l}{\partial y_i} = f'(net_l) \cdot \omega_{li} \tag{10}$$

$$\frac{\partial y_i}{\partial \omega_{ij}} = \frac{\partial y_i}{\partial net_i} \frac{\partial net_i}{\partial \omega_{ij}} = f'(net_i) \cdot x_i \tag{11}$$

The (9) ~ (11) formula can be simplified (8).

Formula (8) can be simplified by using formulas (9) ~ (11):

$$\begin{aligned}\frac{\partial E}{\partial \omega_{ij}} &= -\sum_{l}(t_l - O_l) f^{'}(net_l)\cdot \omega_{li}\cdot f^{'}(net_i)\cdot x_j \\ &= -\sum_{l}\sigma_l \omega_{li}\cdot f^{'}(net_i)\cdot x_i\end{aligned} \tag{12}$$

The adjusted weights are obtained.

$$\omega_{ij} = \omega_{ij} + \eta\frac{\partial E}{\partial \omega_{ij}} \tag{13}$$

From the gradient formula above, we can see that:

(1) Under the premise of guaranteeing convergence, the (13) formula shows that the greater the magnitude of weight adjustment, the greater the gradient changes, and the faster learning speed. According to (7),The gradient $\frac{\partial E}{\partial \omega_{ij}}$ is directly proportional to the derivative $f^{'}(net_i)$ of the excitation function. Because the derivatives of the excitation function are widely used in the learning process, the learning process of the whole network will be greatly affected.

(2) Due to the existence of the saturation area of the S type excitation function, when the output of the neuron falls into the saturation area of the excitation function, it is necessary to make a larger correction of the weight value to escape the processing unit from the saturation area as soon as possible. As the derivative of excitation function is very small in the saturation area, the weight value can only be minor corrected in each learning cycle, and the output unit will work in a flat area for a period of time, so that the root mean square error of the network will remain unchanged or change very little. This produces the so-called paralysis phenomenon, which slows down the speed of convergence of the network.

Based on the above analysis, we should choose the appropriate excitation function and introduce some necessary adjustment parameters. When the output of the neuron falls into the saturation area of the excitation function, the value of the guide function of the excitation function can be adjusted accordingly, which can effectively avoid the emergence of the numbness phenomenon and accelerate the convergence of the network. Speed.

2.3 Improvement Based on Parametrization of Excitation Function

Improving the excitation function can change the error surface and minimize the possibility of local minimum. In principle, any nonlinear function can be used, but for the inversion learning algorithm, a continuous differentiable nonlinear function is required.

In this paper, the Tansig function is chosen as the basic form of the hidden layer node excitation function.

$$f(x) = \frac{1}{1+e^{-x}} \tag{14}$$

The improved BP algorithm introduces new parameters θ_0 θ_1 to the standard S type function, and the function changes to:

$$f(x) = \frac{1}{1+\exp[-(x+\theta_0)/\theta_1]} \tag{15}$$

θ_0 and θ_1 is the adjustable parameter in the excitation function. The parameter θ_0 is used to change the slope of the excitation function. The smaller θ_0 makes the sigmoid function approximate to a step limiting function. However, the larger θ_0 will make the sigmoid function flatter. The function of parameter θ_1 makes the excitation function move along the horizontal direction, and positive θ_1 moves the excitation function horizontally to the left. The introduction of two parameters enables the excitation function to be freely retractable for input X. The following figure shows how the parameter A regulates the excitation function.

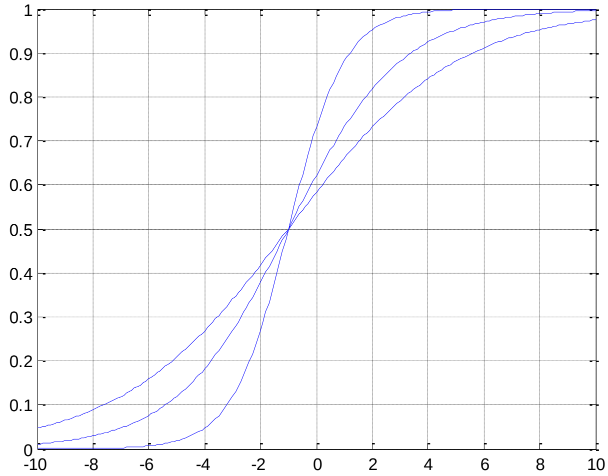

Fig.1 Adjustment of the Excitation Function

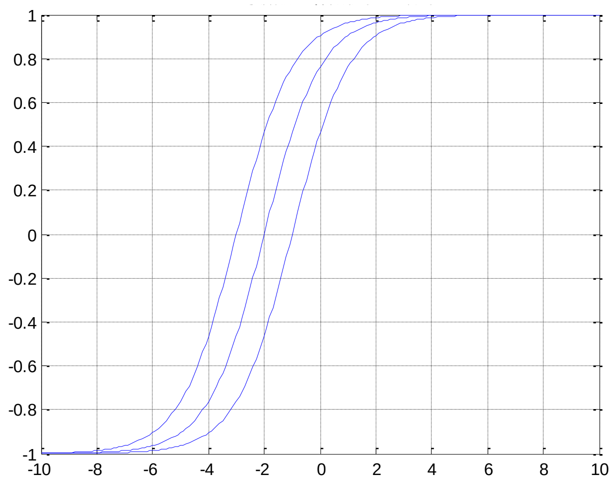

Fig.2 Adjustment of the Excitation Function

The nonlinear functions shown in formula (15) satisfy the following two conditions: First, it is continuous smooth and monotonicity; two, its definition domain is (- infinity, + infinity), its range is (0, 1), so it meets the requirements of the excitation function. Moreover, it makes the curve of the excitation function flat and avoids the local minimum at $y_j \approx 0$ and $1-y_j \approx 0$, so this function has better function approximation ability and fault tolerance ability.

3. Simulation Experiment Data and Result Analysis

3.1 Simulation Data

The data used in this experiment are measured aircraft data of a ISAR radar. The bandwidth of radar transmitting pulse is 400MHz (range resolution is 1m and radar radial sampling interval is 0.5m). Data with different pitch angles are selected as training and testing data respectively. The training samples of three kinds of aircraft safety -26, commendation and Jacques 42 were 50, 50 and 50 respectively, and the test samples were 400, 500 and 400 respectively.

3.2 Simulation Experiment and Result Analysis

BP neural network based on excitation function parameterization is used to recognize three kinds of targets. Table 1 shows the average of 200 simulation results.

Tab.1 Three Types of Target Recognition Results Based on Two Methods

Target	correct recognition rate					
	0 º-30º		0 º-60º		0 º-90 º	
	BP	improved BP	BP	improved BP	BP	improved BP
Target 1	0.872	0.905	0.867	0.897	0.853	0.892
Target 2	0.873	0.913	0.863	0.904	0.854	0.897
Target 3	0.869	0.901	0.856	0.893	0.841	0.885
average recognition rate	0.871	0.906	0.862	0.898	0.849	0.891

The data in Table 1 show that the two methods have better recognition rate for all three types of target in all attitude angles, and the rate of recognition of the three targets in each attitude angle is more than 84%. At the same time, we can see that the average recognition rate will decrease with the increase of the number of samples in three different angles.

The average recognition rate of the improved BP network is 4% higher than the average recognition rate of the standard BP network recognition in the same attitude angle range, compared with the standard BP neural network and the BP neural network based on the parametrization of the excitation function.

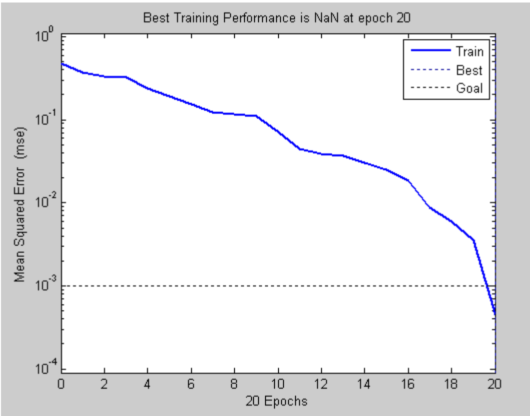

Fig.3 The Identification Error Curves of Standard BP Neural Network for Three Types of Targets.

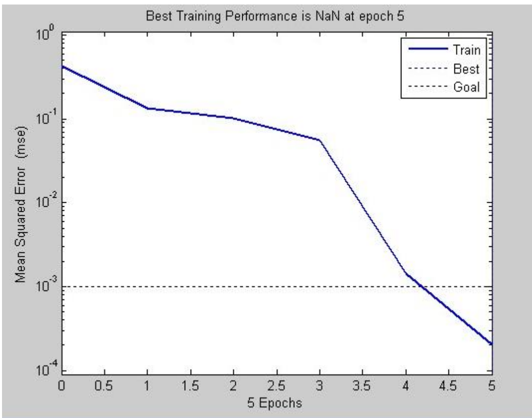

Fig.4 The Identification Error Curves of Improved BP Neural Network for Three Types of Targets.

Figure 3 and Figure 4 show the identification error curves of the standard BP neural network and the improved BP neural network under the same group of samples. Both of them use the same feature extraction method based on the translation invariant KPCA, the same model parameters and the sample noise conditions. It can be seen from the error curves that the convergence rate of the improved BP neural network is obviously higher than that of the standard BP neural network.

4. Summary

In this paper, a BP network algorithm based on the parametrization of excitation function is proposed for the fixed excitation function model of each neuron node in the standard BP algorithm. The adjustable parameter is added to the neuron transfer function, and an exciting function form with adjustable parameters is proposed. The effectiveness of the radar target high resolution distance image recognition experiment is verified. The average recognition rate of the improved BP network is improved compared with the BP network, which provides a certain way of thinking for the subsequent research of BP algorithm. However, the improved algorithm proposed in this paper also has limitations. The improved functions in this paper only aim at S type functions.

References

[1] Li Enyu, Yang Ping first, Sun Xingbo:Based on activation function four parameter adjustable BP neural network improved algorithm , Microelectronics and computer, Vol. 25 (2008) No. 11, p. 89-93.

[2] Gao Lipeng, Zheng Hui: Convolution neural network based on Prelus-Soft plus nonlinear excitation function, Journal of Shenyang University of Technology, Vol. 25 (2018) No.1, p. 89-93.

[3] Zhang Hongyu, Ding Ning, Xu Jiangrong: Research on improvement of BP neural network excitation function, Journal of Hangzhou Dianzi University (NATURAL SCIENCE), Vol. 25 (2017) No.6, p. 89-93.

[4] Tan Yuanshun, Kuang Yang: Multi stability neural network with discontinuous excitation function, Journal of Nanjing Normal University (NATURAL SCIENCE), Vol. 25 (2013) No.3, p. 89-93.

[5] Wei Cao , Hui Zhou, Zhimin Zhou, Zuowei Fu : An Approach for High Resolution Radar Target Recognition Based on BP Neural Network, International Conference on Intelligent Computing, ICIC 2011: Advanced Intelligent Computing ,pp 33-39.

[6] Sun Shaoyuan, Li Linna, Zhao Haitao: Depth estimation of monocular vehicle infrared images using KPCA and BP neural networks , Infrared and laser engineering, Vol 42(2013) No.9 , p.2348-2352.

[7] Dongliang Ma, Tao Zhou, Jie Chen,Shi Qi, Zejun Xiao: Supercritical water heat transfer coefficient prediction analysis based on BP neural network Research article, Nuclear Engineering and Design, Vol 320(2017) No 8, p.400-408.

[8] Mingshun Li, Wencui Chen: Application of BP Neural Network Algorithm in Sustainable Development of Highway Construction Projects Open access, Research article, Physics Procedia, Vol 25(2012) No.2 , p. 1212-1217.

[9] Zhigang Zhang, Cong Wang: The Research of Vehicle Plate Recognition Technical Based on BP Neural Network Open access, Research article, AASRI Procedia, Vol 1(2012) No.2 , p.74-81.