УДК 512.7

I. Didmanidze, G. Kakhiani

ARTIFICIAL NEURAL NETWORK AND WAYS OF ITS OPTIMIZATION FOR SOME PROBLEMS

Training effectiveness of the Multi layer neuron networks depends on the quantity of neuron layers within the network, quantity of neurons in the hidden layer and initialization of synaptic connections' weight coefficients.

Keywords: neuron network, artificial neural network, synaptic connection.

Neuron network's architecture has a great significance for the effectiveness of training. The dimension of the input and output layers depends on the problem for solution and the structure of the data collection. As mentioned earlier, three layer neuron network can approximate any function with the preliminary exactness. Besides, the exactness of the approximation depends on the quantity of the neurons in the hidden layer.

As mentioned earlier, the problem in the algorithm of error's back propagation was that it didn't allow for choosing preferable speed of training, which was needed for the increase of algorithm's rapidity and ensure its collectability.

The given equation cannot be solved with the analytic methods to apply to $\alpha(t)$. Therefore, [1] research suggests to apply axial flow methods in order to find adapt rapidity. However it is related with the large volume calculations. Therefore, we can suggest approximating method for finding α(t) which depends on the expansion of activation function of the neuron elements in the Taylor line. I will discuss it in details: if we calculate the j neuron's output value in the last layer of the neuron networks:

$$y_j(t) = F\left(S_j(t)\right),$$

$$s_j(t) = \sum_i y_i(t) w_{ij}(t) - T_i(t) \quad (1)$$

Where $y_i(t)$ is the output value of the hidden layer's j neuron. To learn the value (1) of the j neuron in the t+1:

$$S_j(t+1) = \sum_i y_i \left(w_{ij} - \alpha \frac{\partial E}{\partial w_{ij}}\right) - T_j + \alpha \frac{\partial E}{\partial T_j} =$$

$$= \sum_i y_i w_{ij} - T_j + \alpha \cdot \left(\sum_j y_j \cdot \frac{\partial E}{\partial w_{ij}} - \frac{\partial E}{\partial T_j}\right). \quad (2)$$

In order to do so, I have to introduce the significations:

$$a_j = \sum_j y_j \cdot \frac{\partial E}{\partial w_{ij}} - \frac{\partial E}{\partial T_j}. \quad (3)$$

Which represents the figure (3) in the following way:

$$S_j(t+1) = S_j(t) - \alpha \cdot a_j. \quad (4)$$

By the (t+1) momentum the j neuron's output value will be equal:

$$y_j(t+1) = F\left(S_j(t+1)\right).$$

If expand be the last figure in the Taylor line with the precision of the first two components:

$$y_j(t+1) = F(0) + F'(0) \cdot F\left(S_j(t+1)\right), \quad (5)$$

Where

$$F'(0) = \frac{\partial F}{\partial s_j},$$

When $S_j = 0$. figure (5) considered in the figure (4) will give:

$$y_j(t) = F(0) + F'(0) \cdot S_j(t) - \alpha F'(0) a_j, \quad (6)$$

because:

$$y_j(t) = F(0) + F'(0) \cdot S_j(t).$$

Then figure (6) can be represented the following way:

$$y_j(t+1) = y_j(t) F(0) - \alpha F'(0) a_j \quad (7)$$

In order to find the adaptive training rapidity the following should be procured:

$$E = \frac{1}{2} \sum_j \left(y_j(t+1) - t_j\right)^2 \to min$$

then

$$\frac{\partial E}{\partial \alpha} = \sum_j \left(y_j(t) - t_j - \alpha F'(0) a_j\right) \cdot \left(-F'(0) a_j\right) = 0$$

if we take $\alpha(t)$ from the last figure, we will compute:

$$\alpha(t) = \frac{\sum_j (y_j(t) - t_j) a_j}{F'(0) \sum_j a_j^2}, \quad (8)$$

if $\frac{\partial^2 E}{\partial \alpha^2} > 0$ then for the given α(t) can be reached minimum of the average square deviation. We have to find depiction for a_j – therefore, we will introduce the definitions:

$$\frac{\partial E}{\partial w_{ij}} = \frac{\partial E}{\partial y_j} \cdot \frac{\partial y_j}{\partial S_j} \cdot \frac{\partial S_j}{\partial w_{ij}} = (y_j - t_j)F'(S_j)y_i,$$

$$\frac{\partial E}{\partial T_j} = \frac{\partial E}{\partial y_j} \cdot \frac{\partial y_j}{\partial S_j} \cdot \frac{\partial S_j}{\partial T_j} = -(y_j - t_j)F'(S_j). \quad (10)$$

If we insert (9) and (10) in the (3) figure, following will be formulated:

$$\alpha_j = (1 + \Sigma_i y_i^2) \cdot (y_j - t_j) \cdot F'(S_j). \quad (11)$$

Due to the principle of the layers independence we can assume that

$$\gamma_j = y_j - t_j \quad (12)$$

with consideration of (11) in (12) and (7) we can obtain formula to calculate approximate value of the adapt rapidity for the neuron network training.

$$\alpha(t) = \frac{\Sigma_j \gamma_j^2 F'(S_j)}{F'(0) \cdot (1 + \Sigma_i y_i^2)(\Sigma_j \gamma_j^2 (F'(S_j))^2)}. \quad (13)$$

Where γ_j – j- is neuron element's error which can be calculated for various layers of the network.

If have hyperbolic tangent in the role of activation function, we can compute the following:

$$y_j' = F'(S_j) = \frac{1}{ch^2(S_j)} = (1 - y_j^2),$$

$$y_j'(0) = F'(0) = 1.$$

Consequently:

$$\alpha(t) = \frac{\Sigma_j \gamma_j^2 (1 - y_j^2)}{(1 + \Sigma_i y_i^2)\Sigma_j \gamma_j^2 (1 - y_j^2)^2}.$$

In which the error of the hidden layer's j neuron element will be:

$$\gamma_j = \sum_i \gamma_i (1 - y_i^2) w_{ij}.$$

Should be mentioned, that α(t) was calculated on separate basis for each layer of the neuron network in the above representations. As the experiment proved it, adapt rapidity of training can be used to obtain major values of α(t). This can lead to the de-synchronization of the training process, when weight coefficients of the synaptic connections change drastically in certain direction. As a result, the average square error's change will acquire time dependent wave mode. It is recommended to limit α(t) according to the absolute value. The above figures, obtained for the adapt rapidity for training can enable us to increase training rapidity significantly and to avoid the problem of random selection of the neuron network's training rapidity. This is significant priority in comparison with the standard algorithm of the error's back propagation. However, in case there is a good selection of the training's constant rapidity, the functioning time of the given algorithm will not exceed the functioning time of the gradient descending algorithm [2].

References

1. К вопросу об использовании искусственной нейронной сети для прогнозирования валютных курсов на высоковолантильных рынках / И. Дидманидзе, З. Мегрелишвили, Г. Кахиани // Материалы I международной научно-практической интернет-конференции «Сучасні тенденції розвитку математики та її прикладні аспекти». – Донецк : [б. в.], 2012. – С. 166-167.
2. Hortz J. Introduction to the theory of neural computation / J. Hortz, A. Krogh, R. Palmer // Addison Wesley Publishing Company, 1991. – 327 p.

Дідманідзе І. Ш., Кахіані Г. О.

НЕЙРОННА МЕРЕЖА ТА ШЛЯХИ ЇЇ ОПТИМІЗАЦІЇ

Ефективність багатошарових нейронних мереж залежить від кількості шарів нейронів у межах мережі, кількості нейронів у прихованому шарі та ініціалізації коефіцієнтів ваги синаптичних зв'язків.

Ключові слова: нейронна мережа, штучна нейронна мережа, синаптичний зв'язок.

Матеріал надійшов 31.07.2013