

# Neural Network Optimization Algorithm Model Combining L1 / 2 Regularization and Extreme Learning Machine

**Anzhi Qi**

Liaoning Jianzhu Vocational College, Liaoyang, Liaoning, 111000

**Keywords:** ELM, L1/2 Regulation, Neural Network Optimization

**Abstract:** Extreme Learning Machine (ELM) is a fast learning algorithm that uses random mechanism to reduce parameter setting and selection, thereby greatly improving learning speed and ensuring generalization ability. Unlike traditional learning methods, ELM The variables are no longer iterative but randomly generated, so that a nonlinear system that expresses a forward neural network can be reduced to a linear system that only needs to compute the output weights, and the least-squares method can be used to solve the linear system directly. Although the ELM algorithm However, the effectiveness of this algorithm in the application of large-scale data needs to be further improved. This thesis is based on L1 / 2 regularization theory and full rank Cholesky matrix decomposition respectively Two Improved ELM Algorithms Applied to Large Scale Data.

## 1. Introduction

In recent years, neural networks as a highly efficient learning method is widely used in various fields. However, the traditional feedforward neural network training algorithm has the problem of long training time. In 2006, Huang Guangbin proposed a new neural network called extreme learning machine (ELM). Because extreme learning machine as a single hidden layer feedforward neural network, its input layer to hidden layer weights and bias values are randomly selected, while the hidden layer to the output layer weight vector obtained directly by solving a linear system , So ELM can greatly improve the learning speed of the network. At the same time, Huang Guangbin gives the theoretical support of extreme learning machine, experimental data show that ELM can not only fast training, but also has good generalization ability. In recent years in many fields of research has been widely used.

## 2. Fast Algorithm Based On L1 / 2 Regularization

Neural networks originated in the biological system, which is an intelligent bionic network formed in an attempt to find a mathematical representation of information processing. Neural networks are parallel processing systems composed of a large number of simple neurons. This processing method is inspired by the biological nervous system. In fact, neural networks have been widely used in different fields, including pattern recognition, image processing, time series analysis and control systems. This section introduces a new neural network model that has emerged in recent years, the extreme learning machine (ELM). Extreme learning machine was first proposed by Huang. In the extreme learning machine, the weights and bias values connecting the input layer to the hidden layer are randomly selected, and the weights connecting the hidden layer to the output layer can be calculated by the least square method. As a result, extreme learning machines learn much faster than traditional training methods.

Extreme learning machine is a single hidden layer feedforward neural network, Figure 1 for the network structure. The activation functions of hidden layer and output layer are respectively sigmoid function and linear function, and the expressions of sigmoid function and linear function are as follows:

$$f(x) = \frac{1}{1 + \exp(-x)}, \quad g(x) = x, \quad (x \in \mathbb{R})$$

Let  $N$ , different random samples  $\{x_i\}_{i=1}^N$ ,  $p \times p = x^T$ , where  $d_{px} \in \mathbb{R}$ ,  $m_{pt} \in \mathbb{R}$ . For the input vector, the corresponding output of the network can be expressed as  $p \times 1$   $(y) = H_N \beta$ ,  $N = \sum \beta$

$w \cdot x + b$  where  $H_N$  is the number of implicit nodes connected to the  $j$ th implicit node and  $w_j$  is the weight vector of the input node, which is the offset value of the  $j$ th implicit node, is the inner product of the sum and is the weight vector connecting the  $j$ th implicit node and the output node.

For extreme learning machines, the weight vectors and bias values that connect the input layer to the hidden layer are chosen randomly, so the nonlinear system of the extreme learning machine can be transformed into a linear system:  $Y = H\beta$ ,

In order to further improve the stability and generalization ability of extreme learning machine, Huang Guangbin proposed an extreme learning machine with equality constraint optimization. The optimal weight expression of network training is:  $\beta = (C^{-1} + I)^{-1} C^{-1} H^T Y$ ,

Where  $C$  is a constant and  $I$  is a unit matrix. Let  $C = \lambda^{-1}$ , then the above formula can be expressed as:  $\beta = (H^T H + I)^{-1} H^T Y$ .

Where  $\lambda$  is the regularization parameter, which plays an important role in controlling the trade-off between training errors and the complexity of the model.

### 3. Neural Network Optimization

Artificial neural network is a mathematical model constructed by using bionics for information processing, which can well simulate the information dissemination mechanism of the brain's nervous system. The network model is interconnected by many hidden nodes (neurons) according to a certain rule. Information processing is completed through the dynamic process of neuronal interaction, each node is equipped with an adder and an Activation Function, the nodes between adjacent hidden layers pass the weights. This network increases the network complexity by increasing the number of hidden layers and the number of neurons per layer and achieves the purpose of processing information by adjusting the corresponding connection weights. In most network models, the value is finally determined by means of iterative optimization with the aid of a specific optimization algorithm. Iteration of the network is usually terminated when a certain training accuracy or a certain upper limit of the number of iterations is reached. At the same time, the connection weight of the network is finally determined, This process can also be considered as a "memory" of the constructed artificial neural network, which achieves the goal of learning with network parameters. Method to simulate the effect of a potential law between a given sample input and output and then use the resulting network to predict other data of that type, also known as the network generalization process.

BP neural network is one of the most widely used neural networks for power load forecasting. However, there are some shortcomings in the standard BP neural network algorithm. Here we introduce some commonly used typical improvement methods.

1) Increase the momentum term. In the general algorithm of BP neural network, the weight of each layer in the model is adjusted in the way of gradient descent according to the error curve at time  $t$ , and the gradient between them is not considered here. Direction, if adjusted in this way will result in unstable training process, prone to oscillation, resulting in slow convergence process results. Therefore, some scholars have accelerated the process of network training in order to speed up the convergence process.

In this formula,  $W$  represents the weight matrix of each layer in the BP neural network,  $O$  represents the output vector matrix of each layer in the neural network,  $\alpha$  is called the momentum coefficient factor of the neural network, and its value In the range of 0 to 1, during the process of training the network, if its error gradient network has a local minimum phenomenon, although the first one here will tend to hug the zero, but this one  $t W, \square \square 0) 1$  (, so this training process will be avoided limited to the local minimum area of the situation, thus accelerating its training speed, so that the neural network convergence speed, so this with a momentum factor BP neural network algorithm applied to a lot of BP network.

Learning speed adaptive adjustment method. Learning speed  $\eta$  in a standard BP neural network is a constant that we also call the step size, and in the actual operation, it is difficult to find a value as the optimal learning speed. We can see from the graph of the error surface that when the area in the surface is in a flat area, we need to set a relatively large value of  $\eta$  so that it can jump out of the

flat area; and when the area in the surface is in a relatively large change Of the area, then the value of  $\eta$  at this time we need to reduce or increase its operation. Self-adaptive adjustment of learning speed  $\eta$  can be adjusted according to the total network error self-adjustment in the network after several adjustments, if  $E$  total continued to rise, then the adjustment here is invalid, and  $\eta = \beta\eta$ ,  $\beta < 1$ ; Often adjusted here  $E$  total decline, it shows that the adjustment here is effective, and  $\eta = \alpha\eta$ .

Introduce a steepness factor (to prevent saturation). During the course of network training, due to the flat surface of its error surface, the S-type activation function has the characteristic of saturation when it is in this area, which will slow down the adjustment of weight and thus affect the speed of adjustment. In the process of training, if the algorithm adjusts into this region, we can reduce the input of neurons so that the output value is rapidly separated from the saturation region of the activation function, and the value of the error function will change accordingly. The value adjustment is also out of the flat area. To achieve the above ideas need to introduce a steepness factor  $\lambda$  in the activation function.

Because ELM learning algorithm randomly selects the center and width of hidden neurons, which can improve the learning speed of algorithm, these centers and widths are not the best choice relative to the input data. In practice, in order to make the neural network have better performance, it may need more hidden layer neurons, thus increasing the network complexity. In order to obtain a relatively simple network structure, reduce the number of neurons in the hidden layer of the network and improve the performance of the neural network, a population intelligent optimization algorithm is adopted to adjust the center and width of the hidden neurons. The group intelligent optimization algorithm is essentially a kind of probability search algorithm, which does not need the gradient information of the problem, and all are optimization algorithms based on the group intelligence theory. The group intelligence guidance and optimization search through the cooperation and competition among individuals in the group is a good global search algorithm. Compared with the general evolutionary algorithms, the DE algorithm retains the global search strategy based on population, which uses real number coding, simple mutation operation based on difference and one-to-one competition survival strategy to reduce the common intelligent optimization algorithm. The complexity of genetic manipulation. At the same time, DE unique memory capabilities make it possible to dynamically track the current search conditions to adjust its search strategy, with strong global convergence and robustness, and does not need to rely on the feature information of the problem.

The design of an artificial neural network applied to practical problems can be attributed to the study of the network's own weight parameters and the rational design of network topology, which are the two major types of optimization problems. Because the training of artificial neural network is a very complex problem, it takes a long time to train using the traditional learning algorithm. The gradient descent method is used in the BP algorithm to find the appropriate parameter values in a certain range Model structure. Therefore, in order to improve the generalization ability of neural networks better and to design the network topology more reasonable, a large number of optimization algorithms about neural networks are successively produced.

#### **4. Conclusion**

Based on the extreme learning machine, this paper proposes an extreme learning machine based on regularized correlation entropy. Due to the extreme learning machine, the input layer to the hidden layer of the weights and bias values randomly assigned way to significantly reduce the network learning time. Therefore, the methods proposed in this paper all adopt the model of extreme learning machine, then only the weight and offset of hidden layer to output layer need to be calculated. However, the MSE criterion of extreme learning machine is sensitive to noise. Therefore, this paper replaces the MSE criterion with the entropy criterion of anti-noise and introduces regularization term to make the network have good anti-noise performance. This paper describes and verifies the regularization-related entropy based on L2 norm and the regularization-related entropy based on L1 norm.

**References**

- [1] Toh K A. Deterministic neural classification [J]. *Neural Computation*. 2008, 20 (6): 1565–1595.
- [2] Huang G B, Zhu Q Y, Siew C K. Universal approximation using incremental constructive feed-forward networks with random hidden nodes [J]. *IEEE Transactions on Neural Networks*. 2006, 17 (4): 879–892.
- [3] Huang G B, Chen L. Enhanced random search based incremental extreme learning machine [J]. *Neuro-computing*. 2008, 71: 3460–3468.
- [4] Feng G R, Huang G B, Lin Q P, et al. Error minimized extreme learning machine with growth of hidden nodes and incremental learning [J]. *IEEE Transaction on Neural Networks*. 2009, 20:1352–1357.
- [5] Huang G B, Ding X, Zhou H. Optimization method based extreme learning machine for classification [J]. *Neuro-computing*. 2010, 74: 155–163.