Multi-objective Model Selection for Extreme Learning Machine

Liyun Wang[1 a], Zhenshen Zhu[1 b] and Bin Sun[1 c]

[1]School of Information Engineering, Zhengzhou University of Industrial Technology, Zhengzhou Henan 451150, China

[a]912725921@qq.com, [b]254337266@qq.com, [c]709579637@qq.com

Keywords: Extreme learning machine; Generalization performance; Multi-objective optimization; Model selection

Abstract. Recently, Extreme Learning Machines (ELMs) have get successful application in the fields of classification and regression. However, the generalization performance of ELM will be decreased if there exits non-optimal input weights and hidden biases. To solve this problem, this paper introduced a new model selection method of ELM based on multi-objective optimization. This method views ELM model selection as a multi-objective global optimization problem, in which the generalization error and output weights are as optimization objectives. To accelerate the optimization speed, a fast Leave-one-out (LOO) error estimate of ELM is introduced to refer to the generalization error. Taking into account the contradiction between these two objectives, multi-objective comprehensive learning particle swarm optimization algorithm is utilized to find non-dominated solutions. Experiment on four UCI regression data sets are conducted.

Introduction

As an important branch of the neural network, a professor Huangguang Bin Nanyang Technological University [1] proposed in 2004 Extreme Learning Machine (ELM), ELM is the former single hidden layer feed-forward neural network (SLFN) learning algorithm, Compared with traditional BP algorithm, ELM simply set the number of neurons in the hidden, it's not need to adjust the input parameters of the network, without iterative solver, with "extreme" fast features.

However, ELM for redundant hidden neurons weaken shortcomings model generalization ability, Mao Wentao [2], who proposed a hidden feature space ELM based model selection algorithm. According incremental learning point of view, Feng [3] and others based on the increased number of hidden neurons to calculate the generalization error, the final choice corresponds to the minimum error of the network structure. Liu apprenticeship [4] constructed a fast left cross validation algorithm, generalization of almost unbiased estimates. In order to improve the accuracy of the model, Wang Jie [5] proposed particle swarm ultimate learning machine algorithm can rely on metadata saphenous nerve fewer higher accuracy.

Especially for incremental learning, time series and other issues, the offline learning stage, the choices are good ELM model, it helps to improve the effectiveness of online learning phase. Zhu [6],who proposed an ELM model selection algorithm, when the number of neurons increases hidden by an improved micro-evolutionary algorithm to select the input weights and bias, which the root mean square error of validation set (RMSE) as a model generalization error estimate. this paper presents[7] a multi-objective optimization ELM model selection policy; ELM model selection problem as a multi-objective optimization problem.

Extreme Learning Machine Introduction

For N any different sample (x_i, y_i), among them $x_j = [x_{j1}, x_{j2}, x_{j3}, \ldots, x_{jn}]^T \in R^n$, $y_i = [y_{i1}, y_{i2}, y_{i3}, \cdots y_{im}]^T \in R^m$, L is having a hidden neuron, activation function for the output of the ELM can be expressed as $G(x)$

$$f(x_j) = \sum_{i=1}^{L} \beta_i G(a_i \cdot x_j + b_i), a_i \in R^n, \beta_i \in R^m \tag{1}$$

them $a_i = [a_{i1}, a_{i2}, a_{i3}, \cdots a_{in}]^T$ Is input to i input weight vector saphenous nerve, b_i is the i hidden neuron bias, $\beta_i = [\beta_{i1}, \beta_{i2}, \beta_{i3}, \cdots \beta_{in}]$ It is connected to the i hidden neuron output weights; $a_i \cdot x_i$ represents the inner product of vectors a_i and x_i.Where in the excitation function is "Sigmoid"[8] function.

If L contains a hidden neuron feed forward neural network can approach zero error that N samples [5]. a_i, b_i, β_i there is, so

$$f(x_j) = \sum_{i=1}^{L} \beta_i G(a_i \cdot x_j + b_i) = y_i, \tag{2}$$

$$x_j \in R^n, a_i \in R^n, \beta_i \in R^m$$

(2) Formula can be simplified

$$H\beta = Y \tag{3}$$

Thus, the output layer parameter β can be represented by the formula (3) was the least-squares solution

$$\beta = H^+Y \tag{4}$$

Based on the above analysis [9] can be obtained, ELM can be summarized in the following

Step 1: randomly generated weights and bias
$(a_i, b_i), i = 1,2,\cdots, L$.

Step 2: Calculate the hidden layer output weighting matrix H.

Step 3: Calculate the output weights $\beta = H^+Y$.

Multi-Objective Model Selection

As can be seen from (1) the equation, the number of hidden neurons directly determines the complexity of the structure and model of the network, it is clear, redundant neurons affects generalization ability of the model.

Optimization Strategy. Generally looking for model training error is zero, but in practice the error is there, in fact, as long as the training error is within an acceptable range, which:

$$\min \sum_{i=1}^{N} \| \beta \cdot h(x_i) - y_i \| \tag{5}$$

$$\min \|\beta\| \tag{6}$$

This paper use of multi-objective comprehensive learning particle swarm optimization (MOCLPSO)[10] find non-dominated solutions. MOCLPSO can effectively use historical information, the use of short-range structure based strategy update frequency and location.

Optimization algorithm to find non-dominated solutions with multi-objective particle swarm optimization, multi-objective optimization problem solving method there are two types: multi-objective optimization problem into a single-objective optimization problem to solve and multiple targets simultaneously direct optimization.

Optimize the Target. Due to the large formula (5) calculation of ratio, so verification algorithm by a left cross to calculate the error of the model. Liu Arts [4] constructed ELM leave a fast cross-validation algorithms can be selected and evaluated ELM model quickly.

$$f_i(x_i) = \frac{H_{x_i} H^+ Y - \left(H_{x_i} H^+\right)_i y_i}{1 - \left(H_{x_i} H^+\right)_i} \tag{7}$$

Under $1-\left(H_{x_i} H^+\right)_i \neq 0$ conditions established, the i iteration condition is established when the error Leaving a cross-validation in Dir iteration condition is established when the error

$$r_i = y_i - f_i(x_i) = \frac{y_i - H_{x_i} H^T Y}{1 - \left(H_{x_i} H^T\right)_i} \tag{8}$$

$$LOO_{ELM} = \frac{1}{N} \sum_{i=1}^{N} (r_i)^2 \tag{9}$$

From the foregoing, the goal of this paper is to optimize the amount of two mutually exclusive, that left a error estimation (9) and output weights mold (6). (8) shows a cross-validation error calculation remain closed-form expression, N times without training ELM model to calculate a one-time complete, you can then by the formula (8) to calculate a left cross validation error vector.

Description of the Algorithm. ELM model selection into looking for the smallest ϕ, that is looking to stay a minimum cross-validation error output and minimum weight, which can be described as the following form:

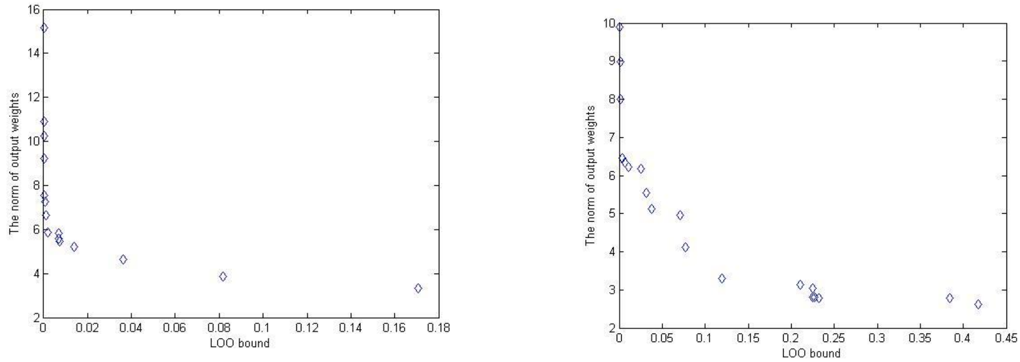
$$\bar{\phi} = \arg\min_{\phi} E(\phi) = \arg\min_{\phi} \left(LOO_{ELM}(\phi), \|\beta\|\right) \tag{10}$$

Algorithm in three steps, step 1 the number of neurons are randomly generated, individual data is normalized to [-1, 1]. Step 2 multi-objective particle swarm optimization algorithm to find non-dominated solutions. Step 3 Increase hidden neuron number, return to step 1, select the best input weights and hidden bias.

Simulation

To verify the model's performance, the paper selection and the traditional BP neural network, ELM and E-ELM comparison. Using Levenberg-Marquardt (LM) algorithm BP neural network, LM algorithm has the fastest speed in the BP training. The paper selection and the traditional BP neural network, LM algorithm has the fastest speed in the BP training.

Fig. 1 depicts a left error estimation and output weights die mutually exclusive. In UCI dataset Boston Housing, Corresponding decision function shown in Fig. 2. The red line shows "sinc" function. the UCI regression data sets ELM-FLOO and LM, ELM, E-ELM comparison test. Table 1 and Table 2 shows the experimental results.



(a) (b)

Figure 1. (a) and (b) different hidden neurons

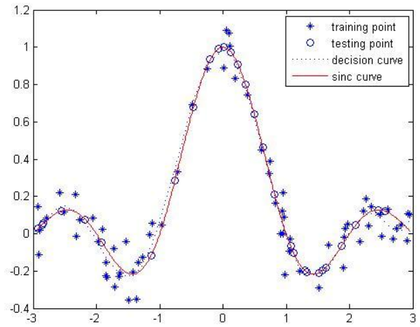


Figure 2. Sinc ELM-FLOO predict the effect on the data set

As it can be seen from Table 1, both in the training set or test set, ELM-FLOO have made less errors than the other three algorithms. As can be seen from Table 2, the Pyrim set, test error ELM-FLOO are smaller than the other three algorithms, in all five data sets, the number of hidden layer neurons ELM-FLOO algorithm than for each data set less other three algorithms. Tables 1 and 2 show, ELM-FLOO than E-ELM, ELM and LM test error is small.

Conclusion

In this paper, ELM model selection method based on multi-objective optimization. The method is to select the model as a global multi-objective optimization problem, which will leave a fast cross-validation error and output weights mold as the optimization target, and the use of multi-objective comprehensive learning particle swarm optimization algorithm to find non-dominated solutions.

Acknowledgements

Funding: National Natural Science Foundation of China (No: U1204609), based in Henan Province and cutting-edge technology research projects (number: 132300410430)

Table 1. Decomposition fixed data comparison test

Dataset	Algorithm	Time /s	Root mean square error		Hidden neurons
			Training	Test	
Housing	LM	185.23	2.8616	3.0288	60
	ELM	0.0706	2.5136	3.2765	75
	ELM-FLOO	348.65	0.1356	0.1126	30
Bodyfat	LM	340.24	6.02e-05	0.0461	50
	ELM	0.4531	0.0012	0.0086	43
	ELM-FLOO	187.59	3.9209e-05	2.3103e-04	29
Mpg	LM	385.17	1.8443	3.4672	45
	ELM	0.0701	2.3566	2.8965	50
	ELM-FLOO	304.28	0.0879	0.0809	29
Triazines	LM	340.60	0.1023	0.5251	30
	ELM	0.0483	0.0621	0.05732	45
	ELM-FLOO	222.91	0.0042	0.0042	24

Table 2 Random data into the comparison test

Dataset	Algorithm	Time /s	Root mean square error		Hidden neurons
			Training	Test	
Housing	LM	520.10	2.6031	4.2305	56
	ELM	0.0343	2.8016	4.0312	75
	ELM-FLOO	348.65	4.4551	4.5072	20
Bodyfat	LM	106.24	0.0021	0.0053	40
	ELM	0.3082	0.0017	0.0032	55
	ELM-FLOO	109.39	0.0026	0.0035	15
Mpg	LM	1024.5	2.3048	2.5481	45
	ELM	0.0741	2.4573	2.4736	53
	ELM-FLOO	151.38	2.8036	2.5641	15
Triazines	LM	332.34	0.0783	0. 3427	48
	ELM	0.0621	0.1346	0.1431	59
	ELM-FLOO	151.62	0.1357	0.1324	20

References

[1] Huang GB, Zhou QY, Siew CK Extreme learning machine: Theory and application [J]. Neurocomputing , 2006,70:489-501.

[2] Mao WT, Zhao ZT, He HH. ELM implicit model selection feature space-based [J]. computer application 2013,33(6):1600-1603.

[3] Feng GR, Huang GB, Lin QP, Gay R Error minimized extreme learning with growth of hidden nodes and incremental learning. IEEE Trans Neural Netw, 2009, 20:1352-1357.

[4] Liu XY, Li Ping, Gao CH. ELM leave a fast cross-validation algorithm [J]. Shanghai Jiaotong University, 2011, 45(8):6-11.

[5] Wang Jie, Bi HY. Based on Particle Swarm Optimization Extreme Learning Machine [J]. Zhengzhou University: Natural Science, 2013, 45(1):100-104.

[6] Zhu QY, Qin AK, Suganthan P, Huang GB Evolutionary extreme learning machine. Pattern Recognit, 2005, 38:1759-1763.

[7] Huang GB, Zhou QY, Siew C (2004) Extreme learning maching: a new learning scheme of feed forward neural networks. In: Proceedings of Internation Joint Conference on Neural Networks, 2004, 2:985-990.

[8] Jun Y, Er MJ, An enhanced online sequential extreme learning machine algorithm [J]. Proceedings of Control and Decision Conference, 2008, 2902-2907.

[9] Turney P. Types of Cost in Inductive Concept Learning [J].The computer Research Repository.2002:15-12.

[10] P Jeatrakul, KW Wong, CC Fung, Classification of Imbalanced Data by Combining the Complementary Neural Network and SMOTE Algorithm[J],Neural Information Processing.2010,6444:152-159.