International Journal of Electrical and Electronics Engineers
Vol. No. 9, Issue No. 01, January-June 2017
ISSN (O) 2321-2055
ISSN (P) 2321-2045

EXTREME LEARNING MACHINE ALGORITHM AND ITS APPLICATION

R.ManjuParkavi[1], M.Shanthi[2], M.C.Bhuvaneshwari[3]

[1]*Student, Department of ECE, Kumaraguru College of Technology (India)*
[2]*Associate Professor, Department of ECE, Kumaraguru College of Technology (India)*
[3]*Associate Professor, Department of EEE, PSG College of Technology, (India)*

ABSTRACT

Extreme learning machine (ELM) is an efficient learning algorithm for the single hidden layer feed forward neural networks. Compared with the other conventional neural network algorithm it has the advantage of over-fitting problems and slow training speed. ELM is based on the empirical risk minimization theory and its learning process needs single iteration only. The algorithm avoids local minimization and multiple iterations. It has been used for various fields and applications because of the better generalization, controllability and robustness and fast learning rate. This paper makes a review of ELM latest research progress about the algorithms and applications. It first analyzes the algorithm ideas of ELM, and then describes the latest progress of ELM in current years, including the modeling and specific applications of ELM, finally state the research and development of ELM in the future.

***Keywords* - Extreme Learning Machine (ELM) · Single-Hidden Layer Feed forward Neural Networks (SLFNS) Local Minimum · Over-Fitting · Least-Squares**

I INTRODUCTION

Early in 1940s, mathematician Pitts and psychologist McCulloch developed a mathematical model (MP model) from the mathematical logic view (McCulloch and Pitts 1943) that opened the introduction of artificial neural network (ANN) research. Neural network has the strong adaptive self-learning and robustness and fault tolerance characteristics and nonlinear mapping ability.

In trading with small samples, nonlinear adaptive performance issues ELM has many more unique advantages. Its unique nonlinear adaptive processing capacity can overcome the other traditional artificial intelligence methods; they are speech recognition, pattern recognition, and unstructured information processing deficiencies. Coupled with its solid theoretical basis and simple network structure model make the neural network in the fields of image processing, pattern recognition, signal processing, sensors and automatic control have significant results (Quteishat and Lim 2008; Zhang and Wang 2009;Ding et al. 2011a,b, 2012; Ding and Jin 2013). It is generally used in the fields of prediction (Hagan et al. 2002) expert system (Markowska- Kaczmar and Trelak 2005),), intelligent control (Ding et al. 2011a,b), pattern recognition (Mohamed 2011) and combinatorial optimization (Kahramanli and Allahverdi 2009.

At present, there are several kinds of neural network model like BP network (Ding et al. 2011a,b; Feng et al. 2009), RBF network (Ding et al. 2011a,b), Hopfield net- work, CMAC cerebellar model, etc. The efficient computing capabilities of the neural network are achieved by the propagation of information between the neurons. By direction of the neural network internal information transfer, the neural network can be separated into two group : feedforward neural network and feedback type neural networks. Extreme learning machine (ELM) described here is for single hidden layer feedforward neural network which is one method of feedforward neural networks.

Feedforward neural network model has been generally used in many fields due to its ability to approximate complex nonlinear mappings directly from the input samples. The majority of studies focused on the input samples, divided into two aspects of the finite set and compact set. Hornik (1991) proved that if the activation function is continuous, nonconstant and bounded, then continuous mappings can be approximated by neural networks over sets of compact input. On the basis of Leshno et al. (1991) improved the results and proved that non-polynomial activation function can approximate the continuous functions. Through further study containing N different instance in a finite training set, Huang and Babri (Huang and Babri 1998; Huang 2003) states that the nonlinear activation function can exactly approximate N distinct instances.

Single hidden layer feed forward neural network has very powerful learning ability, that can approximate a complex nonlinear function and it will solve the problems those cannot be solved by the old traditional parameter learning method. However, due to the lack of fast learning method, often cannot meet the real demand. To solve the above mentioned problem, a new and effective algorithm which is called the ELM has been proposed by Huang et al. (2006). As a new proposed learning algorithm can easily solve the traditional feedforward neural network, such as local minimum problem of BP network. ELM can easily set the hidden layer node number and randomly assigned for the input weights and hidden layer biases, then the output layer weights is determined by using the least square method, the whole process without iteration and improves the learning ability and neural network generalization ability.

The extreme machine learning algorithm proposed has been vastly used and it become the research focus of data mining, image processing, machine learning and other areas. This paper, deals with the review of extreme machine learning latest research development about the algorithms, and applications. This paper is organized as follows. Section II introduces the basic idea and algorithm model of ELM and further summarizes some improved ELM algorithms and its model. The applications of ELM in related fields are presented in last of Sect. II. Conclusions and prospects are given in Sect.III.

II EXTREME LEARNING MACHINE

2.1 ELM Introduction

Extreme learning machine is a fast growing learning algorithm for the single hidden layer feedforward neural networks used in both classification and regression problems. The ELM used for single hidden layer feedforward neural network training can set the node number of hidden layer and assign the input weights and hidden layer biases randomly, the output layer weights is calculated by the least square method, the entire learning process finished through one mathematical change without the requirement of iteration. The training speed compared with the traditional BP algorithm has been remarkably improved (usually 10 times or more) (Deng et al.).

In practical applications, first it trains the ELM and then predict. The data set used for training is mainly combined with some specific issues. The data sets include real results and its related factors. During training process, the influence factors and the results will be put into ELM for training stage, through an iteration to complete learning process. Then, with the trained ELM t, only need to input and the training data set is similar to the influencing factors. ELM model can be determined by the prediction results according to the memory.

Extreme learning machine is easy to implement and powerful algorithm for single hidden layer feedforward neural network. The old traditional neural network learning algorithm (e.g. BP algorithm) needed to set up lots of artificial network training parameters, and can easily lead to local optimal solution.

2.2 ELM Algorithm

For the N distinct samples (x_i, t_i), where $x_i = [x_{i1}, x_{i2}, \ldots, x_{in}]^T \in R^n$ and $t_i = [t_{i1}, t_{i2}, \ldots, t_{im}]^T \in R^m$, what's more $(x_i, t_i) \in R^n \times R^m (i = 1, 2, \ldots, N)$, standard SLFNs with N hidden nodes and activation function f(x) are mathematically modeled

$$\sum_{i=1}^{N} \beta_i f_i(x_j) = \sum_{i=1}^{N} \beta_i f(a_i . x_j + b_i) = t_j, j = 1, \ldots, N \quad (1)$$

here $a_i = [a_{i1}, a_{i2}, \ldots, a_{in}]^T$ is the weight vector connecting the i th node of hidden node and the input nodes, and b_i is the threshold of the i th node of hidden node. $\beta_i = [\beta_{i1}, \beta_{i2}, \ldots, \beta_{im}]^T$ is the weight vector connecting the both i th hidden node and the output nodes. $a_i \cdot x_j$ represents the inner product of a_i and x_i, and the activation function generally choose "Sigmoid", "RBF" and "Sine",

The above Eq. (1) can be written compactly as

$$H\beta = T \quad (2)$$

where $H(a_1, \ldots, a_{\tilde{N}}, b_1, \ldots, b_{\tilde{N}}, x_1, \ldots x_N)$

$$= \begin{bmatrix} f(a_1.x_1 + b_1) & \ldots & f(a_N.x_1 + b_N) \\ \vdots & \vdots & \vdots \\ f(a_1.x_N + b_1) & \ldots & f(a_N.x_N + b_N) \end{bmatrix}, \beta = \begin{bmatrix} \beta_1^T \\ \vdots \\ \beta_N^T \end{bmatrix}, T = \begin{bmatrix} t_1^T \\ \vdots \\ t_N^T \end{bmatrix}$$

H is the output matrix of hidden layer for the neural network; the i th column of H is the i th hidden node output with respect to inputs $x_1, x_2, \ldots, x_N$.

When the input weights and hidden layer biases are determined by random assignment, then the input samples can

International Journal of Electrical and Electronics Engineers
Vol. No. 9, Issue No. 01, January-June 2017
ISSN (O) 2321-2055
ISSN (P) 2321-2045

get the hidden layer output matrix H. Therefore, training SFLN is converted into solving linear equations $H\beta = T$ least squares solution.

$$\left\| H(a_1,\ldots,a_{N^*},b_1,\ldots,b_{N^*})\hat{\beta} \dot{-} T \right\| = min_{\beta} \left\| H(a_1,\ldots,a_{N^*},b_1,\ldots,b_{N^*})\hat{\beta} \dot{-} T \right\| \quad (3)$$

The above Eq. (3) least squares solution of the above liner system is

$$\beta = H^{\dagger} T \quad (4)$$

In the Eq. (4), $H^{\dagger}$ representing then Moore–Penrose (Liang and Huang 2006) generalized inverse of the hidden layer output matrix H .Usually, output weight βcontains the following features:

(1) According to β, the algorithm can get the minimal error in training phase;

(2) Can get the optimal generalization capability of the minimum paradigm of the output connection weights and network;

(3) $\beta$is unique. It can avoid producing the local optimal solution.

In summary, given a training set $(x_i , t_i) \in R^n \times R^m(i = 1, 2,..., N)$,activation function f(x) and hidden node number N, then the ELM algorithm can be evaluated in upcoming three steps:

Step 1: define the node number of hidden layer N, assign input weight and hidden layer biases randomly
Step 2: Calculate the output matrix H *of* hidden layer.

Step 3: Finally calculate the output weight β.

2.3 Implementation of ELM

To further display the performance of ELM, experimentally on ELM, support vector machines (SVMs) and BP neural network to compare the overall performance. ELM and BP execution environment for Matlab7.11 and the SVM execution environment for VC + + 6.0.The activation function used here is a sigmoid function: $f(x) = 1/(1 + \exp(-x))$ and the kernel function of SVM choose radial basis function. Experimental data inputs are normalized to [-1, 1] range.

TABLE 1 Basic Characteristic of The Three Datasets

DATASETS	CLASSIFICATION	NO.OF SAMPLES
State variable filter	9	1853
Sallenkey bandpass filter	7	1443
Polynomial filter	8	1850

International Journal of Electrical and Electronics Engineers
Vol. No. 9, Issue No. 01, January-June 2017
ISSN (O) 2321-2055
ISSN (P) 2321-2045

Table 2 Performance comparison of three kinds of algorithms

Datasets	Algorithm	Time/s		Classification accuracy (in %)		Hidden layer nodes/number of support vector
		Train time	Test time	Training accuracy	Testing accuracy	
State variable filter filter	ELM	1.127	0.0568	96.58	95.13	100
	BP	3438.5	0.0781	95.36	84.01	200
	SVM	0.2641	0.3865	83.07	76.92	100
Sallenkey bandpass filter	ELM	8.7656	0.1875	94.43	87.75	200
	BP	3786.9	0.2548	90.68	83.02	150
	SVM	1.1384	0.4351	82.93	80.45	500
Polynomial filter	ELM	0.0313	0.1563	84.77	75.63	60
	BP	0.9456	0.4581	78.56	69.73	100
	SVM	0.0179	0.0275	64.36	58.49	300

In the experiment we use several classification data sets and then compare experimental results. Datasets used in the experiment include: Sallenkey bandpass filter dataset, State variable filter dataset and Polynomial filter dataset. The fundamental characteristics of the three datasets are as shown in Table 1.For all samples in the experiment, 75% of samples are used for training and 25% of samples is used for testing. ELM and BP network hidden layer nodes and the number of SVM support vectors are the average of several experimental results. As shown in Table 2 are ELM, BP and SVM algorithm performance comparison.

It can be seen that ELM is a simple and very fast neural network learning algorithm just need only one iteration process. Compared with the BP algorithm, ELM randomly assigns input connection weights and hidden layer neurons threshold by just tuning the number of nodes of hidden layer N. As well as the widely used neural network, ELM based on single hidden layer feedforward neural network due to the good learning capability and generalization performance, make it a very good development. However, ELM algorithm also obtains several defects (Deng et al. 2010a,b) such as poor robustness, over fitting problem and poor controllability

2.4 Improvements on Extreme Learning Machine

For improving the speed of network training in ELM and also to avoid the gradient descent learning method for many problems, such as, too much iterative times, local minimal, termination conditions and the definition of learning rate parameter (Huang et al. 2004). However, ELM is based on the principle of empirical risk minimization and it’s randomly select the input weights and hidden layer biases which cause non-optimal or unnecessary input weights and hidden layer biases. Compared with the gradient descent learning algorithm, ELM need more number of hidden layer neurons, it will reduces the computation rate and training effect of ELM. Therefore, in order to increase the response rate of the training network, people put forward a lot of ELM improved algorithm.

2.5 Online Extreme Learning Machine

Extreme learning machine algorithm usually used in solving multiple logistic classification and regression problems. In this, it reflects the fast learning speed and good generalization performance. But in some applications it has some problems, such as small sample data set, prone to over-fitting phenomenon. After the ELM proposed (Huang 2005) put forward sequential modification based on recursive least-squares (RLS) algorithm that is online sequential extreme learning machine (OS-ELM). Based on OS-ELM, online sequential fuzzy extreme learning machine (Fuzzy-ELM) is established to implement zero order and first order TSK model. The performance of OS-ELM and Fuzzy-ELM is compared with other popular sequential learning algorithms shows that the proposed OS-ELM produces better performance at very fast learning speed.

Rong et al. (2009) combined with previous work and find an online sequential fuzzy extreme learning machine (OS-Fuzzy-ELM) developed for function approximation and classification problems. To solve the problem of extreme learning machine (ELM) online training with sequential training samples, an efficient algorithm called selective forgetting extreme learning machine (SF- ELM) is proposed by Zhang (2011) and applied to chaotic time series prediction. The SF- ELM uses the latest training sample and weights the old training samples iteratively to insure that the influence of the old training samples is weakened. The output weight of the SF-ELM is recursively determined during on-line training procedure according to its generalization performance.

2.6 Pruned Extreme Learning Machine

Extreme learning machine is one of the recent successful algorithms in machine learning, particularly for performing pattern classification. One important advantage of ELM is the short training time. The number of hidden layer nodes can be randomly selected and analyzed to determine is to reduce the calculation time while learning speed fast. However, ELM has some disadvantages such as over-fitting problem. Rong et al. (2008) proposed the architectural design of the ELM classifier network, since too few/many hidden nodes employed would lead to under-fitting/over-fitting issues in pattern classification. A pruned-ELM (P-ELM) algorithm was proposed as an automated and systematic approach for designing ELM classifier network. P-ELM uses statistical methods to determine the hidden nodes relevance. The proposed approach leads to compact network classifiers that produce the fast response and robust prediction accuracy on unseen data, comparing with traditional ELM and other popular machine learning approaches.

2.7 Applications of ELM

Neural network is largely used in data mining, artificial intelligence and other applications Extreme learning methods proposed to overcome the disadvantage of a single hidden layer feedforward neural network and to improve the learning ability and generalization performance. Compared with the traditional intelligence techniques, such as BP algorithm ,it needs only one iteration and the training speed is improved more than 10 times.Extreme learning machine is used in the fields such as automatic control (Mao and Huang 2005), signal processing (Mao 2002), image processing (Huang et al. 2011), aviation, market analysis and aerospace and medical diagnosis (Shang et al. 2005), etc. Many researchers using the ELM algorithm for solving the compete phenomena in the whole nature and society, like winner-take-all (WTA) competition which is widely observed in both inanimate and biological media and society.

In order to solve the problem of hidden layer neuron determination, regularized extreme learning machine is applied to chaotic time series prediction, a new algorithm based on Cholesky factorization is proposed by Zhang (2011). Initially RELM with one hidden-layer neuron is constructed and then a new hidden-layer neuron is increased to the prediction model in each training step until the generalization performance of the prediction model reaches its highest value. Thus, the optimal structure of the prediction model is determined by using RELM. In the training procedure, Cholesky factorization is used to determine the output weight of RELM.

Recently, error minimized extreme learning machines (EM-ELMs) was proposed by Romero and Alquezar (2012) as an efficient approach to build SLFNs sequentially. They add random hidden nodes one by one and to update the output weights incrementally for minimizing the sum-of-squares error in the training set. EM-ELMs can also be a pitch as a particular case of SAOCIF. Generally EM-ELMs can easily be extended to test number of random candidates at every step and select the best of them, as SAOCIF does. Moreover, it is demonstrated that the cost of the computation in EM-ELMs can be improved if it is replaced by the one included in SAOCIF.

Pan et al. (2010) used the ELM algorithm for predicting the reservoir permeability, through the analysis with SVM.Yao and Han (2010) use the ELM for the remote sensing image fusion. Cai et al. (2010) use the ELM method for lithology identification, and established the ELM lithology classification model, the results specify that the usage of ELM in the field provide the feasibility and effectiveness. At present, the ELM is applied in many fields and plays a important role. Especially in high dimensional complex data mining has powerful advantage.

III CONCLUSION

Extreme learning machine algorithm is proposed to the local minimum and over-fitting problems, and the network structure has also been greatly improved. However, the ELM algorithm still exists some disadvantages, this need further development and perfection.

In view of the above shortages, ELM needs further improvement. It includes:

(1) Do further refine and improve the existing method of ELM algorithm. The ELM algorithm efficiency of learning and generalization performance has been identified, but the number of neurons in hidden layer is numerous, the application is not very suitable also affect the accuracy of the results. Therefore enhancing the ELM model structure and the generalization performance of the algorithm is necessary.

(2) The integration of other disciplines and other technology with ELM. In recent years, many researchers try to

combine other kinds of disciplines algorithm with the ELM, for better training model. In the future study, how to make the genetic algorithm, online learning, SVM and ELM together will be a very valuable to explore direction.

(3) Stretch out ELM applications. Although ELM has prominent advantages in theory, but the real application field is limited at present. Therefore, how to apply ELM to the daily life effectively is an important aspect in future research. How to use ELM better to address and resolve any multi-classification and regression problem is worth to discover.

REFERENCES

[1] Benqio Y (2009) Learning deep architectures for AI. Found Trends Mach Learn 2(1):1–127

[2] Cai L, Cheng G, Pan H (2010) Lithologic identification based on ELM. Comput Engi Des 31(9):210–

[3] Deng W, Zheng Q, Chen L et al (2010a) Research on extreme learning of neural networks. Chin J Comput 33(2):279–287

[4] Deng W, Zheng Q, Lian S et al (2010b) Ordinal extreme learning machine. Neurocomputing 74(1–3):447–456

[5] Ding S, Xu L, Chunyang SuC et al (2012) An optimizing method of RBF neural network based on genetic algorithm. Neural Comput Appl 21(2):333–336

[6] Hagan MT, Demuth HB, Beale MH (2002) Mechanical Industry Press, Beijing, China

[7] Han F, Huang D (2006) Improved extreme learning machine for function approximation by encoding apriori information. Neurocomputing 69(16–18):2369–2373

[8] Huang G (2003) Learning capability and storage capacity of two-hidden-layer feedforward network. IEEE Trans Neural Netw 14(2):274–281

[9] Huang G, Liang N, Rong HJ et al. (2005) On-line sequential extreme learning machine. In: The IASTED international conference on, computational intelligence, pp. 232–237

[10] Huang G, Wang D, Lan Y (2011) Extreme learning machines: a survey. Int J Mach Learn Cybern 2(2):107–122

[11] Huang G, Zhu Q, Siew CK (2006) Extreme learning machine: theory and applications. Neurocomputing 70(1–3):489–501

[12] Lan Y, Soh YC, Huang G (2010) Two-stage extreme learning machine for regression. Neurocomputing 73(2010):3028–3038

[13] Li S, Chen SF, Liu B et al (2012) Decentralized kinematic control of a class of collaborative redundant manipulators via recurrent neural networks. Neurocomputing 91:1–10

[14] Li S, Liu B, Li YM (2013) Selective positive-negative feedback produces the winner-take-all competition in recurrent neural networks. IEEE Trans Neural Netw Learn Syst 24(2):301–309

[15] Liang N, Huang G (2006) A fast and accurate online sequential learning algorithm for feed forward networks IEEE Trans Neural Netw 17(6):1411–1423

[16] McCulloch W, Pitts W (1943) A logical calculus of the ideas immanent in nervous activity. Bull Math Biophys 5(4):115–133

[17] Mohamed MH (2011) Genetic algorithms. Neurocomputing 74(17):3180–3192

[18] Romero E, Alquezar R (2012) Comparing error minimized extreme learning machines and support vector sequential feed-forward neural networks. Neural Netw 25:122–129

[19] Wang Y, Cao F, Yuan Y (2011) A study on effectiveness of extreme learning machine. Neurocomputing 74(16):2483–2490

[20] Zhu Q, Qin A, Suganthan PN et al (2005) Evolutionary extreme learning machine. Patt Recognit 38(10):1759–1763