Extreme learning machine: algorithm, theory and applications

Shifei Ding · Han Zhao · Yanan Zhang · Xinzheng Xu · Ru Nie

Abstract Extreme learning machine (ELM) is a new learning algorithm for the single hidden layer feedforward neural networks. Compared with the conventional neural network learning algorithm it overcomes the slow training speed and over-fitting problems. ELM is based on empirical risk minimization theory and its learning process needs only a single iteration. The algorithm avoids multiple iterations and local minimization. It has been used in various fields and applications because of better generalization ability, robustness, and controllability and fast learning rate. In this paper, we make a review of ELM latest research progress about the algorithms, theory and applications. It first analyzes the theory and the algorithm ideas of ELM, then tracking describes the latest progress of ELM in recent years, including the model and specific applications of ELM, finally points out the research and development prospects of ELM in the future.

Keywords Extreme learning machine (ELM) · Single-hidden layer feedforward neural networks (SLFNs) · Local minimum · Over-fitting · Least-squares

1 Introduction

As early as in the 1940s, mathematician Pitts and psychologist McCulloch have put forward neurons mathematical model (MP model) from the mathematical logic view (McCulloch and Pitts 1943) which opened the prelude of artificial neural network research. Neural network with parallel and distributed information processing network structure has a strong nonlinear mapping ability and adaptive self-learning, robustness and fault tolerance characteristics.

S. Ding (✉) · H. Zhao · Y. Zhang · X. Xu · R. Nie
School of Computer Science and Technology, China University of Mining
and Technology, Xuzhou 221116, China
e-mail: dingsf@cumt.edu.cn

S. Ding
Key Laboratory of Intelligent Information Processing, Institute of Computing Technology,
Chinese Academy of Science, Beijing 100190, China

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

At present, there are many common kinds of neural network model, such as BP network (Ding et al. 2011a,b; Feng et al. 2009), RBF network (Ding et al. 2011a,b), Hopfield network, CMAC cerebellar model, ART adaptive resonance theory (LeCun and Bengio 1995; Benqio 2009; Carpenter and Grossberg 2003; Li et al. 2013), etc. The powerful computing capabilities of the neural network are achieved through the propagation of information between neurons. According to the direction of the neural network internal information transfer, the neural network can be divided into two categories: feedforward neural network and feedback type neural networks. Extreme learning machine (ELM) described in this paper is for single hidden layer feedforward neural network which is one of feedforward neural networks.

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

Single hidden layer feed forward neural network has very strong learning ability, can approximate a complex nonlinear function and can solve the problems which cannot be solved by the traditional parameter learning method. However, the lack of fast learning method, often cannot meet the actual demand. To solve this problem, based on the above research conclusions of many scholars a new algorithm which is called the ELM has been proposed by Huang et al. (2006). As a new proposed learning algorithm can be easily solve the traditional feedforward neural network, such as BP network local minimum problem. ELM can adaptively set the hidden layer node number and randomly assign for the input weights and hidden layer biases, then the output layer weights obtained by the least square method, the whole process of a complete without iteration and improves the neural network generalization ability and learning ability.

Since the extreme machine learning algorithm proposed, it has been widely used and became research focus of data mining, machine learning, image processing and other areas. In this paper, we make a review of extreme machine learning latest research progress about the algorithms, theory and applications. This paper is organized as follows. Section 2 introduces the basic idea and algorithm model of ELM. Section 3 further summarizes some improved

ELM algorithms and its model. The applications of ELM in related fields are presented in Sect. 4. Conclusions and prospects are given in Sect. 5.

2 Extreme learning machine

2.1 Algorithm thought of ELM

Extreme learning machine is a learning algorithm for the single hidden layer feedforward neural networks used in classification and regression. The ELM used for single hidden layer feedforward neural network training can adaptively set the hidden layer node number and randomly assign for the input weights and hidden layer biases, the output layer weights obtained by the least square method, the whole learning process completed through one mathematical change without iteration. The training speed compared with the traditional BP algorithm based on gradient descent has been significantly improved (usually 10 times or more) (Deng et al. 2010a,b).

In practical applications, firstly training on ELM and then predict. The training data set is mainly combined with the specific issues. The data sets include actual results and its related factors. During training, the influence factors and the corresponding results will be put into ELM for training, through an iteration to complete learning process. Then, with the trained ELM to predict, only need to input and the training data set is similar to the influencing factors. ELM model can be obtained the prediction results according to the memory.

Extreme learning machine is an easy to use and effective algorithm for single hidden layer feedforward neural network. The traditional neural network learning algorithm (e.g. BP algorithm) need to set up lots of artificial network training parameters, and can easily lead to local optimal solution. ELM algorithm only need to set the number of hidden layer nodes, in the algorithm implementation process does not need to adjust the network input weights and hidden biases, and generates a unique optimal solution, with advantages of fast learning speed and generalization performance.

2.2 Model of ELM

For N arbitrary 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 $\tilde{N}$ hidden nodes and activation function $f(x)$ are mathematically models as

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

where $a_i = [a_{i1}, a_{i2}, \ldots, a_{in}]^T$ is the weight vector connecting the i th hidden node and the input nodes, and b_i is the threshold of the ith hidden node. $\beta_i = [\beta_{i1}, \beta_{i2}, \ldots, \beta_{im}]^T$ is the weight vector connecting the ith 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 usually choose "Sigmoid", "Sine", "RBF".

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

$$H\beta = T \tag{2}$$

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

$$= \begin{bmatrix} f(a_1 \cdot x_1 + b_1) & \cdots & f(a_{\tilde{N}} \cdot x_1 + b_{\tilde{N}}) \\ \vdots & \cdots & \vdots \\ f(a_1 \cdot x_N + b_1) & \cdots & f(a_{\tilde{N}} \cdot x_N + b_{\tilde{N}}) \end{bmatrix}_{N \times \tilde{N}}, \beta = \begin{bmatrix} \beta_1^T \\ \vdots \\ \beta_{\tilde{N}}^T \end{bmatrix}_{\tilde{N} \times m}, T = \begin{bmatrix} t_1^T \\ \vdots \\ t_N^T \end{bmatrix}_{N \times m}$$

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

Theorem proving (Huang 2003), as long as the number of hidden nodes is enough, when the activation function $f(x)$ is infinitely differentiable at any interval the parameters of the network does not all need to adjust. When the training starts, SLFN randomly assigns to the input connection weights a and hidden layer node biases b, moreover, while the training process unchanged it can approximate any continuous function. Generally, in order to get good generalization performance, take $\tilde{N} \ll N$.

When the input weights and hidden layer biases are determined in accordance with the random assignment, according to the input samples can get the hidden layer output matrix H. Therefore, training SFLN is converted into solving linear equations $H\beta = T$ least squares solution.

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

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

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

In the Eq. (4), $H^{\dagger}$ represents Moore–Penrose (Liang and Huang 2006) generalized inverse of the hidden layer output matrix H. Usually, the optimal solution $\hat{\beta}$ contains the following features:

(1) According to $\hat{\beta}$, the algorithm can gain the minimal training error;
(2) Can get the optimal generalization capability of the minimum paradigm of the output connection weights and network;
(3) $\hat{\beta}$ is unique. This can avoid producing the local optimal solution.

In summary, given a training set $(x_i, t_i) \in R^n \times R^m (i = 1, 2, \ldots, N)$, activation function $f(x)$ and hidden node number $\tilde{N}$, then the ELM algorithm can be can be summarized as follows 3 steps:

Step 1: Definition of hidden layer node number $\tilde{N}$, randomly assign input weights a_i and hidden layer biases b_i, $(i = 1, 2, \ldots, \tilde{N})$.
Step 2: Calculate the hidden layer output matrix H.
Step 3: According to the Eq. (4), calculate the output weight β.

2.3 Performance of ELM

To further demonstrate the performance of ELM, through experiments on ELM, support vector machines (SVMs) and BP neural network to compare the performance. ELM and BP execution environment for Matlab7.11 and the SVM execution environment for VC + + 6.0. ELM's source code can be downloaded directly from the profile of Huang. BP algorithm has been integrated in the neural network toolbox in Matlab and can directly execute experiment. SVM algorithm using the C language implementation of the SVM package: LibSVM. ELM and BP activation function to select a sigmoid function: $f(x) = 1/(1 + \exp(-x))$ and the

Table 1 Basic characteristic of the three datasets

Datasets	Classification	Number of samples	Number of attributes
Segmen	7	2,310	19
Satimage	7	4,435	36
Diabetes	2	768	8

Table 2 Performance comparison of three kinds of algorithms

Datasets	Algorithm	Time/s		Classification accuracy/%		Hidden layer nodes/number of support vector
		Train time	Test time	Training accuracy	Testing accuracy	
Segmen	ELM	1.007	0.0568	97.58	95.03	100
	BP	2438.5	0.0781	94.36	83.01	200
	SVM	0.2641	0.3865	84.07	77.92	100
Satimage	ELM	9.7656	0.1875	94.63	87.75	200
	BP	4786.9	0.2548	90.68	82.02	150
	SVM	1.1384	0.4351	84.93	80.45	500
Diabetes	ELM	0.0313	0.1563	84.77	77.63	60
	BP	0.9456	0.4581	78.56	69.73	100
	SVM	0.0179	0.0275	65.36	58.49	300

kernel function of SVM choose radial basis function. Experimental data inputs are normalized to [0, 1] range, while the output is normalized to [−1, 1] range.

In the experiment we use several classification data sets in the UCI database and then compare experimental results. Datasets used in the experiment include: Segmen dataset, Satimage dataset and Diabetes dataset. The fundamental characteristics of the three datasets are as shown in Table 1.

For all samples in the experiment randomly divided into two groups for training and testing data sets. 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 very simple and fast neural network learning algorithm just need one iteration process. Compared with the BP algorithm, ELM can randomly initialize input connection weights and hidden layer neurons threshold by just tuning the number of hidden layer nodes $\tilde{N}$. In practical applications, $\tilde{N}$ can be identified through constantly trying. Moreover, in a large number of experiments on standard UCI data sets show that, ELM has faster training speed and better generalization performance (Han and Huang 2006) than the BP algorithm and SVM method.

To further validate the performance of the ELM, the article refers to the experimental results from literature (Huang et al. 2006) to analysis. In the literature, Huang et al, do the experiment with small and medium real classification datasets and then compare ELM, BP and SVM classification results. The basic characteristics of the data set in Table 3.

Performance comparison of ELM, BP and SVM as in Table 4.

As well as the widely used neural network, ELM based on single hidden layer feedforward neural network due to its good learning capability and generalization performance, make it a

Table 3 Basic characteristic of the datasets

Datasets	Classification	Number of samples	Number of attributes
Image segmentation	7	2,310	18
Satellite image	7	6,400	36
Shuttle	7	58,000	9
Banana	2	495,200	2

Table 4 Performance comparison of three kinds of algorithms (continued)

Datasets	Algorithm	Time/s		Classification accuracy/%		Hidden layer nodes/number of support vector
		Train time	Test time	Training accuracy	Testing accuracy	
Satellite image	ELM	14.92	0.34	93.52	89.04	500
	BP	12,561	0.08	95.26	82.34	100
Image segment	ELM	1.40	0.07	97.35	95.01	200
	BP	4,745.7	0.04	96.92	86.27	100
Shuttle	ELM	5.740	0.23	99.65	99.40	50
	BP	6,132.2	0.22	99.77	99.27	50
Banana	ELM	2.19	20.06	92.36	91.57	100
	BP	6,132.2	21.10	90.26	88.09	100
Diabetes	ELM	0.0118	0.0031	78.68	77.57	20
	BP	3.0116	0.0035	86.63	74.73	20
	SVM	0.1860	0.0673	78.76	77.31	317.16
Forest-type	ELM	1.6148	0.7195	92.35	90.21	200
	BP	12 min	N/A	82.44	81.85	100
	SVM	693.6000	347.7833	91.70	89.90	31,806

very good development. However, ELM algorithm also obtains several defects (Deng et al. 2010a,b):

(1) ELM algorithm is based on the empirical risk minimization, without considering the structural risk and this may lead to over-fitting problems.
(2) ELM directly calculates the least squares solution. The users cannot fine-tune according to the characteristics of the data sets. It is also poorly controllability.
(3) When there are outliers in the data sets, the performance of model will be greatly affected, poor robustness.

3 Improvements on extreme learning machine

Since ELM was proposed by Huang et al. (Lan et al. 2010) at 2004, many researchers began to study on ELM and put forward lots of improved algorithms. To make ELM in improving the speed of network training, at the same time also to avoid the based on gradient descent

learning method for many problems, for example, local minimal, too much iterative times, termination conditions as well as the definition of learning rate parameter (Huang et al. 2004).

However, because the ELM is based on the empirical risk minimization principle, randomly select the input weights and hidden layer biases which may lead to non-optimal or unnecessary input weights and hidden layer biases. Compared with the based on gradient descent learning algorithm, ELM may need more hidden layer neurons, which reduces the computation rate and training effect of ELM. Therefore, in order to speed up the response rate of the training network, people put forward a lot of ELM improved algorithm.

3.1 Online extreme learning machine

Extreme learning machine algorithm usually used in solving multiple logistic regression and classification problems. In the applications, it reflects good generalization performance and fast learning speed. But in concrete applications it will also encounter 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, which referred as online sequential extreme learning machine (OS-ELM). Based on OS-ELM, online sequential fuzzy extreme learning machine (Fuzzy-ELM) is also introduced to implement zero order TSK model and first order TSK model. The performance of OS-ELM and Fuzzy-ELM are evaluated and compared with other popular sequential learning algorithms, and compared with other sequential learning algorithm shows that the proposed OS-ELM produces better generalization 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. The equivalence of a Takagi–Sugeno–Kang (TSK) fuzzy inference system (FIS) to a generalized single hidden layer feedforward network is shown first, which is then used to develop the OS-Fuzzy-ELM algorithm. This results in a FIS that can handle any bounded non-constant piecewise continuous membership function. Furthermore, the learning in OS-Fuzzy-ELM can be done with the input data coming in a one-by-one mode or a chunk-by-chunk mode with fixed or varying chunk size. In OS-Fuzzy-ELM, all the antecedent parameters of membership functions are randomly assigned first, and then, the corresponding consequent parameters are determined analytically. Performance comparisons of OS-Fuzzy-ELM with other existing algorithms are presented using real-world benchmark problems in the areas of nonlinear system identification, regression, and classification.

To solve the problem of extreme learning machine (ELM) online training with sequential training samples, a new algorithm called selective forgetting extreme learning machine (SF-ELM) is proposed by Zhang (2011) and applied to chaotic time series prediction. The SF-ELM adopts 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 determined recursively during on-line training procedure according to its generalization performance. In comparison with on-line sequential extreme learning machine, the SF-ELM has better performance in the sense of computational cost and prediction accuracy.

3.2 Pruned extreme learning machine

Extreme learning machine represents one of the recent successful approaches in machine learning, particularly for performing pattern classification. One prominent advantage of ELM is its short training time. The number of hidden layer nodes can be randomly selected and analyzed to determine. This reduces the calculation time while learning speed fast.

However, ELM also exist some shortcomings such as over-fitting problems. Rong et al. (2008) address 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 a systematic and automated approach for designing ELM classifier network. P-ELM uses statistical methods to measure the relevance of hidden nodes. Beginning from an initial large number of hidden nodes, irrelevant nodes are then pruned by considering their relevance to the class labels. As a result, the architectural design of ELM network classifier can be automated. Empirical study of P-ELM on several commonly used classification benchmark problems and with diverse forms of hidden node functions show that the proposed approach leads to compact network classifiers that generate fast response and robust prediction accuracy on unseen data, comparing with traditional ELM and other popular machine learning approaches.

An algorithm for pruning ELM networks by using regularized regression methods was put forward by Martinez-Martinez et al. (2011), thus obtaining a suitable number of the hidden nodes in the network architecture. Beginning from an initial large number of hidden nodes, irrelevant nodes are then pruned using ridge regression, elastic net and lasso methods; hence, the architectural design of ELM network can be automated. Empirical studies on several commonly used regression benchmark problems show that the proposed approach leads to compact networks that generate competitive results compared with the ELM algorithm.

3.3 Some other improved ELM algorithms

ELM may need higher number of hidden neurons due to the random determination of the input weights and hidden biases, a hybrid learning algorithm was proposed by Zhu et al. (2005) which uses the differential evolutionary algorithm to select the input weights and Moore–Penrose (MP) generalized inverse to analytically determine the output weights. This approach is able to achieve good generalization performance with much more compact networks.

Consider about the improved ELM algorithm, many researchers combine ELM and other data processing methods to propose a new learning model which is applied to the related fields, and achieved very good results.

In order to solve the ELM non-optimal performance and the over fitting issue, a hybrid approach is proposed based on group search optimizer strategy to select input weights and hidden biases for ELM algorithm by Silva (2011). In this method, need to evaluate the influence of different forms of handling members that fly out of the search space bounds. And in the benchmark data sets for processing can be obtained than traditional ELM better generalization performance.

Because of the random choice of input weights and biases, the ELM algorithm sometimes makes the hidden layer output matrix H of SLFN not full column rank, which lowers the effectiveness of ELM. An effectiveness improved algorithm of ELM was proposed by Wang and Cao (2011) called EELM that makes a proper selection of the input weights and biases before calculating the output weights, which ensures the full column rank of H in theory. This improves to some extend the learning rate and the robustness property of the networks. The

experimental results based on both the benchmark function approximation and real-world problems including classification and regression applications show the good performances of EELM.

Fernandez-Navarro et al. (2011) propose a methodology for training a new model of artificial neural network called the generalized radial basis function (GRBF) neural network. This model is based on generalized Gaussian distribution, which parameterizes the Gaussian distribution by adding a new parameter τ. For example, when GRBF takes a value of $\tau = 2$, it represents the standard Gaussian radial basis function. The model parameters are optimized through a modified version of the extreme learning machine (ELM) algorithm. The centers of each GRBF were taken randomly from the patterns of the training set and the radius and τ values were determined analytically, taking into account that the model must fulfill two constraints: locality and coverage. A thorough experimental study is presented to test its overall performance.

4 Applications of ELM

Neural network is widely used in artificial intelligence, data mining and other applications. Single hidden layer feedforward neural network has a strong learning ability. Extreme learning methods proposed to improve the disadvantage of a single hidden layer feedforward neural network, improve the learning ability and generalization performance. Due to ELM algorithm does not require large-scale operations, there is no need to adjust the hidden layer of the network and compared with the traditional computational intelligence techniques, such as BP algorithm the whole learning process needs only one iteration and the training speed improved more than 10 times.

The algorithm is different from the traditional neural network methods. It can randomly select the hidden layer neurons number and weights, and analysis to determine the output weights of SLFN. The training speed is fast and with good generalization ability, which makes SLFN training time greatly reduced.

Extreme learning machine is used in the fields such as signal processing (Mao 2002), automatic control (Mao and Huang 2005), image processing (Huang et al. 2011), market analysis, aviation and aerospace and medical diagnosis (Shang et al. 2005), etc.

A new human face recognition algorithm based on bidirectional two dimensional principal component analysis (B2DPCA) and ELM is proposed by Mohammed et al. (2011). The method is based on curvelet image decomposition of human faces and a subband that exhibits a maximum standard deviation is dimensionally reduced using an improved dimensionality reduction technique. Discriminative feature sets are generated using B2DPCA to ascertain classification accuracy. Other notable contributions of the proposed work include significant improvements in classification rate, up to hundred folds reduction in training time and minimal dependence on the number of prototypes.

Many researchers using the ELM to solve the compete phenomena in the whole nature and society, such as the winner-take-all (WTA) competition which is widely observed in both inanimate and biological media and society. Then, Li et al. (2013) presented a simple model which produces the WTA competition by taking advantage of selective positive–negative feedback through the interaction of neurons via p-norm.

In order to solve the hidden layer neuron determination problem of regularized extreme learning machine applied to chaotic time series prediction, a new algorithm based on Cholesky factorization is proposed by Zhang (2011). First, an RELM-based prediction model with one hidden-layer neuron is constructed and then a new hidden-layer neuron is added to the

prediction model in each training step until the generalization performance of the prediction model reaches its peak value. Thus, the optimal network structure of the prediction model is determined. In the training procedure, Cholesky factorization is used to calculate the output weights of RELM. The algorithm can be effectively used to determine the optimal network structure of RELM, and the prediction model trained by the algorithm has excellent performance in prediction accuracy and computational cost.

A sign-bi-power activation function is proposed by Li et al. (2012) to accelerate Zhang neural network to finite-time convergence. The problem of this neural network is with the suggested activation functions never converge to the desired value in finite time, which may limit its applications in real-time processing. In addition, the proposed algorithm is applied to online calculating the pseudo-inverse of a matrix and nonlinear control of an inverted pendulum system. Both theoretical analysis and numerical simulations validate the effectiveness of proposed activation function.

Recently, error minimized extreme learning machines (EM-ELMs) have been proposed by Romero and Alquezar (2012) as a simple and efficient approach to build SLFNs sequentially. They add random hidden nodes one by one and update the output weights incrementally to minimize the sum-of-squares error in the training set. Other very similar methods that also construct SLFNs sequentially had been reported earlier with the main difference that their hidden layer weights are a subset of the data instead of being random. These approaches are referred to as support vector sequential feed-forward neural networks (SV-SFNNs), and they are a particular case of the sequential approximation with optimal coefficients and interacting frequencies (SAOCIF) method. In this paper, it is shown that EM-ELMs can also be cast as a particular case of SAOCIF. In particular, EM-ELMs can easily be extended to test some number of random candidates at each step and select the best of them, as SAOCIF does. Moreover, it is demonstrated that the cost of the computation of the optimal output-layer weights in the originally proposed EM-ELMs can be improved if it is replaced by the one included in SAOCIF.

In addition, there are still a lot of applications about ELM in other areas. An improved ELM method was proposed for soft sensing modeling by Chang et al. (2007). By adding a kind of sorting neurons into the hidden layer of the single hidden layer feedforward neural network (SLFN), a new SLFN structure was proposed. The method provided a new approach to build soft sensing models, and it was successfully applied to the soft sensing modeling the of cell concentration for the Nosiheptied fermentation process to realize the online prediction of the cell concentration.

Pan et al. (2010) use the ELM to predicting reservoir permeability, through the contrast analysis with SVM, found in the reservoir permeability prediction of possibility and advantage.

Yao and Han (2010) use the ELM for remote sensing image fusion. It gets the training sample regression relationship between the remote sensing image data quickly and efficiently while improving the real quality of a thermal infrared image data and provide convenience for the practical application of remote sensing.

Cai et al. (2010) use the ELM method for lithology identification, and the establishment of the ELM lithology classification model, the results indicate that the application of ELM in the field the feasibility and effectiveness of the algorithm.

Li et al. (2012) studies the decentralized kinematic control of multiple redundant manipulators for the cooperative task execution problem. They formulated the problem as a constrained quadratic programming problem and proposed a recurrent neural network with independent modules to solve the problem in a distributed manner. Each module in the neural network

controls a single manipulator in real time without explicit communication with others and all the modules together collectively solve the common task.

At present, the ELM is applied in more and more fields and plays a significant role. Especially in high dimensional complex data mining has prominent advantage. In the future, how to make better combinations between other learning methods and ELM and applications in the broader field are the issues of concern.

5 Conclusions and prospects

Extreme learning machine algorithm was proposed to solve the traditional single hidden layer feed forward neural network easily to appear the local minimum and over-fitting problems, and the network structure has also been greatly improved. However, the ELM algorithm still exists some shortcomings, this need further development and perfection.

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

(1) Do further refine and improve of the existing ELM algorithm. The ELM algorithm efficiency of learning and generalization performance has been recognized, but the number of neurons in hidden layer is numerous, the application is not very convenient also will affect the accuracy of the results. Therefore, further enhancing 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 in view of ELM problems try to combine other kinds of disciplines algorithm with the ELM, presented better training model. In the future study, how to make the online learning, genetic algorithm, SVM and ELM together will be a very worthwhile to explore direction.
(3) Expand ELM applications. Although ELM has obvious advantages in theory, but the actual 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 the ELM better use to address and resolve any regression and multi-classification problem is worth to explore and discover.

References

Benqio Y (2009) Learning deep architectures for AI. Found Trends Mach Learn 2(1):1–127
Cai L, Cheng G, Pan H (2010) Lithologic identification based on ELM. Comput Engi Des 31(9):210–2012
Carpenter G, Grossberg S (2003) Adaptive resonance theory. In: Arbib MA (ed) The handbook of brain theory and neural networks, 2nd edn. MIT Press, Cambridge, pp 87–90
Chang Y, Li Y, Wang F (2007) Soft sensing modeling based on extreme learning machine for biochemical processes. J Syst Simul 19(23):5587–5590
Deng W, Zheng Q, Chen L et al (2010a) Research on extreme learning of neural networks. Chin J Comput 33(2):279–287
Deng W, Zheng Q, Lian S et al (2010b) Ordinal extreme learning machine. Neurocomputing 74(1–3):447–456
Ding S, Jia W, Su C et al (2011a) Research of neural network algorithm based on factor analysis and cluster analysis. Neural Comput Appl 20(2):297–302

Ding S, Su C, Yu J et al (2011b) An optimizing BP neural network algorithm based on genetic algorithm. Artif Intell Rev 36(2):153–162

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

Ding S, Jin F, Zhao X (2013) Modern data analysis and information pattern recognition. Science Press, Beijing

Feng G, Huang G, Lin Q et al (2009) Error minimized extreme learning machine with growth of hidden nodes and incremental learning. IEEE Trans Neural Netw 20(8):1352–1357

Fernandez-Navarro F, Hervas-Martinez C, Sanchez-Monedero J et al (2011) MELM-GRBF: A modified version of the extreme learning machine for generalized radial basis function neural networks. Neurocomputing 74(16):2502–2510

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

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

Hornik K (1991) Approximation capabilities of multilayer feedforward networks. Neural Netw 4(2):251–257

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

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

Huang G, Babri HA (1998) Upper bounds on the number of hidden neurons in feedforward networks with arbitrary bounded nonlinear activation functions. IEEE Trans Neural Netw 9(1):224–229

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

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

Huang G, Zhu Q, Siew CK (2004) Extreme learning machine: a new learning scheme of feedforward neural networks. IEEE Int Jt Conf Neural Netw 1–4:985–990

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

Kahramanli H, Allahverdi N (2009) Rule extraction from trained adaptive neural networks using artificial immune systems. Expert Syst Appl 36(2):1513–1522

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

LeCun Y, Bengio Y (1995) Convolutional networks for images, speech, and time series. MIT Press, Cambridge

Leshno M, Lin V, Pinkus A et al (1991) Multilayer feedforward networks with a non-polynomial activation function can approximate any function. Neural Netw 6(6):861–867

Li S, Chen SF, Liu B (2013) Accelerating a recurrent neural network to finite-time convergence for solving time-varying Sylvester equation by using a sign-bi-power activation function. Neural Process Lett 37:189–205

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

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

Li S, Wang Y, Yu J et al (2013) A nonlinear model to generate the winner-take-all competition. Commun Nonlinear Sci Numer Simul 18(3):435–442

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

Mao K (2002) RBF neural network center selection based on Fisher ratio class separability measure. IEEE Trans Neural Netw 13(5):1211–1217

Mao K, Huang G (2005) Neuron selection for RBF neural network classifier based on data structure preserving criterion. IEEE Trans Neural Netw 16(6):1531–1540

Markowska-Kaczmar U, Trelak W (2005) Fuzzy logic and evolutionary algorithm-two techniques in rule extraction from neural networks. Neurocomputing 63:359–379

Martinez-Martinez JM, Escandell-Montero P, Soria-Olivas E et al (2011) Regularized extreme learning machine for regression problems. Neurocomputing 74(17):3716–3721

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

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

Mohammed AA, Minhas R, Wu Q et al (2011) Human face recognition based on multidimensional PCA and extreme learning machine. Pattern Recognit 44(10–11):2588–2597

Pan H, Cheng G, Cai L (2010) Comparison of the extreme learning machine with the support vector machine for reservoir permeability prediction. Comput Engi Sci 32(2):131–134

Quteishat A, Lim CP (2008) A modified fuzzy min-max neural network with rule extraction and its application to fault detection and classification. Appl Soft Comput 8(2):985–995

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

Rong HJ, Huang G, Sundararajan N et al (2009) Online sequential fuzzy extreme learning machine for function approximation and classification problems. IEEE Trans Syst Man Cybern Part B Cybern 39(4):1067–1072

Rong HJ, Ong YS, Tan AH et al (2008) A fast pruned-extreme learning machine for classification problem. Neurocomputing 72(1–3):359–366

Shang L, Wang J, Yao W (2005) A classification approach based on evolutionary neural networks. J Softw 16(9):1577–1583

Silva DNG, Pacifico LDS, Ludermir TB (2011) An evolutionary extreme learning machine based on group search optimization. IEEE Congress on Evolutionary Computation, pp. 574–580

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

Xu X, Ding S, Shi Z et al (2012) A novel optimizing method for RBF neural network based on rough set and AP clustering algorithm. J Zhejiang University-SCIENCE C 13(2):131–138

Yao W, Han M (2010) Fusion of thermal infrared and multispectral remote sensing images via neural network regression. J Image Graphics 15(8):1278–1284

Zhang D, Wang Y (2009) Rough neural network based on bottom-up fuzzy rough data analysis. Neural Process Lett 30(3):187–211

Zhang X, Wang H (2011) Incremental regularized extreme learning machine based on Cholesky factorization and its application to time series prediction. Acta Physica Sinica 60(11):110201-1–111201-6

Zhang X, Wang H (2011) Selective forgetting extreme learning machine and its application to time series prediction. Acta Physica Sinica, 60(8):080504-1–080504-6

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