

Extreme learning machine and its applications

Shifei Ding · Xinzheng Xu · Ru Nie

Received: 25 June 2013 / Accepted: 20 November 2013
© Springer-Verlag London 2013

Abstract Recently, a novel learning algorithm for single-hidden-layer feedforward neural networks (SLFNs) named extreme learning machine (ELM) was proposed by Huang et al. The essence of ELM is that the learning parameters of hidden nodes, including input weights and biases, are randomly assigned and need not be tuned while the output weights can be analytically determined by the simple generalized inverse operation. The only parameter needed to be defined is the number of hidden nodes. Compared with other traditional learning algorithms for SLFNs, ELM provides extremely faster learning speed, better generalization performance and with least human intervention. This paper firstly introduces a brief review of ELM, describing the principle and algorithm of ELM. Then, we put emphasis on the improved methods or the typical variants of ELM, especially on incremental ELM, pruning ELM, error-minimized ELM, two-stage ELM, online sequential ELM, evolutionary ELM, voting-based ELM, ordinal ELM, fully complex ELM, and symmetric ELM. Next, the paper summarized the applications of ELM on classification, regression, function approximation, pattern recognition, forecasting and diagnosis, and so on. In the last, the paper discussed several open issues of ELM, which may be worthy of exploring in the future.

Keywords Single-hidden-layer feedforward networks · Neural Networks · Extreme learning machine · Classification · Regression

1 Introduction

In the past decades, feedforward neural networks have been widely used in many fields because of its obvious virtues. On the one hand, it could approximate complex nonlinear mappings straightly from the input samples. On the other hand, it can offer models for numerous natural and artificial phenomena, which are hard for classical parametric techniques to handle. However, there exists the dependency between different layers of parameters for that all the parameters of the feedforward network need to be tuned, which render feedforward neural networks time-consuming. Single-hidden-layer feedforward networks (SLFNs), as one of the most popular feedforward neural networks, have been extensively studied from both theoretical and application aspects for their learning capabilities and fault-tolerant abilities [1–6]. However, most popular learning algorithms for training SLFNs are still relatively slow since all the parameters of SFLNs need to be tuned through iterative procedures and these algorithms may also easily get stuck in a local minimum.

Recently, a new fast learning neural algorithm for SLFNs, named extreme learning machine (ELM) [7, 8], was developed to improve the efficiency of SLFNs. Different from the conventional learning algorithms for neural networks (such as BP algorithms), which may face difficulties in manually tuning control parameters (learning rate, learning epochs, etc.) and/or local minima, ELM is fully automatically implemented without iterative tuning, and in theory, no intervention is required from users. Furthermore, the learning speed of ELM is extremely fast compared to other traditional methods. In ELM algorithm,

S. Ding (✉) · 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

the learning parameters of hidden nodes, including input weights and biases, can be randomly assigned independently, and the output weights of network can be analytically determined by the simple generalized inverse operation. The training phase can be efficiently completed through a fixed nonlinear transformation without time-consuming learning process. Moreover, ELM algorithm can achieve a good generalization performance. In addition, the universal approximation ability of the standard ELM with additive or RBF activation function [9–11] has been proved. ELM has been successfully applied to many real-world applications, such as classification and regression problems [12–16].

However, an issue of ELM is that the classification boundary of ELM may not be an optimal one for the learning parameters of hidden nodes are randomly assigned while they remain unchanged during the training phrase [17]. Thus, some samples may be misclassified by ELM, especially for those which are near the classification boundary. It is also found that ELM tends to require more hidden neurons than conventional tuning-based algorithms in many cases [18]. To overcome above-mentioned shortcomings of ELM, some researchers proposed several variants of ELM, such as incremental ELM [9], pruning ELM [12], error-minimized ELM [19], two-stage ELM [20], online sequential ELM [21], evolutionary ELM [18], voting-based ELM [17], ordinal ELM [22], fully complex ELM [23], and symmetric ELM [24].

This paper is organized as follows. Related works are summarized in Sect. 2. Section 3 describes a brief review of ELM. The variants of ELM are then provided in Sect. 4. Section 5 introduces some classical applications of ELM. In Sect. 6, the discussions are given. Finally, Conclusions are drawn in Sect. 7.

2 Related works

As a new learning algorithm, ELM has low computational time requirement for training new classifiers since the weights and biases of the hidden layer are randomly assigned and the output weights are analytically determined by a simple mathematical manipulation. In recent years, ELM had absorbed more and more interests of the researchers, and many variants of ELM were proposed to improve the performance of the ELM algorithm. Furthermore, ELM algorithm has been applied to optimize some problems in the area of computational intelligence, pattern reorganization, machine learning, and so on. Next, we summarized the research results about the variants of ELM.

As is shown in Table 1, we summarized and described briefly the ELM algorithm and its typical variants, including method name, author and year, brief description and applications.

3 Brief review of ELM

ELM, as a novel training algorithm for SLFNs, is very efficient and effective. In this section, we will give a brief review of ELM.

Given N distinct training samples $(x_i, t_i) \in R^n \times R^m (i = 1, 2, \dots, N)$, the output of a SLFN with \tilde{N} hidden nodes (additive or RBF nodes) can be represented by

$$o_j = \sum_{i=1}^{\tilde{N}} \beta_i f_i(x_j) = \sum_{i=1}^{\tilde{N}} \beta_i f(x_j; a_i, b_i), \quad j = 1, \dots, N \quad (1)$$

where o_j is the output vector of the SLFN with respect to the input sample x_i . $a_i = [a_{i1}, a_{i2}, \dots, a_{in}]^T$ and b_i are learning parameters generated randomly of the j th hidden node, respectively. $\beta_i = [\beta_{i1}, \beta_{i2}, \dots, \beta_{im}]^T$ is the link connecting the j th hidden node and the output nodes. $f(x_j; a_i, b_i)$ is the activation function of the original ELM.

Set $a_i \cdot x_j$ be the inner product of a_i and x_j . Equation (1) can be written compactly as

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

where

$$H = \begin{bmatrix} f(a_1 \cdot x_1 + b_1) & \cdots & f(a_{\tilde{N}} \cdot x_1 + b_{\tilde{N}}) \\ \vdots & \cdots & \vdots \\ f(a_1 \cdot x_N + b_1) & \cdots & f(a_{\tilde{N}} \cdot x_N + b_{\tilde{N}}) \end{bmatrix}_{N \times \tilde{N}},$$

$$\beta = \begin{bmatrix} \beta_1^T \\ \vdots \\ \beta_{\tilde{N}}^T \end{bmatrix}_{\tilde{N} \times m}, \quad O = \begin{bmatrix} o_1^T \\ \vdots \\ o_N^T \end{bmatrix}_{N \times m}$$

Here, H is called the output matrix of the hidden layer.

To minimize the network cost function $\|O - T\|$, ELM theories claim that the hidden nodes' learning parameters a_i and b_i can be assigned randomly without considering the input data. Then, Eq. (2) becomes a linear system, and the output weights β can be analytically determined by finding a least-square solution as follows

$$\hat{\beta} = H^\dagger T \quad (3)$$

where H^\dagger is the Moore–Penrose generalized inverse of H . So, calculation of the output weights is done by a mathematical transformation, which avoids any lengthy training phrase where the parameters of the network are adjusted iteratively with some appropriate learning parameters (such as learning rate and iterations).

Thus, the three-step ELM algorithm can be summarized as follows.

ELM algorithm:

Input: a training set $(x_i, t_i) \in R^n \times R^m (i = 1, 2, \dots, N)$, the activation function f , and the hidden node number \tilde{N} .

Table 1 The ELM algorithm and its variants

Method name	References	Brief description	Applications
Original ELM	Huang et al. [7]	An extreme learning algorithm for SLFNs with randomly assigned input weights and biases. The only unknown parameter is the output weights, which can be calculated by a mathematical transformation	Classification and regression problems
Incremental ELM	Huang et al. [9]	An incremental ELM model, in which nodes of the hidden layer were added to the hidden layer one	Several benchmark problems in the function approximation area
Pruning ELM	Rong et al. [12]	A pruned-ELM model began with an initial large number of hidden nodes and then removed the irrelevant or lowly relevant hidden nodes by their relevance to the class labels	Eight real-world classification problems from UCI ML repository
Error-minimized ELM	Feng et al. [19]	An error-minimization-based method for ELM, which can grow hidden nodes one by one to automatically determine the number of hidden nodes in generalized SLFNs	Some real benchmark regression and classification problems
Two-stage ELM	Lan et al. [20]	A systematic two-stage algorithm for ELM with a much smaller network structure by the two-stage adjustment of hidden nodes	Six real regression problems from UCI ML repository
Online sequential ELM	Liang et al. [21]	An online sequential ELM algorithm, which only required the number of hidden nodes to be specified as the conventional ELM	Classification, regression, and time series prediction problems
Evolutionary ELM	Zhu et al. [18]	In Evolutionary ELM, the input weights and hidden biases were optimized by the modified differential evolutionary algorithm	Four real benchmark classification problems
Voting-based ELM	Cao et al. [17]	V-ELM performed multiple independent ELM training instead of a single ELM training and then made the final decision based on the majority voting method	Nineteen real-world datasets from the UCI database and the protein information resource center
Ordinal ELM	Deng et al. [22]	By the ordinal ELM algorithm, the SLFN was redesigned for ordinal regression problems, and the algorithms were trained by the ELM	Artificial data, nine small sample, and five large regression datasets
Fully complex ELM	Li et al. [23]	In fully complex ELM algorithm, the ELM algorithm was extended from the real domain to the complex domain	A complex nonminimum-phase channel model introduced by Cha and Kassam
Symmetric ELM	Liu et al. [24]	Symmetric ELM transformed the original activation function of hidden neurons into a symmetric one with respect to the input variables of the samples	Two toy function approximation problems, and two chaotic time series prediction tasks

Output: the output weights β .

Step 1. Randomly assign the parameters of hidden nodes (a_i, b_i) , $i = 1, \dots, \tilde{N}$.

Step 2. Calculate the output matrix of the hidden layer H .

Step 3. Calculate the output weight $\beta : \beta = H^\dagger T$.

4 Variants of ELM

In this section, several typical variants of ELM were summarized and introduced briefly.

4.1 Incremental ELM

Huang et al. [9] proposed an incremental extreme learning machine (I-ELM) to construct an incremental feedforward network. I-ELM randomly added nodes to the hidden layer one by one and freezed the output weights of the existing hidden nodes when a new hidden node was added. I-ELM

is not only efficient for SLFN with continuous activation functions (including differentiable), but also for SLFNs with piecewise continuous activation functions (such as threshold).

On that basis of I-ELM, convex I-ELM (CI-ELM) and enhance I-ELM (EI-ELM) were presented by Huang et al. Different from I-ELM, CI-ELM [11] recalculated the output weights of the existing hidden nodes after a new hidden node was added. CI-ELM could achieve faster convergence rates and more compact network architectures than I-ELM while retaining the I-ELM's simplicity and efficiency. EI-ELM [25] allowed maximum number of hidden nodes, no control parameters need to be manually set by users. Different from the original I-ELM, EI-ELM picked the optimal hidden, node which led to the smallest residual error at each learning step among several randomly generated hidden nodes. EI-ELM could achieve faster convergence rate and much more compact network architect. In addition, Huang et al. [26] also presented an improved I-ELM with fully complex hidden nodes, which extended I-ELM from the real domain to the complex domain.

4.2 Pruning ELM

In view of too few/many hidden nodes employed would lead to underfitting/overfitting issues in pattern classification, Rong et al. [12] presented a pruned-ELM (P-ELM) algorithm as a systematic and automated approach for designing ELM network. P-ELM began with an initial large number of hidden nodes and then removed the irrelevant or lowly relevant hidden nodes by considering their relevance to the class labels during learning. As a result, the architectural design of ELM can be automated. Simulation results showed that the P-ELM led to compact network classifiers that generate fast response and robust prediction accuracy on unseen data when compared with the standard ELM, BP, and MRAN. P-ELM mainly adapted to pattern classification problems.

4.3 Error-minimized ELM

Feng et al. [19] proposed an error-minimization-based method for ELM (EM-ELM) that can grow hidden nodes one by one or group by group to automatically determine the number of hidden nodes in generalized SLFNs. During the growth of the networks, the output weights were updated incrementally, which significantly reduced the computational complexity. The simulation results on sigmoid type of hidden nodes showed that this approach could significantly reduce the computational complexity of ELM and propose an efficient implementation of ELM.

4.4 Two-stage ELM

To obtain a parsimonious solution for the network structure of preliminary ELM, Lan et al. [20] introduced a systematic two-stage algorithm (named TS-ELM). In the first stage, a forward recursive algorithm was applied to select the hidden nodes from the candidates randomly generated in each step and add them to the network until the stopping criterion was met. Meanwhile, the significance of each hidden node was measured by the net contribution when it was added to the network. In the second stage, the selected hidden nodes were reviewed to eliminate the insignificance nodes from the network, which drastically reduced the network complexity. Empirical studies on the six cases showed that TS-ELM with a much smaller network structure could achieve better or similar performance as that of EM-ELM.

4.5 Online sequential ELM

When the conventional ELM is used, all the training data should be available for training. However, in real applications, the training data may be obtained chunk by chunk or one by one. Liang et al. [21] presented a sequential learning algorithm referred to as online sequential extreme

learning machine (OS-ELM), which can handle both additive and RBF nodes in a unified framework. In OS-ELM with additive nodes, the input weights linking the input nodes to hidden nodes and biases were randomly generated, and then, the output weights were analytically determined based on the output of hidden nodes. Unlike other sequential learning algorithms, OS-ELM only required the number of hidden nodes to be specified as the conventional ELM. To improve the performance of OS-ELM and introduce the sequential learning mode into the ensemble networks, Lan et al. [27] proposed an integrated network structure, called ensemble of online sequential extreme learning machine (EOS-ELM). EOS-ELM was composed by several OS-ELM networks, whose final measurement of network performance was calculated by the average value of outputs of each OS-ELM in the ensemble. Moreover, to reflect the timeliness of training data in the process of learning, Zhao et al. [28] introduced an improved EOS-ELM, called online sequential extreme learning machine with forgetting mechanism (FOS-ELM), which can retain the advantages of EOS-ELM and improve the learning effects by discarding the outdated data quickly in the process of learning to reduce their bad affection to the following learning.

4.6 Evolutionary ELM

Generally, the number of hidden neurons is determined randomly when ELM is applied. However, ELM may need higher number of hidden neurons due to the random determination of the input weights and hidden biases. A novel learning algorithm named evolutionary extreme learning machine (E-ELM) was proposed by Zhu et al. [18], to optimize the input weights and hidden biases and determine the output weights. In E-ELM, the modified differential evolutionary (DE) algorithm was used to optimize the input weights and hidden biases. And Moore–Penrose (MP) generalized inverse was used to analytically determine the output weights. Experimental results show that E-ELM is able to achieve good generalization performance with much more compact networks, superior to other algorithms including BP, GALS, and the original ELM.

4.7 Voting-based ELM

Since learning parameters of hidden nodes in ELM are randomly assigned and remain unchanged during the training procedure, ELM may not obtain the optimal classification boundary, which leads to those samples near the classification boundary may be misclassified. So, Cao et al. [17] proposed an improved algorithm called voting-based extreme learning machine (V-ELM) to reduce the

number of those misclassified samples near the classification boundary. The main idea in V-ELM is to perform multiple independent ELM training instead of a single ELM training and then make the final decision based on the majority voting method [17]. V-ELM not only enhanced the classification performance and reduced the number of misclassified samples, but also lowered the variance among different realizations. Simulations on many real-world classification datasets indicated that V-ELM generally outperformed the original ELM algorithm as well as several recent classification algorithms.

4.8 Ordinal ELM

In order to further study the ELM algorithm for ordinal regression problems, Deng et al. [22] presented an encoding-based ordinal regression framework and three ELM-based ordinal regression algorithms. The paper designed an encoding-based framework for ordinal regression which included three encoding schemes: single multi-output classifier, multiple binary classifications with one-against-all decomposition method, and one-against-one method. Based on the framework, the SLFN was redesigned for ordinal regression problems, and the algorithms were trained by the extreme learning machine. Widely experiments on three kinds of datasets showed that ordinal ELM can obtain extremely rapid training speed and good generalization ability.

4.9 Fully complex ELM

To extend the application of the ELM algorithm, Li et al. [23] proposed a fully complex extreme learning algorithm (named C-ELM). In C-ELM, the ELM algorithm was extended from the real domain to the complex domain. Similar to ELM, the input weights and hidden layer biases of C-ELM were randomly chosen based on some continuous distribution probability, and then, the output weights were simply analytically calculated instead of being iteratively tuned. Then, C-ELM is used for equalization of a complex nonlinear channel with QAM signals.

4.10 Symmetric ELM

Liu et al. [24] presented a modified ELM algorithm, called symmetric ELM (S-ELM), which transformed the original activation function of hidden neurons into a symmetric one with respect to the input variables of the samples. In theory, S-ELM can preserve the capability of approximating N arbitrary distinct samples with zero error. Simulation results showed that S-ELM can obtain better generalization performance, faster learning speed, and more compact network architecture with the help of the prior knowledge of symmetry.

Besides the above-mentioned models of ELM, there are some other modified methods used to improve the performance of ELM, such as PCA-ELM [29], fuzzy ELM [30], robust ELM [31], parallel ELM [32], regularized ELM [33], and weighted ELM [34]. Due to limited space, we do not describe these methods in detail.

5 Applications of ELM

Recently, ELM algorithm has been applied to many areas. This section lists some classical applications of ELM.

5.1 Classification

Wang et al. [35] proposed a novel architecture of mobile object index, where R-tree was used to index the occupied regions instead of the mobile objects themselves and ELM was used to classify the region dynamically to adapt to changes in environment. Zheng et al. [36] applied the regularization extreme learning machine (RELM) to text categorization in which RELM algorithm was developed including the uni-label and multi-label situations. Karpagachelvi et al. [37] used ELM to classify ECG signals, which was an electrical recording of the heart and was used in the investigation of heart disease. Kim et al. [38] proposed an arrhythmia classification algorithm using ELM in ECG, which showed effective accuracy performance with a short learning time. Lee et al. [39] used ELM to classify machine control commands out of time series of spike trains of ensembles of CA1 hippocampus neurons ($n = 34$) of a rat.

5.2 Regression

To solve the regression problem by ELM on very large-scale datasets, He et al. [32] designed and implemented an efficient parallel ELM (PELM) for regression. Experiments demonstrated that PELM not only could process large-scale dataset, but also had a good speedup, scaleup, and sizeup performance. To avoid the adverse effects caused by the perturbation or the multi-collinearity, Li and Niu [40] proposed an enhanced ELM based on ridge regression (RR-ELM) for regression. In RR-ELM, the output weight matrix was calculated analytically by the method of ridge regression estimator. Balasundaram and Kapil [41] proposed the study of ELM for e-insensitive regression formulated in 2-norm as an unconstrained optimization problem in primal variables. Feng et al. [42] addressed a novel ELM framework based on the evolutionary algorithm for regression. In this framework, two ELM networks were generated, which have L and $L/2$ hidden nodes separately, and then, natural selection strategy was used to

ensure the better hidden nodes to survive in next generation.

5.3 Pattern recognition

Zong and Huang [43] studied the performance of the one-against-all (OAA) and one-against-one (OAO) ELM for classification in multi-label face recognition applications, and the performance was verified through four benchmarking face image datasets. Mohammed et al. [44] introduced a human face recognition algorithm based on bidirectional two-dimensional principal component analyses (B2DPCA) and ELM. Minhas et al. [45] proposed a recognition framework for human actions using ELM based on visual vocabularies. Chacko et al. [46] applied wavelet energy and ELM to handwritten character recognition where ELM was used to classify the features of handwritten characters to accelerate the speed of learning algorithms. Lan et al. [47] used ELM to examine on the text-independent speaker verification task. Nian et al. [48] presented a method based on the geometrical topology model and ELM for 3D object recognition, which can identify the inherent distribution and the dependence structure for each 3D object. Besides, ELM is also applied in other aspects, such as surface reconstruction [49], face gender recognition [50], fingerprint matching [51], and text categorization [36].

5.4 Forecasting and diagnosis

Chen and Ou [52] presented the Gray extreme learning machine (GELM), integrated by Gray relation analysis and ELM with Taguchi method, to construct a forecasting model in the retail industry, which not only provided smaller predicting errors but also improved the training speed more than other forecasting models. Sun et al. [53] applied ELM to investigate the relationship between sales amount and some significant factors, which affect demand, and outperformed other methods based on BPNN. Hu et al. [54] proposed a multi-stage ELM to improve the accuracy of clustering and used it in hydraulic tube tester data. Daliri [55] presented a hybrid automatic diagnosis system combining genetic algorithm (GA) and fuzzy ELM for the lung cancer, which can be used for clinical applications. Xu et al. [56] developed an ELM-based predictor for real-time frequency stability assessment (FSA) of power systems.

5.5 Image processing

Zhou et al. [49] used an improved ELM called polyharmonic extreme learning machine (P-ELM) to reconstruct a smoother surface with a high accuracy and robust stability. Chen Pan et al. [57] presented a fast and simple framework for leukocyte image segmentation by learning with ELM and sampling via simulating visual system. In the framework, ELM classifier was trained online to simulate visual

neuron system and then extracted pixels of object from image. Pan et al. [58] proposed an iterative framework for figure-ground segmentation by sampling learning via simulating human vision. In this framework, ELM was used to train the pixels classifier based on the RGB color to extract object regions and provide a reference boundary of objects.

5.6 Other applications

Malathi et al. [59] proposed a new approach based on combined wavelet transform-extreme learning machine (WT-ELM) technique for fault section identification, classification, and location in a series-compensated transmission line. Zhao et al. [60] present a partial least-squares-based extreme learning machine (called PLS-ELM) to enhance the estimate performance of effluent quality in terms of accuracy and reliability. Li et al. [61] developed an efficient ELM-based model for evaluating unit generation strategies in RTS games, by which both the unit interactions and the production sequence can be implicitly and simultaneously handled. Li et al. [62] present an effective computer-aided diagnosis (CAD) system based on principle component analysis (PCA) and ELM to assist the task of thyroid disease diagnosis.

6 Discussion

In the variants of ELM, incremental ELM and pruning ELM are two basic methods adjusting the number of nodes in the hidden layer, which are mainly proposed by Huang and his research team. The purpose of these methods is to find the appropriate number of nodes in the hidden layer. In addition, error-minimized ELM and two-stage ELM also focus on adjusting the number of nodes in the hidden layer in essence. Different from above methods, the online sequential ELM provided a fast and accurate online learning method for ELM, which makes that the ELM algorithm can learn data one by one or chunk by chunk (a block of data) with fixed or varying chunk size. Then, evolutionary ELM used DE algorithm to optimize the input weights and hidden biases, which may be a time-consuming method for the constant iteration of DE algorithm. After that, fully complex ELM extended the ELM algorithm from the real domain to the complex domain. Besides, voting-based ELM, ordinal ELM, and symmetric ELM also improved the ELM algorithm to a certain degree.

The type applications of ELM include classification and regression problems. In these problems, ELM has lower computational time, better performance, and generalization ability than the conventional classifiers, such as BP neural networks and LS-LVM. In addition, ELM was also successfully applied on pattern recognition, forecasting and diagnosis, image processing, and other areas.

7 Conclusions

In this paper, we have demonstrated an overall review of the ELM algorithm, especially emphasizing on its variants and applications. Our goal is to introduce a valuable tool for applications to the researches, which can provide more accurate results and spend less the calculation time in the classification or regression problems than the conventional methods, such as BP neural networks and LS-LVM.

There are also some open problems of the ELM algorithm to be solved. The following issues remain open and may be worth absorbing the attentions of researchers in the future.

1. How to determine the appropriate number of neurons in the hidden layer according to the different datasets. In fact, experimental studies demonstrate that the performance of basic ELM is stable in a wide range of number of hidden nodes. Thus, how to find the range of the optimum solution and how to prove it in theory remain open.
2. Compared to the conventional learning algorithms, the generalization performance of ELM turns out to be more stable. How to estimate the oscillation bound of the generalization performance of ELM remains open too [63].
3. How to effectively solve the classification problems of mass data. Existing experiments proved that ELM has better performance and generalization ability than the conventional models of neural networks. However, when processing mass data or big data, the ELM algorithm needs to be tested and verified.
4. Parallel and distributed computing of ELM [63] will become the next focus of the ELM theories, which will broaden the applications of ELM. So, how to adjust the ELM algorithm to improve its ability of parallel and distributed computing of ELM remains open too.
5. More applications may be needed to check the generalization ability of ELM, especially in some areas with mass data.

Acknowledgments This work is supported by the National Natural Science Foundation (No. 61379101), the 973 Program (No. 2013CB329502), the Basic Research Program (Natural Science Foundation) of Jiangsu Province of China (No. BK20130209), the Opening Foundation of the Key Laboratory of Intelligent Information Processing of Chinese Academy of Sciences (No. IIP2010-1), and the Opening Foundation of Beijing Key Lab of Intelligent Telecommunications Software and Multimedia, Beijing University of Posts and Telecommunications.

References

1. Xu XZ, Ding SF, Shi ZZ, Zhu H (2012) Optimizing radial basis function neural network based on rough set and AP clustering algorithm. *J Zhejiang Univ Sci A* 13(2):131–138
2. Chen Y, Zheng WX (2012) Stochastic state estimation for neural networks with distributed delays and Markovian jump. *Neural Netw* 25:14–20
3. Ding SF, Su CY, Yu JZ (2011) An optimizing BP neural network algorithm based on genetic algorithm. *Artif Intell Rev* 36(2):153–162
4. Francisco FN, César HM, Gutiérrez PA, Carbonero-Ruz M (2011) Evolutionary q -Gaussian radial basis function neural networks for multiclassification. *Neural Netw* 24(7):779–784
5. Ding SF, Jia WK, Su CY, Zhang LW (2011) Research of neural network algorithm based on factor analysis and cluster analysis. *Neural Comput Appl* 20(2):297–302
6. Razavi S, Tolson BA (2011) A new formulation for feedforward neural networks. *IEEE Trans Neural Netw* 22(10):1588–1598
7. Huang GB, Zhu QY, Siew CK (2004) Extreme learning machine: a new learning scheme of feedforward neural networks. In: *Proceedings of international joint conference on neural networks (IJCNN2004)*, vol 2, no 25–29, pp 985–990
8. Huang GB, Zhu QY, Siew CK (2006) Extreme learning machine: theory and applications. *Neurocomputing* 70(1–3):489–501
9. Huang GB, Chen L, Siew CK (2006) Universal approximation using incremental constructive feedforward networks with random hidden nodes. *IEEE Trans Neural Netw* 17(4):879–892
10. Huang GB, Chen L (2008) Enhanced random search based incremental extreme learning machine. *Neurocomputing* 71:3060–3068
11. Huang GB, Chen L (2007) Convex incremental extreme learning machine. *Neurocomputing* 70:3056–3062
12. Rong HJ, Ong YS, Tan AH, Zhu Z (2008) A fast pruned-extreme learning machine for classification problem. *Neurocomputing* 72:359–366
13. Huang GB, Ding X, Zhou H (2010) Optimization method based extreme learning machine for classification. *Neurocomputing* 74:155–163
14. Lim JS, Lee S, Pang HS (2013) Low complexity adaptive forgetting factor for online sequential extreme learning machine (OS-ELM) for application to nonstationary system estimations. *Neural Comput Appl* 22(3–4):569–576
15. Huang GB, Zhou H, Ding X, Zhang R (2012) Extreme learning machine for regression and multiclass classification. *IEEE Trans Syst Man Cybern Part B Cybern* 42(2):513–529
16. Wang L, Huang YP, Luo XY, Wang Z, Luo SW (2011) Image deblurring with filters learned by extreme learning machine. *Neurocomputing* 74:2464–2474
17. Cao JW, Lin ZP, Huang GB, Liu N (2012) Voting based extreme learning machine. *Inf Sci* 185(1, 15):66–77
18. Zhu QY, Qin AK, Suganthan PN, Huang GB (2005) Evolutionary extreme learning machine. *Pattern Recognit* 38:1759–1763
19. Feng GR, Huang GB, Lin QP, Gay R (2009) Error minimized extreme learning machine with growth of hidden nodes and incremental learning. *IEEE Trans Neural Netw* 20(8):1352–1357
20. Lan Y, Soh YC, Huang GB (2010) Two-stage extreme learning machine for regression. *Neurocomputing* 73(16–18):3028–3038
21. Liang NY, Huang GB, Saratchandran P, Sundararajan N (2006) A fast and accurate on-line sequential learning algorithm for feed-forward networks. *IEEE Trans Neural Netw* 17(6):1411–1423
22. Deng WY, Zheng QH, Lian SG, Chen L, Wang X (2010) Ordinal extreme learning machine. *Neurocomputing* 74(1–3):447–456
23. Li MB, Huang GB, Saratchandran P, Sundararajan N (2005) Fully complex extreme learning machine. *Neurocomputing* 68:306–314
24. Liu XY, Li P, Gao CH (2013) Symmetric extreme learning machine. *Neural Comput Appl* 22(3–4):551–558
25. Huang GB, Chen L (2008) Enhanced random search based incremental extreme learning machine. *Neurocomputing* 71:3460–3468

26. Huang GB, Li MB, Chen L, Siew CK (2008) Incremental extreme learning machine with fully complex hidden nodes. *Neurocomputing* 71:576–583
27. Lan Y, Soh YC, Huang GB (2009) Ensemble of online sequential extreme learning machine. *Neurocomputing* 72:3391–3395
28. Zhao JW, Wang ZH, Park DS (2012) Online sequential extreme learning machine with forgetting mechanism. *Neurocomputing* 87(15):79–89
29. Castano A, Fernandez-Navarro F, Hervás-Martínez C (2013) PCA-ELM: a robust and pruned extreme learning machine approach based on principal component analysis. *Neural Process Lett* 37(3):377–392
30. Zhang WB, Ji HB (2013) Fuzzy extreme learning machine for classification. *Electron Lett* 49(7):448–449
31. Horata P, Chiewchanwattana S, Sunat K (2013) Robust extreme learning machine. *Neurocomputing* 102(SI):31–44
32. He Q, Shang TF, Zhuang FZ (2013) Parallel extreme learning machine for regression based on MapReduce. *Neurocomputing* 102(SI):52–58
33. Yu Qi, Miche Yoan, Eirola Emil (2013) Regularized extreme learning machine for regression with missing data. *Neurocomputing* 102(SI):45–51
34. Zong WW, Huang GB, Chen YQ (2013) Weighted extreme learning machine for imbalance learning. *Neurocomputing* 101:229–242
35. Wang BT, Wang GR, Li JJ, Wang B (2012) Update strategy based on region classification using ELM for mobile object index. *Soft Comput* 16(9):1607–1615
36. Zheng WB, Qian YT, Lu HJ (2013) Text categorization based on regularization extreme learning machine. *Neural Comput Appl* 22(3–4):447–456
37. Karpagachelvi S, Arthanari M, Sivakumar M (2012) Classification of electrocardiogram signals with support vector machines and extreme learning machine. *Neural Comput Appl* 21(6):1331–1339
38. Kim J, Shin HS, Shin K, Lee M (2009) Robust algorithm for arrhythmia classification in ECG using extreme learning machine. *Biomed Eng*. doi:[10.1186/1475-925X-8-31](https://doi.org/10.1186/1475-925X-8-31)
39. Lee Y, Lee H, Kim J, Shin HC, Lee M (2009) Classification of BMI control commands from rat's neural signals using extreme learning machine. *Biomed Eng*. doi:[10.1186/1475-925X-8-29](https://doi.org/10.1186/1475-925X-8-29)
40. Li GQ, Niu PF (2013) An enhanced extreme learning machine based on ridge regression for regression. *Neural Comput Appl* 22(3–4):803–810
41. Balasundaram S (2013) On extreme learning machine for e-insensitive regression in the primal by Newton method. *Neural Comput Appl*. doi:[10.1007/s00521-011-0798-9](https://doi.org/10.1007/s00521-011-0798-9)
42. Feng GR, Qian ZX, Zhang XP (2012) Evolutionary selection extreme learning machine optimization for regression. *Soft Comput* 16(9):1485–1491
43. Zong WW, Huang GB (2011) Face recognition based on extreme learning machine. W. Zong, G.-B. Huang *Neurocomput* 74:2541–2551
44. Mohammed AA, Minhas R, Jonathan WuQM, Sid-Ahmed MA (2011) Human face recognition based on multidimensional PCA and extreme learning machine. *Pattern Recognit* 44:2588–2597
45. Minhas R, Baradarani A, Seifzadeh S, Jonathan WuQM (2010) Human action recognition using extreme learning machine based on visual vocabularies. *Neurocomputing* 73:1906–1917
46. Chacko BP, Vimal Krishnan VR, Raju G, Babu Anto P (2012) Handwritten character recognition using wavelet energy and extreme learning machine. *J Mach Learn Cyber* 3:149–161
47. Lan Y, Hu ZJ, Soh YC, Huang GB (2013) An extreme learning machine approach for speaker recognition. *Neural Comput Appl* 22(3–4):417–425
48. Nian R, He B, Lendasse A (2013) 3D object recognition based on a geometrical topology model and extreme learning machine. *Neural Comput Appl* 22(3–4):427–433
49. Zhou ZH, Zhao JW, Cao FL (2013) Surface reconstruction based on extreme learning machine. *Neural Comput Appl* 23(2):283–292
50. Yang JC, Jiao YB, Xiong NX (2013) Fast face gender recognition by using local ternary pattern and extreme learning machine. *KSII Trans Intern Inf Syst* 7(7):1705–1720
51. Yang JC, Xie SJ, Yoon S (2013) Fingerprint matching based on extreme learning machine. *Neural Comput Appl* 22(3–4):435–445
52. Chen FL, Ou TY (2011) Sales forecasting system based on Gray extreme learning machine with Taguchi method in retail industry. *Expert Syst Appl* 38:1336–1345
53. Sun ZL et al (2008) Sales forecasting using extreme learning machine with applications in fashion retailing. *Decis Support Syst* 46:411–419
54. Hu XF, Zhao Z, Wang S, Wang FL, He DK, Wu SK (2008) Multi-stage extreme learning machine for fault diagnosis on hydraulic tube tester. *Neural Comput Appl* 17:399–403
55. Daliri MR (2012) A hybrid automatic system for the diagnosis of lung cancer based on genetic algorithm and fuzzy extreme learning machines. *J Med Syst* 36:1001–1005
56. Xu Y, Dai YY, Dong ZY, Zhang R, Meng K (2013) Extreme learning machine-based predictor for real-time frequency stability assessment of electric power systems. *IET Gener Transm Distrib* 7(4):391–397
57. Pan C, Park DS, Yang Y, Yoo HM (2012) Leukocyte image segmentation by visual attention and extreme learning machine. *Neural Comput Appl* 21(6):1217–1227
58. Pan C, Park DS, Lu HJ, Wu XP (2012) Color image segmentation by fixation-based active learning with ELM. *Soft Comput* 16(9):1569–1584
59. Malathi V, Marimuthu NS, Baskar S, Ramar K (2011) Application of extreme learning machine for series compensated transmission line protection. *Eng Appl Artif Intell* 24:880–887
60. Zhao LJ, Wang DH, Chai TY (2013) Estimation of effluent quality using PLS-based extreme learning machines. *Neural Comput Appl* 22(3–4):509–519
61. Li YJ, Li Y, Zhai JH, Shiu S (2012) RTS game strategy evaluation using extreme learning machine. *Soft Comput* 16(9):1627–1637
62. Li LN, Ouyang JH, Chen HL, Liu DY (2012) A computer aided diagnosis system for thyroid disease using extreme learning machine. *J Med Syst* 36(5):3327–3337
63. Huang GB, Wang DH, Lan Y (2011) Extreme learning machines: a survey. *Int J Mach Learn Cyber* 2:107–122