ASTES

Advances in Science, Technology and Engineering Systems Journal Vol. 2, No. 1, 69-75 (2017)

www.astesj.com

Special Issue on Computer Systems, Information Technology, Electrical and Electronics Engineering

ASTESJ

ISSN: 2415-6698

Recent Trends in ELM and MLELM: A review

R. Manju Parkavi*[,1], M. Shanthi[1], M.C. Bhuvaneshwari[2]

[1]*Department of ECE, Kumaraguru College of Technology, 641035, India*

[2]*Department of EEE, PSG College of Technology, 641004, India*

ARTICLE INFO

Article history:
Received: 08 December, 2016
Accepted: 15 January, 2017
Online: 28 January, 2017

Keywords :
Extreme learning machine (ELM)
Artificial Neural Network (ANN)
MLELM
Deep Learning

ABSTRACT

Extreme Learning Machine (ELM) is a high effective learning algorithm for the single hidden layer feed forward neural networks. Compared with the existing neural network learning algorithm it solves the slow training speed and over-fitting problems. It has been used in different fields and applications such as biomedical engineering, computer vision, remote sensing, chemical process and control and robotics. It has better generalization stability, sparsity, accuracy, robustness, optimal control and fast learning rate This paper introduces a brief review about ELM and MLELM, describing the principles and latest research progress about the algorithms, theories and applications. Next, Multilayer Extreme Learning Machine (MLELM) and other state-of-the-art classifiers are trained on this suitable training feature vector for classification of data. Deep learning has the advantage of approximating the complicated function and mitigating the optimization difficulty associated with deep models. Multilayer extreme learning machine is a learning algorithm of an Artificial Neural Network (ANN) which takes to be good for deep learning and extreme learning machine. This review presents a comprehensive view of these advances in ELM and MLELM which may be worthy of exploring in the future.

1. Introduction

As early as in the 1940s, arithmetician Pitts and psychologist McCulloch have put forward neurons mathematical model (MP model) from the arithmetical logic view (McCulloch and Pitts 1943) which opened the prelude of artificial neural network experimental work. Neural network with parallel and distributed information processing network framework has a strong nonlinear mapping ability and robust self-learning, adaptive and fault tolerance characteristics. Due to the recent popularity of deep learning, two of the most widely studied artificial neural networks these days are auto encoders and Boltzmann machines. An auto encoder with a single hidden layer as well as a structurally restricted version of the Boltzmann machine, called a restricted Boltzmann machine, have become popular due to their application in layer-wise pretraining of deep Multilayer perceptron's. Table.1 describes about Difference in Conventional learning method and Biological learning method. The progress of machine learning and artificial intelligence is based on the coexistence of three necessary conditions: superlative computing environments, rich and large data, and very effective learning techniques (algorithms). The Extreme Learning Machine as an evolving learning technique gives efficient unified solutions to generalize feed-forward networks including but not limited to (each supports single- and multi-hidden-layer) neural networks, Radial Basis Function (RBF) networks and kernel learning. Single hidden layer feed forward networks (SLFNs), is one of the most accepted feed forward neural networks, have been greatly studied from each theoretical and application aspect for their learning capabilities and fault-tolerant dependability [1–6]. However, most accepted learning algorithms for training SLFNs are still relatively not quick since all the parameters of SFLNs need to be tuned through repetitive procedures and these algorithms may also stuck in a local minimum.

Recently, a new rapid learning neural algorithm for SLFNs, named extreme learning machine (ELM) [7, 8], was developed to enhance the efficiency of SLFNs. Different from the existing

*R.ManjuParkavi, 4 Mahendrakumar building, Burgur road, anthiyur, erode. 9894768880, vprmanju@gmail.com

https://dx.doi.org/10.25046/aj020108

learning algorithms for neural networks (such as Back Propagation (BP) algorithms), which may face problems in manually tuning control parameters (learning rate, learning epochs, etc.) and/or local minima, ELM is automatically employed without iterative tuning, and in theory, no intervention is required from users. Furthermore, the learning speed of ELM is exceptionally fast compared to other traditional methods. Unlike current single-hidden-layer ELM methods, in this paper, we indicate that with general hidden nodes (or called sub network nodes) added to existing networks, MLELM can be used for classification, dimensions reduction, and feature learning.

In addition, Autoencoder (AE) based on deep network also can get better features and classification results. From this point of view, deep learning is suitable for feature extraction of images. However, the learning speed of deep learning is not fast. Due to this reason, similar to deep learning, MLELM stacks Extreme Learning Machine Autoencoder (ELM-AE) to create a Multilayer neural network which has much faster learning speed with better performance [14-16].

Table 1. Difference in Conventional learning method and Biological learning method

Conventional Learning Methods	Biological Learning
Very sensitive to network size	Stable in a wide range (tens to thousands of neurons in each module)
Difficult for parallel implementation	Parallel implementation
Difficult for hardware implementation	“Biological” implementation
Very sensitive to user specified parameters	Free of user specified parameters
Different network types for different type of applications	One module possibly for several types of applications
Time consuming in each learning point	Fast in micro learning point
Difficult for online sequential learning	Nature in online sequential learning
“Greedy” in best accuracy	High speed and high accuracy
“Brains (devised by conventional learning methods)” are chosen after applications are present	Brains are built before applications

2. Extreme Learning Machine

Extreme learning machine was planned for generalized single hidden layer feed forward neural network [9] where the hidden layer need not be neuron alike. Unlike other neural networks with back propagation [10], the hidden nodes in ELM are randomly generated, whereas the activation functions of the neurons are nonlinear piecewise continuous. The weights between the hidden layer and the output layer are calculated analytically. The general architecture of single layer ELM is shown in Figure 1.

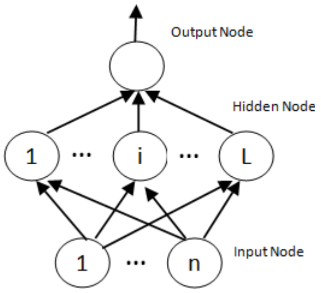

Figure 1.Structure of single layer feed-forward neural network

There are two training phases in ELM algorithm: feature mapping and output weights solving. ELM feature mapping: Given input data x, the output function of ELM for generalized SLFNs is

$$f(x) = \sum_{i=1}^{L} \beta_i h_i(x) = h(x)\beta \quad (1)$$

Where $h(x) = [h_1(x), \ldots, h_L(x)]$is the output vector of the hidden layer and $\beta = [\beta_1, \ldots, \beta_L]^T$ denotes the output weights between the hidden layer of L hidden nodes and the output layer. The procedure of getting h is called ELM feature mapping which maps the input data from RD to the feature space RL. In real applications, **h** can be described as

$$h_i(x) = g(a_i, b_i, X), a_i \epsilon R^D, b_i \epsilon R \quad (2)$$

Where g(a,b,x) is an activation function satisfying ELM universal approximation capability theorems [11]. In fact, any nonlinear piecewise continuous functions (e.g. Gaussian, Sigmoid, etc.) can be used as activation function h. In ELM, the parameters of h are randomly produced based on a continuous probability distribution.

ELM output weights solving: In the second phase, given a training sample set, $(\mathrm{X_i}, \mathrm{t_i})_{\mathrm{i}}^{\mathrm{n}} = 1$ with $t_i = [0, \ldots 0, 1_{\mathrm{i}}, 0, . \ldots, 0_{\mathrm{m}}]^{\mathrm{T}}$the class indicator of xi, ELM aims to minimize both the training error and the Frobenius norm of output weights. This objective function used for both binary and multi-class classification tasks, can be expressed as follows,

$$\underset{\beta,\xi}{\mathrm{Min}} \frac{\omega}{2} \sum_{i=1}^{n} \|\xi\|_2^2 + \frac{1}{2} \|\beta\|_F^2,$$

$$s.t. \quad \beta h(X_i) = t_i - \xi_i, \ \forall \in 1,2, \ldots, \mathrm{n} \quad (3)$$

where n is the number of samples and denotes the training error of the i-th sample, ω is a regularization parameter which trades off the average of output weights and training error and denotes the Frobenius norm. The optimization problem in (3) can be efficiently solved. Specifically, according to the Wood bury identity [12] the optimal β can be analytically obtained as

$$\beta^* = \begin{cases} \left(H^T H + \frac{I_L}{\omega}\right) H^T T & if\ L \leq k \\ H^T (H H^T + \frac{I_n}{\omega})^{-1} T & otherwise \end{cases} \quad (4)$$

Where In and IL are identity matrices, and H is the hidden layer output matrix (randomized matrix) which is defined in (5)

$$H = \begin{bmatrix} h(x_1) \\ . \\ . \\ . \\ h(x_n) \end{bmatrix} = \begin{bmatrix} h_1(x_1) & \dots & h_L(x_1) \\ . & . & . \\ . & . & . \\ . & . & . \\ h_1(x_n) & \dots & h_L(x_n) \end{bmatrix} \quad (5)$$

2.1. ELM-Based Autoencoder

Apart from the ELM-based SLFNs, the ELM theory has also been applied to build an autoencoder for an MLP. Concept wise, the autoencoder functions are of feature extractor in a multilayer learning framework. It uses the encoded outputs to approximate the original inputs by minimizing the reconstruction errors.

Mathematically, the autoencoder maps the input data x to a higher level representation, and then uses latent representation y through a deterministic mapping $y = h\theta(X) = g(A * X + b)$, parameterized by $\theta = \{A, b\}$, where g(□) is the activation function, A is a d’ × d weight matrix and b is a bias vector. The resulting latent representation y is then mapped back to a reconstructed vector z in the input space

z = hθ ‘(y) = g(A’□ y + b) with θ’ = {A’, b’}.

Using randomly mapped outputs as inherent representation y, one can easily build the ELM-based autoencoder as shown in [10]. Then, the reconstruction of x can be regarded as an ELM learning problem, in which A is obtained by solving regularized least mean square optimization. However, due to the use of penalty in the original ELM, the extracted features by the ELM autoencoder in [10] tend to be dense and may have redundancy. In this case, a more sparse solution is preferred.

In classifier design, used for classification of heart stroke level with the help of 13 inputs, is derived from the standard database and 5 target values of kind of disease, each of which represents a class Algorithm, describes the steps of the proposed solving system [17]. As we can see from the algorithm steps, the output binary values are changed to [-0.9 +0.9] make the output more observable, the two selected parameters that effect the performance of the network are the constant value C, which is selected here to be 10000, and number of hidden neurons L, which is set to 100 for most of the works. The biggest computation effort will be done in the inversion of the matrix G. Whenever the dimensions are big, the system needs more time and even specific inverse computation algorithms, like the Moore-Penrose generalized inverse which has been worked by Samet and Miri to increase the speed of computation.

2.2. ELM Variants

Though achieving great victory on both theoretical and practical aspects, basic ELM cannot efficiently handle large-scale learning tasks because of the limitation of memory and the intensive computational cost of the inverse of larger matrices. To decrease the runtime memory, many variants including Online Sequential ELM (OS-ELM) [20] and Incremental ELM (I-ELM) [21] have been proposed. OS-ELM can decrease the requirement of runtime memory because the model is trained based on each chunk. I-ELM can diagnose memory problem by training a basic network with some hidden nodes and then adding hidden node to the conventional network one by one. To rest the computational cost incurred by these operations, many variants containing partitioned ELM [22] and parallel ELM [23], have been recommended. Recently, a variant of OS-ELM also called parallel OS-ELM (POS-ELM) [24] and parallel regularized ELM (PR-ELM) [25] are recommended to reduce training time and memory requirement instantaneously.

3. Multilayer Extreme Learning Machine

Multilayer ELM is an advanced machine learning approach based on the architecture of artificial neural network and is inspirited by deep learning and extreme learning machine. Deep learning was first suggested by Hinton and Salakhutdinov (2006) who in their process used deep structure of multilayer autoencoder and recognized a multilayer neural network on the unsupervised data. In their method, first they used an unsupervised training to obtain the parameters in each layer. Then, the network is fine-tuned by supervised learning. Hinton et al. (2006), who suggested the deep belief network, outperforms the conventional multilayer neural network, SLFNs, Support Vector Machines (SVMs), but it has slow learning speed. Working in this way, recently Kasun et al. (2013) proposed multilayer ELM which performs unsupervised learning from layer to layer, and it is not necessary to iterate during the training process and hence, it does not spend a long time in the training phase. Compared to other existing deep networks, it has a better or comparable performance.

Figure 2 shows the structure of MLELM with ELM-AE, output weights β^1 of (a) ELM-autoencoder (AE) is denotedwith respect to input data x are the layer 1 weights of MLELM . (b) Output weight matrix is denoted as β^{i+1} in ELM-AE, with respect to i-th hidden layer and output hi of MLELM are the i + 1th layer weights of MLELM; (c) Regularized least squares are used for output layer weight calculation of MLELM.

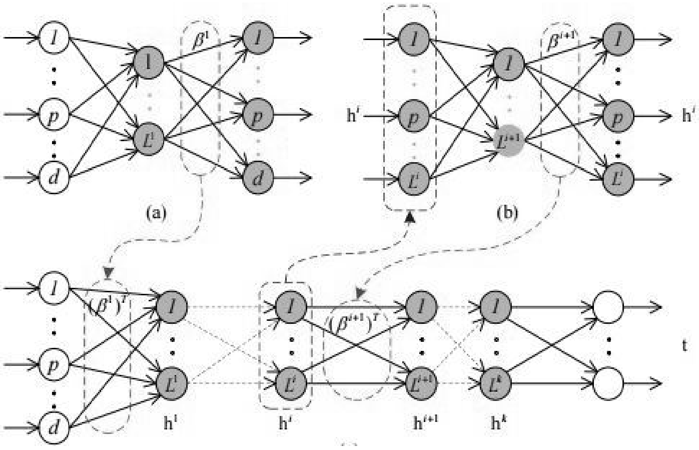

Figure 2. ELM-AE and Multilayer ELM

3.1. Extreme Learning Machine- Autoencoder

Autoencoder is an unsupervised neural network. The outputs and inputs of the autoencoder are similar. Like ELM, ELM-AE has n input layer nodes, single hidden layer of L nodes and n output layer nodes. In spite of many resemblances between these two, there are two major differences that exist between them which are as regards,

i. . ELM is a supervised neural network while ELM-AE is an unsupervised one and its output is similar as the input.

ii. Input weights and biases of the hidden layer are random in case of ELM, but they are orthogonal in ELM-AE.

Depends upon the number of hidden layer nodes, the ELM-AE [18] can be classified and shown as the following three categories,

(i) Compressed representation (n > L): In compressed representation, features of training dataset need to be presented from a higher-dimensional input signal space to a lower-dimensional feature space.

(ii) Equal dimension representation (n = L): In this module representation of features, the dimension of input signal space and feature space needs to be equal.

(iii) Sparse representation (n < L): It is just the reverse of compressed representation where features of training dataset need to be applied from a lower-dimensional input signal space to a higher dimensional feature space.

The multilayer ELM is extensively faster than deep networks because iterative tuning mechanism is not require in case of MLELM and got better or similar performance compared to deep networks. It is also known that in ELM, for L hidden nodes and N training examples (x j,y j), the following (6) holds

$$g_L = (x_j) = \sum_{i=1}^{L} \beta_i g_i \left(x_j, w_i, b_i\right) = y_j, j = 1, \dots, N \qquad (6)$$

In order to perform unsupervised learning, few alterations have been done in ELM-AE whose working principle is similar to regular ELM, which are represented as follows,

1) The output data and the input data remain same for every hidden layer. Hence, for every input data X: Y = X

(2) To improve the performance of ELMs, we need to consider the weights and the biases of the random hidden nodes to be present orthogonal and can be represented as follows

$$H = g(w \cdot x + b), wT \cdot w = I \text{ and } bT \cdot b = 1 \qquad (7)$$

(3) The output weight ***β*** is decided based on the following conditions,

i. if ***n*** > ***L*** then,

$$\beta = \left(\frac{I}{C} + H^T H\right)^{-1} H^T X \qquad (8)$$

ii. if n = L then,

$$\beta = H^{-1} X \qquad (9)$$

iii. if ***n*** < ***L*** then,

$$\beta = H^T \left(\frac{I}{C} + H^T H\right)^{-1} X \qquad (10)$$

Where C is a scale parameter which adjusts analytical and experiential risk. ELM-AE is used for training the parameters in each layer of MLELM. The general formula representing MLELM is described as follows,

$$H^n = g((\beta^n)^T H^{n-}) \qquad (11)$$

For n = 0, the 0th hidden layer or the first layer is considered to be the input layer X. The final output matrix Y can be acquired by computing the results between the last hidden layer and the output layer using the regularized least squares model work.

3.2. Hierarchical Extreme Learning Machine Variants

Currently, ELM is also extended to Multilayer structures, i.e. Multilayer ELM (MLELM) [19], hierarchical ELM [26] and another reference work hierarchical local receptive fields ELM (H-LRF-ELM) [27]. Similar to other deep networks, MLELM [19] performs unsupervised learning in layer-by-layer. In different to deep networks, MLELM does not require fine-tuning using back propagation which will decrease the computational cost in learning process. H-ELM [26] is a new ELM-based hierarchical learning frame work for Multilayer perception. For feature extraction, H-ELM work as unsupervised Multilayer encoding. Unlike the demanding layer-wise training of deep learning, the layers of H-ELM are learned in a forward manner. Therefore, it has better learning efficiency than the deep learning methods. LRF-ELM was first suggested by Huang et al. [27].

In this model, the connections between the input and hidden nodes are sparse and bounded by local receptive fields (LRF). LRF-ELM is learning the local structures and generates more meaningful representations at the hidden layer when dealing with image processing and like tasks. LRF-ELM can be enhanced to Multilayer architecture called hierarchical LRF-ELM (H-LRF-ELM). The layers of H-LRF-ELM can be classified into two parts, namely, the feature extractor and the ELM classifier. After all, ELM-based Multilayer networks seem to provide better accuracy, performance and efficiency than other deep networks.

3.3. Hierarchical Extreme Learning Machine

The proposed H-ELM is built in a multilayer manner, H-ELM training architecture is structurally divided into two separate phases: 1) unsupervised hierarchical feature representation and 2) supervised feature classification. H-ELM, as well as its advantages against the existing DL and multilayer ELM (MLELM) algorithms. Before unsupervised feature learning, the input raw data should be transformed into an ELM random feature space, which can help to exploit hidden information among training samples. The universal approximation capability of the ELM is exploited for the design of autoencoder and moreover, sparse constraint is added upon the autoencoder optimization and therefore, we term it as ELM sparse auto encoder.

4. Discussion

The field of deep neural networks, or deep learning, is expanding rapidly, and it is impossible to discuss everything in this thesis. Single layer ELM, Multilayer ELM, autoencoder and Boltzmann machines, which are main topics of this survey, are certainly not the only neural networks in the field of deep neural networks. However, as the aim of this survey is to provide a brief overview and introduction to deep neural networks.

Although deep neural networks have shown extremely competitive performance in various machine learning tasks, the theoretical motivation for using them is still debated. Compared to the feed forward neural networks such as autoencoders and Multilayer ELM, it is difficult to evaluate Boltzmann machines.

Even when the structure of the network is highly restricted, the existence of the intractable normalization constant requires using a sophisticated sampling-based estimation method to evaluate Boltzmann machines.Table.2 presented the performance comparison of MLELM method with other deep learning algorithm. MLELM achieves best training accuracy and Training Time.

Table 2. Performance comparison of MLELM and other method

Learning method	Testing Accuracy	Training Time
Multilayer ELM [Huang, et al 2013]	99.03±0.04	444.7sec
Deep Belief Networks (DBN)	98.87	20580sec
Deep Boltzmann Machines	99.05	68246sec
Stacked Auto Encoders (SAE)	98.6	>17 hours
Stacked Denoising Auto Encoders (SDAE)	98.72	>17 hours

5. Applications

Neural network is extensively used in artificial intelligence, data mining, image classification and other applications. Single hidden layer feed forward neural network has a strong training ability. Extreme learning methods is proposed to overcome the disadvantage of a single hidden layer feed forward neural network, improve the learning ability and generalization efficiency.

These recent success of deep neural networks in both academic research and commercial applications may be attributed to several recent breakthroughs Layerwise pretraining of a Multilayer [13] showed that a clever way of initializing parameters can easily overcome the difficulties of optimizing a large, deep neural network. As ELM algorithm does not require large-scale operations, there is no need to adjust the hidden layer of the network and compared with the conventional computational intelligence techniques, such as BP algorithm the whole learning process is necessary only one iteration and the training speed improved more than 10 times.

5.1. ELM for fault detection in Analog Integrated Circuits

Test generation algorithm based on ELM, similar algorithm is cost effective and low-risk for analog device under test (DUT). This method uses test patterns from the test generation algorithm to support DUT and then samples output responses of the DUT for fault classification and detection. Active test generation algorithms cannot save time effectively and it lags in accuracy when the number of impulse-response samples decreases. Due to this computational complexity and classification theory of methods, ELM-based algorithm is used [28].

5.2. Image Processing

In H-ELM framework, several H-ELM-based feature extraction and classification algorithms are developed for practical computer vision applications, such as object(image) detection,object recognition, and object tracking. The obtained results are quite promising, and further confirm the generality and capability of the H-ELM.

Applications of MLELM have recently been presented in sparse learning and classification, computer vision, clustering learning, image processing, high-dimensional and large-data applications, etc.

5.3. Car Detection

Car detection is a classical application in computer vision. The proposed MLELM-based car detection algorithm and one can see that a sliding window is used to extract a fixed-size image patch, which are grouped and input into the H-ELM for training and testing.

5.4. Fault Diagnosis of Power Transformer

The power transformer, one of the key power equipment, is mainly composed of hydrocarbon oil and paper insulation. Various methods are applied to interpret dissolved gas-in-oil analysis (DGA) results, they may sometimes fail to diagnose exactly. The incipient fault identification accuracy of artificial intelligence (AI) is varied with input variable variation. Therefore, selection of input variable is prime research area. Principle component analysis using RapidMiner software is adapted to IEC TC10 and related datasets is used to identify most relevant input variables for incipient fault classification. Extreme learning machine (ELM) is implemented to classify the incipient faults of power transformer.New application of ELM to the real-time fault solving system for rotating machinery. The structure is successfully applied on recognizing fault patterns coming from the Windmill [29].

5.5. Gesture Recognition

Gesture recognition is also an important topic in computer vision due to its wide ranges of applications, such as human–computer interfaces, sign language interpretation, and visual surveillance. The different images are directly fed into the AIOS-ELM, H-ELM-based classifier to recognize each of the gesture [30].

5.6. Online Incremental Tracking

Online incremental tracking refers to the tracking method that aims to learn and adapt a representation to reflect the appearance changes of the target. The appearance changes of the specified target include pose variation and shape deformation. Apart from these, extrinsic illumination change, camera motion, viewpoint, and occlusions inevitably cause large appearance variation. Modelling such appearance variation is always one of the nature problems of tracking. For this purpose H-ELM is used for incremental tracking method.

5.7. Multiclass Applications

SVM is used only for binary classification. In order to solve problem in multiclass classification, ELM is preferred. ELM, which is of higher scalability and low computational complexity, not only unifies other popular learning algorithms but also provides a unified solution to different practical applications (e.g., regression, binary, and multiclass classifications).

5.8. Medical Applications

Extreme Learning Machine (ELM) algorithm is used to model these factors such as age, sex, serum cholesterol, blood sugar, etc. The proposed system can replace costly medical checkups with a warning system for patients of the probable presence of heart based disease. Brain-computer interface (BCI) is a technology that is used for communication with a computer or to control devices with EEG signals .The main technologies of BCI areto extract the preprocessed EEG features and to classify ready processed EEG. BCI is used for many fields, such as medicine and military.

Currently, many methods have been proposed for EEG classification, including decision trees, local Backpropagation algorithm, Bayes classifier, K-nearest neighbors (KNN), support vector machine (SVM, and ELM . Most of them are shallow neural network algorithms where the capabilities achieve approximating the complex functions have certain restrictions, and there is no such restriction in deep learning. Deep learning [31] is a multilayer perceptron artificial neural network algorithm. DELM in EEG classification and makes use of two BCI competition datasets to test the performances of DELM.

5.9. 3D Applications

3D shape features play a critical role in graphics applications like 3D shape matching, recognition, and retrieval. Various 3D shape descriptors have been developed over the last two decades.

The Convolutional Autoencoder Extreme Learning Machine (CAE-ELM) can be used in practical graphics applications, such as in 3D shape completion. Optical acquisition devices most often generate incomplete 3D shape data because of occlusion and unfavorable surface reflectance properties.

5.10. Security

Privacy protection is a significant issue such that confidential images must be encrypted in corporations. Nevertheless, decryption and then classifying millions of encrypted images becomes a heavy burden to computation. Multilayer Extreme Learning Machine is proposed for encrypted image classification framework that is able to precisely classify encrypted images without decryption.

5.11. Fault Diagnosis of Wind Turbine Equipment

Fault diagnosis is critical in the wind turbine generator system to reduce the maintenance cost and to avoid unplanned interruption. The data generated from wind turbine generator system is always of high-dimensional and dynamical. To avoid these problems, a new fault diagnosis method of multiple extreme learning machines is proposed. It enables both representational feature learning and fault classification. Feature learning includes data preprocessing, feature extraction and dimensional reduction. Unlike the other greedy layer wise learning scheme adopted in back propagation; it does not need iterative fine-tuning of parameters. The results show that the evolved diagnostic framework accomplishes the best performance among the compared approaches in terms of accuracy and effectiveness in several faults detection in wind turbines [32].

6. Conclusion

This paper demonstrates the overall review of ELM and MLELM algorithm, especially emphasizing on its variants and applications. Our aim is to introduce a valuable tool for applications to the researches, which provides more accurate results and spend less calculation time in the classification or regression problems than the existing methods, such as BP neural networks and LS-LVM. MLELM training scheme, which is based on the universal approximation capability of the original ELM. The proposed H-ELM achieves high level representation with layerwise encoding, and outperforms the original ELM in various simulations. Moreover, compared with other ELM training methods, the training of MLELM is much faster and achieves higher learning accuracy.There are also some open problems of the ELM algorithm to be diagnosed. The following problems remain open and may be worth absorbing the attentions of researchers in the future,

1. ELM with sub network nodes architecture [33] has not attracted much research attentions. Recently, many works have been proposed for supervised/unsupervised dimension reduction or representation learning, one type of problem.

2. More applications may be needed to check the generalization ability of MLELM, especially in some areas with mass data.

References

[1] Adamos, D. A., Laskaris, N. A., Kosmidis, E. K., &Theophilidis, G. (2010). NASS: an empirical approach to spike sorting with overlap resolution based on a hybrid noise-assisted methodology. *Journal of neuroscience methods*, *190*(1), 129-142.

[2] An, L., &Bhanu, B. (2012, September). Image super-resolution by extreme learning machine. In *2012 19th IEEE International Conference on Image Processing* (pp. 2209-2212). IEEE.

[3] Avci, E. (2013). A new method for expert target recognition system: Genetic wavelet extreme learning machine (GAWELM). *Expert Systems with Applications*, *40*(10), 3984-3993.

[4] Avci, E., & Coteli, R. (2012). A new automatic target recognition system based on wavelet extreme learning machine. *Expert Systems with Applications*, *39*(16), 12340-12348.

[5] Bai, Z., Huang, G. B., Wang, D., Wang, H., & Westover, M. B. (2014). Sparse extreme learning machine for classification. *IEEE transactions on cybernetics*, *44*(10), 1858-1870.

[6] Balbay, A., Avci, E., Şahin, Ö.,&Coteli, R. (2012). Modeling of drying process of bittim nuts (Pistaciaterebinthus) in a fixed bed dryer system by using extreme learning machine. *International Journal of Food Engineering*, *8*(4).

[7] Balbay, A., Kaya, Y., & Sahin, O. (2012). Drying of black cumin (Nigella sativa) in a microwave assisted drying system and modeling using extreme learning machine. *Energy*, *44*(1), 352-357.

[8] Baradarani, A., Wu, Q. J., &Ahmadi, M. (2013). An efficient illumination invariant face recognition framework via illumination enhancement and DD-DTCWT filt.

[9] Barea, R., Boquete, L., Ortega, S., López, E., & Rodríguez-Ascariz, J. M. (2012). EOG-based eye movements codification for human computer interaction. *Expert Systems with Applications*, *39*(3), 2677-2683.

[10] Huang, G. B., Zhou, H., Ding, X., & Zhang, R. (2012). Extreme learning machine for regression and multiclass classification. *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics)*, *42*(2), 513-529.

[11] Basu, A., Shuo, S., Zhou, H., Lim, M. H., & Huang, G. B. (2013). Silicon spiking neurons for hardware implementation of extreme learning machines. *Neurocomputing*, *102*, 125-134.

[12] Bazi, Y., Alajlan, N., Melgani, F., AlHichri, H., Malek, S., &Yager, R. R. (2014). Differential evolution extreme learning machine for the classification of hyperspectral images. *IEEE Geoscience and Remote Sensing Letters*, *11*(6), 1066-1070.

[13] Belkin, M., &Niyogi, P. (2003). Laplacianeigenmaps for dimensionality reduction and data representation. *Neural computation*, *15*(6), 1373-1396.

[14] Wang, Y., Xie, Z., Xu, K., Dou, Y., & Lei, Y. (2016). An efficient and effective convolutional autoencoder extreme learning machine network for 3d feature learning. *Neurocomputing*, *174*, 988-998.

[15] Roul, R. K., Asthana, S. R., & Kumar, G. (2016). Study on suitability and importance of multilayer extreme learning machine for classification of text data. *Soft Computing*, 1-18.

[16] Wang, W., Vong, C. M., Yang, Y., & Wong, P. K. (2016). Encrypted image classification based on multilayer extreme learning machine. *Multidimensional Systems and Signal Processing*, 1-15.

[17] Ismaeel, S., Miri, A., &Chourishi, D. (2015, May). Using the Extreme Learning Machine (ELM) technique for heart disease diagnosis. In *Humanitarian Technology Conference (IHTC2015), 2015 IEEE Canada International* (pp. 1-3). IEEE.

[18] Decherchi, S., Gastaldo, P., Leoncini, A., & Zunino, R. (2012). Efficient digital implementation of extreme learning machines for classification. *IEEE Transactions on Circuits and Systems II: Express Briefs*, *59*(8), 496-500.

[19] Decherchi, S., Gastaldo, P., Leoncini, A., & Zunino, R. (2012). Efficient digital implementation of extreme learning machines for classification. *IEEE Transactions on Circuits and Systems II: Express Briefs*, *59*(8), 496-500.

[20] Lemme, A., Freire, A., Barreto, G., &Steil, J. (2013). Kinesthetic teaching of visuomotor coordination for pointing by the humanoid robot icub. *Neurocomputing*, *112*, 179-188.

[21] Kavukcuoglu, K., Ranzato, M. A., &LeCun, Y. (2010). Fast inference in sparse coding algorithms with applications to object recognition. *arXiv preprint arXiv:1010.3467*.

[22] Tang, J., Deng, C., & Huang, G. B. (2016). Extreme learning machine for multilayer perceptron. *IEEE transactions on neural networks and learning systems*, *27*(4), 809-821.

[23] Le, Q., Sarlós, T., & Smola, A. (2013, June). Fastfood-approximating kernel expansions in loglinear time. In *Proceedings of the international conference on machine learning*.

[24] Liang, N. Y., Huang, G. B., Saratchandran, P., &Sundararajan, N. (2006). A fast and accurate online sequential learning algorithm for feedforward networks. *IEEE Transactions on Neural networks*, *17*(6), 1411-1423.

[25] Li, B., Li, Y., &Rong, X. (2013). The extreme learning machine learning algorithm with tunable activation function. *Neural Computing and Applications*, *22*(3-4), 531-539.

[26] Li, L., Liu, D., & Ouyang, J. (2012). A new regularization classification method based on extreme learning machine in network data. *J. Inf. Comput. Sci*, *9*(12), 3351-3363.

[27] Li, Y., Li, Y., Zhai, J., & Shiu, S. (2012). RTS game strategy evaluation using extreme learning machine. *Soft Computing*, *16*(9), 1627-1637.

[28] Zhou, J., Tian, S., Yang, C., & Ren, X. (2014). Test generation algorithm for fault detection of analog circuits based on extreme learning machine. *Computational intelligence and neuroscience*, *2014*, 55.

[29] Malik, H., & Mishra, S. (2015, December). Extreme learning machine based fault diagnosis of power transformer using IEC TC10 and its related data. In *2015 Annual IEEE India Conference (INDICON)* (pp. 1-5). IEEE.

[30] Cambria, E., Huang, G. B., Kasun, L. L. C., Zhou, H., Vong, C. M., Lin, J., ... & Leung, V. C. (2013). Extreme learning machines [trends & controversies]. *IEEE Intelligent Systems*, *28*(6), 30-59

[31] Ding, S., Zhang, N., Xu, X., Guo, L., & Zhang, J. (2015). Deep extreme learning machine and its application in EEG classification. *Mathematical Problems in Engineering*, *2015*

[32] Yang, Z. X., Wang, X. B., & Zhong, J. H. (2016). Representational Learning for Fault Diagnosis of Wind Turbine Equipment: A Multi-Layered Extreme Learning Machines Approach. *Energies*, *9*(6), 379.

[33] Yang, Y., & Wu, Q. J. (2016). Multilayer extreme learning machine with subnetwork nodes for representation learning.