Neural Networks 61 (2015) 32–48

Contents lists available at ScienceDirect

Neural Networks

journal homepage: www.elsevier.com/locate/neunet

Review

Trends in extreme learning machines: A review

Gao Huang [a], Guang-Bin Huang [b,*], Shiji Song [a,*], Keyou You [a]

[a] *Department of Automation, Tsinghua University, Beijing 100084, China*

[b] *School of Electrical and Electronic Engineering, Nanyang Technological University, Nanyang Avenue, 639798, Singapore*

ARTICLE INFO

Article history:
Received 20 March 2014
Received in revised form 25 August 2014
Accepted 2 October 2014
Available online 16 October 2014

Keywords:
Extreme learning machine
Classification
Clustering
Feature learning
Regression

ABSTRACT

Extreme learning machine (ELM) has gained increasing interest from various research fields recently. In this review, we aim to report the current state of the theoretical research and practical advances on this subject. We first give an overview of ELM from the theoretical perspective, including the interpolation theory, universal approximation capability, and generalization ability. Then we focus on the various improvements made to ELM which further improve its stability, sparsity and accuracy under general or specific conditions. Apart from classification and regression, ELM has recently been extended for clustering, feature selection, representational learning and many other learning tasks. These newly emerging algorithms greatly expand the applications of ELM. From implementation aspect, hardware implementation and parallel computation techniques have substantially sped up the training of ELM, making it feasible for big data processing and real-time reasoning. Due to its remarkable efficiency, simplicity, and impressive generalization performance, ELM have been applied in a variety of domains, such as biomedical engineering, computer vision, system identification, and control and robotics. In this review, we try to provide a comprehensive view of these advances in ELM together with its future perspectives.

Contents

1. Introduction ... 33
2. Classical ELM ... 34
   2.1. ELM hidden nodes, feature mappings and feature space ... 34
   2.2. Basic ELM ... 34
   2.3. Essences of ELM ... 35
   2.4. Regularized ELM and ELM kernels ... 35
   2.5. Learning principles of ELM ... 36
3. ELM theories ... 36
   3.1. Interpolation theory ... 36
   3.2. Universal approximation capability ... 36
   3.3. Generalization bound ... 37
4. ELM for classification and regression ... 37
   4.1. Improving the stability of ELM ... 37
   4.2. Improving the compactness of ELM ... 37
   4.3. ELM for online sequential data ... 38
   4.4. ELM for imbalanced data ... 38
   4.5. ELM for noisy/missing data ... 39
   4.6. Incremental ELM ... 39

* Corresponding authors.
E-mail addresses: huang-g09@mails.tsinghua.edu.cn (G. Huang), egbhuang@ntu.edu.sg (G.-B. Huang), shijis@mail.tsinghua.edu.cn (S. Song), youky@mail.tsinghua.edu.cn (K. You).

http://dx.doi.org/10.1016/j.neunet.2014.10.001

4.7. ELM ensembles 39
4.8. ELM for ranking 40
4.9. ELM for semi-supervised learning 40
4.10. Other variants of ELM 41
5. ELM for unsupervised learning 41
5.1. ELM for embedding and clustering 41
5.2. ELM for representational learning 42
6. ELM for feature selection 42
7. Implementation of ELM 42
7.1. Parallel implementation of ELM 42
7.2. Hardware implementation of ELM 42
8. Empirical study of ELM and applications 43
8.1. Comparison with SVM and its variant 43
8.2. Comparison with deep learning 43
8.3. ELM for medical/biomedical application 43
8.4. ELM for computer vision 43
8.5. ELM for image processing 43
8.6. ELM for system modeling and prediction 43
9. Conclusions 44
Acknowledgments 44
References 45

1. Introduction

Feedforward neural networks (FNN) have been well studied and widely used since the introduction of the well-known back-propagation (BP) algorithm (Rumelhart, Hinton, & Williams, 1986). Traditional BP algorithm is essentially a first order gradient method for parameter optimization, which suffers from slow convergence and local minimum problem. Researchers have proposed various ways to improve the efficiency or optimality in training FNN, such as second order optimization methods (Hagan & Menhaj, 1994; Wilamowski & Yu, 2010), subset selection methods (Chen, Cowan, & Grant, 1991; Li, Peng, & Irwin, 2005) or global optimization methods (Branke, 1995; Yao, 1993). Though leading to faster training speed or better generalization performance compared to the BP algorithm, most of these methods still cannot guarantee a global optimal solution.

Recently, extreme learning machine (ELM) has been proposed for training single hidden layer feedforward neural networks (SLFNs). In ELM, the hidden nodes are randomly initiated and then fixed without iteratively tuning. Actually, the hidden nodes in ELM are even not required to be neuron alike. The only free parameters need to be learned are the connections (or weights) between the hidden layer and the output layer. In this way, ELM is formulated as a *linear-in-the-parameter* model which boils down to solving a linear system. Compared to traditional FNN learning methods, ELM is remarkably efficient and tends to reach a global optimum. Theoretical studies have shown that even with randomly generated hidden nodes, ELM maintains the universal approximation capability of SLFNs (Huang & Chen, 2007, 2008; Huang, Chen, & Siew, 2006). With commonly used activation functions, ELM can attain the almost optimal generalization bound of traditional FNN in which all the parameters are learned (Lin, Liu, Fang, & Xu, 2014; Liu, Lin, Fang, & Xu, 2014). The advantages of ELM in efficiency and generalization performance over traditional FNN algorithms have been demonstrated on a wide range of problems from different fields (Huang, Zhou, Ding, & Zhang, 2012; Huang, Zhu, & Siew, 2006). It is worth noting that ELM is generally much more efficient than support vector machines (SVMs) (Cortes & Vapnik, 1995), least square support vector machines (LS-SVMs) (Suykens & Vandewalle, 1999) and other state-of-the-art algorithms. Empirical studies have shown that the generalization ability of ELM is comparable or even better than that of SVMs and its variants (Fernández-Delgado, Cernadas, Barro, Ribeiro, & Neves, 2014; Huang, Song, Gupta, & Wu, 2014; Huang, Zhou, et al., 2012; Huang, Zhu, et al., 2006). Detailed comparisons of ELM and SVM can be found in Huang (2014) and Huang, Zhou, et al. (2012).

During the past decade, theories and applications of ELM have been extensively studied. From learning efficiency point of view, the original design objects of ELM have three-folds: least human invention, high learning accuracy and fast learning speed (as shown in Fig. 1). Various extensions have been made to the original ELM model to make it more efficient and suitable for specific applications. A literature survey on ELM theories and applications was given by Huang, Wang, and Lan (2011). Since then, research on ELM has become even more active. From theoretical aspect, the universal approximation capability of ELM has been further studied in Huang, Zhou, et al. (2012). The generalization ability of ELM has been investigated in the framework of statistical learning theory (Lin et al., 2014; Liu, Gao, & Li, 2012; Liu et al., 2014) and the initial localized generalization error model (LGEM) (Wang, Shao, Miao, & Zhai, 2013). Many variants of ELM have been proposed to meet particular application requirements. For example, in *cost sensitive* learning, the test time should be minimized, which requires a compact network to meet test time budget. To this end, ELM has been successfully adapted to achieve high compactness in network size (Bai, Huang, Wang, Wang, & Westover, 2014; Deng, Li, & Irwin, 2011; Du, Li, Irwin, & Deng, 2013; He, Du, Wang, Zhuang, & Shi, 2011; Lahoz, Lacruz, & Mateo, 2013; Li, Li & Rong, 2013; Martinez-Martinez et al., 2011; Wang, Er, & Han, 2014a; Yang, Wang, & Yuan, 2013, 2012; Yu & Deng, 2012). We also witness the extensions of ELM for online sequential data (Lan, Soh, & Huang, 2009; Liang, Huang, Saratchandran, & Sundararajan, 2006; Rong, Huang, Sundararajan, & Saratchandran, 2009; Ye, Squartini, & Piazza, 2013; Zhao, Wang, & Park, 2012), noisy/missing data (Horata, Chiewchanwattana, & Sunat, 2013; Man, Lee, Wang, Cao, & Miao, 2011; Miche et al., 2010; Yu, Miche, et al., 2013), imbalanced data (Horata et al., 2013; Huang et al., 2014; Zong, Huang, & Chen, 2013), etc. Additionally, apart from being used for traditional classification and regression tasks, ELM has recently been extended for clustering, feature selection and representational learning (Benoit, van Heeswijk, Miche, Verleysen, & Lendasse, 2013; Huang et al., 2014; Kasun, Zhou, & Huang, 2013). In this review, we provide a snapshot assessment of these new developments in the ELM theories and applications.

It is worth noting that the ELM learning frameworks' randomized strategies for nonlinear feature construction have drawn great interests in the computational intelligence and machine learning community (Le, Sarlos, & Smola, 2013; Rahimi & Recht, 2007, 2008a, 2008b; Saxe et al., 2011; Widrow, Greenblatt, Kim, & Park,