Expert Systems With Applications 140 (2020) 112875

Contents lists available at ScienceDirect

Expert Systems With Applications

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

Comparing of deep neural networks and extreme learning machines based on growing and pruning approach

Kemal Akyol

Department of Computer Engineering, Kastamonu University, Kastamonu 37100, Turkey

ARTICLE INFO

Article history:
Received 8 June 2019
Revised 30 July 2019
Accepted 15 August 2019
Available online 16 August 2019

Keywords:
Deep Neural Networks
Extreme Learning Machines
Growing and pruning
Parkinson
Self-care activities

ABSTRACT

Recently, the studies based on Deep Neural Networks and Extreme Learning Machines have become prominent. The models of parameters designed in these studies have been chosen randomly and the models have been designed in this direction. The main focus of this study is to determine the ideal parameters i.e. optimum hidden layer number, optimum hidden neuron number and activation function for Deep Neural Networks and Extreme Learning Machines architectures based on growing and pruning approach and to compare the performances of the models designed. The performances of the models are evaluated on two datasets; Parkinson and Self-Care Activities Dataset. Multi experiments have verified that the Deep Neural Networks architectures present a good prediction performance and this architecture outperforms the Extreme Learning Machines.

1. Introduction

Artificial Neural Network (ANN) is computational tool inspired from biological neural network system (Parveen Kumar & Pooja Sharma, 2014). It is used in many fields such as engineering sciences, specially computer sciences like medical diagnosis (Jafari-Marandi, Davarzani, Soltanpour Gharibdousti, & Smith, 2018), feature extraction based on image classification (Aytaç Korkmaz & Binol, 2018), time series prediction (Panigrahi & Behera, 2017) etc. A considerable amount of research based on the Deep Neural Networks (DNN) and Extreme Learning Machines (ELM) models have been proposed in recent years, such as a DNN model combined with the discrete wavelet transform and principal components analysis (Mohsen, El-Dahshan, El-Horbaty, & Salem, 2018); a DNN model combined with signal processing (Sannino & De Pietro, 2018), two-hidden-layer ELM (Qu, Lang, Liang, Qin, & Crisalle, 2015), two-stage ELM (Lan, Soh, & Huang, 2010), a weighted ELM for imbalanced class distribution (Li, Kong, Lu, Wenyin, & Yin, 2014), face recognition (Mohammed, Minhas, Wu, & Sid-Ahmed, 2011), handwritten character recognition (Chacko, Vimal Krishnan, Raju, & Babu Anto, 2012), image classification (Jun, Shitong & Chung, 2011), multiclass classification (Eirola et al., 2015).

The main focus of this study is to compare and evaluate the performances of the Multi-layer ELM architectures and Multi-layer DNN architectures based on growing and pruning approach. Also, the main contribution of this study is to find optimum hidden layer number and hidden neuron number and determine the ideal activation function for ELM model and ideal couple of activation function and optimization function for DNN model. The reason why these two architectures are preferred is that these two architectures are suitable for this approach.

The rest of this study is organized as follows; Section 2 presents the related works for ELM end DNN architectures briefly. Section 3 presents some background on the learning algorithms we have used. Section 4 addresses the experiments and results of the DNN and ELM models carried out on two datasets. Finally, Section 5 draws discussion and conclusion.

2. Literature review

An overview of some of the previous studies related with DNN and ELM is presented below. Lee et al. tried to estimate LDL-cholesterol by utilizing the DNN model including three input values of total cholesterol, HDL cholesterol, and triglyceride. The model, which consists of six hidden layers with 30 nodes, was trained on the dataset collected from Korean National Health and Nutrition Examination Survey. The performance of the model was tested on another dataset collected from Wonju Severance Christian Hospital. The model presented better performance compared to other existing methods (Lee, Kim, Uh, & Lee, 2019).

Feng et al. presented a DNN regression in order to predict solidification defects on the small dataset which consists of 487 instances. According to this study, pre-trained and fine-tuned DNN outperform neural network, support vector machine, and DNN

E-mail address: kakyol@kastamonu.edu.tr

https://doi.org/10.1016/j.eswa.2019.112875
0957-4174/

trained by conventional methods (Feng, Zhou, & Dong, 2019). Kudugunta and Ferrara presented a deep neural networks based on long short-term memory architecture in order to detect bots on tweets. For this purpose, contextual features obtained from user metadata were sent to the DNN architecture. Based on results, proposed architecture presents high classification accuracy in recognizing the bots from humans (Kudugunta & Ferrara, 2018). Qi et al. presented a deep convolutional neural networks with multi-scale kernels and skip connections to diagnose breast ultrasonography images. Firstly, malignant tumors on the image were detected and then solid nodules were recognized (Qi et al., 2019). Another deep convolutional neural networks, which includes multiple-layer perceptrons and convolutional neural networks, were presented by Yang et al. They proposed a novel regulator named Structured Decorrelation Constraint in order to tackle both the generalization and optimization of deep neural networks (Yang, Xiong, Li, & Xu, 2019).

Cheng and Xiong presented an ELM model in order to improve the accuracy of dam displacement prediction (Cheng & Xiong, 2017). Yeom and Kwak focused on the prediction of the short-term electricity-load using a Takagi-Sugeno-Kang-based ELM. They achieved superior prediction performance and knowledge information with four activation functions such sigmoid, sine, radial basis function, and rectified linear unit (Yeom, Kwak, Yeom, & Kwak, 2017). Lu et al. proposed a novel adaptive weight online sequential extreme learning machine for predicting time series problems. The proposed study has good performance with respect to generalization performance, stability, and prediction ability (Lu et al., 2017). Men et al. developed a paraffin odor analysis system. The performances of the Support Vector Machine, Random Forest, and ELM algorithms on original feature set and optimized datasets are evaluated. Based on results, the ELM based model outperformed others (Men et al., 2018). Hosseinioun used the wavelet transform and adaptive ELM for prediction of the outlier occurrence in stock market time series (Hosseinioun, 2016). Lastly, Huang focused on the ELM theories such as hidden nodes and hidden neurons that need to be tuned in learning, and proved that it is good performance (Huang, Zhu, & Siew, 2006).

3. Background

3.1. Extreme Learning Machines and Deep Neural Networks

With the purpose of offering the information about the methods used in this study, the basic concepts of ELM and DNN architectures are briefly introduced in this section. Theoretical foundations of these architectures are well rooted from the classical neural networks architecture and they have been quite popular lately in the machine learning and data mining studies.

ELM proposed by Huang et al. (2006) is a new feed-forward neural network method which is presented high classification accuracy, good generalization ability rapidly (Cheng & Xiong, 2017). The basic of ELM is generalized as single hidden layer feed-forward networks where input weights and hidden biases are selected randomly. During the training process, the hidden layer of single hidden layer feed-forward networks need not to be tuned. The output weights are stated by using Moore-Penrose generalized inverse of the hidden-layer output matrix (Yeom et al., 2017). The structure of the ELM is shown in Fig. 1.

$$f_L(x) = \sum_{i=1}^{L} \beta_i G_i(w_i, b_i, x) \quad (1)$$

$$H(x) = [G_1(w_1, b_1, x), \ldots, G_L(w_L, bL_L, x)] \quad (2)$$

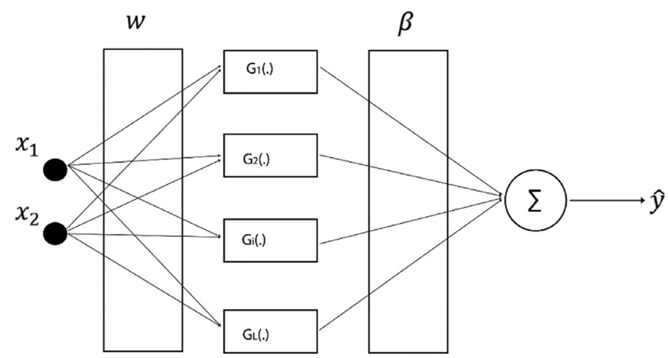

Fig. 1. An overview of conventional ELM architecture (Yeom et al., 2017).

Eq. (1) presents the output function of a generalized single hidden layer feed-forward networks.

Eq. (2) presents the output function in the hidden-layer mapping. w_i and b_i indicate weights and biases between the input layer and the hidden layer, respectively.

Huang revealed that hidden layer parameters can be assigned randomly and then the output weight can be calculated analytically (Huang, Zhu, & Siew,). Also, the execution time of an ELM model is very low and it provides more successful performance than other algorithms (Huang et al., 2006). As a known subject in the field of machine learning, of course, these parameters will vary for any dataset to be studied. Recently, there are many studies based on multi-layers ELM: a) the numbers of multi-layer and multi neurons assigned randomly b) average performance was calculated by utilizing multiple tests (i.e. the program run 1000 times). For example, Li et al. presented the number of optimum neurons (Li et al., 2014) and Xiao et al. found the best transfer function (Xiao, Li, & Mao, 2017). Deng et al. used random parameters and evaluated the averaged results in order to reduce the effect of these parameters given randomly (Deng, Zheng, & Chen, 2009).

The theoretical foundations of Deep Learning (DL) are well rooted from the classical neural networks architecture. In other words, DL is an up-to-date ANN architecture, which has been developed continuously and rapidly with different algorithms and approaches since the first day of its emergence, will continue to be popular for a long time in the computer science and other many fields. DNN which is a general deep framework covers the classification or regression analysis applications such as pattern recognition, data mining, image recognition and natural language processing etc. It is a powerful architecture and very popular in machine learning achieving successful results by making inferences from a dataset (Ravi et al., 2017).

An overview of the proposed DNN is given in Fig. 2. Input layer consists of input parameters for the input, hidden layers consist of hidden neurons and output layer consists of target class parameters.

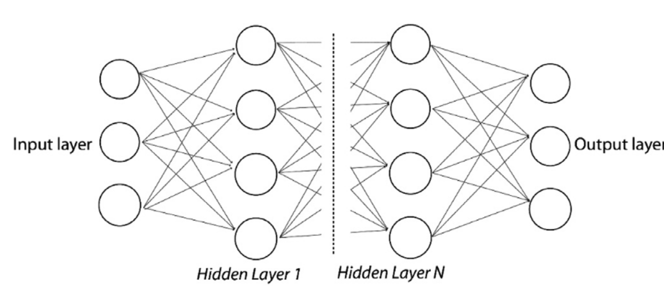

Fig. 2. An overview of the DNN architecture (Ravi et al., 2017).