Swarm and Evolutionary Computation 24 (2015) 25–49

Contents lists available at ScienceDirect

Swarm and Evolutionary Computation

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

Regular Paper

An improved cuckoo search based extreme learning machine for medical data classification

P. Mohapatra [a], S. Chakravarty [b], P.K. Dash [c,*]

[a] IIIT, Bhubaneswar, India
[b] Orissa Engineering College, Bhubaneswar, India
[c] Siksha 'O' Anusandhan University, Bhubaneswar, India

ARTICLE INFO

Article history:
Received 5 October 2014
Received in revised form
29 May 2015
Accepted 31 May 2015
Available online 10 June 2015

Keywords:
Extreme learning machine (ELM)
Online sequential extreme learning machine (OSELM)
Cuckoo search (CS)
Cuckoo search based extreme learning machine (CSELM)
Improved cuckoo search based extreme learning machine (ICSELM)

ABSTRACT

Machine learning techniques are being increasingly used for detection and diagnosis of diseases for its accuracy and efficiency in pattern classification. In this paper, improved cuckoo search based extreme learning machine (ICSELM) is proposed to classify binary medical datasets. Extreme learning machine (ELM) is widely used as a learning algorithm for training single layer feed forward neural networks (SLFN) in the field of classification. However, to make the model more stable, an evolutionary algorithm improved cuckoo search (ICS) is used to pre-train ELM by selecting the input weights and hidden biases. Like ELM, Moore–Penrose (MP) generalized inverse is used in ICSELM to analytically determines the output weights. To evaluate the effectiveness of the proposed model, four benchmark datasets, i.e. Breast Cancer, Diabetes, Bupa and Hepatitis from the UCI Repository of Machine Learning are used. A number of useful performance evaluation measures including accuracy, sensitivity, specificity, confusion matrix, Gmean, *F*-score and norm of the output weights as well as the area under the receiver operating characteristic (ROC) curve are computed. The results are analyzed and compared with both ELM based models like ELM, on-line sequential extreme learning algorithm (OSELM), CSELM and other artificial neural networks i.e. multi-layered perceptron (MLP), MLPCS, MLPICS and radial basis function neural network (RBFNN), RBFNNCS, RBFNNICS. The experimental results demonstrate that the ICSELM model outperforms other models.

1. Introduction

Classification of the exponentially growing complex, continuous data with large number of records and features has become the most challenging data mining tasks of human activity. In the last 20 years, it is being applied in the field of pattern classification like optical character recognition [1], text and image classification [2], machine vision [3], fraud detection [4], natural language processing [5], market segmentation [6,7], bioinformatics [8], protein sequence classification [9], biomedical image classification [10] and real world data classification [11]. The research community has given increasing attention in developing fast and accurate classifiers with good generalization capability. Even though a lot of classifiers [12] are already available, there is ample scope for improving the performance of the classifiers or to design good classifiers to handle more complex datasets and to gain more accuracy.

As of now various statistical and soft computing based classifiers have been proposed in the literature. The traditional statistical techniques like Euclidean minimum distance (EMD), quadratic minimum distance (QMD) and k-nearest neighbor (KNN) classifiers [1] and Bayesian decision theory [13] are used to build different classifiers. One of the drawbacks of the statistical method [13,14] is that it depends on the correctness of the underlying assumptions for its successful application. Unlike soft computing methods, in statistical method user needs to have thorough grasp over the properties of dataset for successful application of the model. However, it is not always possible. As a result statistical technique based classifiers generally give less accuracy as compared to the soft computing based classifiers.

Multilayer perceptron (MLP) [15], radial basis function neural networks (RBFNNs) [16], fuzzy rule based systems [17], adaptive neuro-fuzzy systems (ANFIS) [18], support vector machine (SVM) [19,20], k-nearest neighbor classifier (KNNs) [21], Naive Bayes classifier [22], polynomial classifiers [23], CART [24], decision tree classifier [25] and random forest [26] have been used as classifiers in many applications. However, amongst all the classifiers, ANN has been chosen by the researchers most frequently and extensively. While considering ANN as a classifier, the major things that

* Corresponding author. Tel.: +91 674 2727336.
E-mail address: pkdash.india@gmail.com (P.K. Dash).

http://dx.doi.org/10.1016/j.swevo.2015.05.003
2210-6502/

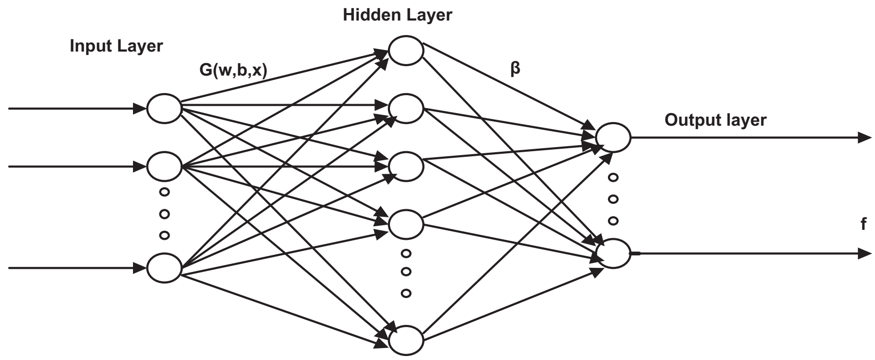

Fig. 1. Architecture of SLFN model.

are taken into consideration are the number of hidden layers, the values of the weights between input to hidden layer, hidden to output layer and the learning algorithms [27] i.e. backpropagation (BP), recursive least square (RLS) and different evolutionary learning algorithms. So the performance of ANN as a classifier mostly depends on the right combination of the structures and learning algorithms. However, the major disadvantage of ANN is considered to be its association with gradient descent learning algorithm that makes the performance of the model time consuming and increases the computational overhead [16]. In the gradient descent learning algorithm, due to the initial random choice of parameters, the convergence rate becomes very slow and most often it gets trapped in the local minima. To avoid the above said limitations, Huang et al. proposed a novel learning algorithm i.e. extreme learning machine (ELM) [11]. Researchers have shown that ELM owes its origin to random vector functional link (RVLF) [28–35]. However Huang [36] has shown its newness in Ref. [36]. The two major advantages of ELM are faster learning speed and good generalization ability. Literature survey reveals that ELM has been extensively used in many applications [37–43]. Although several variants of extreme learning machines [28–43] are now available for multiclass classification there remains several problems like the optimal choice of network size requiring a large number of hidden nodes for better generalization and choice of activation functions. Besides the randomness of ELM causes an additional uncertainty in regression and classification problems with regards to universal approximation and learning.

In the last seven years, a lot of research has already been done using ELM in the field of designing filters in image applications [44], sales forecasting [45], time series prediction [46], power system economic dispatch [47], electricity load forecasting [48,49], target recognition, aircraft recognition, clustering [50], real time fault diagnosis, end point prediction model, neural architecture design [51], disease diagnosis [52], mobility prediction in mobile ad-hoc networks, corporate life cycle prediction, system identification, breast tumor detection, etc. Especially for classification problems, ELM has been successfully employed in the various fields of classification like gene expression classification [53], binary class and multiclass data classification [11,54].

In ELM, the input weights and hidden biases are chosen randomly from which the output weights are computed. During this process, ELM tries to minimize the training error and determine the smallest norm of the output weights. Due to the random choice of the input weights and biases in ELM, many a times the output matrix does not show full column rank and this leads to ill-conditioning [55] of the system that generates non-optimal solutions. So, to improve the conditioning of ELM and to ensure the optimal solutions, evolutionary ELM [30,56–62] is used. The evolutionary ELM not only gives better accuracy but also ensures system stability. Amongst the evolutionary learning algorithms,

Table 1
Specification of the datasets used for classification.

Dataset name	#Trg samples	#Tst samples	#Features	#Classes	Random permutation
Breast Cancer	499	200	9	2	Yes
Diabetes	576	192	8	2	Yes
Bupa	199	146	6	2	Yes
Hepatitis	80	75	19	2	Yes

GA based ELM [27], PSO based ELM [57,58,60], DE based ELM [61], group search based ELM [62] have already been used in the literature. In recent years meta-heuristic algorithms have been successfully applied for solving practical and difficult optimization problems.

Thus in this paper, the recently developed meta-heuristic algorithm i.e. cuckoo search (CS) proposed by Yang and Deb [63–68] is used to pre-train the ELM ensuring optimal solutions. This algorithm is based on the combination of breeding behavior of cuckoos with the Lévy flight seen in some species of birds [66]. The CS algorithm has been developed to provide better performance than other established meta-heuristic algorithms. Civicioglu [68] proves that CSA yields better and robust solution as compared to particle swarm optimization (PSO), differential evolution (DE) and artificial bee colony (ABC). Again, to further improve the accuracy and stability of cuckoo search based extreme learning algorithm (CSELM), the improved cuckoo search algorithm [69,70] is combined with ELM i.e. improved cuckoo search extreme learning machine (ICSELM) model is proposed and experimented to classify four binary class datasets, Breast Cancer, Diabetes, Bupa and Hepatitis. Both CSELM and ICSELM choose the input weights and biases before calculating the output weights and they ensure the full column rank of the hidden layer output matrix. A fast batch learning algorithm i.e. online sequential ELM (OSELM) [71–73] and other two ANN based models, MLP and RBFNN with BP, CS and ICS learning algorithms are also taken to classify the above said binary datasets and compared with the proposed model. The performance of all the classifiers discussed in this paper is measured using various performance evaluation measures like Overall Accuracy, sensitivity, specificity, confusion matrix, Gmean, *F*-score and ROC analysis [54,74].The system complexity for CSELM and ICSELM are found to be O(sf^2) where 's' represents the number of samples and 'f' represents the number of features in the dataset.

The paper is organized as follows: The ELM and OSELM algorithms are briefly introduced in Section 2. The cuckoo search (CS), improved cuckoo search (ICS) and cuckoo search based algorithms such as CSELM and ICSELM are explained in Section 3. Section 4 introduces all the benchmark datasets. All the performance evaluation measures used in this study are discussed