Biomedical Signal Processing and Control 38 (2017) 400–410

Contents lists available at ScienceDirect

Biomedical Signal Processing and Control

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

Research paper

An adaptive kernel-based weighted extreme learning machine approach for effective detection of Parkinson's disease

Yang Wang [a,b,*], An-Na Wang [a], Qing Ai [a], Hai-Jing Sun [a]

[a] College of Information Science and Engineering, Northeastern University, Shenyang 110819, Liaoning, China
[b] School of Computer and Communication Engineering, Liaoning Shihua University, Fushun 113001, Liaoning, China

ARTICLE INFO

Article history:
Received 28 February 2017
Received in revised form 24 May 2017
Accepted 30 June 2017

Keywords:
Parkinson's disease
Imbalanced data
Extreme learning machine
Artificial bee colony
Feature selection

ABSTRACT

Imbalanced data appear in many real-world applications, from biomedical application to network intrusion or fraud detection, etc. Existing methods for Parkinson's disease (PD) diagnosis are usually more concerned with overall accuracy (ACC), but ignore the classification performance of the minority class. To alleviate the bias against performance caused by imbalanced data, in this paper, an effective method named AABC-KWELM has been proposed for PD detection. First, based on a fast classifier extreme learning machine (ELM), weighted strategy is used for dealing with imbalanced data and non-linear mapping of kernel function is used for improving the extent of linear separation. Furthermore, both binary version and continuous version of an adaptive artificial bee colony (AABC) algorithm are used for performing feature selection and parameters optimization, respectively. Finally, PD data set is used for evaluating rigorously the effectiveness of the proposed method in accordance with specificity, sensitivity, ACC, G-mean and F-measure. Experimental results demonstrate that the proposed AABC-KWELM remarkably outperforms other approaches in the literature and obtains better classification performance via 5-fold cross-validation (CV), with specificity of 100%, sensitivity of 98.62%, ACC of 98.97%, G-mean of 99.30%, and F-measure of 99.30%.

1. Introduction

After Alzheimer's disease, Parkinson's disease (PD) has become the second most common degenerative diseases of the central nervous system now. Because of the loss of dopamine-producing brain cell, patients with PD (PWP) are generally characterized by motor system disorders including bradykinesia, rigidity, tremor, and posture instability [1]. PD has influenced most of worldwide population, and the disease prevalence is increasing dramatically as people live longer. However, the cause of the disease has still unknown. It is reported that it is possible to alleviate symptoms remarkably in the early diagnosis of PD [2]. Therefore, the early diagnosis and treatment of PD is crucial. Research has shown that about 90% of PWP exhibit vocal impairment symptom known as dysphonia [3]. Dysphonia measurements have been a reliable diagnostic tool for PD.

Many researchers have handled PD diagnosis problem based on various machine learning techniques. Little et al. [4] has conducted a remarkable research on dysphonia measurements to discriminate healthy people from PWP. They employed kernel support vector machine (SVM) with feature selection method to detect PD. The classification accuracy of 91.4% was achieved using only four dysphonic features. Das [5] presented a comparative study of Decision Trees, Regression, DMneural and Neural Networks (NN) to detect PD. NN classifier yielded the best classification accuracy of 92.9%. Guo et al. [6] proposed the combination of genetic programming and expectation maximization algorithm for detecting PD. It improved data representation by creating feature functions and achieved the classification accuracy of 93.1%. Ozcift and Gulten [7] constructed 30 classifier ensembles based on rotation forest (RF) in combination with correlation based feature selection method for detecting PD, and it produced average classification accuracy of 87.13%. Aström and Koker [8] used a parallel feed-forward neural network structure to reduce the possibility of decision with error, and it achieved the classification accuracy of 91.20%. Luukka [9] introduced fuzzy entropy measures based feature selection method with similarity classifier for detecting PD. Mean classification accuracy of 85.03% was obtained using only two dysphonic features. Li et al. [10] used fuzzy-based transformation method to increase the information available with SVM for detecting PD, and it yielded the classification accuracy of 93.47%. Polat [11] introduced fuzzy

* Corresponding author at: School of Computer and Communication Engineering, Liaoning Shihua University, Fushun 113001, Liaoning, China.
E-mail address: wangyang0531@163.com (Y. Wang).

http://dx.doi.org/10.1016/j.bspc.2017.06.015
1746-8094/

c-means clustering-based feature weighting method to increase the distinguishing performance between classes. It combined k-nearest neighbor (KNN) classifier to detect PD and achieved the classification accuracy of 97.93%. Chen et al. [12] combined fuzzy k-nearest neighbor (FKNN) with the principle component analysis to construct the most discriminative features for detecting PD, and it obtained the classification accuracy of 96.07%. Zuo et al. [13] constructed an automatic diagnostic system to detect PD. Mean classification accuracy of 97.47% was obtained by using an adaptive FKNN approach based on particle swarm optimization (PSO). Gök [14] developed a discriminative model which applied RF ensemble KNN classifier algorithm, and it achieved the classification accuracy of 98.46%. From these works, existing methods has integrated feature reduction method with the efficient classifiers to further improve the performance of PD diagnosis. However, they are more concerned with overall accuracy (ACC) and designed based on the assumption that the size of each class is relatively balanced. Therefore, these methods ignore the minority class and tend to be biased against the majority class in dealing with imbalanced data [15,16]. In other words, they may achieve higher misclassification accuracy of the minority class than that of the majority class.

There are two methods dealing with imbalanced data i.e. resampling technique and algorithmic technique [17]. Resampling technique includes oversampling which duplicates some minority class samples randomly or creates new samples in the neighborhood of minority class samples and undersampling which removes some majority class samples randomly to balance the size of each class [18,19]. In the algorithmic technique, cost-sensitive learning method is widely used to cope with imbalanced data [20]. It assigns a different misclassification cost for each sample. Generally, minority class samples are assigned high misclassification cost, while majority class samples are assigned low misclassification cost to improve the classification performance. In this study, cost-sensitive learning method is of particular interest.

Very recently, extreme learning machine (ELM) [21,22] has achieved the excellent performance on the disease diagnosis problems, for instance, hepatitis disease diagnosis [23] and PD diagnosis [24]. In this study, kernel-based weighted extreme learning machine (KWELM) is presented to perform PD diagnosis. Weighted strategy is used to alleviate the bias against performance caused by imbalanced data. An extra weight is designed for each sample to strengthen the relative impact of the minority class. In addition, the kernelized version of ELM [25] is also applied, and its advantage is that only kernel parameter γ and penalty parameter C need to be adjusted.

Previous studies [4,9–14] for PD detection have proven that performing feature selection prior to classification can improve the classification accuracy. Artificial bee colony (ABC) [26] is one of the newest global optimization techniques. In view of its simplicity and robustness, ABC has successfully been used to handle various real-world optimization problems [27–29]. However, ABC is good at exploration but poor at exploitation of solutions researching [30]. Consequently, new dynamic search strategies to generate candidate solutions are proposed in order to balance exploitation and exploration. In this study, both binary version and continuous version of an adaptive artificial bee colony (AABC) are used for performing feature selection and parameters optimization, respectively. Binary AABC is used to identify the most informative features as a feature selection method. Then the reduced feature subsets are used as the input of the trained KWELM classifier whose kernel parameter γ and penalty parameter C are specified by continuous AABC. Finally, experimental results illustrate that the proposed method named AABC-KWELM is effective and robust on PD diagnosis problem.

The rest of this paper is organized as follows. In Section 2, relevant preliminary knowledge is reviewed briefly and new strategies are proposed. Section 3 describes the detailed implementations of the proposed method. In Section 4, the experimental design is described and many experiments are completed to demonstrate that the proposed method presents better performance than that achieved by some existing methods. Finally, conclusions are summarized in Section 5.

2. Preliminaries

In this section, ELM and ABC algorithms are reviewed briefly and new strategies based on them are presented.

2.1. Kernel-based weighted extreme learning machine (KWELM)

ELM is a kind of single-hidden layer feed-forward neural networks (SLFNs). Given a training data set consisting of N arbitrary samples (x_j, t_j), where $t_j = [t_{j1}, t_{j2}, \cdots, t_{jm}]^T \in R^m$ and $x_j = [x_{j1}, x_{j2}, \cdots, x_{jn}]^T \in R^n$. The jth sample t_j is an $m \times 1$ target vector, and x_j is an $n \times 1$ feature vector. Given hidden nodes $L \ll N$ and activation function g(x), then the standard mathematical model of SLFNs is as follows:

$$\sum_{i=1}^{L} \beta_i g(a_i \cdot x_j + b_i) = t_j, j = 1, 2, \cdots, N \tag{1}$$

where $\beta_i = [\beta_{i1}, \beta_{i2}, \cdots, \beta_{im}]^T$ is the output weight vector connecting the ith hidden node and the output nodes, $a_i = [a_{i1}, a_{i2}, \cdots, a_{in}]^T$ is the input weight vector connecting input nodes and the ith hidden node, $a_i \cdot x_j$ is the inner product of a_i and x_j, and b_i is the bias of the ith hidden node.

SLFNs can approximate the training samples with zero error if the number of hidden nodes L is equal to the number of training samples N. Eq. (1) can compactly be reformulated as

$$H\beta = T \tag{2}$$

$$H = \begin{bmatrix} h(x_1) \\ \vdots \\ h(x_N) \end{bmatrix} = \begin{bmatrix} g(a_1 \cdot x_1 + b_1) & \cdots & g(a_L \cdot x_1 + b_L) \\ \vdots & \cdots & \vdots \\ g(a_1 \cdot x_N + b_1) & \cdots & g(a_L \cdot x_N + b_L) \end{bmatrix}_{N \times L} \beta$$
$$= \begin{bmatrix} \beta_1^T \\ \vdots \\ \beta_L^T \end{bmatrix}_{L \times m}, \text{ and } T = \begin{bmatrix} t_1^T \\ \vdots \\ t_N^T \end{bmatrix}_{N \times m} \tag{3}$$

where H is the hidden layer output matrix, and the jth column of H represents the jth hidden node output vector on all the inputs. T is the output matrix, and β is the output weight matrix.

However, in most cases, it is $L \ll N$ and there may not exist a β that satisfies Eq. (2). The hidden layer biases and input weights need not be tuned at all and can be randomly generated, so the output weights can be determined by finding the Least Square solution $\hat{\beta} = H^+T$ of $H\beta = T$, where H^+ is the Moore- Penrose generalized inverse of matrix H. In short, ELM algorithm is summarized as follows.

1) Generate randomly input weights a_i and biases b_i, $i = 1, 2, \ldots, L$.
2) Calculate the hidden layer output matrix H.
3) Calculate the output weight $\beta = H^+T$.

According to Bartlett's theory [31], not only is ELM to minimize the training error but also the norm of the output weights. Meanwhile, an extra weight is designed for each sample to bet-