Expert Systems with Applications 40 (2013) 1519–1529

Contents lists available at SciVerse ScienceDirect

Expert Systems with Applications

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

Parkinson's disease prediction using gene expression – A projection based learning meta-cognitive neural classifier approach

G. Sateesh Babu, S. Suresh *

School of Computer Engineering, Nanyang Technological University, Singapore

ARTICLE INFO

Keywords:
Meta-cognitive learning
Radial basis function network
Parkinson disease
Microarray gene expression
Dysphonia
Gait analysis

ABSTRACT

In this paper, we propose a gene expression based approach for the prediction of Parkinson's disease (PD) using 'projection based learning for meta-cognitive radial basis function network (PBL-McRBFN)'. McRBFN is inspired by human meta-cognitive learning principles. McRBFN has two components, a cognitive component and a meta-cognitive component. The cognitive component is a radial basis function network with evolving architecture. In the cognitive component, the PBL algorithm computes the optimal output weights with least computational effort. The meta-cognitive component controls the learning process in the cognitive component by choosing the best learning strategy for the current sample and adapts the learning strategies by implementing self-regulation. The interaction of cognitive component and meta-cognitive component address the *what-to-learn*, *when-to-learn* and *how-to-learn* of human learning principles efficiently.

PBL-McRBFN classifier is used to predict PD using micro-array gene expression data obtained from ParkDB database. The performance of PBL-McRBFN classifier has been evaluated using Independent Component Analysis (ICA) reduced features sets from the complete genes and selected genes with two different significance levels. Further, the performance of PBL-McRBFN classifier is statistically compared with existing classifiers using one-way repeated ANOVA test. Further, it is also used in PD prediction using the standard vocal and gait PD data sets. In all these data sets, the performance of PBL-McRBFN is compared against existing results in the literature. Performance results clearly highlight the superior performance of our proposed approach.

1. Introduction

Parkinson's disease (PD) is characterized by progressive degeneration of dopaminergic neurons in the pars compacta of the substantia nigra (Manciocco et al., 2009). According to the global declaration for PD, 6.3 million people are affected by this disease worldwide, and affects all races and cultures. Most important symptoms of PD include muscle rigidity, tremors, and changes in speech and gait (Manciocco et al., 2009; Cho, Chao, Lin, & Chen, 2009). The exact causes of PD are not known, but research studies show that both environmental triggers and genetic influences play an important role. At present there is no cure for PD, the diagnosis of PD is based on medical history and neurological examination conducted by interviewing and observing the patient in person using the disease rating scales. Unified Parkinson's disease Rating Scale (UPDRS), Hoehn and Yahr scale, Schwab and England Activities of Daily Living (ADL) scale, PDQ-39, PD Non-motor symptoms (NMS) questionnaire, NMS survey are most commonly used PD rating scales. The reliable diagnosis of PD using these scales is difficult, especially in its early stages (Lee & Lim, 2012). As the symptoms of PD are comorbid with other neurological disorders, only 75% of clinical diagnoses of PD are confirmed to be idiopathic PD at autopsy. Thus, automatic approaches based on machine learning are needed to increase the diagnosis accuracy and to help physicians to make better decisions. Most machine learning approaches for PD classification were undertaken using vocal and gait features (Little, McSharry, Hunter, Spielman, & Ramig, 2009; Das, 2010; Caglar, Cetisli, & Toprak, 2010; Sakar & Kursun, 2010; Strom & Koker, 2011; Polat, 2012; Jeon, Han, Yi, Jeon, & Park, 2008; Tahir & Manap, 2012; Lee & Lim, 2012; Eskidere, Ertas, & Hanilci, 2012; Chen et al., 2012; Rodriguez-Solano & Laita, 2010; Banaie, Pooyan, & Mikaili, 2011; Wu et al., 2010; Engin et al., 2007; Pan, Iplikci, Warwick, & Aziz, 2012).

Machine learning approaches for PD classification by detecting dysphonia using acoustic measurements are experimented with a data set created by Little, McSharry, Roberts, Costello, and Moroz (2007). Further, Little et al. (2007) collected a data set consisting of sustained vowel phonations from 31 people of whom 23 with PD, and used kernel-support vector machine for PD classification. Here, exhaustive search process is implemented to select the best kernel width and penalty value. Sakar and Kursun (2010) used Sup-

* Corresponding author. Tel.: +65 6790 6185.
E-mail address: ssundaram@ntu.edu.sg (S. Suresh).

0957-4174/$ - see front matter © 2012 Elsevier Ltd. All rights reserved.
http://dx.doi.org/10.1016/j.eswa.2012.08.070

port Vector Machine (SVM) classifier with selected features from maximum-relevance-minimum-redundancy (mRMR) criterion. For mRMR, all the available samples are used in mutual information computations. Also, the above two approaches uses all the available data samples to optimize the SVM parameters which is inevitable when working with such a small data set. Das (2010) compared four independent classification approaches (neural networks, data mining neural, logistic regression and decision trees) for diagnosis of PD. Among the four approaches, neural networks classifier (Multi-layer feed-forward neural network with Levenberg–Marquardt algorithm) yields the best performance. The drawback in the above neural network approach is random initialization of weights and heuristic determination of number of hidden neurons, which affects the classification performance significantly. Also, it requires retraining when the training samples changes with time. Strom and Koker (2011) used parallel neural networks approach for prediction of PD. The training time and complexity of the parallel network approach do increase as the number of parallel networks increases. Caglar et al. (2010) used adaptive neuro-fuzzy classifier with linguistic hedges for feature selection and classification. Linguistic hedges feature selection will require the optimization search from huge set of theoretically possible encoding combinations for each feature and hence computationally intensive. Polat (2012) has presented a fuzzy c-means clustering-based feature weighting with *k*-NN classification approach. The choice of number of clusters or selection of k-neighbors affect the performance significantly. In addition, increasing the number of training samples may complicate the choice of number of clusters further. In (Sateesh Babu, Suresh, Uma Sangumathi, & Kim, 2012b) paper, we have presented a meta-cognitive radial basis function network approach and shown a better performance.

Machine learning approaches for PD classification based on gait analysis are also reported in literature. Jeon et al. (2008) created an image data with plantar pressure measurements of right foot during heel to motion from 17 control and 21 PD patients, and SVM is applied to distinguish gait pattern. Here, other important basic, kinetic and kinematic features are not used in gait analysis. Tahir and Manap (2012) used a data set from 20 control and 12 PD patients consists of basic spatiotemporal, kinematic and kinetic gait features, and the ability of ANN and SVM classifiers is discussed. The above two studies on gait features use their own proprietary data, with fewer number of subjects. Lee and Lim (2012) used a database provided by PhysioBank (Hausdorff et al., 2007) consisting of data from sensors located under the feet of 73 controls and 93 PD patients. In their approach, wavelet transform has been employed to extract the relevant features and neural networks with weighted fuzzy membership has been used to approximate the functional relationship between the extracted features and class label.

Recent studies on gene expression analysis found that there is a profound change in gene expression for individuals affected by PD (Scherzer et al., 2007). These studies discovered that, diagnosis of early stage PD using vocal and gait features is impossible because tremor and slow movements develop in PD patients only after approximately 70% of vulnerable dopaminergic neurons in the substantia nigra have already died (Scherzer et al., 2007). However, machine learning approaches for PD classification based on gene analysis have been limited. Therefore, there is a need for devising a new machine learning approach for PD classification based on gene analysis.

On the other hand, classification algorithms used in literature for prediction of PD use all the samples in the training data set to address *how to learn* the functional relationship between the input features and their target class labels. In other words, they possess information-processing abilities of humans, including perception, learning, remembering, judging, and problem-solving, and these abilities are cognitive in nature. However, recent studies in human learning (Wenden, 1998; Rivers, 2001; Isaacson & Fujita, 2006) have revealed that the learning process is effective when the learners adopt self-regulation in learning process using meta-cognition. Meta-cognition means *cognition about cognition*. In a meta-cognitive framework, human-beings think about their cognitive processes, develop new strategies to improve their cognitive skills and evaluate the information contained in their memory. Hence, there is a need to develop a meta-cognitive machine learning network that analyzes its cognitive processes and chooses suitable strategies to improve its cognitive skills. Such a machine learning network must be capable of deciding *what-to-learn*, *when-to-learn* and *how-to-learn* the decision function from the training data by emulating the human self-regulated learning principles.

In machine learning literature, Self-adaptive Resource Allocation Network (SRAN) (Suresh, Dong, & Kim, 2010) and Complex-valued Self-regulating Resource Allocation Network (CSRAN) (Suresh, Savitha, & Sundararajan, 2011) address the *what-to-learn* component of meta-cognition by selecting significant samples using misclassification error and hinge loss function. It has been shown in SRAN and CSRAN, that the selecting appropriate samples for learning and removing repetitive samples helps in improving the generalization performance. The subsequently developed, Meta-cognitive Neural Network (McNN) (Sateesh Babu & Suresh, 2012) and Meta-cognitive Neuro-Fuzzy Inference System (McFIS) (Subramanian & Suresh, 2012) address the three components of meta-cognition. However, McNN and McFIS update the network parameters using extended Kalman filter algorithm which increases computational burden for large data sets. Similar meta-cognitive learning studies in complex domain are reported in Savitha, Suresh, and Sundararajan (2012a) and Savitha, Suresh, and Sundararajan (2012b). In (Sateesh Babu, Savitha, & Suresh, 2012a) paper, we introduced a meta-cognitive radial basis function network (McRBFN) that addresses the three components of meta-cognition with least computational effort simultaneously.

Unlike the existing batch learning algorithms that require the number of hidden neurons to be fixed a priori, the Projection Based Learning (PBL) of McRBFN begins with zero hidden neurons and adds neurons during the learning process to obtain an optimum network structure. When a neuron is added to the cognitive component, the input/hidden layer parameters are fixed based on the input of the sample and the output weights are estimated by minimizing an error function given by the hinge-loss error (Suresh, Sundararajan, & Saratchandran, 2008a). The McRBFN using the PBL to obtain the network parameters is referred to as, 'projection based learning algorithm for a meta-cognitive radial basis function network (PBL-McRBFN)'.

In this paper, we use PBL-McRBFN to predict PD using gene expression features. For this purpose, we obtain the microarray gene expression data set from ParkDB database (Taccioli et al., 2011). The obtained complete microarray gene expression data set consists of 22283 genes expression information from 50 PD patients at early stage and 22 controls. Since the complete gene expression data set consists of large number of redundant genes, we selected the most informative genes with *p*-value selection at values less than 0.05/0.01. After *p*-value selection, two data sets are formed, one data set is with 1594 genes at significance level less than 0.05 and another data set is with 412 genes at significance level less than 0.01. However, as the feature space of data sets with the complete and selected genes is high compared to the number of samples, it will be difficult to predict PD accurately. Hence, we employ Independent Component Analysis (ICA) (Hyvarinen, 1999) to reduce the feature dimension. The complete and selected gene expression features are reduced to 10, 25, 50 features using ICA. The performance of PBL-McRBFN classifier is evaluated by conducting 10 random trials of experiments on ICA reduced fea-