Journal of King Saud University – Computer and Information Sciences (2014) **26**, 1–6

King Saud University

Journal of King Saud University – Computer and Information Sciences

www.ksu.edu.sa
www.sciencedirect.com

ORIGINAL ARTICLE

Classification of EEG Signals using adaptive weighted distance nearest neighbor algorithm

E. Parvinnia [a,*], M. Sabeti [a], M. Zolghadri Jahromi [b], R. Boostani [b]

[a] *Department of Computer Engineering, Shiraz Branch, Islamic Azad University, Shiraz, Iran*
[b] *Department of Computer Science and Engineering, Shiraz University, Shiraz, Iran*

Received 3 March 2012; revised 31 December 2012; accepted 2 January 2013
Available online 8 January 2013

KEYWORDS

Nearest neighbor;
Noisy training data;
EEG signal classification;
Band power;
Fractal dimension;
Autoregressive coefficients

Abstract Electroencephalogram (EEG) signals are often used to diagnose diseases such as seizure, alzheimer, and schizophrenia. One main problem with the recorded EEG samples is that they are not equally reliable due to the artifacts at the time of recording. EEG signal classification algorithms should have a mechanism to handle this issue. It seems that using adaptive classifiers can be useful for the biological signals such as EEG. In this paper, a general adaptive method named weighted distance nearest neighbor (WDNN) is applied for EEG signal classification to tackle this problem. This classification algorithm assigns a weight to each training sample to control its influence in classifying test samples. The weights of training samples are used to find the nearest neighbor of an input query pattern. To assess the performance of this scheme, EEG signals of thirteen schizophrenic patients and eighteen normal subjects are analyzed for the classification of these two groups. Several features including, fractal dimension, band power and autoregressive (AR) model are extracted from EEG signals. The classification results are evaluated using Leave one (subject) out cross validation for reliable estimation. The results indicate that combination of WDNN and selected features can significantly outperform the basic nearest-neighbor and the other methods proposed in the past for the classification of these two groups. Therefore, this method can be a complementary tool for specialists to distinguish schizophrenia disorder.

1. Introduction

Electroencephalogram (EEG) signals (Sanei and Chambers, 2007) are brain activities recorded using electrodes placed on the scalp. Although several methods for the brain function analysis such as megnetoencephalography (MEG), functional magnetic resonance imaging (fMRI) and positron emission tomography (PET) have been introduced, the EEG signal is still a valuable tool for monitoring the brain activity due to its relatively low cost and being convenient for the patient.

There have been several EEG classification studies within the recent years. These studies used different classification techniques, compared their performance, and evaluated different combinations of feature sets. Among these classifiers, k-nearest neighbor (k-NN), linear discriminant analysis (LDA), support vector machine (SVM), artificial neural network

* Corresponding author. Tel/fax: +98 711 6191656.
E-mail addresses: parvinnia@iaushiraz.ac.ir, parvinn@shirazu.ac.ir, eparvinnia@gmail.com (E. Parvinnia).

Peer review under responsibility of King Saud University.

ELSEVIER Production and hosting by Elsevier

1319-1578 © 2013 Production and hosting by Elsevier B.V. on behalf of King Saud University.
http://dx.doi.org/10.1016/j.jksuci.2013.01.001

(ANN) have been popular. Boostani et al. (2008) used five different classification algorithms including LDA, Boosted version of direct LDA (BDLDA), Adaboost, SVM, and fuzzy SVM to classify two schizophrenic and normal groups. Their result showed the BDLDA method achieved slightly better performance than the other classification methods. Hazarika et al. (1997) applied the three-layered ANN using wavelet transform as a feature extraction method for classifying of three groups: normal, schizophrenia, and obsessive compulsive disorder. Their results showed the wavelet transform can be used as a powerful technique for preprocessing EEG signals prior to classification. Li and Fan, 2005 studied the classification of three kinds of subjects (10 schizophrenic patients, 10 depressive patients and 10 normal controls) with EEG rhythms used as feature vectors. They used two ANN approaches, BP ANN and self-organizing competitive ANN for classification. Their results showed that BP ANN has a better comprehensive performance than the self-organizing competitive ANN technique.

Hornero et al. (2006) used three nonlinear methods of time series analysis for analyzing the time series generated by 20 schizophrenic patients and 20 control subjects. Their results show that the ability of generating random time series between schizophrenic subjects and controls is different. The patient group is characterized by less complex neurobehavioral and neuropsychologic measurements. Rosenberg et al. (1990) studied a random number generation experiment. They asked the participant to choose a random number in interval [1..10] without any generative rule. They found that schizophrenic patients tended to be more repetitive. AlZoubi et al. (2009) evaluated three different classifier techniques to classify the EEG signals in a 10-class emotion experiment. Their results showed using the adaptive algorithm can improve the performance of the classification task.

We believe that the main problem in the classification of EEG signals is the quality of the recorded signal, which can be different during the experiment. These unwanted disturbances cannot be controlled since many activities are going on at the same time in the brain. Existence of artifacts at the time of recording the EEG signal, directly affects the reliability of the recorded signal. It seems that using adaptive classifiers can be useful for the biological signals such as EEG. In this paper, a general adaptive method named weighted adaptive nearest neighbor (WDNN) (Zolghadri et al., 2009) is applied for EEG signal classification. This classifier assigns a weight to each training sample that controls its influence in classifying test samples. When a large weight is assigned to a training sample, it will increase its influence in classifying many samples. On the other hand, reducing the weight of a training sample will decrease its influence in the classification task. The most important ability of this classifier is determining the quality of each EEG segment by assigning different weights for the classification task. Therefore if the training samples are changed, the weights of these samples will be recalculated.

To assess the performance of the WDNN classifier, EEG signals of thirteen schizophrenic patients and eighteen normal subjects are analyzed for the classification of the two groups. The EEG signals are recorded in the Center for Clinical Research in Neuropsychiatry, Perth, Western Australia.

This paper is structured as follows. Section 2 presents nearest neighbor (NN) classification with weighted training samples. In Section 3, feature extraction techniques are illustrated. Experimental results are discussed in Section 4 and Section 5 presents our conclusion.

2. Weighted adaptive nearest-neighbor classification

This method, by assigning a weight to each training sample, attempts to improve the performance of the 1-NN. WDNN tries to minimize the leave one out (LV1) classification error on the given training set by assigning the weights of training samples. These weights are used in the test phase for finding the nearest neighbor of a query sample. By assigning small weights to low quality training samples, their influence in feature space can be reduced.

Assume there is a problem with a set of training samples like (A_i, C_i) where $i = 1, \ldots, n$, A_i has f features, and C_i has M-classes. Different types of distance functions have been introduced by Wilson and Martinez (2000) for measuring the distance between two patterns for identifying the NN of a query pattern. Euclidean distance has been suggested, in most situations, for the distance between two samples A_i and A_j:

$$distance(A_i, A_j) = \sqrt{\sum_{k=1}^{f} (A_{ik} - A_{jk})^2} \tag{1}$$

The similarity measure can be used instead of using the distance function as follows:

$$\lambda(A_i, A_j) = \frac{1}{distance(A_i, A_j)} \tag{2}$$

The sample A_r that has the most similarity to a query sample Q can be mentioned as follows by using (2):

$$r = \underset{1 \leqslant i \leqslant n}{\arg\max}\{\lambda(Q, A_i)\} \tag{3}$$

The assumption of NN classifier is all of the training samples have the same weight. The WDNN believes that the quality of the stored samples is not equal. This is especially true when each sample represents an EEG sample recording. To take this into account, a weight w_k is allocated to each training sample A_k. In the test phase, these weights are used for finding the sample A_p that has the most similarity to a query sample Q.

$$p = \underset{1 \leqslant i \leqslant n}{\arg\max}\{w_i.\lambda(Q, A_i)\} \tag{4}$$

2.1. Learning algorithm for weighting training samples

The WDNN is a greedy method that tries to minimize the LV1 error rate of classification on the given training set by specifying the weights of training samples. Note that, a training sample with a large weight can increase its influence in classifying many samples in LV1 test. On the other hand, a training sample having zero weight is not used to classify any test samples and can be removed from the data set.

The main part of the WDNN learning method is a procedure that specifies the best weight for a training sample with respect to all other samples having fixed weights.

WDNN starts with an initial set of weights equal to one ($w_j = 1.0$). The weight of each training sample is adjusted in turn. Assuming a training sample A_k belongs to a sample class that is denoted by *ClassT*, the algorithm tries to specify the best weight w_k, that is a real number in the interval $[0, \infty]$, as follows: