Expert Systems with Applications 39 (2012) 9072–9078

Contents lists available at SciVerse ScienceDirect

Expert Systems with Applications

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

Use of principal component analysis for automatic classification of epileptic EEG activities in wavelet framework

U. Rajendra Acharya [a], S. Vinitha Sree [b,*], Ang Peng Chuan Alvin [a], Jasjit S. Suri [c,d]

[a] Department of Electronics and Communication Engineering, Ngee Ann Polytechnic, Singapore 599489, Singapore
[b] Global Biomedical Technologies, CA, USA
[c] CTO, Global Biomedical Technologies, CA, USA
[d] Biomedical Engineering Department, Idaho State University (Aff), ID, USA

ARTICLE INFO

Keywords:
Epilepsy
Principal component analysis
Eigenvalues
Classification
Non-linear analysis
Wavelet Packet Decomposition
Ictal
Interictal
Electroencephalogram

ABSTRACT

Electroencephalogram (EEG) signals are used to detect and study the characteristics of epileptic activities. Owing to the non-linear and dynamic nature of EEG signals, visual inspection and interpretation of these signals are tedious, time-consuming, error-prone, and subjected to inter-observer variabilities. Therefore, several Computer Aided Diagnostic (CAD) based studies have adopted non-linear techniques to study the normal, interictal, and ictal activities in EEGs. In this paper, we present a novel automatic technique based on data mining for epileptic activity classification. In order to compare our study with the results of relative studies in the literature, we used the widely used benchmark dataset from Bonn University for evaluation of our proposed technique. Hundred samples each in normal, interictal, and ictal categories were used. We decomposed these segments into wavelet coefficients using Wavelet Packet Decomposition (WPD), and extracted eigenvalues from the resultant wavelet coefficients using Principal Component Analysis (PCA). Significant eigenvalues, selected using the ANOVA test, were used to train and test several supervised classifiers using the 10-fold stratified cross validation technique. We obtained 99% classification accuracy using the Gaussian Mixture Model (GMM) classifier. The proposed technique is capable of classifying EEG segments with clinically acceptable accuracy using less number of features that can be extracted with less computational cost. The technique can be written as a software application that can be easily deployed at a low cost and used with almost no expert training. We foresee that this software can, in the future, evolve into an efficient adjunct tool that cannot only classify epileptic activities in EEG signals but also automatically monitor the onset of seizures and thereby aid the doctors in providing better and timely care for the patients suffering from epilepsy.

1. Introduction

Epilepsy is a common neurological disorder that is characterized by occurrence of recurrent seizures. World Health Organization statistics (WHO, 2011) indicate that every year between 40 and 70 per 100,000 people are diagnosed with epilepsy in developed countries, and this figure is almost twice in the case of developing countries. Generally, Electroencephalogram (EEG) signals are used to detect seizures (Thakor & Tong, 2004). Typically, the physicians analyze the EEG segments for three types of activities: ictal, which is usually characterized by continuous rhythmical activity that has a sudden onset when the patient is exhibiting a seizure; interictal, which is characterized by small spikes and subclinical seizures that generally occur during the time between seizures in epileptic patients; and normal EEG segments. Characterization of EEG segments into these three classes will help the physicians in studying the underlying cause of these changes, in monitoring seizures, and also in administering appropriate seizure management protocols in order to improve the quality of life of epileptic patients (Osorio & Frei, 2009; Shoeb, Guttag, Pang, & Schachter, 2009). EEG signal recordings are generally long, and hence, the resulting signal to be analyzed is voluminous. Visual inspection, therefore, becomes tedious, time-consuming, error-prone, and subjected to inter-observer variabilities. To address these limitations, Computer Aided Diagnostic (CAD) tools have been developed for several medical diagnostic applications.

Many automated CAD techniques extract linear time-domain and frequency-domain based features from the EEG signal to detect epileptiform discharges. EEG signals are by nature non-linear (Kannathal et al., 2005a; Lehnertz, 2008; Pijn et al., 1997; Subha, Joseph, Acharya, & Lim, 2010) and seizures are characterized by non-linear transitions of an epileptic brain from its less ordered interictal state to a more ordered ictal state. Therefore, many

* Corresponding author.
E-mail address: vinitha.sree@gmail.com (S. Vinitha Sree).

0957-4174/$ - see front matter © 2012 Elsevier Ltd. All rights reserved.
doi:10.1016/j.eswa.2012.02.040

recent studies (Acharya, Chua, Lim, Dorithy, & Suri, 2009; Acharya, Faust, Kannathal, Chua, & Laxminarayan, 2005; Acharya, Vinitha Sree, Chattopadhyay, Ng, & Suri, 2011b; Acharya, Vinitha Sree, Chattopadhyay, Yu, & Alvin, 2011a; Adeli, Ghosh-Dastidar, & Dadmehr, 2005a, 2005b, 2007, 2008; Ahmadlou & Adeli, 2010, 2011; Ahmadlou, Adeli, & Adeli, 2010a, 2010b, 2011; Bai, Qiu, & Li, 2007; Chua, Chandran, Acharya, & Lim, 2009a; Chua, Chandran, Acharya, & Lim, 2009b, 2010; Faust, Acharya, Lim, & Sputh, 2010; Ghosh-Dastidar & Adeli, 2007, 2009; Ghosh-Dastidar, Adeli,& Dadmehr, 2007, 2008; Good, Sabesan, Marsh, Tsakalis, & Iasemidis, 2009; Guler, Ubey, & Guler, 2005; Kannathal, Lim, Acharya, & Sadasivan, 2005b; Ocak, 2009; Sadati, Mohseni, & Magshoudi, 2006; Sankari & Adeli, 2011a; Sankari, Adeli, & Adeli, 2011b; Srinivasan, Eswaran, & Sriraam, 2007; Subasi, 2007; Subasi & Gursoy, 2010) focussed on extracting non-linear features and using them in classifiers. Details on the techniques and results of these studies are presented and compared with our technique in the discussion section of this paper. The literature survey indicated that there is still a need for classification systems that present high accuracies in classifying all three classes.

The key objective of this work is to develop a CAD technique that extracts a small set of highly significant features from the EEG segments and uses them in simple classifiers to accurately classify them into normal, interictal or ictal classes. In our work, we perform Wavelet Packet Decomposition (WPD) of the EEG segments to obtain the wavelet coefficients at different levels. We determine the eigenvalues of these coefficients using Principal Component Analysis (PCA). We observed that these eigenvalues had significantly different ranges for all the three classes, thereby deeming them fit candidates for training classifiers. Even though there are a few studies that extract features from the wavelet transform applied signals, these studies either focused on classifying only the normal and ictal classes (Ocak, 2009; Sadati et al., 2006; Subasi, 2007) or they used time frames or sub-bands of data (Adeli et al., 2007; Ghosh-Dastidar et al., 2007, 2008; Ghosh-Dastidar & Adeli, 2007, 2009) which demanded additional pre-processing work and time. In this work, we use the entire EEG segments (each of length 23.6 s) and use simple classifiers and easily extractable features to classify all the three classes of EEG segments, namely normal, interictal, and ictal. The developed technique can be used as an automated, simple, objective, fast, and cost-effective efficient secondary diagnostic tool that provides additional confidence to the clinician's initial diagnosis of the class of the EEG segment.

The proposed technique is illustrated by a block diagram in Fig. 1. Part of the dataset is used for building the classifier and the rest is used for evaluating the classifier. The offline training system comprises of the blocks outside the dotted rectangular box that indicate the steps followed in developing the classifier for EEG segment classification. First, we apply WPD to the EEG segments, and then we extract eigenvalues from the resultant WPD coefficients. Significant eigenvalues are selected using the Analysis of Variance (ANOVA) test. These eigenvalues and the Ground Truth (GT) of whether the segment is normal, interictal, or ictal as determined by the physicians are used to train several supervised learning based classifiers to determine the optimum classifier parameters for real-time use. In the real-time online system (blocks indicated inside the dotted rectangular box), the eigenvalues that were deemed significant by the offline system are extracted from the test segment whose class label is to be determined. The classifier parameters are applied on these eigenvalues to determine the class label. The class labels of the test dataset that were predicted by the real-time system are used to determine performance measures such as accuracy, sensitivity, and specificity.

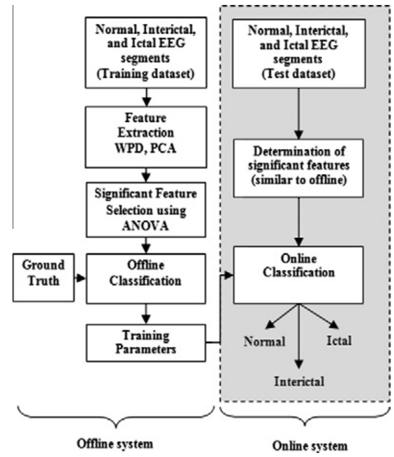

Fig. 1. Block diagram of the proposed system for EEG classification; the blocks outside the dotted shaded rectangular box represent the flow of offline training system, and the blocks within the dotted box represent the online real-time system.

The paper is organized as follows: Section 2 contains a description of the data used in this work. Section 3 describes the WPD and PCA algorithms. It also presents brief descriptions of the classifiers used in this work and the ANOVA test. The resultant significant features and the classification results are presented in Section 4. Section 5 presents a discussion on related studies and compares our results with other published work. The paper is concluded in Section 6.

2. Data

The EEG dataset used in this work was taken from the artifact free EEG time series data available at the University of Bonn (Andrzejak et al., 2001). EEG segments, each of length 23.6 s, were taken from five healthy subjects and five epileptic patients. One hundred segments of data in each of the three categories: normal, interictal, and ictal/epileptic were selected. The normal segments were taken from the five healthy subjects. The standard surface electrode placement scheme (the international 10–20 system) was used to obtain the EEG from the healthy cases. Both the interictal and ictal segments were obtained from epilepsy patients. The interictal segments were recorded during seizure free intervals from the depth electrodes that were implanted into the hippocampal formations. The ictal segments were recorded from all sites exhibiting ictal activity using depth electrodes and also from strip electrodes that were implanted into the lateral and basal regions of the neocortex (Andrzejak et al., 2001). All the segments were recorded using a 128-channel amplifier system, digitized with a sampling rate of 173.61 Hz and 12-bit A/D resolution, and filtered using a 0.53–40 Hz (12 dB/octave) band pass filter. Typical normal, interictal, and epileptic (ictal) EEG signals are illustrated in Fig. 2.

3. Methodology: feature extraction and classification

This section contains brief descriptions of the WPD and PCA techniques. It also provides a description of the ANOVA test used to select the significant features. It presents brief details about the six classifiers that were evaluated.

3.1. Feature extraction

3.1.1. Wavelet Packet Decomposition (WPD)

A wavelet transform uses wavelets, which are scaled and translated copies of a basic wavelet shape called the 'mother wavelet', to