Available online at www.sciencedirect.com

ScienceDirect

Expert Systems with Applications 36 (2009) 1329–1336

Expert Systems with Applications

www.elsevier.com/locate/eswa

Cross-correlation aided support vector machine classifier for classification of EEG signals

Suryannarayana Chandaka, Amitava Chatterjee *, Sugata Munshi

Department of Electrical Engineering, Jadavpur University, Kolkata 700 032, India

Abstract

Over the last few decades pattern classification has been one of the most challenging area of research. In the present-age pattern classification problems, the support vector machines (SVMs) have been extensively adopted as machine learning tools. SVM achieves higher generalization performance, as it utilizes an induction principle called structural risk minimization (SRM) principle. The SRM principle seeks to minimize the upper bound of the generalization error consisting of the sum of the training error and a confidence interval. SVMs are basically designed for binary classification problems and employs supervised learning to find the optimal separating hyperplane between the two classes of data. The main objective of this paper is to introduce a most promising pattern recognition technique called cross-correlation aided SVM based classifier. The idea of using cross-correlation for feature extraction is relatively new in the domain of pattern recognition. In this paper, the proposed technique has been utilized for binary classification of EEG signals. The binary classifiers employ suitable features extracted from crosscorrelograms of EEG signals. These cross-correlation aided SVM classifiers have been employed for some benchmark EEG signals and the proposed method could achieve classification accuracy as high as 95.96% compared to a recently proposed method where the reported accuracy was 94.5%.

Keywords: Support vector machines; Cross-correlation; Electroencephalogram signals; Optimal separating hyperplane

1. Introduction

Epileptic seizures are due to temporary electrical disturbances of the brain. Unfortunately, the occurrence of epileptic seizure is unpredictable (Subasi, 2007). The neuroscientists and biological psychiatrists prefer to study the electrical activity of the brain with the help of electroencephalographic records for diagnosis of neurological disorders (Gular, Ubeyli, & Güler, 2005). Basically, the electroencephalogram (EEG) is a complex and aperiodic time series containing information of the electrical activity generated by cerebral cortex nerve cells. Earlier the analysis of the electroencephalograph (EEG) records was restricted to visual inspection. It has been shown that such visual inspection is insufficient to closely study the minute observational variations in EEG signals, thus necessitating computer based automation tools (Subasia, Alkana, Koklukayab, & Kiymika, 2005).

Spectral analysis can be recommended for EEG signals, which can give information regarding brain activities. Nowadays artificial neural networks (ANNs) may offer a superior performance for analysis of EEG signals, compared to the spectral analysis methods (Guler, Ubeyli, & Guler, 2005). As compared to the conventional spectral analysis methods, ANNs can effectively make a decision regarding the class of the signal. Neural networks have been successfully adopted for so many medical applications (Baxt, 1990). The invention of neural network models produces a great revolution in the area of pattern recognition. Support vector machines constitute one of the relatively contemporary branches of neural network models. Nowadays in the machine learning technologies for pattern classification problems, the support vector machines have been extensively adopted for machine learning tools. In contrast

* Corresponding author.
E-mail address: cha_ami@yahoo.co.in (A. Chatterjee).

0957-4174/$ - see front matter © 2007 Elsevier Ltd. All rights reserved.
doi:10.1016/j.eswa.2007.11.017

to traditional neural networks SVM achieves higher generalization performance, due to the utilization of structural risk minimization principle. The SRM principle seeks to minimize the upper bound of the generalization error consisting of the sum of the training error and a confidence interval, where as incase of traditional neural networks the empirical risk minimization principle only minimizes the training error (Cao & Tay, 2003).

Several works on the classification of EEG signals have been reported. In Guler et al. (2005), the researchers evaluated the classification accuracy of the recurrent neural networks (RNNs) using Lyapunov exponents trained with Levenberg–Marquardt algorithm on the electroencephalogram (EEG) signals. The database used for their work is EEG database (EEG time series), which is developed by department of Epileptology, University of Bonn, California. In this work three classes of EEG signals have been considered. They are from healthy volunteers with eyes open, epilepsy patients in the epileptogenic zone during a seizure-free interval, and epilepsy patients during epileptic seizures. An overall classification accuracy of 96.79% was obtained in this work. In Subasi (2007), mixture of experts model (ME) was utilized for classification of EEG signals. A double-loop expectation-maximization (EM) algorithm was introduced to the ME network structure for detection of epileptic seizure and wavelet feature extraction was employed. The database used in this work was with University of Bonn EEG time series, California. Healthy volunteers with eyes open and epilepsy patients during epileptic seizures were used for classification and overall accuracy was 94.5%. Up to now no study has been reported in literature related to cross-correlation based feature extraction in the analysis of EEG signals.

Correlation is a mathematical operation that is very similar to convolution. Cross-correlation of two signals measures the extent of similarity between these signals. In the present work, a healthy subject is considered as reference and the EEG signal of the reference subject is correlated with the EEG signal of each of the other subjects. From each of resulting cross-correlation sequences a set of five features is extracted. These feature vectors called patterns are segregated into training and testing dataset and the training set is utilized for training SVM classifier. The LS-SVM package, available in svm toolbox has been used for this purpose. The generalization capability of the classifier is then tested by utilizing it for testing dataset. The structure of the paper is organized as follows. Section 2 briefly reviews the sets of the EEG signals used in our work. Section 3 reviews the cross-correlation of discrete-time signals and the computation of various features of crosscorrelograms utilized in this work. Section 4 is devoted to provide necessary background and basic idea behind support vector machine. Section 5 presents how the proposed method is applied for classification of EEG signals and how the cross-correlation is performed and features are extracted for each EEG signals is explained in detail. Finally how the SVM is utilized for classification is discussed. Chapter 6 highlights the conclusions of the work and discuses directions for future investigations.

2. EEG signals

In this work, a publicly available database (EEG time series; Andrzejak et al., 2001) has been used. All EEG signals recorded were with the same 128-channel amplifier system, using an average common reference. The data were digitized at 173.61 samples per second using 12-bit resolution. Band-pass filter settings were 0.53–40 Hz (12 dB/oct). The EEG segments in this database were cut out from continuous multi-channel EEG recordings after visual inspection for artifacts, e.g., due to muscle activity or eye movements. Fig. 1 depicts the electrode placement for recording of EEG signal. The complete dataset consists of five sets (denoted A–E), each containing 100 single-channel EEG signals of 23.6 s. Fig. 2 depicts example of EEG signals of each of the five sets. Sets A and B have been taken from surface EEG recordings of five healthy volunteers with eyes open and closed, respectively. Signals in two sets have been measured in seizure-free intervals from five patients in the epileptogenic zone (D) and from the hippocampal formation of the opposite hemisphere of the brain (C). Set E contains seizure activity, selected from all recording sites exhibiting ictal activity. Sets A and B have been recorded extra cranially, whereas sets C, D, and E have been recorded intracranially. In the present study we classified only two (A and E) of the complete dataset.

3. Cross-correlation of signals

Correlation is a mathematical operation that is very similar to convolution. In correlation, a cross-correlation sequence between two energy signals measures the extent of similarity between these two signals (Proakis & Manolakis, 1997). If a signal is correlated with itself, the resulting

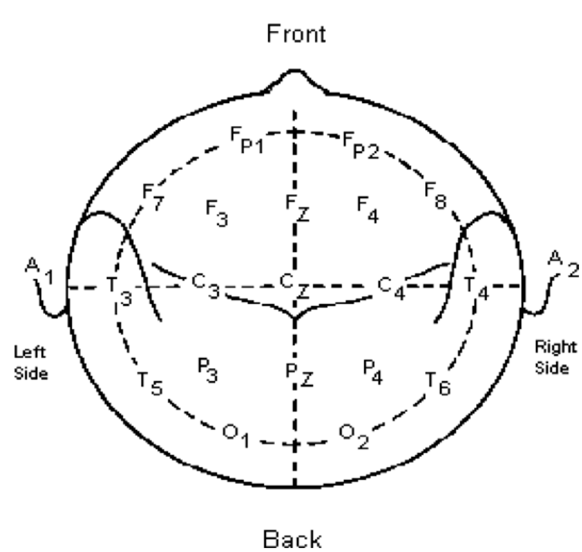

Fig. 1. The 10–20 electrode placement for recording a EEG pattern.