Biomedical Signal Processing and Control 38 (2017) 148–157

Contents lists available at ScienceDirect

Biomedical Signal Processing and Control

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

Research paper

Automated identification of epileptic seizures in EEG signals based on phase space representation and statistical features in the CEEMD domain

Jian Jia [a,b], Balaji Goparaju [c], JiangLing Song [a,b], Rui Zhang [a,b], M. Brandon Westover [c,*]

[a] School of Mathematics, Northwest University, Xi'an, Shaanxi 710127, China
[b] Medical Big Data Research Center, Northwest University, Xi'an, Shaanxi 710127, China
[c] Massachusetts General Hospital, Neurology Department, Boston, MA 02114, USA

ARTICLE INFO

Article history:
Received 27 December 2016
Received in revised form 16 May 2017
Accepted 29 May 2017

Keywords:
EEG
Epileptic seizure detection
Complete ensemble empirical mode decomposition
Random forest classifier

ABSTRACT

Epileptic seizure detection based on visual inspection by expert physicians is burdensome, and subject to error and bias. In this work, we present a novel method for the automated identification of epileptic seizure using a single-channel EEG signal. We utilize the complete ensemble empirical mode decomposition with adaptive noise (CEEMDAN) technique to devise an effective feature extraction scheme for physiological signal analysis, and construct the corresponding growth curve. Then, various statistical features are extracted from the growth curve as the feature set, and this is fed to the random forest classifier for completing the detection. The suitability of the extracted features is established through statistical measures and graphical analysis. The proposed method is evaluated for the well-known problem of classifying epileptic seizure and seizure-free signals using a publically available EEG database from the University of Bonn. To assess the performance of the classification method, 10-fold cross-validation is performed. Compared to state-of-the-art algorithms, the numerical results confirm the superior algorithm performance of the proposed scheme in terms of accuracy, sensitivity, specificity, and Cohen's Kappa statistics.

1. Introduction

Epilepsy, a common neurological disorder, raises the risk of death and causes a social barrier, which degrades the quality of life of the patient [1]. More than 150,000 new cases of epilepsy are reported every year in the United States, and over 50 million people have epilepsy worldwide; these numbers are expected to rise in the coming years [2]. Therefore, diagnosis of epilepsy is an essential tool needed to combat this widespread public health problem. Traditionally, seizure detection is performed by visual inspection of electroencephalogram (EEG) signals by an expert physician, who classifies the report into different classes, specifically, seizure (ictal), seizure free (inter-ictal), and healthy. Manual detection can be a troublesome and time-consuming process. Thus, there is a need for an automated seizure detection algorithm that can analyze high-capacity data and accelerate diagnosis, alleviating the load on the physician.

The EEG is an efficient method of obtaining neural activity in the brain. It is often used in the diagnosis of epilepsy, as deviations like spikes or spindles can be observed in the EEG signals during epileptic seizures. Various algorithms have been proposed to automatically detect and identfy epileptic seizures from EEG recordings. A novel methodology based on key-point local binary pattern and support vector machine (SVM) was proposed in [3]. Most algorithms use domain transformation techniques, such as wavelet transforms [4,5] and multi-wavelet transform [6], as the first step in EEG signal detection. Peker et al. [7] extracted features at different levels of granularity by using the dual-tree complex wavelet transformation, followed by complex-valued methods for the classification of EEG data. Swami et al. [8] extracted the feature sets using the dual-tree complex wavelet transformation, following which the feature sets were fed into a general regression neural network to detect seizures in electroencephalography. Bhati et al. [5] developed a time-frequency optimal biorthogonal three-band wavelet filter bank for EEG signal classification. Sharma et al. [9] used the analytic time-frequency flexible wavelet transform coupled with the fractal dimension for the analysis of EEG signals. Following this, feature vectors were passed through a least squares

* Corresponding author.
E-mail address: MWESTOVER@mgh.harvard.edu (M.B. Westover).

http://dx.doi.org/10.1016/j.bspc.2017.05.015
1746-8094/© 2017 Published by Elsevier Ltd.

support vector machine (LS-SVM) classifier to discriminate among the various EEG classes.

Apart from discrete wavelet transform and its variants, empirical mode decomposition (EMD) is one of the most commonly used methods for time-frequency transformation. This useful tool describes the behavior of non-stable and nonlinear signals by decomposing signals into intrinsic mode functions (IMF) to obtain instantaneous frequency data. In [10], features like ellipse area were extracted from the second-order difference plot of IMFs and used to perform classification of epileptic seizure and seizure-free EEG signals. In additional, the fluctuation index and the variation of IMFs coefficient were estimated from the EEG signals and found appropriate for the identification of modes of ictal EEG signals [11]. Subasi and Gursoy [12] performed epileptic EEG signal classification using principal component analysis (PCA), independent component analysis (ICA), and SVM. Bajaj and Pachori [13] computed modulation bandwidth features in the EMD domain and used LS-SVM to perform classification. Alam and Bhuiyan [14] used the EMD and the artificial neural network to classify the epileptic seizures. Sharma and Pachori [15] extracted the interquartile range of the Euclidean distance parameters and the confidence ellipse area from the two-dimensional and the three-dimensional phase space representation (PSR) of IMFs to detect the epileptic EEG signals. Hassan and Bhuiyan [16] proposed an EMD and statistical feature based feature generation scheme and used adaptive boosting as a classification model for single channel automatic sleep scoring.

All the methods mentioned above suffer from the mode-mixing problem (mixed frequencies) as they are based on the EMD method. Attempts have been made to resolve this issue. Riaz et al. [17] utilized temporal and spectral statistics in the EMD domain and SVM to devise an automated seizure detection scheme. Fu et al. [18] implemented Hilbert marginal spectrum analysis and SVM for epilepsy seizure detection from EEG data. Hassan et al. [19] employed feature extraction from complete ensemble empirical mode decomposition with adaptive noise (CEEMDAN). This technique provided good separation of spectrum in the IMFs and faithful reconstruction of the original signal. Then, classification was performed using linear programming boosting classifier to detect epilepsy seizure data.

In our work, we mainly focus on the features extracted from the PSR of IMFs in the complete ensemble empirical mode decomposition domain of EEG signals. The PSR is calculated with different parameters, such as time lag and embedding dimension (d). After extraction of the two-dimensional (d = 2) and three-dimensional (d = 3) PSRs of the IMFs from CEEMDAN, we transform the PSRs by averaging the paired-distances of each point in the PSR and the midpoint of the PSR. Then, we calculate the sorted differences between the transformed phase space data to generate the growth curve, and finally use these as features to detect epileptic seizure EEG signals. In our work, CEEMDAN is applied to decompose the signals associated with IMFs of the EEG signals. Next, the growth curve of the data in two-dimensional and three-dimensional PSRs of the extracted IMFs of EEG signals are analyzed. Additionally, we compute the spectral features used in the field of physiological signal classification and normal inverse Gaussian (NIG) parameters as the feature set. This data is fed to a random forest (RF) classifier to perform classification of the epileptic seizure and seizure-free EEG signals.

The outline of the paper is as follows. Section 2 describes the EEG dataset, the CEEMDAN method, PSR, and growth curve computation from two-dimensional and three-dimensional PSRs of IMFs, and RF classifier. The experimental results are presented in Section 3 and analyzed in Section 4. Finally, Section 2 concludes the paper.

2. Methodology

2.1. Dataset

A dataset from the University of Bonn is used as the benchmark dataset in this paper. This dataset is available online publicly and described in [20]. It consists of five sets of packet data, namely Z, O, N, F, and S. Each set has 100 single-channel EEG segments belonging to individuals, and each single-channel segment lasts for 23.6 s. Based on the international standard 10–20 EEG electrode placement system, sets Z and O record the signal of the awake and relaxed state of surface EEG segments of five healthy volunteers when they open and close their eyes. Sets F and N comprise signals generated when the electrodes are placed in the epileptogenic zone and the hippocampal formation of the opposite hemisphere, respectively. The data in set S consist of signals from electrodes placed intra-cranially, as well as those implanted in the temporal and basal regions of the neocortex. The EEG data in set S correspond to seizure attacks, whereas seizure-free epochs are present in sets F and N. Each EEG signal segment is sampled using a 128-channel amplifier system with a 12-bit resolution and digitized at 173.61 Hz. Finally, each segment is sampled at a length of $173.61 \times 23.6 \approx 4097$, and the corresponding bandwidth is 86.8 Hz.

In this paper, we utilize the EEG signals in subsets S, F, and N to assess the performance of the proposed method. In the first step of the experiment, the seizure-free EEG signals are derived from packaged sets N and F, and the epileptic seizure EEG signals are taken from set S. In the second step of the experiment, sets F and S are used as the seizure-free class and epileptic seizure class of EEG signals, respectively.

2.2. Complete ensemble empirical mode decomposition with adaptive noise

Using the EMD technique [21], the signal can be decomposed into a set of IMFs. Only one mode of oscillation exists in each IMF and this satisfies two essential principles [13,21]. The first principle states that in the total data set, the number of zero crossings and the number of extreme points should either be equal or differ at most by one. According to the second, at any point of IMF, the average value of the envelopes defined by local minima and local maxima is zero.

CEEMDAN was preferred over EMD for use in the algorithm, owing to its advantages over commonly used signal processing techniques. EMD exhibits mode mixing as various kinds of oscillations can be found in the same mode and different modes. CEEMDAN, on the other hand, can be reconstructed exactly by adding all the IMFs and terminal residue, and provides a better spectral separation of the modes [22]. CEEMDAN is a powerful tool for processing highly non-stable and nonlinear signals such as epilepsy signals because of its data-driven nature, low computational cost [23], and lack of need for a prior basis function. To obtain a good performance, traditional time-frequency transforms need to make a choice among many basis functions. Depending on the chosen option, the performance would change from case to case, and even between datasets [22].

Let $w^i(n)$ with $i = 1, 2, \ldots, I$ represent the different realizations of white Gaussian noise with zero mean and unit variance. The standard deviation of the white Gaussian noise is denoted by ε_0. The operator $E_j(\cdot)$ is defined as producing the j-th mode of EMD. The CEEMDAN algorithm comprises the following steps [24].

Step 1. Calculate $x(n) + \varepsilon_0 w^i(n)$.

Step 2. Utilize the EMD method to decompose the mentioned I signals and obtain their first modes.