Biomedical Signal Processing and Control 39 (2018) 360–377

Contents lists available at ScienceDirect

Biomedical Signal Processing and Control

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

Research paper

Fuzzy distribution entropy and its application in automated seizure detection technique

Tao Zhang, Wanzhong Chen*, Mingyang Li

College of Communication Engineering, Jilin University, 130012 Changchun, China

ARTICLE INFO

Article history:
Received 12 February 2017
Received in revised form 16 July 2017
Accepted 3 August 2017
Available online 30 August 2017

Keywords:
Electroencephalogram (EEG)
Automated seizure detection
Fuzzy distribution entropy (fDistEn)
Wavelet packet decomposition (WPD)
k-nearest neighbor (*k*-NN)

ABSTRACT

Visual inspection of Electroencephalogram (EEG) records is the conventional diagnostic method of epilepsy but it is expensive, time-consuming and tedious. Therefore, it is necessary to develop automated seizure detection technologies. In this paper, a new entropy named fuzzy distribution entropy (fDistEn) was first put forward and then a seizure detection scheme combining wavelet packet decomposition (WPD), fDistEn, Kruskal-Wallis nonparametric one-way analysis of variance and *k*-nearest neighbor (*k*-NN) classifier was proposed. In the proposed scheme, WPD was firstly implemented to decompose the filtered EEG into several wavelet sub-bands. Subsequently, fDistEn values of all nodes in every level were calculated and followed by selecting significant features using Kruskal-Wallis test. Finally, *k*-NN was employed to classify ten kinds of EEG combinations. Experimental results show fDistEn can measure the complexity of signals and our proposed scheme is qualified to detect seizure automatically with not less than 98.338% accuracy in all cases. Compared with existing methods, our scheme outperforms most of state-of-the-art articles and it indicates the effectiveness of the proposed seizure detection scheme.

1. Introduction

Epilepsy is characterized by intermittent abnormal neuronal firing and the recurrent occurrence of excessive spontaneous discharge is the so-called "epileptic seizures" [1]. Being one of the most common neurological diseases, approximately 50 million people all over the world are suffering from epilepsy and nearly four-fifths of epilepsy patients come from the developing countries [2,3]. In China, over 6 million people have epilepsy and the proportion continues to grow rapidly [4]. Epilepsy can be detected noninvasively by electroencephalogram (EEG) which is the external manifestation of a very large number of neuronal membrane potentials and can be measured through electrodes placed on the scalp [5,6]. As the most frequently used diagnostic approach of epilepsy, by visual inspection of EEG records, professional neurologists are able to judge whether the subjects are suffering from epilepsy and further to provide proper therapeutic protocol for diagnosed patients. However, dozens hours of continuous EEG record is required which makes epilepsy diagnosis expensive, time-consuming and tedious. What is more, the diagnosis results of the same record from different doctors may be discrepant due to lacking of consistent reference standard and being prone to personal experience [3]. Compared with manual diagnosis of epilepsy, automated computer aided diagnostic (CAD) techniques combining signal processing and machine learning algorithms are of lightened weight of artificial workload, unified processing procedures and evaluation indexes as well as low financial burden. The advanced signal processing algorithms provide the possibility of capturing minute but helpful changes in EEGs. During the past several decades, numerous promising CAD techniques have been developed to achieve the goal of automated seizure detection [7,8].

The overwhelming majority of seizure detection technologies consist of two major procedures: feature extraction and classification, and the former is the key element in achieving satisfactory detection result. In recent years, an increasing attention is payed to nonlinear analytical techniques like fractal dimension, correlation dimension, Lyapunov exponent and so on, as the fact that researches have testified EEG is a nonlinear and nonstationary signal [7]. Entropy is an index which is capable of reflecting the complexity of discrete sequence and the internal dynamic characteristic of a system like brain. Taking advantage of embedding entropies (for example, permutation entropy (PE) [9–11], approximate entropy (ApEn) [4,12–15], sample entropy (SampEn) [16–19]), spectral entropies (for instance, normalized bispectrum entropy, phase entropy (PhEn S1 and S2) [8,20]) and other entropies (such as wavelet entropy [2], recurrence quantification analysis

* Corresponding author.
E-mail address: chenwz@jlu.edu.cn (W. Chen).

http://dx.doi.org/10.1016/j.bspc.2017.08.013
1746-8094/

(RQA)-based entropy [21]), a series of seizure detection schemes have been put forward. For example, a previous study presented by Song and Zhang showed SampEn value of EEG from preictal period is lower than that of EEG from interictal state [22]. The extracted SampEn features of preictal and interictal EEGs were classified using extreme learning machine (ELM), the sensitivity and specificity are 86.75% and 83.80%, respectively. With the application of RQA-based maximum entropy ratio (MER), a similar finding that MER of ictal state is lower than that of interictal phase is reported [23]. In literature [9], utilizing the character that normal EEG possesses a higher PE than seizure EEG, a combination of PE and support vector machine (SVM) was investigated for the first time. Experimental results showed the proposed technique can differentiate epileptic and seizure-free EEGs, with 94.38% average sensitivity and 93.23% average specificity. Based on ApEn and artificial neural networks (ANNs), an automated seizure detection system was designed to classify normal and epileptic EEGs in [14]. A clear discrimination of ApEn between the normal and epileptic EEG signals was found when the length of EEG was 256 and the presented system yielded high overall accuracies of 99.35% to 100%. In addition to merely applying entropies to characterize EEG, many fusion schemes of wavelet and entropies had also been raised. To name a few, authors in [12] proposed a novel seizure detection technique based on fifth-level discrete wavelet transform (DWT), ApEn and ANN. They declared their approach can discriminate normal and epileptic EEGs with a high accuracy of 100%. In one of our previous works [4], frequency slice wavelet transform was used to extract five rhythms of epileptic EEGs, their ApEn and fluctuation index values were then computed and finally fed into a genetic algorithm-SVM (GA-SVM) classifier for classification. A classification accuracy of 98.33%, a sensitivity of 99%, a specificity of 99% and a positive predictive value of 99.5% were obtained in classifying normal, interictal and ictal EEGs. A comparative study of seizure detection exploiting wavelet packet decomposition (WPD)-based ApEn and SampEn was carried out by Zhang et al. [24]. They concluded that “db2” wavelet function with five-level WPD was the optimum choice to analyze epileptic EEGs. Furthermore, the WPD+SampEn+ELM strategy achieved better performance than WPD+SampEn+SVM approach. These studies mentioned above all result in remarkable accuracies and they manifest the entropy-based approaches are able to provide the feasibility of revealing the hidden peculiarity of epileptic EEGs.

Recently, based on the principle of SampEn and fuzzy theory, Chen et al. [25] introduced the improved SampEn named fuzzy entropy (FuzzyEn) to represent the regularity of surface electromyography (EMG) signal. FuzzyEn, in nature, is the negative natural logarithm of the conditional probability aiming at measuring the degree that two vectors similar for m points remain similar for the next $m+1$ points [25]. Unlike ApEn and SampEn in which the similarity of two vectors is defined by Heaviside function, instead, an exponential function was adopted to estimate the similarity of reconstructed vectors in FuzzyEn. Compared with Heaviside function, exponential function is a continuous and convex curve by which the similarity of any two vectors changes gradually and the self-similarity is the maximum. In view of these advantages, Xiang et al. [26] presented a FuzzyEn-based seizure detection scheme and this scheme was tested on two epileptic EEG datasets. They concluded that the hybrid method of FuzzyEn and SVM outperformed the SampEn-based method, the former obtained 98.31% and 100% accuracies for the CHB-MIT and Bonn EEG databases, respectively. In [27], the double-density discrete wavelet transform-based FuzzyEn and Hurst exponent features of normal, interictal and ictal EEGs were extracted, GA-SVM classifier was employed to classify various kinds of EEG combinations and the accuracies were equal to or higher than 99.36% for all the combinations.

More recently, Li et al. [28] developed another new entropy called distribution entropy (DistEn) to measure the complexity of heartbeat interval series. Depending on the definition of DistEn, information of a time series hidden in the state space is quantified by evaluating the distribution character of intervector distances [28]. Differing from ApEn, SampEn and FuzzyEn in which two adjacent state spaces are reconstructed, only the m-dimensional state space requires to be reconstructed in DistEn so that the calculation amount of DistEn is almost 50% lower than that of ApEn. Straight after the success of manifesting the descending complexity of healthy aging and heart failure patients using DistEn, they also investigated the possibility of applying DistEn to distinguish normal, interictal, and ictal EEG signals [29]. Moreover, to highlight the superiority of DistEn, SampEn was employed to handle the same task as a comparison. In their experiment, the original EEG sequence was segmented into 5 s length (868 points), the DistEn and SampEn were extracted, respectively. To explore the influence of parameters of DistEn on final classification performance, they had conducted a suite of experiments using 20 kinds of parameter combinations. In these tasks, DistEn shows superior performance than SampEn and it possesses better robustness in dealing with epileptic EEGs.

In the light of the foregoing, DistEn is a productive indicator in characterizing the complexity of biomedical signals especially short time sequences on one hand; on the other hand, FuzzyEn can estimate the complexity of longish segment accurately, but the computational complexity of FuzzyEn is almost twice than that of DistEn. Integrating the empirical probability density function (ePDF) of DistEn and the fuzzy membership function as well as the baseline removal operation of FuzzyEn, a new embedding entropy named fuzzy distribution entropy (fDistEn) is developed to characterize the nonlinear characteristic of epileptic EEG signals in this paper. Additionally, a novel automated seizure detection scheme based on WPD, fDistEn, Kruskal-Wallis nonparametric one-way analysis of variance (ANOVA) and k-nearest neighbor (k-NN) is also proposed. The original EEG signal is first filtered by a low-pass filter, followed by implementing WPD to decompose it into multiple wavelet sub-bands. Subsequently, the fDistEn features of all nodes of all levels are extracted and Kruskal-Wallis nonparametric one-way analysis of variance (ANOVA) is then carried out to select the significant features. At last, those significant features are passed into k-NN for classification. Effects of parameters of fDistEn, wavelet basis, wavelet decomposition level and number of nearest neighbor of k-NN on final classification performance are all explored.

The rest of this paper is structured as follows. In Section 2, EEG database used in this study is first briefly introduced, then algorithm of fDistEn and the proposed automated seizure detection scheme are described in the following two sub-sections. Experimental results are presented in Section 3. Section 4 discusses the results and then a comparison between the proposed technique and existing methods is made. Finally, Section 5 concludes this work.

2. Material and methods

2.1. EEG database

The EEG database, obtained from the Epilepsy Center of the Bonn University Hospital of Freiburg [30], is employed to verify the effectiveness of proposed seizure detection scheme. The database consists of five EEG subsets and they are denoted as Sets A, B, C, D and E, respectively. In each subset, there are 100 single-channel EEG records and each record has 23.6 s in duration captured by the international 10–20 electrode placement scheme. Sets A and B are taken from five healthy volunteers with eyes open and closed,