Biomedical Signal Processing and Control 30 (2016) 86–97

Contents lists available at ScienceDirect

Biomedical Signal Processing and Control

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

Sleep staging from the EEG signal using multi-domain feature extraction

Zhiyong Liu[a], Jinwei Sun[a,*], Yan Zhang[a], Peter Rolfe[a,b]

[a] School of Electrical Engineering and Automation, Harbin Institute of Technology, Harbin 150001, China
[b] Oxford BioHorizons Ltd, Oxford, UK

ARTICLE INFO

Article history:
Received 27 October 2015
Received in revised form 11 April 2016
Accepted 22 June 2016

Keywords:
EEG
MF-DFA
VGA
LS-SVM

ABSTRACT

The analysis of the electroencephalogram (EEG) can yield much useful information about brain function, including indications of sleep stage. During the process of EEG analysis, feature extraction is one of the most critical technical aspect. Traditional EEG feature extraction methods are mainly based on single domain analysis. However, due to the highly non-stationary and nonlinear characteristics of the EEG, it is difficult to extract comprehensive information only from single domain analysis. In the present study, a novel feature extraction method was proposed based on the multi-domain analysis of the EEG. Fifteen characteristic parameters were extracted based on the multifractal detrended fluctuation analysis (MF-DFA), visibility graph algorithm (VGA), frequency analysis and nonlinear analysis. Ten optimal parameters of the fifteen parameters were selected by the genetic algorithm (GA). Then the Least Squares-Support Vector Machines (LS-SVM) were used to classify the sleep states. The cross validation results demonstrated that multi-domain feature extraction method can obtain more useful information in the EEG signal. Compared to the frequency domain parameters, nonlinear parameters and time domain parameters, the predictive accuracy of sleep staging classification with optimal multi-domain parameters improved 11.08%, 10.76% and 6.40% respectively.

1. Introduction

Electroencephalographic (EEG) signals can be used to reflect the electrical potential change between brain cells in the surface of head and provide abundant information about brain function. EEG is a weak signal with the amplitude of microvolt level. The amplitude of EEG is about 50–100 μV, and the frequency is lower than 50 Hz. The EEG signal possesses obvious rhythmicity. The information related to sleep states in EEG is mainly focused on δ rhythm (1–4 Hz), θ rhythm (4–8 Hz), α rhythm (8–13 Hz), and β rhythm (14–30 Hz) [1]. Table 1 shows the frequency range and the relationship between EEG rhythms and brain states.

In the process of EEG analysis, feature extraction is implemented by utilizing a series of transformations so that the characteristics can be observed easily in the transform domain to provide the best input to the feature classifier [2]. Since the information related to sleep states is mainly focused on the four rhythms, it can be used to evaluate the quality of sleep or secondary diagnosis of sleep disorders by extracting the suitable parameters from these rhythms.

Prerau et al. researched the physiological and behavioral dynamics of sleep [3]. Reinhard et al. analyzed the sleep EEG signals and found the relationship between the anterior cingulate cortex and dreaming by using the method of power spectral analysis [4]. Huang et al. presented a novel sleep stage classification system, which can be easily used at home. The two forehead EEG signals are analyzed and the accuracy of different sleep states are nearly 76.7% [5]. Imtiaz et al. detected the REM phases by the edge frequency of 8–16 Hz [6]. Immanuel et al. determined the changes of respiratory cycle-related EEG by the scored event-free(SEF) respiratory cycles with different sleep states [7]. Brett et al. demonstrated that the recommended EEG montage (F4-M1, C4-M1, O2-M1) and acceptable EEG montage (Fz-Cz, C4-M1, Oz-Cz) have similar EEG signal [8]. Reinke et al. introduced a novel ICU depth of sleep index from single channel EEG signal, which has excellent ability in estimating the depth of sleep [9]. Zhen et al. utilized the first principal component of EEG singular value with support vector machine method to exploit the classification of sleep states and achieved the accuracy rate up to 84.40% [10]. Jiang et al. analyzed the relationship between heart rate variability and sleep states, and obtained the comprehensive sleep state with multi-parameters [11]. The wavelet multifractal algorithm was employed to realize sleep states fitting to the slp01 sample in MIT-BIH sleep database

* Corresponding author.
E-mail address: jwsun@hit.edu.cn (J. Sun).

http://dx.doi.org/10.1016/j.bspc.2016.06.006
1746-8094/

based on the nonlinear analysis of EEG [12]. The EEG complexity and approximate entropy with support vector machine were adopted to automatically distinguish the sleeping mode and the accuracy rate was up to 85.67% [13].

Recently, researchers found that the nonlinear dynamic parameters could well reflect the information contained in these kinds of non-stationary and nonlinear signals like EEG [14]. There are many kinds of nonlinear dynamic parameters. And among them, the following parameters are commonly used: correlation dimension [15], Kolmogorov entropy [16], Lyapunov exponent [17], Approximate Entropy [18,19], L-Z complexity [20,21], and fractal dimension [22], etc.

There are also some EEG time series analysis methods used to sleep staging, such as detrended fluctuation analysis (DFA) and visibility graph (VG). The multifractal detrended fluctuation analysis (MF-DFA) is developed from the original DFA [23–25]. Matic et al. introduced that the DFA index is highly robust when the time scales is 10–60 s [26]. Lee et al. researched the EEG in sleep apnea using DFA method, and found that the mean scaling exponents of EEG can discriminate the Non-REM, REM (Rapid Eye Movement) and waken stage[27]. Stam et al. fount that the cognitive dysfunction in alzheimer's disease is highly related to the abnormal instinctive fluctuations of EEG synchronization levels [28]. Besides, the DFA has been used in Hurst exponent estimating [29], earth's gravity [30], seismicity [31], high-energy nuclear collisions [32], fast event detection on massive phasor measurement unit data [33], etc.

The visibility algorithm is a simple and fast computational method, which converts a time series into a graph proposed by Lacasa at 2008 [34]. The visibility graph remains invariant under some transformation of the time series. The horizontal visibility graph is proposed by Luque based on the natural visibility at 2009 [35]. Recently, the visibility algorithm has been researched deeply and applied in many fields. The phase locking value and visibility graph similarity are combined together to stage the sleep states, the accurate rate can reach 86.7% [36,37]. A fast weighted horizontal visibility algorithm is also researched to identify the seizure of epileptic [38]. The parametric natural visibility graph algorithm, which presented by Bezsudnov is a modification of natural visibility graph by adding a view angle, has been used for the mapping of the time series to the complex networks [39]. A visibility graph averaging aggregation operator has been proposed based on the natural visibility graph [40].

Accordingly, the EEG signal has been researched deeply in many domains, and many useful results have been achieved. However, due to the highly non-stationary and nonlinear characteristics of the EEG, the comprehensive information could not be extracted completely from EEG only through one-domain feature parameters. In this paper, we explored a novel sleep staging method based on multi-domain feature extraction technique. The remaining of the paper is organized as follows: the experimental data is briefed in the next section. The method is fully described in Section 3. Section 4 investigates the results of each feature extraction method and the performance of different classification methods.

2. Materials

The data analyzed is from the MIT-BIH polysomnography database. The database contains 18 groups of sleep EEG data collected from 16 different volunteers. All the 16 volunteers are male Records slp01a and slp01b which have one hour interval are collected from the same person. slp02a and slp02b are also recorded from the same volunteers. The interval is about ten minutes. The other 14 samples are all from different volunteers.

The sample rate is 250 Hz. Every 30 s, one sleep state is given in this database, and thus every 7500 points of EEG data corresponded to one sleep state. The sleep states included wake(W), sleep 1(S1), rapid eye movement(R), sleep 2(S2), sleep 3(S3) and sleep 4(S4), being represented by 0–5 in this paper. The higher the number is, the deeper the sleep state level is. Among all the 18 samples, the slp01, slp02, slp14, slp16, slp45 and slp59 covered all the six sleep states. So, the six samples are analyzed in this paper, and the details of the samples are shown in Table 2.

Table 1
Relationship between EEG rhythms and brain states.

EEG rhythm	Frequency range (Hz)	States
δ	1–4	Deep sleep, NREM sleep, Unconsciousness
θ	4–8	Reminiscence, Fantasy, Imagine, Light sleep
α	8–13	Relaxed but not sleepy, Calmness, Conscious
β	14–30	Vigilance, Conscious of self and surroundings

Table 2
MIT-BIH polysomnography database.

Sample	Reference voltage (mv)	Position	Time (h)
slp01	−6430	C4-A1	5:00
slp02	−8300	O2-A1	5:15
slp14	9310	C3-O1	6:00
slp16	−13215.4	C3-O1	6:00
slp45	−9653.33	C3-O1	6:20
slp59	6284.92	C3-O1	4:00

3. Methods

Fig. 1 is the process diagram of sleep staging from the EEG signal using multi-domain feature extraction.

From Fig. 1, the new sleep staging algorithm contains four steps: (1) Pretreatment. Generally, the EEG data collected from volunteers always contains many kinds of noise. This step is to get the pure EEG signal by removing noise as much as possible. (2) Feature extraction. Based on the analysis of the EEG characteristics, nine time domain parameters, four frequency domain parameters, two nonlinear parameters, a total of 15 parameters are extracted to classify the sleep states. (3) Feature selection. The purpose of feature selection is to get the most optimal EEG features used to classification. (4) Feature classification. Classify the sleep states by putting the most optimal EEG features into the LS-SVM classifier.

3.1. Pretreatment

The information related to sleep states in EEG is mainly focused on δ rhythm (1–4 Hz), θ rhythm (4–8 Hz), α rhythm (8–13 Hz), and β rhythm (14–30 Hz), thus the MIT-BIH database EEG signal are filtered by a band pass FIR filter. The passband is 1–30 Hz. Besides, there are also some inevitable artifacts in the original EEG, for instance, electrooculography (EOG), electromyogram (EMG) and electrocardiogram (ECG). While the EEG in the MIT-BIH polysomnography database are collected in the sleep state, there are only a few EMG and ECG artifacts in the original EEG. The EOG artifacts generated by blinks and eye movements appeared in the wake state, the others which generated by eye movements also accrues in the REM (Rapid eye movement) sleep state. The two kinds EOG artifacts have different frequency feature. We use the EOG artifacts as the useful signal in the EEG to classify the sleep state.