Biomedical Signal Processing and Control 8 (2013) 624–633

Contents lists available at SciVerse ScienceDirect

Biomedical Signal Processing and Control

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

Sleep stages classification based on heart rate variability and random forest

Meng Xiao*, Hong Yan, Jinzhong Song, Yuzhou Yang, Xianglin Yang

China Astronaut Research and Training Center, P.O. Box 5132-16, Beijing 100094, China

ARTICLE INFO

Article history:
Received 5 February 2013
Received in revised form 17 May 2013
Accepted 5 June 2013

Keywords:
Sleep stage
Heart rate variability
Random forest
Feature importance

ABSTRACT

An alternative technique for sleep stages classification based on heart rate variability (HRV) was presented in this paper. The simple subject specific scheme and a more practical subject independent scheme were designed to classify wake, rapid eye movement (REM) sleep and non-REM (NREM) sleep. 41 HRV features extracted from RR sequence of 45 healthy subjects were trained and tested through random forest (RF) method. Among the features, 25 were newly proposed or applied to sleep study for the first time. For the subject independent classifier, all features were normalized with our developed fractile values based method. Besides, the importance of each feature for sleep staging was also assessed by RF and the appropriate number of features was explored. For the subject specific classifier, a mean accuracy of 88.67% with Cohen's kappa statistic κ of 0.7393 was achieved. While the accuracy and κ dropped to 72.58% and 0.4627, respectively when the subject independent classifier was considered. Some new proposed HRV features even performed more effectively than the conventional ones. The proposed method could be used as an alternative or aiding technique for rough and convenient sleep stages classification.

1. Introduction

As an important physiological activity of people, sleep contributes to self-repairing and self-recovering. However, as living rhythm quickens and lifestyle changes, sleepiness and sleep structure disorder have become severe factors threatening people's normal daily life and public safety. Thus it is of great importance to monitor sleep and analyze sleep structure.

Nowadays, the 'gold standard' method of evaluating sleep structure is overnight polysomnography (PSG). It is a multi-parametric system which routinely records many kinds of biological signals synchronously, such as electroencephalogram (EEG), electrooculogram (EOG), electromyogram (EMG), blood oxygenation, airflow, and respiratory effort. Based on the golden manual sleep classification criterion, an overnight sleep was divided into 30-s epochs, and every epoch was categorized as wake, rapid eye movement (REM) sleep or an approximate continuum of depth (stages 1–4) during non-REM (NREM) sleep. In 2007, American Academy of Sleep Medicine (AASM) released the latest edition of sleep staging criterion [1], where NREM3 and NREM4 were combined into a single stage and some other details were changed. To overcome the tedious manual work, many automatic PSG systems have appeared. And most of them can give many annotations such as respiratory events and body movements. However, the PSG system is rather expensive and too cumbersome for using by untrained persons. And it is very intrusive for normal sleep because of too many electrodes. Therefore, PSG is restrictively used in specialized hospital-based sleep laboratory and suffers difficulty in wider application like home nursing.

There has been considerable interest in the development of some alternative reliable low cost sleep staging techniques. In fact, due to the modulation of autonomic nervous system (ANS), some other biological signals such as heart rate variability (HRV), respiratory effort, and oxygen saturation also represent characteristic behaviors that vary according to sleep type and depth. Specially, HRV is generally derived from RR intervals of electrocardiogram (ECG) that can be acquired from some ambulatory devices with only a few electrodes. It shows the advantages of convenience and low-cost, and is widely used as noninvasive method to get insight into ANS functioning. Numerous investigations have demonstrated that heart rate (HR) decreased in associated with decreased variability in NREM sleep, while HR increased, with increased variability in wake and REM sleep [2–5]. A decreased ratio of the power in low frequency (LF, 0.04–0.15 Hz) and high frequency (HF, 0.15–0.4 Hz) band (LF/HF) is associated with NREM sleep and significantly increased LF/HF is shown in REM sleep [6]. In addition, some nonlinear measures like fractal component [7] and detrended fluctuation analysis (DFA) [8] during different sleep stages show significant differences as well.

* Corresponding author. Tel.: +86 13811942507; fax: +86 01066362295.
E-mail addresses: xiaomeng16@126.com, xiaomeng16@foxmail.com (M. Xiao), hholden@sina.com (H. Yan), sjinzhongkx@163.com (J. Song), dove_yyz@126.com (Y. Yang), xianglinyang@163.com (X. Yang).

1746-8094/$ – see front matter © 2013 Elsevier Ltd. All rights reserved.
http://dx.doi.org/10.1016/j.bspc.2013.06.001

Thus the technique of HRV based sleep staging has attracted much attention in the past decades. Nason et al. [9] applied the non-decimated wavelet packet transform to model HRV of infants. Through the linear discriminate analysis, classification accuracy between sleep and wake was reported as 75–90% for the infants at different stages of development. Adnane et al. [10] showed that analyzing HRV in time domain, frequency domain, and by DFA and windowing DFA one can classify wake and sleep to an average accuracy of 79.31% and Cohen's κ statistic [11,12] of 0.41 using supporting vector machine.

But the above literatures were restricted in subject specific scheme, where the training and testing set come from the same subject. In order to develop more practical solution, the subject independent scheme, where the training and testing group are selected from different subjects should also be discussed. Researchers have recognized this problem and made some significant explorations.

A comparison between HRV and actigraphy for sleep-wake identification in infants was made by Lewicke et al. [13]. With 13 infants combined into training data and the other 12 served as testing set, the performance with accuracy of 79.7% was reported. Mendez et al. [14] used varying autoregressive model to classify REM and NREM sleep by hidden Markov model. With training and testing set including 12 different subjects respectively, their result agreed with R&K standard criteria of average 79.3%.

These investigations strode an important step to more practical study. However, after decades of investigation, it is still an open question and many problems still need to be solved yet. Further exploration employing more effective HRV features and classification methods should be made to improve classification accuracy. To the authors' knowledge, although large amount of HRV features have been extracted for sleep stages classification, few literatures have evaluated the importance of single feature.

In this paper, a total of 41 comprehensive HRV features including time domain measures, frequency domain measures, and nonlinear parameters were extracted. 25 features were proposed or applied to sleep study for the first time. Then the classification method of random forest (RF) was applied to distinct wake, REM sleep and NREM sleep within two schemes, *i.e.* subject specific and subject independent scheme. Additional, the importance of every single feature was assessed and the appropriate number of features was explored as well.

2. Materials and methods

2.1. Database

The public database Sleep Heart Rate and Stroke Volume Data Bank (SHRSV) [15] was used in our study. It was provided to aid research leading to development of automatic classification of sleep stages from heart rate related data. Between 1999 and 2005, overnight PSG records corresponding to 45 healthy subjects, aged 16–61, with 28 men and 17 women, were acquired in the Sleep Center Institute of Psychophysiology and Rehabilitation of Kaunas University of Medicine. The whole sleep was classified as wake, REM sleep and stage 1, 2, 3 and 4 of NREM sleep at 30-s epochs according to sleep scoring manual by experienced doctors.

The full database only contains RR sequence and stroke volume (SV) instead of the entire PSG signals. RR intervals were extracted from ECG signals at a sampling frequency of 500 Hz by automated rhythmus analysis with automatic and manual review and correction. Besides, the stationarity of RR records was also described. For this study, only data labeled with 'stationary' were analyzed, while the others labeled with 'artifact' or 'non-stationary', and the records with else abnormal phenomenon such as premature beats, atrial fibrillation, paroxysmal tachycardia, sleep apnea, *etc.*, were excluded. In addition, the subtle stages 1, 2, 3 and 4 of NREM sleep were combined into NREM sleep, meanwhile wake and REM sleep remained as the same. After records filtering and stages combination, the time length of whole sleep for each individual were 5.09 ± 1.35 h (mean ± standard deviation). The percentage of each stage takes in the whole night data was listed in Table 1.

Table 1
Percentage of each sleep stage takes in the whole night data (mean ± standard deviation).

Stage	Percentage
Wake	15.73% ± 7.55%
REM	17.03% ± 8.77%
NREM	67.24% ± 9.54%

2.2. RR preprocessing and section

RR preprocessing is a prerequisite for extracting effective HRV features. In order to remove out the influence of outliers, each RR value was compared to the mean value (*mRR*) within a 21 points rectangle window centered around the tested value. If RR was less than *0.5*mRR* or lager than *1.5*mRR*, it was replaced by *mRR*, otherwise it remained the same. This obtained sequence was called RR_{norm}.

Afterwards, each RR sequence was sectioned. At first, each RR record was divided into 30-s epochs synchronizing in time with sleep stages classification. And then the whole record was further segmented into sections with a duration of 5-min centered around every 30-s epoch [16,17]. Every overlapping 5-min section belongs to a specific sleep stage and it is the data to be analyzed. For convenience, those 5-min sections were still called RR_{norm}.

2.3. Feature extraction

The assembly of HRV features examined in this study included traditional time domain features, frequency domain features, and some nonlinear analysis measures. To the authors' knowledge, some of the features were proposed or applied to sleep study for the first time in our paper.

2.3.1. Time domain features

Time domain features, which analyze the variation of RR intervals through statistical methods, are the simplest and most intuitive measures to characterize HRV. The time domain features adopted in our study were listed in Table 2.

2.3.2. Frequency domain features

Frequency domain features of HRV were important indicators to reflect the activity of ANS. The power in LF (0.04–0.15 Hz) and HF (0.15–0.4 Hz) band were related to the regulation of sympathetic (SNS) and para-sympathetic (PNS) nervous system, respectively. Some literatures also demonstrated that medium frequency (MF, 0.1–0.15 Hz) power was related to baroreflex activity [18,19]. To make full use of spectral information of HRV, LF was further divided into true LF (TLF, 0.04–0.1 Hz) and MF in our study. The combination of LF and HF, denoted total frequency (TF) in our study was also investigated. And more comprehensive measures, *i.e.* spectrum power, mean frequency and spectral entropy corresponding to different spectral bands were determined. In addition, peak in HF and fractal dimension reflected in very low frequency (0.0033–0.04 Hz, VLF) were discussed as well. The partition of a typical HRV power spectrum was shown in Fig. 1.

RR intervals were interpolated at a sampling frequency of 4 Hz with the help of cubic spline function [4]. Then the auto regression