Computer Methods and Programs in Biomedicine 140 (2017) 201–210

Contents lists available at ScienceDirect

Computer Methods and Programs in Biomedicine

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

Automated identification of sleep states from EEG signals by means of ensemble empirical mode decomposition and random under sampling boosting

Ahnaf Rashik Hassan*, Mohammed Imamul Hassan Bhuiyan

Department of Electrical and Electronic Engineering, Bangladesh University of Engineering and Technology, Dhaka 1205, Bangladesh

ARTICLE INFO

Article history:
Received 21 May 2016
Revised 21 December 2016
Accepted 26 December 2016

Keywords:
RUSBoost
Sleep stage classification
EEMD
EEG
Statistical features

ABSTRACT

Background and objective: Automatic sleep staging is essential for alleviating the burden of the physicians of analyzing a large volume of data by visual inspection. It is also a precondition for making an automated sleep monitoring system feasible. Further, computerized sleep scoring will expedite large-scale data analysis in sleep research. Nevertheless, most of the existing works on sleep staging are either multichannel or multiple physiological signal based which are uncomfortable for the user and hinder the feasibility of an in-home sleep monitoring device. So, a successful and reliable computer-assisted sleep staging scheme is yet to emerge.

Methods: In this work, we propose a single channel EEG based algorithm for computerized sleep scoring. In the proposed algorithm, we decompose EEG signal segments using Ensemble Empirical Mode Decomposition (EEMD) and extract various statistical moment based features. The effectiveness of EEMD and statistical features are investigated. Statistical analysis is performed for feature selection. A newly proposed classification technique, namely – Random under sampling boosting (RUSBoost) is introduced for sleep stage classification. This is the first implementation of EEMD in conjunction with RUSBoost to the best of the authors' knowledge. The proposed feature extraction scheme's performance is investigated for various choices of classification models. The algorithmic performance of our scheme is evaluated against contemporary works in the literature.

Results: The performance of the proposed method is comparable or better than that of the state-of-the-art ones. The proposed algorithm gives 88.07%, 83.49%, 92.66%, 94.23%, and 98.15% for 6-state to 2-state classification of sleep stages on Sleep-EDF database. Our experimental outcomes reveal that RUSBoost outperforms other classification models for the feature extraction framework presented in this work. Besides, the algorithm proposed in this work demonstrates high detection accuracy for the sleep states S1 and REM.

Conclusion: Statistical moment based features in the EEMD domain distinguish the sleep states successfully and efficaciously. The automated sleep scoring scheme propounded herein can eradicate the onus of the clinicians, contribute to the device implementation of a sleep monitoring system, and benefit sleep research.

1. Introduction

In clinical practice, sleep stage annotation is typically performed by an expert scorer on the basis of visual examination of polysomnographic (PSG) measurements which is composed of electroencephalogram (EEG), electromyogram (EMG) and electrooculogram (EOG). In this respect, Rechtschaffen's and Kales's (R&K) recommendations [1] are widely followed. The annotation of 8-h (whole night) recording requires approximately 2–4 h which is not pragmatic in current clinical settings [2]. Moreover, visual inspection of this gargantuan volume of data not only makes this process onerous for clinicians but also makes sleep scoring subject to human error, monotonous and dependent on expensive human resources. Furthermore, even among experts, the inter scorer agreement is less than 90% [2]. So manual sleep scoring is a subjective process as well. Automatic sleep scoring, on the other hand can eradicate all the aforementioned problems and ensure early detection and rapid diagnosis of various neurological disorders.

* Corresponding author.
E-mail address: ahnaf.hassan0@gmail.com (M.I.H. Bhuiyan).

http://dx.doi.org/10.1016/j.cmpb.2016.12.015

Most prior computer-aided sleep staging algorithms are either multichannel or multiple physiological signal based. Both of these approaches have a number of caveats which make them practically unsuitable. Multichannel based sleep scoring imposes limitations on the subject's movements [3]. Multiple physiological signal based methods on the other hand, complicates the preparation procedure for the subject [4]. Furthermore, it requires more electrodes, leading to more inferences and thus degrading the quality of the recordings. As a result, single channel based sleep scoring using only EEG signal is garnering attention of sleep research community.

We now cite some of the prior pertinent works in the literature. Various signal processing and transformation techniques, such as-time-frequency distributions [5,6], graph theory [3], signal modeling [7], wavelet transform [8], data-adaptive EMD [9] etc. have been used in the literature for computerize sleep scoring. For the classification part, various classification models have been used as well. These include- support vector machine [3,10], neural network based classifiers [6,11], ensemble learning based classifiers [12,13], discriminant analysis [4,14,15], partial least squares [7], relevance vector machine [16] etc. Karkovska et al. [14] extracted many features such as average amplitude, variance, spectral powers, coherence, fractal exponent etc. from data collected from six EEG channels, two EOG channels and one EMG channel and classified using quadratic discriminant analysis. Charbonnier et al. [17] utilized multiple physiological signals (EEG, EMG and EOG) to devise a multichannel-based two stage classification scheme. Liang et al. [4] used multiscale entropy and autoregressive model parameters as features and linear discriminant analysis as classifier for single-channel automatic sleep scoring. Renyi's entropy based features extracted from various time-frequency distributions were used in [5] for sleep stage identification from single channel EEG. In [11], six energy features obtained from single channel EEG were used an Elman neural network classifier for sleep classification. Zhu et al. [3] generated difference visibility graph (VG) and horizontal VG from single channel EEG signal and extracted nine features from them to classify using support vector machine. Koch et al. [18] put forward a Latent Dirichlet Allocation topic model based method using four-channel multiple physiological signals (EEG and EOG) for sleep staging. Long et al. [15] computed various respiratory amplitude, depth and volume based features from respiratory effort signals and used linear discriminant classifier to perform sleep classification. Lajnef et al. [10] employed various features such as linear prediction error energy, variance, skewness, kurtosis, permutation entropy and multi-class support vector machine to perform automatic sleep scoring based on multichannel EEG, EOG and EMG signals. Kayikcioglu et al. [7] proposed an AR coefficient-based feature extraction scheme and utilized partial least squares (PLS) algorithm to classify sleep stages. Dong et al. [9] employed EMD for computer-assisted sleep staging. Tsinalis et al. [6] performed time-frequency analysis for feature extraction and employed stacked sparse autoencoders for sleep stage classification.

This work uses six sleep states in accordance with R&K standard: Awake (AWA), Non-Rapid Eye Movement stages 1–4 (S1–S4) and Rapid Eye Movement (REM). In this study, 2–6 stage sleep state classification problems are considered in our experiments. These classes are described in Table 1. Fig. 1 gives a schematic outline of our method. In this paper, we propound a data-driven single channel based algorithm for computerized sleep scoring. First, we decompose EEG signal segments using Ensemble Empirical Mode Decomposition (EEMD). We then extract statistical moment based features from the resulting mode functions. Statistical analyses are performed to establish the efficacy of the selected features. Finally, we perform classification using a newly proposed hybrid sampling/boosting classification algorithm, namely – Random under sampling boosting (RUSBoost). Until now this is the first time RUSBoost is implemented for automatic sleep staging to the best of our knowledge.

Table 1
Description of various classes considered in this work.

Class	Sleep states
6	AWA, S1, S2, S3, S4, REM
5	AWA, S1, S2, SWS (S3–S4), REM
4	AWA, S1–S2, SWS (S3–S4), REM
3	AWA, NREM (S1–S4), REM
2	AWA, Sleep (REM & NREM)

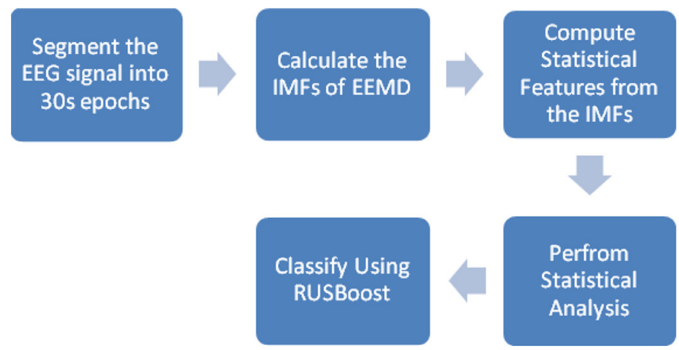

Fig. 1. A schematic outline of the proposed automatic sleep stage classification approach.

The remainder of the paper is organized as follows. Section 2 describes our experimental data, elucidates our feature extraction scheme and analyzes its performance. We also explicate RUSBoost in this section. The experimental results are expounded in Section 3. Section 4 discusses about the significance of the results. Finally, Section 5 suggests some of the future directions of this work and concludes the paper.

2. Materials and methods

We initiate this section with a description of our experimental data. We then elucidate our feature extraction scheme and expound its efficacy. Statistical analysis and the description of our classifier – RUSBoost are also presented in this section.

2.1. Experimental data description

To conduct the experiments, we have used three publicly available and widely used benchmark EEG data-sets, namely – Sleep-EDF database, St. Vincent's University Hospital/University College Dublin sleep apnea database, and DREAMS Subjects database.

2.1.1. Sleep-EDF database

The recordings used for evaluation of the proposed scheme have been obtained from Caucasian males and females (21–35 years old) without any medication. The data can be accessed in Physionet Data Bank's Sleep-EDF Database which is a publicly available benchmark sleep-EEG database [19]. There are eight recordings in two subsets (marked as *sc** and *st**). The first four recordings (sc4002e0, sc4012e0, sc4102e0, sc4112e0) were obtained in 1989 from ambulatory healthy volunteers during 24 h in their normal daily life. The last four data recordings (st7022j0, st7052j0, st7121j0, st7132j0) were obtained in 1994 from subjects who had mild difficulty falling asleep but were otherwise healthy. They contain horizontal EOG, Fpz-Cz and Pz-Oz EEG data, each sampled at 100 Hz. More details on the recordings can be found in [19]. Various prior studies suggest that EEG signal from Pz-Oz