Computers in Biology and Medicine 59 (2015) 42–53

Contents lists available at ScienceDirect

Computers in Biology and Medicine

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

A two-step automatic sleep stage classification method with dubious range detection

Teresa Sousa *, Aniana Cruz, Sirvan Khalighi, Gabriel Pires, Urbano Nunes

Institute of Systems and Robotics (ISR-UC), Electrical and Computer Engineering Department, University of Coimbra, Portugal

ARTICLE INFO

Article history:
Received 31 July 2014
Accepted 21 January 2015

Keywords:
Automatic sleep scoring
Misclassifications detection
Subjects' variability
Dubious range
Clinical applications

ABSTRACT

Background: The limitations of the current systems of automatic sleep stage classification (ASSC) are essentially related to the similarities between epochs from different sleep stages and the subjects' variability. Several studies have already identified the situations with the highest likelihood of misclassification in sleep scoring. Here, we took advantage of such information to develop an ASSC system based on knowledge of subjects' variability of some indicators that characterize sleep stages and on the American Academy of Sleep Medicine (AASM) rules.

Methods: An ASSC system consisting of a two-step classifier is proposed. In the first step, epochs are classified using support vector machines (SVMs) spread into different nodes of a decision tree. In the post-processing step, the epochs suspected of misclassification (dubious classification) are tagged, and a new classification is suggested. Identification and correction are based on the AASM rules, and on misclassifications most commonly found/reported in automatic sleep staging. Six electroencephalographic and two electrooculographic channels were used to classify wake, non-rapid eye movement (NREM) sleep – N1, N2 and N3, and rapid eye movement (REM) sleep.

Results: The proposed system was tested in a dataset of 14 clinical polysomnographic records of subjects suspected of apnea disorders. Wake and REM epochs not falling in the dubious range, are classified with accuracy levels compatible with the requirements for clinical applications. The suggested correction assigned to the epochs that are tagged as dubious enhances the global results of all sleep stages.

Conclusions: This approach provides reliable sleep staging results for non-dubious epochs.

1. Background

The manual sleep stage classification is a labor-intensive task that involves the interpretation, by an expert, of polysomnographic (PSG) signals captured from a subject's overnight sleep session. The PSG includes electroencephalographic (EEG), electrooculographic (EOG) and electromyographic (EMG) records, respiratory effort and other physiological characteristics while a patient is asleep [1,2].

Sleep stage classification is the first step in modern diagnosis of sleep disorders. The identification of REM (rapid eye movements) sleep, NREM sleep – N1, N2 and N3 stages (non-rapid eye movement) and wake stage is performed manually by experts based on the rules of the American Academy of Sleep Medicine (AASM) reproduced in Table 1 [5,6].

Many different methods for automatic sleep staging have been proposed. ASSC algorithms consist of: data pre-processing, feature extraction and classification. The features are extracted from PSG signals and then are used as input for a classifier that provides sleep scoring. Methods for feature extraction rely mainly on frequency domain techniques, such as the spectral power of frequency bands [7]. Notwithstanding, time domain analysis [8], and more recently time-frequency domain analysis, such as wavelet [9], have been also successfully used. A wide range of machine learning techniques were already tested including linear discriminate analysis (LDA) [10], artificial neural networks (ANN) [4,11], fuzzy logic [12], decision tree classification [13], hidden Markov models (HMM) [14], clustering approaches [15] and support vector machine (SVM) [16,17]. Table 2 summarizes relevant ASSC studies and respective classification techniques. As far as we know, just a few works [7,15,17] were validated in clinical datasets. Therefore, the assessment of ASSC systems in clinical context is still very incipient.

1.1. ASSC challenges

There is no consensus about the best features and the best classification models for ASSC [11,13]. However, it is commonly accepted that the existing automatic techniques are not accurate and reliable

* Corresponding author. Tel.: +351 239 796 201; fax: +351 239 406 672.
E-mail addresses: tsousa@isr.uc.pt (T. Sousa), anianabrito@isr.uc.pt (A. Cruz), skhalighi@isr.uc.pt (S. Khalighi), gpires@isr.uc.pt (G. Pires), urbano@deec.uc.pt (U. Nunes).

http://dx.doi.org/10.1016/j.compbiomed.2015.01.017
0010-4825/

Table 1
American Academy of Sleep Medicine rules for sleep scoring.

Sleep stage	AASM rules
Wake	A. Score epochs as stage Wake when more than 50% of the epoch has alpha rhythm over the occipital region. B. Score epochs without visually discernable alpha rhythm as stage Wake if any of the following are present: (1) Eye blinks at frequency of 0.5–2 Hz; (2) reading eye movements and (3) irregular conjugate rapid eye movements associated with normal or high chin muscle tone.
N1	A. In subjects who generate alpha rhythm, score stage N1 if alpha rhythm is attenuated and replaced by low amplitude, mixed frequency activity for more than 50% of the epoch. B. In subjects who do not generate alpha rhythm, score stage N1commencing with the earliest of any of the following phenomena: (1) Activity in range of 4–7 Hz with slowing of background frequencies by ≥ 1 Hz from those of stage W; (2) vertex sharp waves; and (3) slow eye movements.
N2	A. Begin scoring stage N2 (in an absence of criteria for N3) if 1 or both of the following occur during the first half of that epoch or the last half of the previous epoch: (a) one or more K complexes unassociated with arousals and (b) one or more trains of sleep spindles. B. Continue to score epochs with low amplitude, mixed frequency EEG activity without K complexes or sleep spindles as stage N2 if they are preceded by K complexes unassociated with arousals or sleep spindles. C. End stage N2 sleep when one of the following events occurs: (1) transition to stage W; (2) an arousal (change to stage N1 until a K complex unassociated with an arousal or a sleep spindle occurs); (3) a major body movement followed by slow eye movements and low amplitude mixed frequency EEG without non-arousal associated K complexes or sleep spindles (score the epoch following the major body movements as stage N1; score the epoch as stage N2 if there are no slow eye movements); (4) transition to stage N3; and (5) transition to stage REM.
N3	A. Score stage N3 when 20% or more of an epoch consists of slow wave activity, irrespective of age.
REM	A. Score stage REM sleep in epochs with all the following phenomena: (1) low amplitude, mixed frequency EEG; (2) low chin EMG tone; and (3) rapid eye movements. B. Continue to score stage REM sleep, even in the absence of rapid eye movements, for epochs following one or more epochs of stage REM as defined in A, if the EEG continues to show low amplitude, mixed frequency activity without K complexes or sleep spindles and the chin EMG tone remains low. C. Stop scoring stage REM when one or more of the following occur: (1) there is a transition to stage Wake or N3; (2) an increase in chin EMG tone above the level of stage REM and criteria for stage N1 are met; (3) an arousal occurs followed by low amplitude, mixed frequency EEG and slow eye movements (score as stage N1; if no slow eye movements and chin EMG tone remains low, continue to score stage REM); (4) a major body movement followed by slow eye movements and low amplitude mixed frequency EEG without non-arousal associated K complexes or sleep spindles (score the epoch following the major body movement as stage N1; if no slow eye movements and the EMG tone remains low, continue to score as stage REM); and (5) one or more non-arousal associated K complexes or sleep spindles are present in the first half of the epoch in the absence of rapid eye movements(score as stage N2).

Table 2
Relevant ASSC studies. MLP: multilayer perceptron, MT: movement time, PS: paradoxical sleep, SAS: sleep apnea syndrome, SWS: slow wave sleep.

Text reference	Tested dataset	Subjects age (average)	Classification method	Classified sleep stages	Agreement *(%)
[4]	13 healthy subjects	33	MLP	Wake, N1, N2, N3, REM, MT	34; 43; 51; 82; 82; 13
[7]	4 with different pathologies from a total of 12	42.3	Clustering	Wake, N1, N2, N3, N4, REM	84; 39; 24; 81; 93; 73
[8]	47 healthy subjects	33	MLP	Wake, N1, N2, N3, PS	85; 65; 86; 93; 73
[9]	32 PSG records from MIT-BIH database	–	Regression trees	Wake, N1, N2, N3, N4, REM	93;46; 76; 58;86;77
[10]	10 healthy subjects	–	LDA	Wake, N1, N2, N3, REM	61 (ambiguous sleep), 90 (epochs with high agreement between experts)
[11]	8 healthy subjects	28	ANN	Wake, N1, N2, N3, N4, REM	84; 31; 90; 29; 77; 82
[12]	4 healthy subjects	27.5	Genetic Fuzzy	Wake, shallow sleep, deep sleep, REM	86; 84; 84; 86
[13]	–	41.8	K-NN and Decision tree	Wake, N1, N2, N3, REM, MT	80; 7; 89; 65; 82
[14]	PhysioNetdatabase	28	Hidden Markov Models	Wake, N1, N2, N3, N4, REM	51; 5; 69; 64; 92; 86
[15]	15 subjects with different severity of SAS	42	Clustering	Wake, N1, N2, N3	60 (average)
[17]	28 subjects suspected of SAS	45.5	An ensemble of five binary SVM classifiers	Wake, N1, N2, SWS, REM	96; 96; 94; 96; 95
[20]	8 healthy subejcts	28	Recurrent neural classifier	Wake, N1, N2, SWS, REM	71; 37; 97; 90; 90
[22]	20 healthy subjects	52.5	Quadratic discriminant analysis	Wake, N1, N2, SWS, REM	86; 61; 75; 93; 86
[23]	8 healthy subjects	28	Multiclass least squares SVM	Wake, N1, N2, SWS, REM	88; 76; 97; 92; 93
[25]	20 healthy subjects	21.2	Heuristic Rules	Wake, N1, N2, N3, REM	88; 35; 87; 91; 91

* The performance values here presented are the measures described in the articles. The statistical performance measures used varies from study to study.

enough to be routinely used [11]. State-of-the-art shows that sleep epochs of healthy subjects, free of sleep stage transitions and without scoring disagreement between experts, can be automatically classified using a low number of features. However, in ambiguous epochs (e.g., epochs with sleep-stage transitions) the agreement level between the results of ASSC systems and human experts is only around 60% [10].

The comparison between the studies involving patients and studies involving healthy subjects show that PSG records of patients are usually more contaminated with artifacts, have a