Expert Systems With Applications 114 (2018) 414–427

Contents lists available at ScienceDirect

Expert Systems With Applications

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

Knowledge-based decision system for automatic sleep staging using symbolic fusion in a turing machine-like decision process formalizing the sleep medicine guidelines

Adrien Ugon[a,b,1,*], Amina Kotti[b,2], Brigitte Séroussi[c,d,1,2], Karima Sedki[c,1], Jacques Bouaud[c,f,1], Jean-Gabriel Ganascia[b,1], Patrick Garda[b,1], Carole Philippe[e,1,2], Andrea Pinna[b,1]

[a] ESIEE-Paris, Cité Descartes BP99, Noisy-le-Grand, 93162, France
[b] Sorbonne Université, CNRS, Laboratoire d'Informatique de Paris 6, Paris, F-75005, France
[c] Sorbonne Université, INSERM, Sorbonne Paris Cité, Université Paris 13, LIMICS, UMR S 1142, Paris, France
[d] Assistance Publique-Hôpitaux de Paris (AP-HP), Hôpitaux Universitaires Est Parisien, Département de Santé Publique, Paris, France
[e] Assistance Publique-Hôpitaux de Paris (AP-HP), Groupe Hospitalier Pitié-Salpêtrière Charles Foix, Service de Pathologies du Sommeil (Département R3S), Paris, France
[f] Assistance Publique-Hôpitaux de Paris (AP-HP), Department of Clinical Research and Innovation, Paris, France

ARTICLE INFO

Article history:
Received 23 January 2018
Revised 6 June 2018
Accepted 9 July 2018

Keywords:
Sleep stages
Symbolic fusion
Knowledge based systems
Automated decision support systems
Turing machine
Polysomnography

ABSTRACT

Automatic sleep staging is challenging since several issues need to be addressed. Traditional approaches from literature do not satisfy medical experts since they do not reflect the cognitive process they perform when scoring polysomnographic curves. We propose a new approach that is based on the implementation of medical knowledge by symbolic fusion. Medical knowledge coming from the international clinical practice guidelines for sleep medicine is formalized as a five-layer framework dedicated to data abstraction in order to deliver local and global propositions and support the interpretation of polysomnographic curves. Firstly, features are extracted from raw curves. Then these features are combined to recognize sleep events in accordance with guidelines. Sleep events are then fused into the criteria required to recognize the different sleep stages. Sleep is not homogeneous through the night. The physiological events observed during the night follow a dynamic that needs to be included into an automatic sleep staging system. In order to take this into account, decision rules are selected and applied to recognize a sleep stage according to the current context. Thereby, transitions are considered with interest. In this paper, we propose to use a Turing Machine-like decision process to handle transitions. To interpret the local observations and properly score a given state, the previous state which has been stored in a specific register is used as a context. One of the advantages of following the principles of symbolic fusion is to benefit from the full traceability of the decision. Hence, it makes possible to discuss each final — or intermediate — decision with an expert and check for relevance. The preliminary results are encouraging since agreement rates provided between decisions taken by our automatic approach and human experts are similar to those measured between human experts (average agreement rate = 54.60% / average Cohen's kappa = 0.40) on a dataset of 131 full polysomnographic recordings.

1. Introduction

1.1. Sleep disorders

Sleep is a fundamental physiological activity that helps the body and the mind to recover different functions (e.g. energy conservation, nervous system recuperation, localized recuperative processes, emotional regulation) (Siegel, 2005). Sleep disorders, among which the sleep apnea syndrome (SAS), have many negative

* Corresponding author.
E-mail addresses: adrien.ugon@esiee.fr (A. Ugon), amina.kotti@hotmail.fr (A. Kotti), brigitte.seroussi@aphp.fr (B. Séroussi), karima.sedki@univ-paris13.fr (K. Sedki), jacques.bouaud@aphp.fr (J. Bouaud), Jean-Gabriel.Ganascia@lip6.fr (J.-G. Ganascia), Patrick.Garda@upmc.fr (P. Garda), carole.philippe@aphp.fr (C. Philippe), andrea.pinna@lip6.fr (A. Pinna).
[1] PhD
[2] MD

https://doi.org/10.1016/j.eswa.2018.07.023
0957-4174/

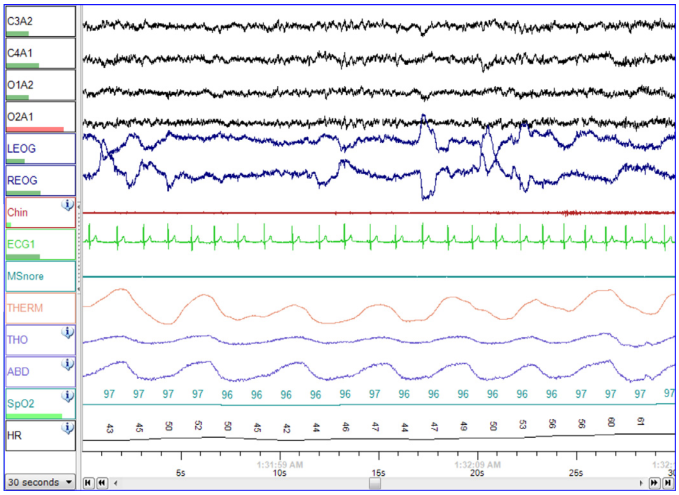

Fig. 1. Polysomnographic curves.

Table 1
Sleep stage degree of agreement (*kappa* measure) between scorers.

Sleep stage	Degree of agreement (kappa)
W	0.8608
N1	0.4608
N2	0.7188
N3	0.7285
R	0.9054

consequences on well-being and health. The prevalence of SAS is estimated between 3 and 7% (Marin, Carrizo, Vicente & Agusti, 2005). It is associated with an increased cardio- and cerebrovascular morbidity and mortality (Arzt, Young, Finn, Skatrud & Bradley, 2005; Punjabi, 2008) which makes SAS a significant public health problem.

The polysomnography is the gold standard diagnosis test for sleep apnea syndrome and for other sleep disorders (Kushida, Littner, Morgenthaler, Alessi, Bailey, Coleman, Friedman, Hirshkowitz, Kapen, Kramer, Lee-Chiong, Loube, Owens, Pancer & Wise, 2005). It consists in recording during a whole night different physiological parameters: brain activity, eye movements, muscle tone, heart activity, airflow, respiratory effort, pulse, blood saturation in oxygen, legs movements, ...(cf Fig. 1). After the acquisition of data, resulting curves need to be scored by a sleep expert according to clinical practice guidelines. This step is known to be tedious and time-consuming (Magalang, Chen, Cistulli, Fedson, Gislason, Hillman, Penzel, Tamisier, Tufik, Phillips & Pack, 2013). In addition, it is subject to a degree of subjectivity and inter- and intra-variability exists between scorers (Danker-Hopfe, Anderer, Zeitlhofer, Boeck, Dorn, Gruber, Heller, Loretz, Moser, Parapatics, Saletu, Schmidt & Dorffner, 2009; Kuna, Benca, Kushida, Walsh, Younes, Staley, Hanlon, Pack, Pien & Malhotra, 2013). Table 1 lists the degree of agreement between scorers for each sleep stage measured by H. Danker-Hopfe et al. (Danker-Hopfe et al., 2009). Although partially solved by the regular update of international guidelines for sleep medicine (Berry, Brooks, Gamaldo, Harding, Lloyd, Marcus & Vaughn, 2017), it is still acknowledged that scoring practices need to be improved (Penzel, Zhang & Fietze, 2013).

1.2. Clinical practice guidelines for sleep medicine

In 1968, Allan Rechtschaffen and Anthony Kales were the first authors to publish guidelines for sleep medicine (Kales & Rechtschaffen, 1968). These guidelines were followed during 40 years before the *American Academy of Sleep Medicine* published new guidelines in 2007 (Iber, Ancoli-Israel, Chesson & Quan, 2007), with regular updates. The current version is the version 2.4 and was published in April 2017 (Berry et al., 2017).

Sleep Staging is the first step of the scoring process of polysomnographic curves. It produces the hypnogram, i.e. the sequence of sleep stages throughout the night. This requires the night-long recording to be first split into 30-second segments called *epochs*. Then, on the basis of observation of electroencephalographic (EEG) for brain activity, electrooculographic (EOG) for eye movements, and submental electromyographic (EMG) curves for muscle tone, one out of the six sleep stages defined by the *American Academy of Sleep Medicine* (AASM) in international clinical practice guidelines for sleep medicine (Berry et al., 2017) should be assigned to each epoch: W (Wakefulness), N1 (NON-REM1), N2 (NON-REM2), N3 (NON-REM3), R (REM) and MOVEMENT; REM stands for *Rapid Eye Movement*. Usually, the MOVEMENT sleep stage is very rarely — or never — assigned and only the five other sleep stages are really used. It is obvious to note that there exists a hierarchical classification of these events. First, N1 and N2 can be grouped as *shallow sleep*. Then, N1, N2, and N3 can be grouped as *NON-REM sleep*. Finally, N1, N2, N3, and R can be grouped as *sleep*, in opposition to W (Wakefulness).

1.3. Objective of this work

In this work, we propose to extend the decision-framework defined by Belur Dasarathy by using five abstraction layers: *Data, Features, Events, Criteria*, and *Classes*. Rules applied to abstract data and information are all inspired from the medical guidelines for sleep medicine (Berry et al., 2017), which aim at standardizing the practice of sleep medicine. Thus, we should be consistent with the cognitive background sleep physicians use when scoring visually the polysomnographic recordings, independently of their experience.

Section 2 will present the state of the art. Section 3 gives the details of our method to integrate knowledge from the guidelines, from the Dasarathy to a five layer symbolic framework, integrating time-reasoning with a Turing-machine like approach. Results are given in Section 4. Discussion follows in Section 5. Finally, conclusion is made in Section 6.

2. State of the art: Existing automated sleep staging methods

In the early 1990s, the automation of computer-based sleep analysis became a topic of interest (Kemp, 1993). Since then, many researchers have worked on the automation of the scoring of sleep stages (Berthomier, Prado & Benoit, 1999). The common approach is a process that can be divided into five steps: (1) pre-processing of a single raw EEG signal to remove noise and artifacts, (2) segmentation of the signal into 30-second epochs, (3) extraction of features from the signals for each epoch, (4) reduction of the features space, (5) conception of a classifier using a learning step (6) evaluation of the classifier performance on an evaluation set.

Nevertheless, the automatic sleep stages scoring functionalities implemented in the different softwares installed in sleep laboratories are not satisfactory according to sleep physicians (Escourrou, Meslier, Raffestin, Clavel, Gomes, Hazouard, Paquereau, Simon & Orvoen Frija, 2010) and they prefer doing it visually rather than using the full automatic analysis or even using the automatic analysis as a pre-launched preliminary analysis without any further visual controls of the results. This non-adoption of existing systems by end-users proves that needs are unsatisfied and that there is room for new systems that could answer these needs and be routinely used.