



HAL
open science

Automatic Sleep Stages Classification Combining Semantic Representation and Dynamic Expert System

Adrien Ugon, Carole Philippe, Amina Kotti, Marie-Amélie Dalloz, Andrea Pinna

► **To cite this version:**

Adrien Ugon, Carole Philippe, Amina Kotti, Marie-Amélie Dalloz, Andrea Pinna. Automatic Sleep Stages Classification Combining Semantic Representation and Dynamic Expert System. MED-INFO 2019: Health and Wellbeing e-Networks for All, Aug 2019, Lyon, France. pp.848-852, 10.3233/SHTI190343 . hal-02314668

HAL Id: hal-02314668

<https://hal.science/hal-02314668>

Submitted on 13 Oct 2019

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Automatic Sleep Stages Classification Combining Semantic Representation and Dynamic Expert System

Adrien Ugon^{a,b}, Carole Philippe^c, Amina Kotti^{b,c}, Marie-Amélie Dalloz^{b,c}, Andrea Pinna^b

^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 Assistance Publique-Hôpitaux de Paris (AP-HP), Groupe Hospitalier Pitié-Salpêtrière Charles Foix, Service de Pathologies du Sommeil (Département R3S), Paris, France

Abstract

Interest in sleep has been growing in the last decades, considering its benefits for well-being, but also to diagnose sleep troubles. The gold standard to monitor sleep consists of recording the course of many physiological parameters during a whole night. The human interpretation of resulting curves is time consuming. We propose an automatic knowledge-based decision system to support sleep staging. This system handles temporal data, such as events, to combine and aggregate atomic data, so as to obtain high-abstraction-levels contextual decisions. The proposed system relies on a semantic representation of observations, and on contextual knowledge base obtained by formalizing clinical practice guidelines. Evaluated on a dataset composed of 131 full night polysomnographies, results are encouraging, but point out that further knowledge need to be integrated.

Keywords:

Sleep Stages, Expert Systems, Semantics

Introduction

In the last decades, sleep has been considered more and more seriously either for wellness reasons, or for diagnosing sleep troubles. Its impact on quality of life and on health is henceforth well known. To diagnose a sleep trouble, the gold standard sleep exam is the polysomnography, which consists in recording, during sleep, the course of a set of physiological signals and then observe the brain activity, recorded by electroencephalography (EEG), the eye movements, recorded using electrooculography (EOG), and the muscle tone recorded using electromyography (EMG).

Sleep staging is a fundamental preliminary step to the diagnosis of sleep troubles. During this task, a sleep expert visually browses the polysomnographic curves in 30-second epochs, to assign one of the five sleep stages defined, since 2007, by the American Academy of Sleep Medicine in their international guidelines [1]. A sleep stage is assigned by considering different criteria observed on EEG, EOG, and EMG curves, and by considering the dynamics of sleep. Guidelines for visual scoring of sleep staging define, firstly, for each sleep stage, a set of criteria that need to be met for a sleep stage to start to be scored ("transition rules"). Secondly, another set of criteria is defined for a sleep stage to continue to be scored ("continuity rules"). Five sleep stages are defined: W (Wakefulness), N1 (Non-REM 1), N2 (Non-REM 2), N3 (Non-REM3) and R (REM). N1 and N2 correspond to light sleep; N3 corresponds to deep sleep; REM stands for Rapid Eye Movements. Even if software dedicated to sleep scoring include automatic scoring

functionalities, they are not used in routine practice, since sleep physicians are not satisfied by their results [2].

Common approaches found in literature for automatic sleep stages are based on machine learning techniques. Features, extracted from the acquired signals, are used to feed a classifier, that will be able, after a learning step, to make a decision on new samples. Many of these approaches use a single channel – generally a single EEG channel – to make the decision. Using an open dataset, they might not be compliant with the current guidelines, but with older ones, defined by Retschaffen & Kales in 1968 [3]. Machine learning approaches mainly ignore the domain of knowledge; moreover, sleep dynamics are insufficiently integrated into the decision, since epochs are considered independently from each other (through the segmentation of the analysis and decision).

Alternatively, semantic approaches are seen as satisfactory solutions to have a formal description of knowledge on a domain without ambiguity. Based on a formal modelisation of concepts and relationships of the field, they allow reasoning. This formalization results in semantic networks, conceptual graphs or, more often, in ontologies. They might be seen as a description of a state of the universe at one time, focusing on one interest domain, listing concepts and specifying relationships between them. However, the dynamic aspects that make the universe change from one state to another remain an issue to be modelised. From one time to another time, the world changes; thus concepts and relationships between them change also. The description of the universe at one time is different to the description of the universe at another time.

Expert systems are decision support systems composed of an observations base and a knowledge base. The observations base contains facts and events observed in the current state of the universe. Knowledge base contains inference rules allowing the system to make sense or assemble existing observations and facts and generate new events or new facts that will then be added to the observations base. Expert systems are knowledge-based systems and can be designed without any data. In the medical field, we benefit from existing knowledge that needs to be formalized into compliant inference rules.

In this paper, we present a dynamic expert system, using a formal representation of observations using conceptual graphs and formalizing rules governing sleep-stages transitions and sleep events occurrences. The next section will focus on methods. Then, results will be presented and be followed by a discussion. Finally, we will give our conclusions in the last section.

Methods

An expert system with four modules

Our expert system is composed of four different modules: (1) the *events vector*, a time vector containing all events observed during sleep; (2) the *sleep stages vector*, a time vector containing all sleep stages assigned to each epoch during sleep; (3) the *events fusion knowledge base*, a knowledge base containing fusion strategies to fuse events (4) *five contextual sleep stage assignment knowledge bases*, five knowledge bases containing inference rules to apply in five different sleep contexts.

All these modules can be considered as an expert system. The observation base is composed of the *events vector* and *sleep stages vector*, composed each of facts of, respectively, low and high abstraction level. The knowledge base is composed of the *events fusion knowledge base*, and of the *five contextual sleep stages assignment knowledge bases*.

Events vector

All events observed during sleep are gathered in a time vector, where each event is identified by an identifier and a label. The vector captures the semantic type of each event, given by the identifier, but also its start and end. If necessary, other useful information can be added to the event; for instance, the lowest value of the saturation can be specified and attached to a desaturation. Each event is then formalized as a conceptual graph, giving its label, its start, its end and all other information that might be useful for the final decision.

Each event is represented by a conceptual graph. Conceptual graphs are a formalism that was introduced by John F. Sowa in 1984 [4]. Semantic concepts are linked by labelled relationships. Fusion algorithms have been defined to combine several conceptual graphs that share concepts.

For example, Figure 1 shows an event vector with all events occurring during the epoch. Representations of event 1 and event 2 are given on Figure 2.



Figure 1 – Events vector

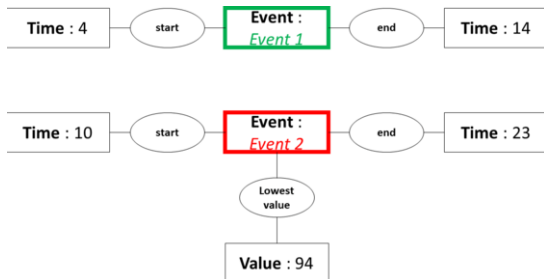


Figure 2- Conceptual graphs of Event 1 and Event 2

Sleep Stages vector

All sleep stages assigned by the expert system are stored in a sleep stages vector. Sleep stages are considered as high-abstraction-level information, obtained by combining and fusioning atomic information observed on other channels. This vector is initialized by specifying that, just before the start of the exam, i.e. the last epoch before the start of the exam, the patient was awake. Thus, we initialize the sleep stages vector

with the W sleep stage assigned to the last epoch preceding the start of the exam.

An example sleep stages vector is given in Figure 3. Each cell represents a 30-s epoch. The vector starts with seven epochs of N2, followed by 15 epochs of N3, followed by eight epochs of R, followed by 2 epochs of W and ends with 2 epochs of N1.

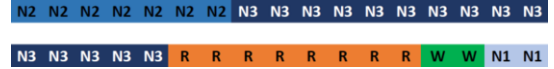


Figure 3 – Sleep stages vector

Events fusion knowledge base

Considered separately, events observed on each acquired channel may not be sufficient to assign a sleep stage. However, they may be different observations of the same physiological event. To be combined, it is hence necessary to fuse conceptual graphs to get a new conceptual graph representing a higher-abstraction-level event. Initially, they are separated; it is necessary to formalize the causal relationships that link them. Abstraction rules, combining all linked events to generate the physiological fact, can be applied to add the physiological fact to the observation base (the events vector).

Each rule of the events fusion knowledge base is formalized as a conceptual graphs fusion strategy. Fusion strategies follow the principles of the maximal join operator. Criteria required to fuse two conceptual graphs are defined. The resulting conceptual graph is entirely defined, on the basis of the concepts – and their values – of the fused conceptual graphs.

Figure 4 illustrates an example of fusion rules using conceptual graphs. Event 1 and event 2 are fused into Event 12, starting at starting date of Event 1, and ending at ending date of Event 2. Criteria to fuse these two events could be defined on start dates, end dates, or on the lowest value of event 2.

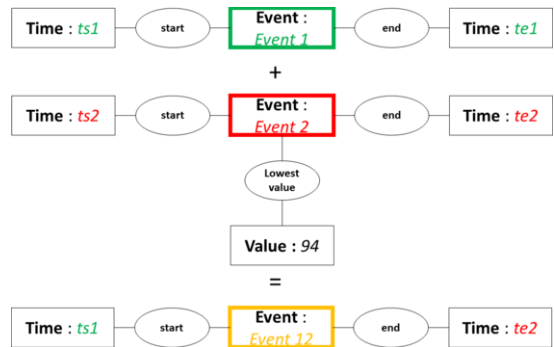


Figure 4 – Example of events fusion

Five contextual sleep-stage assignment knowledge bases

Depending on the current sleep stage in the previous epoch being considered as a sleep context, rules to assign a sleep stage to the current epoch are different, when applying the guidelines of the AASM. As a consequence, five different knowledge bases were defined, each containing the rules to apply in a given sleep stage. Depending on the current stage, each knowledge base contains rules to start a new sleep stage and rules to apply to continue scoring the current sleep stage. For each knowledge base, sleep stages are ordered. The inference step evaluates whether rules of a given sleep stage are met or not, testing all sleep stages successfully until one given sleep stage can be assigned. Ordering sleep stages allows to formalize priority rules, defined in the guidelines.

Rules to assign a sleep stage are based on binary criteria. Each criterion is met whether a high-abstraction-level event is met, or not.

Figure 5 illustrates an example of the five contextual knowledge bases. Depending on the sleep stage assigned to the previous epoch, a knowledge base is selected. Then, criteria are evaluated until one is met. It allows to assign a sleep stage to the current epoch.

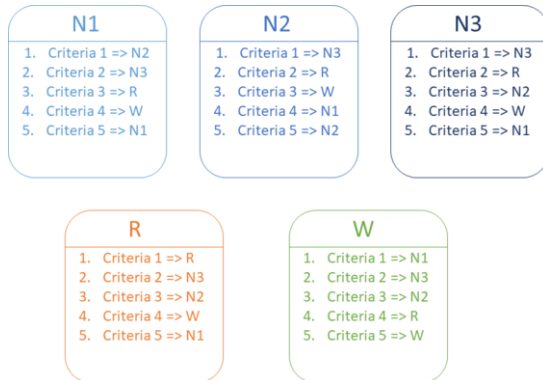


Figure 5 – Contextual Decision knowledge bases

Decision algorithm

Previously of the decision process, it is required that following steps have been done : (1) split the recording into 30-second epochs, (2) extraction and recognition of low-abstraction-level atomic events, (3) representation of all extracted events by conceptual graphs, consistently to a defined terminology.

Figures 6 shows a flowchart of decision process. It is composed of four steps.

1 - Move to the next epoch

A decision is made for each epoch of the recording, starting from the first to the last, moving from one epoch forward at each step, once a decision has been made and a sleep stage has been assigned to the epoch. At each step of the decision process, a “reading head” is moving to the next epoch.

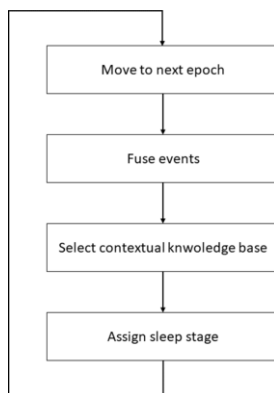


Figure 6– Flowchart of decision process

2 – Fuse events

We assume that events extracted previously to the decision process are low-abstraction-level events. To be able to make a decision, we need high-abstraction-level events. These events

are obtained by events fusion using fusion strategies [5,6]. Fusion strategies used to fuse events are defined and formalized into the *events fusion knowledge* base. Representing events by conceptual graphs allows to fuse delayed events observed in different channels, which is necessary in physiological processes. Using events fusion, it is possible to abstract events from a given abstraction level to a higher abstraction level; it is also possible to integrate literal data to get a semantic event; it is finally possible to aggregate, or combine, events. This step is based on an expert system.

3 – Select contextual knowledge base

Depending on the sleep stage assigned to the previous epoch, we know the sleep context of the subject at the start of the current epoch. Regarding this context, the right knowledge base can be chosen, in order to select the right rules to apply to make the decision and assign the sleep stage.

4 – Assign a sleep stage

Having identified which knowledge base to use, and regarding events of the low- and high- abstraction levels observed and inferred on the events vector, the system is able to assign a sleep stage to the current epoch. To achieve this task, ordered criteria are evaluated. Each criterion is defined as a logical combination of high-abstraction-level events observed. Rules are evaluated following a predefined order until a criterion is met. This allows to assign a sleep stage. Criteria are ordered according to priority rules.

Example

According to the guidelines, the N2 sleep stage should be assigned to an epoch, if a K-Complex, not associated to an arousal, is the first half of the current epoch.

Firstly, K-complexes and arousals need to be scored (manually or automatically). Then, K-complexes and arousals are fused, to classify K-Complexes into 2 categories: K-complexes associated to an arousal, and K-Complexes not associated to an arousal. In the next step, given an epoch, the right time period is observed to check whether a “K-complex non associated to an arousal” was observed. For those epochs where this criterion is met, the N2 sleep stage is assigned. Otherwise, the criteria of other sleep stages are tested.

Results

Description of the dataset

The dataset used for evaluating our method was already used in [7]; Gathering 131 full night polysomnographies, it is composed of 148,407 epochs (1,237 hours). The first subset is composed of 101 polysomnographic recordings of patients suspected of suffering from Sleep Apnea Syndrome; the second subset is composed of polysomnographic recordings of ten control subjects having spent three consecutive nights at the hospital. Polysomnographic recordings were performed in the sleep pathologies unit of La Pitié-Salpêtrière Hospital (AP-HP, Paris, France) using the Graef HD- PSG TM device which is produced by Compumedics Limited® from Australia. All recordings have been fully scored by a sleep expert in accordance with AASM guidelines [1] using Profusion Sleep TM Software from Compumedics®. All 131 recordings were visually scored by two experienced sleep experts, in accordance with the clinical practice guidelines defined by the AASM, without using the automatic pre-analysis functionality. This scoring is used as gold-standard for our work.

Each recording is composed of more than 30 channels for the full-night polysomnography. In this study, we only need some

channels to be compliant with the clinical practice guidelines defined by the AASM: the EEG, the EOG and the submental EMG. Six EEG channels were recorded, following the international 10–20 system with a sampling rate of 256Hz: O1, C3, C4, Fp1, A1, A2. 2 EOG were recorded with a sampling rate of 256Hz: Left EOG and Right EOG; One channel of the submental EMG was recorded with a sampling rate of 256Hz. As recommended in the clinical practice guidelines defined by the AASM, EEG and EOG channels were filtered using a 0.3–35 Hz bandpass filter; EMG submental channel was filtered using a 10–100 Hz bandpass filter.

This study was approved by The Committee for the Protection of Human Research Participants, Paris VI (Comité de Protection des Personnes Ile-de-France VI, Paris, France).

Control subjects were received by a physician to be informed in detail about the purpose and the procedure of the research study. Informed consent was provided and signed by all subjects. A sleep physician performed a physical examination and a questionnaire to validate that they had no suspicion of a sleep disorder.

All subjects signed an informed consent form. All individuals included in our database followed instructions to refrain from alcohol and caffeine ingestion and to avoid engaging in prolonged and/or strenuous exercise before sleeping. All polysomnographies were recorded in a quiet, darkened room.

Measures

To evaluate the performance of our automatic sleep slating tool, we used different measures.

Agreement rate is a measure used to assess the results of the automatic analysis global, i.e. including all sleep stages. It is defined as the ratio of epochs scored with the same sleep stage by the expert and the automatic analysis for each recording.

$$agreement\ rate = \frac{Nb\ epochs\ with\ same\ sleep\ stage}{Total\ number\ of\ epochs}$$

Because sleep stages have an unbalanced number of epochs, Cohen’s kappa was also used. This measure takes into account that some epochs are scored identically by the automatic analysis and by the expert by chance

$$\kappa = \frac{p_0 - p_e}{1 - p_e}$$

Where p_0 is the relative observed agreement among raters (identical to agreement rate); and p_e is the hypothetical probability of chance agreement.

Other measures are defined to assess results obtained for each sleep stage. These measures are defined for binary classifications. All of them are defined considering the following confusion matrix

Table 1 – Confusion Matrix

| | | Automatic analysis | |
|-----------------|-------|--------------------|-------|
| | | C | Not C |
| Expert analysis | C | TP | FN |
| | Not C | FP | TN |

TP refers to True Positive elements. Both expert and automatic analysis have classified these events in the class of interest C.

TN refers to True Negative elements. Both expert and automatic analysis classified these events as not belonging to the class of interest C.

FP refers to False Positive elements. Classified as belonging to the class of interest C by the automatic analysis, they were classified as not belonging by the expert.

FN refers to False Negative elements. These events were classified as not belonging to the class of interest by the automatic analysis, whereas they were classified as belonging by experts.

Recall and precision are widely used in the field of machine learning. Recall is the ratio of elements belonging to the class of interest, that were correctly identified by the automatic analysis. Precision is the ratio of elements belonging effectively to the class of interest among all elements that were considered as belonging to the class of interest by the automatic analysis. Recall and precision are defined by following formulas:

$$\left\{ \begin{array}{l} recall = \frac{TP}{TP + FN} \\ precision = \frac{TP}{TP + FP} \end{array} \right.$$

Sensitivity and specificity are used in the medical field to evaluate a diagnosis test. Sensitivity allows to know the performance of the test to identify sick individuals (=positive elements). It is equal to the predefined recall. Specificity allows to know the performance of the test to identify healthy individuals (=negative elements). Sensitivity and specificity are defined by following formulas:

$$\left\{ \begin{array}{l} sensitivity = \frac{TP}{TP + FN} \\ specificity = \frac{TN}{TN + FP} \end{array} \right.$$

To balance recall and precision, the F Measure was proposed, defined by the following formula:

$$F\ Measure = 2 \times \frac{recall \times precision}{recall + precision}$$

To give more weight to recall than to precision, the F2-Measure was proposed as an extension of the F-Measure, with the following formula:

$$F_2 - Measure = \frac{(2^2 + 1) \times recall \times precision}{2^2 \times precision + recall}$$

Results

On the dataset of 101 polysomnographic recordings of patients suspected to suffer from sleep apnea syndrome, the obtained average agreement rate was 51.5; the average kappa was 0.36. All measures for all sleep stages are detailed in Table 2. Considering the control subjects, we obtained an average agreement rate of 56.55 and an average κ of 0.43 (see Table 3).

Table 2 – Measures obtained on Sleep Apnea Syndrome patients dataset

| Measure | W | N1 | N2 | N3 | R | MVT |
|-----------------------------|-------|-------|-------|------|------|------|
| Recall / Sensitivity | 54.3 | 30.3 | 58.6 | 44.4 | 54.4 | 0.0 |
| Precision | 75.3 | 21.1 | 54.0 | 51.6 | 53.4 | 0.0 |
| Specificity | 92.64 | 86.26 | 73.86 | 91.4 | 91.6 | 99.5 |
| F Measure | 58.8 | 22.3 | 53.4 | 42.9 | 50.3 | 0.0 |
| F2 Measure | 54.5 | 25.3 | 55.8 | 42.5 | 51.6 | 0.0 |

Table 3 – Measures obtained on control subjects dataset

| Measure | W | N1 | N2 | N3 | R | MVT |
|----------------------|------|------|------|------|------|------|
| Recall / Sensitivity | 54.9 | 32.7 | 63.2 | 54.4 | 61.8 | 0.0 |
| Precision | 72.8 | 24.1 | 59.9 | 50.4 | 69.5 | 0.0 |
| Specificity | 92.7 | 86.4 | 74.9 | 86.1 | 91.6 | 96.7 |
| F Measure | 60.4 | 25.1 | 60.5 | 49.0 | 63.7 | 0 |
| F2 Measure | 56.6 | 28.1 | 61.9 | 51.4 | 62.2 | 0 |

As we can see, there are some differences between results obtained on the patients dataset and on the subjects dataset. This can be explained by the quality of the signals to process. Because of sleep troubles, signals from patients are very noisy, which make signal processing methods less efficient. Moreover, sleep troubles generated a highly disturbed sleep with many transitions; and as discussed by Thomas Penzel [8], transitions between sleep stages present the biggest source of differences between human sleep scorers.

Discussion

Our results may appear to be less performant than other results published in recent literature. All these works are mainly based on machine learning algorithms, and ignore, during the evaluation step, the intra- and inter-raters concordance. It was assessed by Danker-Hopfe et al. in 2009 [9]. Inter-raters rates are given by sleep stage in Table 4.

Table 4 – Inter-rater agreement rate (κ)

| W | R | N1 | N2 | z | N3 |
|--------|--------|--------|--------|---|--------|
| 0.8608 | 0.9054 | 0.4608 | 0.7188 | | 0.7285 |

As we see, most of results given in recent published works claim to have a higher concordance than what another human sleep expert would obtain. It means that their results are obtained by overfitting, which is not suitable.

On the contrary, our approach follows rules of practice guidelines. Our results show that these rules still need to be improved and expanded. Knowledge base might be enriched with transitions rules or experience knowledge.

The performance of identifying atomic events remains also to be evaluated.

Weaknesses of our approach include a reliance on the need to formalizing a complete knowledge base of rules to be applied to score sleep stages; medical knowledge is often incomplete, and knowledge acquired from experience is hard to formalize.

Strengths of this approach include its robustness, its flexibility, and its upgradeability. Its ability to use easy time-based rules helps contextualize decisions. Semantic reasoning is fully understandable and customizable by experts. Furthermore, in absence of formalized knowledge it can be hybridized with machine learning approaches.

Conclusion

In this paper, we present an approach representing information by semantic concepts. Temporal aspects are taken into account by reasoning on a dedicated framework, inspired by the Turing machine. Sleep stages are assigned by using different expert systems, one for each sleep context considered.

Acknowledgement

This work was performed within the Institut Universitaire d'Ingénierie en Santé de Sorbonne Université and the Labex SMART supported by French state funds managed by the ANR within the Investissements d'Avenir programme under reference ANR-11-IDEX-0 0 04-02.

References

- [1] B.R. Brooks, R. Gamaldo, C. Harding, R. Lloyd, R. Marcus, and C. Vaughn, The AASM Manual for the Scoring of Sleep and Associated Events, American Academy of Sleep Medicine, Darien, Illinois, 2017.
- [2] P. Escourrou, N. Meslier, B. Raffestin, R. Clavel, J. Gomes, E. Hazouard, J. Paquereau, I. Simon, and E. Orvoen Frija, Quelle approche clinique et quelle procédure diagnostique pour le SAHOS ?, *Revue des Maladies Respiratoires*. **27** (2010) S115–S123. doi:10.1016/S0761-8425(10)70017-6.
- [3] A. Rechtschaffen, and A. Kales, A Manual of Standardized Terminology, Techniques, and Scoring System for Sleep Stages of Human Subjects, US Government Printing Office, Washington Public Health Service, Washington DC, 1968.
- [4] J.-F. Sowa, Conceptual structures: Information processing in mind and machine, United States, 1982.
- [5] A. Ugon, J.-G. Ganascia, C. Philippe, H. Amiel, and P. Lévy, How to Use Symbolic Fusion to Support the Sleep Apnea Syndrome Diagnosis, in: M. Peleg, N. Lavrač, and C. Combi (Eds.), Artificial Intelligence in Medicine, Springer Berlin Heidelberg, Berlin, Heidelberg, 2011: pp. 45–54. doi:10.1007/978-3-642-22218-4_6.
- [6] A. Ugon, J.-G. Ganascia, C. Philippe, H. Amiel, and P. Lévy, Multi-layers Symbolic Fusion with conceptual graphs to support the Sleep Apnea Syndrome Diagnosis, in: Proceedings of Fusion 2013 Conference, Istanbul, Turkey, 2013.
- [7] A. Ugon, A. Kotti, B. Séroussi, K. Sedki, J. Bouaud, J.-G. Ganascia, P. Garda, C. Philippe, and A. Pinna, Knowledge-based decision system for automatic sleep staging using symbolic fusion in a turing machine-like decision process formalizing the sleep medicine guidelines, *Expert Systems with Applications*. **114** (2018) 414–427. doi:10.1016/j.eswa.2018.07.023.
- [8] T. Penzel, X. Zhang, and I. Fietze, Inter-scorer Reliability between Sleep Centers Can Teach Us What to Improve in the Scoring Rules, *JCSM*. **9** (n.d.) 89–91. doi:10.5664/jcsm.2352.
- [9] H. Danker-Hopfe, P. Anderer, J. Zeitlhofer, M. Boeck, H. Dorn, G. Gruber, E. Heller, E. Loretz, D. Moser, S. Parapatics, B. Saletu, A. Schmidt, and G. Dorffner, Interrater reliability for sleep scoring according to the Rechtschaffen & Kales and the new AASM standard, *Journal of Sleep Research*. **18** (2009) 74–84. doi:10.1111/j.1365-2869.2008.00700.x.

Address for correspondence

Dr Adrien UGON
 ESIEE-Paris
 Département Santé, Energie, Environnement
 2 boulevard Blaise Pascal
 BP 99
 93162 Noisy-le-Grand Cedex
 France
 Tel : +33 (0)1.45.92.65.36
 Mail: adrien.ugon@esiee.fr