Continuous unsupervised sleep staging based on a single EEG signal

Arthur Flexer, Georg Gruber, Georg Dorffner

The Austrian Research Institute for Artificial Intelligence
Schottengasse 3, A-1010 Vienna, Austria
arthur@ai.univie.ac.at

Abstract. We report improvements on automatic continuous sleep staging using Hidden Markov Models (HMM). Our totally unsupervised approach detects the cornerstones of human sleep (wakefulness, deep and rem sleep) with around 80% accuracy based on data from a single EEG channel. Contrary to our previous efforts we trained the HMM on data from a single sleep lab instead of generalizing to data from diverse sleep labs. This solved our previous problem of detecting rem sleep.

1 Introduction

Sleep staging is one of the most important steps in sleep analysis and is usually done using the traditional Rechtschaffen & Kales [7] (R&K) rules. It is a very time consuming task consisting of classifying all 30 second pieces of an approximately 8 hour recording into one of six sleep stages (wake, rem (rapid eye movement) sleep, S1, S2, S3, S4 (deep sleep)). A sleep recording is made with a minimum setting of four channels: electro-encephalogram (EEG) from electrodes C3 and C4, electro-myogram (EMG) and electro-oculogram (EOG). In order to classify each 30 second segment of sleep, the human scorer looks for defined patterns of waveforms in the EEG, for rapid eye movements in the EOG and for EMG level.

There is however a considerable dissatisfaction within the sleep research community concerning the very basics of R&K sleep staging: R&K is based on a predefined set of rules leaving much room for subjective interpretation; it is a very time consuming and tedious task; it is designed for young normal subjects only; it has a low 30 second temporal resolution; it is defined in terms of six stages neglecting the micro-structure of sleep; it cannot be automized reliably due to the large inter-scorer variability and insufficient rules for staging.

Our aim is to built an automatic continuous sleep stager, based on probabilistic principles which overcomes the known drawbacks of traditional R&K sleep staging. In previous efforts [1] we tried to find a new description of human sleep which is based on the comparably unambiguous "extreme" cornerstones of traditional sleep staging rather than merely automating and replicating R&K sleep staging. We used a Hidden Markov Model (HMM) to produce three continuous probability traces P(wake), P(deep) and P(rem) with a one second resolution.

The newly obtained continuous sleep profiles were compared to traditional R&K scoring. The two "extreme" R&K stages "wake" and "deep" could be detected very satisfactory with an accuracy of above 80%. However, we had great problems discriminating rem sleep from wakefulness and stages 1, 2 and even 3. The mean accuracy for detection of rem sleep was as low as 26%.

This paper reports about some improvements in continuous sleep staging which allow us to detect all three cornerstones of human sleep with satisfactory accuracy (around 80%). Reviewing our old results lead us to the hypothesis that the problems we encountered so far might be due to problems in the data base. Concentrating on data from a single sleep lab solved the problem of detecting rem sleep.

2 Data

In our previous efforts [1] our data base consisted of nine whole night sleep recordings from a group of healthy adults (total sleep time = 70.5h, age ranges from 20 to 60, 5 females and 4 males). We used reflection coefficients and stochastic complexity computed for EEG channels C3 and C4 and a measure of EMG level (altogether five features) for analysis with an HMM. The nine recordings had been recorded in five different European sleep laboratories during the SIESTA project [3]. The results were as described in Sec. 1: satisfactory detection of wakefulness and deep sleep, but accuracy as low as 26% for rem sleep.

The SIESTA project resulted in yet another sleep stager [9], which is based on a semi-supervised approach using Gaussian kernels plus sensor fusion to fuse information from different channels. Without giving any further detail it should suffice to say that its outputs are again three continuous probability traces. This sleep stager has been evaluated on data from eight different sleep labs. Plotting the mean entropy of the probability traces per sleep lab reveals clear lab effects (see Fig. 1). High entropy values indicate that the probability traces show little variation, i.e. that all three of them stay around .33 and therefore contain only little information. Obviously, the sleep stager seems to work quite well for some of the sleep labs and quite bad for others.

To gather further evidence for this hypothesis, we decided to concentrate our analysis on data from a single sleep lab (the one with the best results according to the entropy plot). Our new data base consists of 40 whole night sleep recordings from a group of healthy adults (total sleep time = 326.4h, age ranges from 20 to 80, 22 females and 18 males). We use only EEG channel C3 for further analysis. Twenty recordings are used to train our automatic sleep stager (training set), twenty are set aside to evaluate it (test set). Both sets are matched for sex and age.

3 Methods

HMMs [6] allow analysis of non-stationary multi-variate time series by modeling, both, the probability density functions of locally stationary multi-variate data

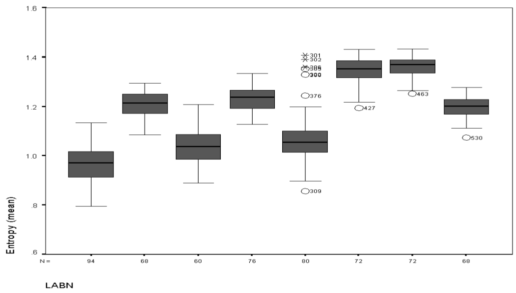

Fig. 1. Mean of the entropy of the probability traces given per sleep lab. Depicted are the means and the 25% and 75% percentile per sleep lab.

and the transition probabilities between these stable states. In the context of sleep analysis, the locally stable states can be thought of as sleep stages.

Following the classical text by Rabiner and Juang [6], an HMM can be characterized as (i) having a finite number of N states; (ii) a new state is entered based upon a transition probability distribution A which depends on the previous state (the Markovian property); (iii) after each transition an observation output symbol is produced according to a probability distribution B which depends on the current state. Although the classical HMM uses a set of discrete symbols as observation output, Rabiner and Juang [6] already discuss the extension to continuous observation symbols. Such a Gaussian Observation HMM (GOHMM) [5] has already been proposed as a model for EEG analysis and the model we now describe is the same we used for our previous work on sleep staging [1]. We use a GOHMM where the observation symbol probability distribution for state j is given by $B = \{b_j(x)\}, b_j(x) = \mathcal{N}[x, \mu_j, U_j]$, where $\mathcal{N}$ is the normal density and μ_j and U_j are the mean vector and covariance matrix associated with state j. Please note that this a simple version of the Gaussian M-component mixture given in [6] with M equal one. The Expectation-Maximization (EM) algorithm is used to train the GOHMM thereby estimating the parameter sets A and B as well as the μ_j and U_j. Viterbi decoding is used to identify most likely state sequences corresponding to a particular time series and enables the computation of the probabilities of being in any of the N states at each point in time. Full details of the algorithms can be found in [6].

A GOHMM is defined over the first reflection coefficient of the EEG channel at C3. Reflection coefficients are the coefficients of the order recursive representation of autoregressive (AR) processes [4]. We used a lattice filter representation of an AR process. The inferred a-posteriori distribution over model coefficients are the reflection coefficients (see [9] for full detail). The reflection coefficient is computed with a one second resolution.

Our aim is not to replicate R&K scoring but to find a new description of human sleep which is based on the comparably unambiguous “extreme” cornerstones of traditional sleep staging. Since R&K sleep staging is based on a predefined set of rules which leave much room for subjective interpretation there can be considerable disagreement between human scorers analyzing the same sleep recording. The three R&K sleep stages "wake", "deep" and "rapid eye movement (rem)" sleep are the sleep stages that can be detected most reliably by human scorers.

We therefore model the human sleep as a mixture of three different processes: wakefulness, deep sleep and rem sleep. The other three stages (S1, S2 and S3) can be seen as mixtures of the three basic processes. Consequently, we use a fully connected 3-state GOHMM to build our sleep stager. The GOHMM is trained on all available data from the training set. The probabilities of being in any of the 3 states are computed at each point in time using the posterior state probabilities. Thereby we obtain 3 continuous probability plots (P(wake),P(deep),P(rem)) which indicate the amount of wakefulness, rem and deep sleep with a one second resolution[1].

Although it is not the purpose of our approach to replicate R&K sleep staging, we nevertheless like to compare our results to the R&K standard. We construct a classifier for 30sec sections as suggested in [8] by computing mean values of P(wake), P(rem) and P(deep) for each of the six human scored R&K stages of all the recordings in the training set. For classification of recordings from the test set we find the minimum Euclidean distance between these mean values and the current probabilities.

4 Results

The trained GOHMM is evaluated using twenty whole night recordings from our test set. The newly obtained continuous sleep profiles are compared to traditional R&K scoring. R&K scores are taken as true scores and for each sleep stage separately the percentages of GOHMM classification into each of the 6 stages are given in Tab. 1. We expected that the GOHMM would be able to correctly classify data from the unambiguous "extreme" R&K stages "wake", "rem" and "deep". As can be seen in Tab. 1, this is indeed the case. The accuracies are 79% for wake, 82% for deep sleep and 68% for rem sleep. This is a clear improvement over our previous results which read as 86% (wake), 81% (deep) and 26% (rem). Probability plots plus R&K and HMM scoring for one whole night recording (subject from the test data group) are given in Fig. 2. The overall structure of sleep plus short periods of wakefulness are clearly visible in the probability plots. Note the respective high values of P(wake) and P(deep) aligned with R&K stages wake and S3 or S4 respectively. There still is some mix up between rem sleep and S1 and S2 at the end of the night.

Coming back to Tab. 1, we again note the mix up between S1, S2 and rem. Three possible explanations come to mind: (i) it is known that detection of rem is difficult from EEG alone; (ii) classification of S2 according to R&K is done based on very short events (spindles and K-complexes) and although the rest

[1] Please note that these probability plots are being smoothed with a moving average window. The length of the window is 361 seconds.

Table 1. R&K scores vs. GOHMM classification; GOHMM classification is given in percentages, separately for each sleep stage. Data is first reflection coefficient at C3.

		GOHMM					
		wake	S 1	S 2	S 3	deep	REM
R & K	wake	**79**	10	4	0	0	7
	S 1	21	**24**	19	4	1	31
	S 2	3	8	**36**	16	8	29
	S 3	0	0	11	**35**	54	0
	deep	0	0	2	16	**82**	0
	REM	14	13	4	0	1	**68**

of a 30 second segment might look like S1, S3 or rem, it is nevertheless judged as S2; (iii) S2 has already been described as a "compound" state not easily discriminable from other states [8]. As for the other sleep stages, S3 is mainly mixed up with deep sleep, which is as expected.

Additional experiments using more than a single feature (first reflection coefficient at C3) did either not change the results (including the second reflection coefficient at C3) or even worsen them considerably (including a feature of EMG level or EOG activity). Including data from C4 does not seem to make any sense since it is highly correlated with data from C3 anyway. The same holds for stochastic temporal complexity compared to the first reflection coefficient.

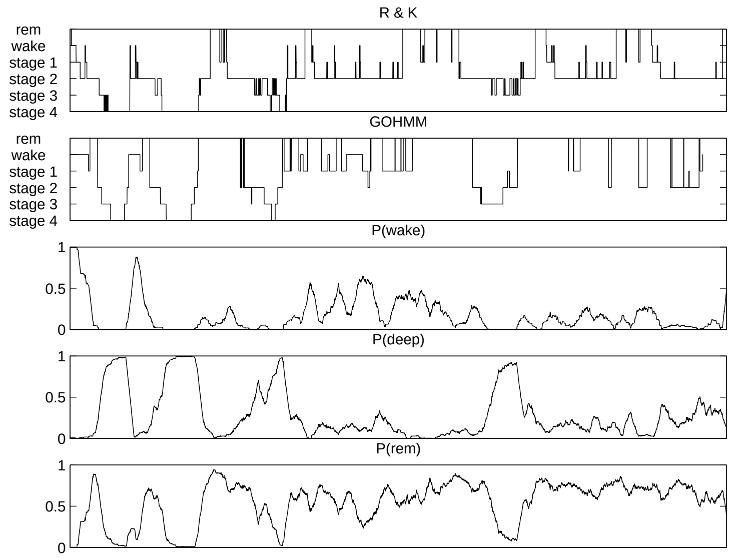

Fig. 2. Whole night results for one test subject; from top to bottom: R&K scoring, GOHMM scoring, P(wake), P(deep), P(rem).

5 Discussion

We presented an approach towards automatic sleep staging that goes beyond mere replication of the traditional R&K standard and offers a new continuous description of human sleep which is based on probabilistic principles. It is therefore in line with previous recommendations [2] and work [8] on continuous sleep staging. The approach is based on Hidden Markov Models, is totally unsupervised and uses only a single channel of EEG. It is superior in performance compared to our own previous results [1]. The output in the form of three continuous probability traces captures the three main processes in human sleep: wakefulness, deep sleep and rem sleep.

The improvement in performance has been made possible by realizing that there exist clear lab effects in our data base of sleep recordings. Although a lot of effort had been put into harmonization of sleep labs and recording protocol, there seem to be differences in hardware and maybe also in filter settings which are still visible in EEG and other signals.

Acknowledgements: The recordings for this work were done within the BIOMED-2 BMH4-CT97-2040 project SIESTA, funded by the EC DG XII. The Austrian Research Institute for Artificial Intelligence is supported by the Austrian Federal Ministry of Education, Science and Culture.

References

1. Flexer A., Sykacek P., Rezek I., Dorffner G.: An automatic, continuous and probabilistic sleep stager based on a Hidden Markov Model, to appear in: Applied Artificial Intelligence, 16(3), 2001.
2. Kemp B.: A proposal for computer-based sleep/wake analysis, Journal of Sleep Research, 2, 179-185, 1993.
3. Kloesch G., Kemp B., Penzel T., Schloegl A., Rappelsberger P., Trenker E., Gruber G., Zeitlhofer J., Saletu B., Herrmann W.M., Himanen S.-L., Kunz D., Barbanoj M., Roeschke J., Vaerri A., Dorffner G.: The SIESTA Project Polygraphic and Clinical Database, IEEE Eng. in Medicine & Biology Magazine, 20(3)51-57, 2001.
4. Ljung L.: System Identification, Theory for the User, Prentice-Hall, Englewood Cliffs, New Jersey, 1999.
5. Penny W.D., Roberts S.J.: Gaussian Observation Hidden Markov Models for EEG analysis, Technical Report, Imperial College, London, TR-98-12, 1998.
6. Rabiner L.R., Juang B.H.: An Introduction To Hidden Markov Models, IEEE ASSP Magazine, 3(1):4-16, 1986.
7. Rechtschaffen A., Kales A.: A Manual of Standardized Terminology, Techniques and Scoring System for Sleep Stages of Human Subjects, U.S. Dept. Health, Education and Welfare, National Institute of Health Publ. No.204, Washington, 1968.
8. Roberts S., Tarassenko L.: New Method of Automated Sleep Quantification, Medical and Biological Engineering and Computing, (5), 509-517, 1992.
9. Sykacek P., Roberts S., Rezek I., Flexer A., Dorffner G.: A Probabilistic Approach to High-Resolution Sleep Analysis, in Dorffner G. et al. (eds.), Artificial Neural Networks - ICANN 2001, International Conference, Vienna, Austria, Lecture Notes In Computer Science 2130, Springer, pp. 617-624, 2001.