An automatic, continuous and probabilistic sleep stager based on a Hidden Markov Model

A. Flexer*, P. Sykacek**, I. Rezek**, G. Dorffner*

*The Austrian Research Institute for Artificial Intelligence
Schottengasse 3, A-1010 Vienna, Austria
**Robotics Research Group, Department of Engineering Science
University of Oxford, Oxford OX1 6PJ, UK

arthur@ai.univie.ac.at

Abstract

We report about an automatic continuous sleep stager which is based on probabilistic principles employing Hidden Markov Models (HMM). Our sleep stager offers the advantage of being objective by not relying on human scorers, having much finer temporal resolution (one second instead of 30 seconds), and being based on solid probabilistic principles rather than a predefined set of rules (Rechtschaffen & Kales). Results obtained for nine whole night sleep recordings are reported.

1 Introduction

Our aim is to built an automatic continuous sleep stager, based on probabilistic principles which overcomes the known drawbacks of traditional Rechtschaffen & Kales [15] (R&K) sleep staging. Sleep staging is one of the most important steps in sleep analysis. 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 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. Quite some work has already been done in trying to replicate R&K sleep staging with automatic methods (see [2] and [12] for overviews) including neural networks (e.g. [13] and [18]).

There is however a considerable dissatisfaction within the sleep research community concerning the very basics of R&K sleep staging [12]: 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 therefore 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. We use a Hidden Markov

Model (HMM) initialized with information from the unambiguous R&K sleep stages "wake", "deep" and "rem" to produce three continuous probabilities P(wake), P(deep) and P(rem) with a one second resolution. Our sleep stager overcomes the known drawbacks of traditional sleep staging by being objective and automatic and having a 30-fold increased temporal resolution compared to R&K. Results for nine whole night recordings are compared to R&K standard. Problems with detection of REM sleep are pointed out and discussed.

2 Data

Data 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 use only channels C3, C4 and EMG for further analysis. The nine recordings have been recorded in five different European sleep laboratories during the EU-funded project SIESTA [5]. Five recordings are used to train our automatic sleep stager (training set), four are set aside to evaluate it (test set). Both sets are matched for sex and age. Some of the recordings in the test set have been made in sleep laboratories distinct from those in the training set.

3 Methods

HMMs [14] allow analysis of non-stationary multi-variate time series by modeling, both, the probability density functions of locally stationary multivariate data 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 [14], an HMM can be characterized as having a finite number of N states Q:

$$Q = \{q_1, q_2, \ldots, q_N\} \tag{1}$$

A new state q_j is entered based upon a transition probability distribution A which depends on the previous state (the Markovian property):

$$A = \{a_{ij}\}, a_{ij} = P(q_j(t+1) \mid q_i(t)) \tag{2}$$

where $t = 1, \ldots, T$ is a time index with T being the length of the observation sequence.

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 [14] already discuss the extension to continuous observation symbols. Such a Gaussian Observation HMM (GOHMM) [11] has already been proposed as a model for EEG analysis. We use a GOHMM where the observation symbol probability distribution for state j is given by

$$B = \{b_j(x)\}, b_j(x) = \mathcal{N}[x, \mu_j, U_j] \tag{3}$$

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 [14] with M equal one. The Expectation-Maximization (EM) algorithm [1] 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 [14].

A GOHMM is defined over the first reflection coefficients and stochastic temporal complexity measures, computed for EEG signals (electrodes C3 and C4), and a measure of EMG power.

Reflection coefficients are the coefficients of the order recursive representation of autoregressive (AR) processes [8]. We used a lattice filter representation of an AR process. The inferred a-posteriori distribution over model coefficients are the reflection coefficients (see [20] and [19] for full detail).

Stochastic temporal complexity [16] is computed using the method of delays [21] to construct embedding matrices for the temporal EEG signals. The numbers of significant singular values of these matrices obtained via singular value decomposition serve as measures of complexity by quantifying the temporal information content of the signals.

The EMG signals are reduced to frequencies between 20 and 45 Hz via FIR-filtering and absolute values of EMG are summed up for non-overlapping one second windows (following an approach given in [22]). These EMG measures are normalized for each subject by subtracting the lower 10% percentile and dividing through the interquartile range to minimize differences in EMG level between subjects. All EEG and EMG features are 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 analysing 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. Some studies on inter-scorer reliability report overall good agreement of around 90% for all sleep stages (see [7]). Others report high inter-scorer reliability of around 80% only for stages "wake", "deep" and "rem" and low reliability of 40% to 60% for all other R&K stages (S1, S2, S3) (see [3]). Previous related research [17] on continuous sleep staging also confirmed that only the three "extreme" R&K sleep stages are relatively unambiguous.

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. We use only data labelled as "extreme" R&K stages "wake", "rem" and "deep"to train a 3-state Gaussian Mixture Model which is needed to initialize the Gaussian kernels ($\mu_j, U_j; j = 1, \ldots, 3$ see Equ. 3) of the 3-state GOHMM. 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 probablities, i.e. the probability that an observation x_t came from state k given the observed sequence x:

$$P_t(wake) = P(q_t = k \mid x). \tag{4}$$

Computation of $P_t(\text{rem})$ and $P_t(\text{deep})$ follows the same principles. Thereby we obtain 3 continuous probability plots 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 as suggested in [17] by computing mean values of

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

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:

$$\bar{P}_i(wake) = \frac{1}{\sum_{t=1}^{T} \delta_{it}} \sum_{t=1}^{T} \delta_{it} \; P_t(wake) \tag{5}$$

$$\delta_{it} = 1 \;\; if \;\; R\&K(t) = Stage_i \;\; else \;\; \delta_{it} = 0 \tag{6}$$

where $R\&K(t)$ is the $R\&K$ sleep stage at time point t and $i = 1, \ldots, 6$ indicates the respective R&K sleep stage. Computation of $\bar{P}_i$(rem) and $\bar{P}_i$(deep) follows the same principles. We therefore compute six sets of mean values $\bar{P}_i$(wake), $\bar{P}_i$(rem) and $\bar{P}_i$(deep). For classification of recordings from the test set we find the minimum Euclidean distance between these mean values and the current probabilities:

$$\begin{aligned} GOHMM(t) = \min_i \quad & \Big[\Big(P_t(wake) - \bar{P}_i(wake)\Big)^2 + \Big(P_t(rem) - \bar{P}_i(rem)\Big)^2 + \\ & \Big(P_t(deep) - \bar{P}_i(deep)\Big)^2\Big]^{\frac{1}{2}} \end{aligned} \tag{7}$$

where $GOHMM(t)$ is the sleep stage as classified by the GOHMM at time point t and the minimum is taken over the $i = 1, \ldots, 6$ sets of mean values $\bar{P}_i(wake \mid rem \mid deep)$.

4 Results

The trained GOHMM is evaluated using 4 whole night recordings from our test set. The newly obtained continuous sleep profiles (using Equ. 7) 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" which we used during initialization. Whereas this is true for "wake" and "deep" (86% and 81%) it is not for "rem" (26%). Probability plots plus R&K and HMM scoring for one whole night recording (subject from the test data group) are given in Fig. 1. 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)

Table 1: R&K scores vs. GOHMM classification; GOHMM classification is given in percentages, separately for each sleep stage.

		GOHMM					
		wake	S 1	S 2	S 3	deep	REM
	wake	**86**	11	0	0	0	3
R	S 1	52	**22**	6	6	0	13
&	S 2	13	12	**14**	14	11	37
K	S 3	2	2	17	**20**	51	8
	deep	1	0	4	14	**81**	0
	REM	32	16	13	12	1	**26**

aligned with R&K stages wake and S3 or S4 respectively. There is substantial mix up between rem sleep and S2 at the end of the night. Whereas the R&K scoring indicates the subject being in stages S1, S2 and rem sleep, high values of P(rem) prevail during the last third of the recording.

We also evaluated a GOHMM with random initialization of the Gaussian kernels instead of initialization with data from the "extreme" R&K stages "wake", "rem" and "deep". Results in terms of agreement with R&K stages were only little worse than with proper initialization and the obtained kernels were very similar in terms of mean vectors μ and covariance matrices U. This further strengthens our belief that human sleep indeed is a mixture of three processes.

5 Discussion

We presented an approach towards automatic sleep staging that is innovative in two major respects. It 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 [4] and work [17] on continuous sleep staging. It applies the method of Hidden Markov Models to analysis of EEG which seems to be a natural choice given the multi-variate temporal nature of the data. However, very little work has been done on applying HMMs to EEG analysis until now (see [11], [6], [9] and [10]).

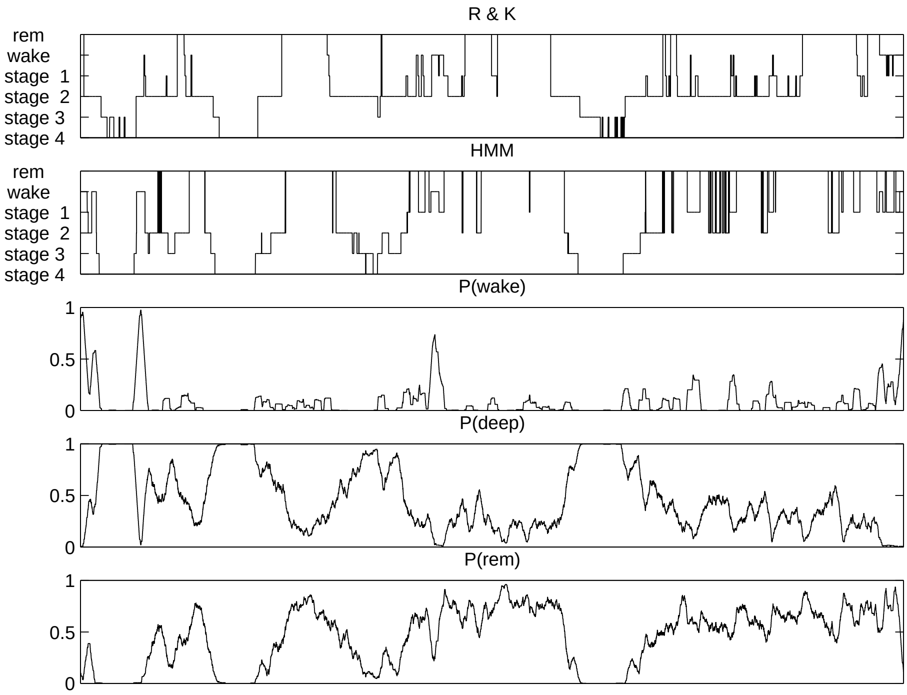

Figure 1: Whole night results for one test subject; from top to bottom: R&K scoring, GOHMM scoring, P(wake), P(deep), P(rem).

Our GOHMM has great problems discriminating rem sleep from wakefulness and stages 1, 2 and even 3. Whereas it is known that detection of rem is difficult from EEG alone, EMG should help in this respect. Close inspection of our EMG recordings reveals that discrimination even within subjects is very difficult. This seems to be due to a too coarse quantification resolution during recording which does not allow to detect the often very small drop in muscle tone in rem sleep. Histogram plots for all sleep stages for the EMG feature for all training data (five persons) are given in Fig. 2. Note the big overlap of EMG values for all sleep stages, but especially for rem sleep and S1, S2, S3 and S4.

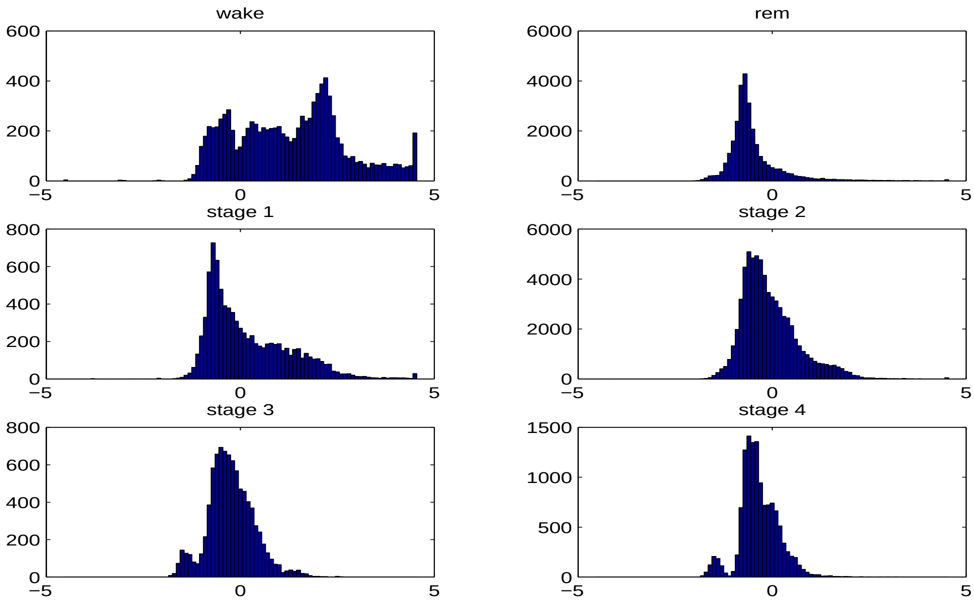

Figure 2: Histogram plots for the EMG feature; from top left to bottom right: wake, rem, stage 1, stage 2 ,stage 3, stage 4 (deep).

The two other "extreme" R&K stages "wake" and "deep" can be detected very satisfactory. There is only minor mix up between wake and S1, and deep and S3. S1 (light sleep) is mainly mixed up with wake, S2 with rem but also with all other stages. Both phenomena might in part be due to the EMG problem described above. S2 has already been described as a "compound" state not easily discriminable from other states (see [17]). S3 is mainly mixed up with deep sleep, which is as expected.

When judging the results concerning the comparison between R&K and GOHMM one should bear in mind that (i) our model is unsupervised and uses class information only partly during initialization, (ii) R&K scoring is done for 30sec sections and therefore not necessarily true for each single second of such epochs. Averaging the GOHMM class labels for each 30sec section would maybe allow fairer comparison. Another possible prroblem could be the fact that data has been recorded in five different sleep laboratories. Subtle differences in recording hardware might have an influence on comparability of sleep recordings. Last but not least it was not our intention to replicate R&K sleep staging but rather to develop a new description of human sleep which is only loosely based on traditional standards.

We conclude that, apart from our EMG problem, our results confirm manual scoring of the "extreme" states of sleep but do so automatically and at 30-fold increased time resolution. Future work will focus on getting around the EMG problem by including EOG, as well as on applying the new method to disturbed sleep in order to demonstrate the advantage of the increased temporal resolution.

Acknowledgements: This work was done within the BIOMED-2 BMH4-CT97-2040 project SIESTA, funded by the EC DG XII. Our HMM code is a modified version of code from Will Penny, Department of Engineering Science of the University of Oxford, which is a modified version of code from Zoubin Ghahramani, Gatsby Computational Neuroscience Unit. The Austrian Research Institute for Artificial Intelligence is supported by the Austrian Federal Ministry of Education, Science and Culture.

References

[1] Dempster A.P., Laird N.M., Rubin D.B.: Maximum likelihood from inclompete data via the EM algorithm, Journal of the Royal Statistical Society, B, 39:1-38, 1977.

[2] Hasan J.: Differentiation of normal and disturbed sleep by automatic analysis, Acta Physiologica Scandinavia, Supplementum 526, 1983.

[3] Kelley J.T., Reilly E.L., Overall J.E., Reed K.: Reliability of Rapid Clinical Staging of All Night Sleep EEG, Clinical Electroencephalography, Vol. 16, No. 1, 16-20, 1985.

[4] Kemp B.: A proposal for computer-based sleep/wake analysis, Journal of Sleep Research, 2, 179-185, 1993.

[5] 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 Engineering in Medicine and Biology Magazine, 20(3)51-57, 2001.

[6] Kohlmorgen J., Mueller K.-R., Pawelzik K.: Analysis of Drifting Dynamics with Neural Network Hidden Markov Models, in Jordan M.I., et al., Advances in Neural Information Processing Systems 10, MIT Press/Bradford Books, Cambridge/London, pp.735-741, 1998.

[7] Kubicki S., Hoeller L., Berg I., Pastelak-Price C., Dorow R.: Sleep EEG Evaluation: A Comparison of Results Obtained by Visual Scoring and Automatic Analysis with the Oxford Sleep Stager, Sleep, 12(2):140-149, 1989.

[8] Ljung L.: System Identification, Theory for the User, Prentice-Hall, Englewood Cliffs, New Jersey, 1999.

[9] Obermaier B., Guger C., Pfurtscheller G.: Hidden Markov Models used for the offline classification of EEG data, Biomedizinische Technik, Vol. 44(6), pp.158-162, 1999.

[10] Penny W.D., Everson R., Roberts S.J.: Hidden Markov Independent Components Analysis, in: Independent Components Analysis, (Ed M. Girolami),

Kluwer Academic Publishers, 2000.
[11] Penny W.D., Roberts S.J.: Gaussian Observation Hidden Markov Models for EEG analysis, Technical Report, Imperial College, London, TR-98-12, 1998.
[12] Penzel T., Stephan K., Kubicki S., Herrmann W.M.: Integrated Sleep Analysis, with emphasis on automatic methods, in Degen R., Rodin E.A. (eds): Epilepsy, Sleep and Sleep Deprivation, Elsevier, pp. 177-203, 1991.
[13] Principe J.C., Tome A.M.P.: Performance and Training Strategies in Feedforward Neural Networks: An Application to Sleep Scoring, in IEEE International Conference On Neural Networks, Washington D.C., IEEE, Volume I, pp.341-347, 1989.
[14] Rabiner L.R., Juang B.H.: An Introduction To Hidden Markov Models, IEEE ASSP Magazine, 3(1):4-16, 1986.
[15] 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.
[16] Rezek I.A., Roberts S.J.: Stochastic complexity measures for physiological signal analysis, IEEE Transactions on Biomedical Engineering, Vol. 44, No.9, 1998.
[17] Roberts S., Tarassenko L.: New Method of Automated Sleep Quantification, Medical and Biological Engineering and Computing, (5), 509-517, 1992.
[18] Schaltenbrand N., Lengelle R., Macher J.-P.: Neural Network Model: Application to Automatic Analysis of Human Sleep, Computers and Biomedical Research, 26, pp. 157-171, 1993.
[19] Sykacek P.: Bayesian Inference for Reliable Biomedical Signal Processing, Institut fuer Med.Kybernetik u. AI, Universitaet Wien, Dissertation, 2000.
[20] Sykacek P., Roberts S.J., Rezek I.A., Flexer A., Dorffner G.: Reliability in preprocessing - Bayes rules SIESTA, Medical and Biological Engineering and Computing, Supplement 2, Proceedings of EMBEC '99, p.1656-1657, 1999.
[21] Takens F.: Detecting strange attractors in turbulence, in Rand D. & Young L.S.(eds.), Dynamical Systems and Turbulence, Lecture Notes in Mathematics, Spinger, Vol. 898, pp.366-381, 1981.
[22] Vaerri A., Hirvonen K., Hasan J., Loula P., Haekkinen V.: A computerized analysis system for vigilance studies, Computer Methods and Programs in Biomedicine, Vol. 39, pp. 113-124, 1992.