Hidden Markov models to improve the performance of an artificial neural network in sleep stage scoring.

A.J. SERRANO, R. MAGDALENA, J.D. MARTÍN, M. BATALLER, M. MARTÍNEZ, E. SORIA.

Department of Electronics Engineering University of Valencia C/. Doctor Moliner, 50. 46100 Burjassot (Valencia) SPAIN

Abstract: - It is common to use classifiers on polisomnographic records in order to determine the different stages during sleep. Most of the times the results yielded by this systems are not coherent with physiological aspects of the sleep. This work uses the Hidden Markov Models as a modeller of the physiological act of sleeping, and uses it as a corrector of the classification yielded by an artificial neural network. It has been tested on polisomnographic records from the MIT database. Results confirm an improvement of $0,17\pm0,05$ in the Kappa coefficient of agreement and an improvement of $12,51\pm4,09\%$ in success during test set.

Key-Words: - Sleep stage scoring, neural networks, hidden markov models, clinical aid decision.

1 Introduction

The classification of sleep stages is an important issue in medical diagnosis because some serious diseases are accompanied by typical sleep disorders. physiological Different signals such as electroencephalogram (EEG), electrooculogram (EOG), electromyogram (EMG), etc., have to be recorded during night sleep [1]. These data are visually scored/annotated by an expert and the results displayed in the form of a hypnogram, sampled in intervals of 30 seconds

Sleep scoring and the construction of a hypnogram is a very laborious and time-consuming task: the paper from an 8 h EEG recording is over 400 m long [2]. It can require several hours of hard work by highly qualified specialists. Moreover, human expertise, which can vary between individual experts, is also mostly inconsistent over such periods of continuous work. Kemp et al. [3] report that the agreement between 6 human experts does not exceed 75%. Hence, a program capable of automatic sleep staging is highly desirable [4].

The sleep stage scoring (sleep classification) is based on the rules of Rechtschaffen and Kales (RK) [5] for adults. It is possible to distinguish three distinct conditions of consciousness. Conciousness can be classified into wakefulness and sleep. Sleep is subdivided into REM sleep and non REM sleep. This NREM sleep is again subdivided into stage 1 (S1), stage 2 (S2), stage 3 (S3) and stage 4 (S4). This sleep classification can be aided with mathematical models and usually classical or statistical methods are used [6]. However, reactions that occur in the human body are much more complex than those simulated by theoretic equations, which often do not match the underlying hypothesis.

These limitations can be alleviated by using artificial neural networks (ANNs); they are nonlinear regression models in which no previous knowledge of the problem is needed and it is not strictly necessary to assume any specific relationship between variables [7]. These features provide higher versatility and generalization performance than classical methods do. For the last years, a great amount of applications have been developed [6].

The multilayer perceptron (MLP) has been widely used in the last years. MLP is, in fact, a nonlinear generalization of the Logistic Regression [8] and achieves, in general, better results [6] An MLP is composed of a layered arrangement of artificial neurons in which each neuron of a given layer feeds all the neurons of the next layer. The first layer contains the input nodes, which are usually fullyconnected to hidden neurons and these, in turn, to the output layer [7].

The sleep classification should be, using whatever automatic classifier, as accurate as posible and, a key condition, be consistent with medical criteria. Markov processes are used to model human sleep structure [9]. Markov models are part of the mathematical procedures used for formal decision analysis. A Markov process is a system that consists of a finite number of mutually exclusive and collectively exhaustive states, meaning that at any given time each person must be in one of those states, and cannot be in more than one state at the same time[10]. The sleep process itself fulfils several properties that facilitate the application of Markov models for the modelling of human sleep. Markov is a state transition models (Fig. 1).



Fig. 1. Representation of a Markov Model in which every state represents a sleep stage. The sleep is divided into 30 second segments and each segment is classified into one of six mutually excluding states: Wake, S1, S2, S3, S4 and REM.

In this paper the use of Hidden Markov Model (HMM) to model the dynamics of sleep stages is proposed. This modelling will be used to verify the consistence of the ANN output with the medical model. This function is achieved due to the characteristic of the HMM to assign every state with an output value.

2 Data and preprocessing

Amongst the common biosignals that can be acquired from a patient, it is considered that EEG, EMG and EOG are those that contain more interesting information for a sleep analysis [11]. Therefore, it is very convenient to work with records that contain these three types of signals for sleep processing. Data used in this paper are polisomnographic records from the MIT-BIH database [12]. These records are the ones labelled SLP32, SLP37, SLP37, SLP37, which contain EEG, EMG and EOG, and besides signals of blood pressure, breath signal sensed with nose thermistor and breath effort signal sensed with pletimography on the chest.

Every record is annotated by experts with 7 annotations $S=\{W, 1, 2, 3, 4, R, MT\}$, corresponding respectively with the states of subject is awake, sleep stage 1, sleep stage 2, sleep stage 3, sleep stage 4, REM stage and movement time.

2.1 Feature selection

The human expert, when assigning a sleep stage to a data time interval, knows perfectly which are the feature to be observed before making the classification. In a similar way, the automatic classifier should observe a set of characteristics, which enable it to assign sleep stages to epochs. The result of this selection process is called feature vector.

There are many possible features to choose to implement a classifier of sleep stages. The following one have been selected for this work [4]:

EEG:

•Relations between the power of the bands d [0-4Hz], ? [4-8Hz], a [8-13Hz], ß1 [13-22Hz], ß2 [22-35Hz] and all the band [0-35Hz] in EEG.

- •Overall power in the band [0-35Hz] in EEG.
- •Relation between the power of the d and ? bands.
- •Relation between the power of the a and ? bands
- •Average frequency of the spectral density of EEG.
- •Standard deviation of the spectral density of EEG. EOG:

•Relations between the power of the band [0-4Hz] and all the band [0-30Hz] in EOG.

- •Average frequency of the spectral density of EOG.
- •Deviation of the spectral density of EOG. EMG:
- •Overall power in the band [10-45Hz] in EMG.
- •Average frequency of the spectral density of EMG.
- •Deviation of the spectral density of EMG.

The features related to EEG help to determine the predominant rates. Characteristics regarding EOG help to determine the ocular speed to detect REM (Rapid characteristic Eyes Moviment) stage. Characteristics regarding EMG detect if exists muscular activity, and thus distinguish between REM stage and wakefulness.

2.2 Preprocessing

The medical experts follow the RK classification criterion settled for 30 seconds epochs [5]; thus the data of the record to be classified by the automatic classifier must be of the same length:

1) Signal is split in parts of 2 seconds length for ensure stationarity.

2) Spectrum of EEG, ECG and EOG is calculated for every time interval.

3) Features formerly selected are obtained from the biosignal spectrum and they are average in order to obtain a feature vector representative of the 30 second length.

3 Method

3.1 Sleep stage detection

It has been decided to use an artificial neuronal networks, namely multilayer perceptron, as the classifier for the discrimination of the sleep stages, because its well-proved excellent performance [7]. In order to verify the performance of the classifiers the data set has been divided in sections of 20 samples; 2/3 of the sections have been selected for training, and the remaining third, has been used solely as test, not for cross-validation.

The layout of the ANN consists of 17 inputs neuron, 12 hidden neurons and 7 output neurons. Learning in the ANN has been carried out using the optimization algorithm of Levenberg-Marquardt [13]. Training has been stopped at epoch 100.

3.2 Hidden Markov Models.

A Markov chain contains the following elements:

•A set of states S={S1, S2, ..., SM}

•An M-by-M transition matrix T(i,j) whose i, j entry is the probability of a transition from state Si to state Sj.

•A set of possible outputs, or emissions, O={O1, O2,...,ON}

•An M-by-N emission matrix E(i,k) whose i,k entry gives the probability of emitting symbol Ok given that the model is in state Si.

When the model is in state Si, it emits an output Ok with probability E(i,k). The model then makes a transition to state Sj with probability T(i,j), and emits another symbol.

In this paper, 7 states have been defined, corresponding to the annotations of the experts $S=\{W, 1, 2, 3, 4, R, M\}$ in the database, and the ANN outputs are mapped to classes $O=\{W, 1, 2, 3, 4, R, M\}$. The transition matrix T contains the probabilities of stage changes in the patient records. Once the ANN has been fully trained, the emission matrix E reflects its performance, because it contains the probabilities of true and false classification for each class.

An estimation of the probabilities of stage transition has been used to figure out T, using the annotations of the four records in the database. The calculation of matrix E is easier, because it coincides with the ANN confusion matrix (normalized for every class).

Once a HMM is selected, it is possible to retrieve, by estimation, the most probable status given an output sequence. Therefore, the output sequence of the ANN during the test set can be associated to the most probable annotation sequence, that besides will fulfil the probability requisites required by the HMM. In this way it is guaranteed that the final annotation sequence in the file is coherent with the physiological act of sleep.

4 Results

The calculation of the probabilities between stage transition confirms the physiological restrictions of sleep. Table 1 shows that stage repetition (diagonal) and smooth transitions (adjacent stages) are quite usual.

Table 1. Transition probability between different stages in the 4 selected records. W; wake, R:REM sleep, MT: movement time (-<0,01)

	Subsequent Stage						
Previous Stage	w	1	2	3	4	R	МТ
W	0,92	0,07	0,02	-	-	-	-
1	0,15	0,52	0,33	-	-	-	-
2	0,03	0,07	0,85	0,04	0,01	0,01	0,00
3	0,01	0,01	0,46	0,39	0,11	-	0,02
4	-	-	0,10	0,14	0,74	-	0,01
R	0,02	0,03	0,01	-	-	0,95	-
MT	-	0,25	0,75	-	-	-	-

Table 2 shows that the ANN obtains good results (greater than 50%) in the most important stages (in bold). It is also valuable to note that arises high error probabilities (in italic) that would be desirable to avoid.

Table 2. Emission probability for the ANN trained for record SLP45 of the MIT database (-<0.01).

	Expert Annotations						
MLP	W	1	2	3	4	R	MT
W	0.93	-	0.02	0.01	0.02	0.03	-
1	0.20	<u>0.20</u>	<u>0.41</u>	-	-	0.17	0.02
2	0.02	0.01	0.91	0.01	0.02	0.03	0.01
3	0.04	-	<u>0.27</u>	0.65	0.04	-	-
4	0.02	-	0.08	0.27	0.63	-	-
R	0.06	0.01	0.02	-	-	0.89	0.01
MT	<u>0.25</u>	-	-	-	-	<u>0.25</u>	0.50
Number of Annotations	86	29	292	33	15	63	2

Once the sequence of classes produced by the ANN is obtained, the algorithm searches the compatible sequence of the most probable states with the hidden model of Markov defined by the calculated probabilities. This corrected sequence of classes improves the results in the test set and also improves the results in the total calculation of samples in 3 of the 4 records.

Table 3. Percentage of success of the stage detection by MLP and the correction by the HMM.

% Success	Comp	lete set	Test set		
	MLP	+HMM	MLP	+HMM	
SLP32	91.09	88.59	73.50	80.50	
SLP37	86.57	91.00	69.09	80.91	
SLP41	52.69	63.72	47.69	63.08	
SLP45	82.24	86.05	55.83	71.67	

In the case of record SLP41 the success improves a 15.39% in the test set and, in the worst case, also an improvement of 7% in record SLP32 is obtained. Considering the total of samples, it is also achieved an improvement of 11.03% in record SLP41 and a worsening, only of 2.5%, in record SLP32.

Table 4. Kappa coefficient in the stage detection by MLP and the correction by the HMM.

Agreement	Comp	olete set	Test set		
	MLP	+HMM	MLP	+HMM	
SLP32	0,83	0,78	0,54	0,65	
SLP37	0,73	0,82	0,37	0,59	
SLP41	0,22	0,38	0,24	0,38	
SLP45	0,73	0,79	0,40	0,61	

5 Conclusions

In the present work the use of the physiological restrictions that exist in any temporary biological process is proposed to improve the efficiency of the classification, in this case, of the sleep stages. The hidden models of Markov offer an excellent frame work to implement in an efficient way the dynamics of systems. The modelling confirms that the transitions between stages are smooth and the abrupt changes are not probable. This fact, together with the high rate of successes of the multilayer perceptron $(78,15\pm17,35\%)$ on the total set of samples allows a better estimation of the real sequence of sleep stages in the patients. Improvements of $0,17\pm0,05$ in the

Kappa coefficient and of $12,51\pm4,09\%$ in success in the test set have been achieved on 4 records of the MIT-BIH database. The improvements also reach to the training sequence, obtaining finally an increase of the Kappa index of $0,06\pm0,08$ and $4,19\pm5,53\%$ in success on the total set of samples.

Acknowledgments

This work has been partially funded by proyect UV-AE-2005 0968 from the University of Valencia and project GV06/248 de la Conselleria d'Empresa Universitat i Ciencia de la Generalitat Valencia.

References:

[1] T. Penzel, K. Stephan, S. Kubicki and W.M. Herrmann, Integrated sleep analysis, with emphasis on automatic methods, *in: R. Degen and E.A. Kodin, eds., Epilepsy, Sleep and Sleep Deprivation*, Elsevier, Amsterdam, 1999, pp.177-203.

[2] J.R. Smith, Automated analysis of sleep EEG data, *in: Handbook of Electroencephalography and Clinical Neurophyiology*, Vol. 2 of revised series, Elsevier, Amsterdam, 1986, pp. 131-147.

[3] B. Kemp, E.W. Gröneveld, A.J.M.W. Janssen and J.M. Franzen, A model-based monitor of human sleep stages, *Biol. Cybernet*. Vol. 57, 1987, pp.365-378.

[4] R. Agarwal, J. Gotman, Computer-Assisted Sleep Staging. *IEEE Transactions on Biomedical Engineering*, Vol. 48, No. 12, 2001.

[5] A. Rechtschaffen and A. Kales. A manual of standardised technology; techniques and scoring systems for sleep stages of human objects, *Public Health Service*, US Government Printing Office, Washington DC, 1986.

[6] P.J.G Lisboa, A review of evidence of health benefit from artificial neural networks in medical intervention. *Neural Networks*, Vol. 15, 2002, pp. 11-39.

[7] C. M. Bishop, *Neural Networks for Pattern Recognition*. Clarendon Press, Oxford, 1996.

[8] J.F. Hair, R.E. nderson, R.L. Tatham, W.C. Black, *Multivariate Data Analysis*. Upper Saddle River, NJ: Prentice-Hall, 1998.

[9] B. Kemp, H.A. Kamphuisen, *Simulation of human hypnograms using a Markov chain model*. Sleep, Vol. 9, No. 3, 1986, pp. 405-414.

[10] I. Koprinska et al., Sleep classification in infants by decision tree-based neural networks. *Artificial Intelligence in Medicine*, Vol. 8, 1996, pp. 387-401.

[11] N. Schaltenbrand, R. Legenlle, J. P. Mancher, Neural Network Model: Application to Automatic Analysis of Human Sleep, *Computers and Biomedical Research, Vol.* 26, 1993, pp. 157-171. [12] MIT-BIH Polysomnographyc DataBase (second edition).*Harvard-MIT Division of Health Sciences and Technology. Biomedical Engineering Center.*MIT room 20A-113, Cambridge, MA 021139 USA.
[13] M. T. Hagan, M. Menhaj, Training feedforward networks with the Marquardt algorithm, *IEEE Transactions on Neural Networks*, Vol. 5, No. 6, 1994, pp. 989-993.