KNOWLEDGE EXTRACTION ABOUT SLEEP/WAKE STAGES USING DATA DRIVEN METHODS

Lukáš Zoubek[1,2], Sylvie Charbonnier[2], Suzanne Lesecq[2], Alain Buguet[3], Florian Chapotot[4,5]

[1] *VSB-Technical University of Ostrava, Department of Measurement and Control, 17.listopadu 15/2172, Ostrava-Poruba, 708 33, Czech Republic*
[2] *Laboratoire d'Automatique de Grenoble, Ecole Nationale Supérieure d'Ingénieurs Electriciens de Grenoble, rue de la Piscine, BP 46, 38402 Saint Martin d'Hères Cedex, France*
[3] *Neurobiologie des états de vigilance, EA 3734,Université Claude-Bernard Lyon 1, Lyon, France*
[4] *Centre de Recherches du Service de Santé des Armées, 24 Avenue des Maquis du Grésivaudan, BP 87, 38702 La Tronche Cedex, France*
[5] *PhiTools - www.phitools.com, 1 rue du Général de Castelnau, Strasbourg, France*

Abstract: This paper describes the process of sleep/wake stage classification. It mainly focuses on the problem of selection of relevant features extracted from the polysomnographic recordings. Iterative features selection methods were applied on a large database composed of several night recordings from different healthy adults. The results showed that the use of relative power of EEG in five frequency bands enables to correctly classify 71% of the whole data base. The addition of features extracted from EOG enables to reach 75% of agreement with the expert classification, but no significant improvement was obtained when adding features extracted from EMG.

Keywords: decision making, diagnosis, medical applications, pattern recognition, signal processing

1. INTRODUCTION

Polysomnography consists in the analysis of electroencephalogram (EEG), electromyogram (EMG) and electroocular (EOG) signals to recognize the different sleep/wake stages: wake, NREM sleep stages 1 to 4 and Paradoxical Sleep. Recognition of sleep stages from physiological signals recorded during a night sleep enables to build a hypnogram, a temporal succession of sleep/wake stages. Hypnogram is a tool used to diagnose sleep disorders, which are rather common (about 5% of people suffer from a sleep disorder). Hypnogram is an overall representation of the sleep architecture. Until now, the analysis of a night sleep has been made visually by the physician, who scores every 20s of the recording (named epoch) into one of 6 sleep/wake stages (wake, NREM I, NREM II, NREM III, NREM IV, Paradoxical Sleep). This manual classification is a tedious and time-consuming task. Thanks to the development of computer technology, automated systems to built hypnograms have emerged, either using classical algorithms or artificial intelligence methods, such as neural networks

(Robert, *et al*., 1998; Oropesa, *et al*., 1999). Features used for classification are extracted from each epoch. They are obtained using signal processing techniques operating in the time domain or in the frequency domain. Studies are still in progress to improve the performance of these automatic sleep/wake classifiers. The aim of this paper is to show how data mining methods can be used to extract knowledge about sleep stages classification. Knowledge extraction is performed from a large database composed of 47 night sleep recordings from 41 healthy subjects. Feature selection methods are applied so as to determine which physiological signals are the most relevant, and to decide which techniques are the most suitable to process them, in order to perform accurate sleep/wake stage classification. EEG signal is processed in the frequency domain using Fourier Transform and Wavelet Transform. EEG, EOG and EMG signals are processed in the time domain using functions of the first moments of the signals, i.e. the standard deviation, the skewness and the kurtosis.

The outline of the paper is the following. The whole database and the features extracted are presented in the second section. Features selection methods and classifiers are described in section 3. Results obtained are presented and discussed in section 4.

2. MATERIALS

2.1 Presentation of the database

In this study, a large database of polysomnographic recordings has been used. The full database contains 47 whole night recordings from 41 healthy subjects (19 – 47 years old, 39 males and 2 females). Recordings have been made continuously during the whole night (8 hours). Four EEG channels (C3-A2, P3-A2, C4-A1, and P4-A1), one diagonal electro-oculogram (EOG) and one chin electromyogram (EMG) have been registered and digitized with the sampling frequency f_s = 128 Hz. The EEG signals were measured on the scalp according to the International 10-20 EEG System of Electrodes Placement.

All the 47 polysomnographic recordings were separately and visually classified by two experts. Visual classification was performed on each epoch of the whole night sleep, according to the classical sleep stage classification manual (Rechtschaffen and Kales, 1968). Each epoch was classified into one of 5 different stages (wake, NREM I, NREM II, NREM III&IV, and Paradoxical Sleep) defined for sleep stage classification. Only the epochs classified in the same stage by both experts have been considered in this project. They represent 84% of the whole recordings and form the data base. The first line of Table 1 presents the number of epochs classified in each sleep stage for the database. The total number of epochs is 63,254.

As it can be seen in Table 1, during a night sleep, the number of epochs classified in a stage is not the same for every stage. Stage NREM II lasts a long time, whereas stage NREM I is rather short. In this study, the database has been reduced to a smaller one where each class is composed of about the same number of epochs, so as to avoid errors in the classification results that could be induced by this difference in classes representation. The numbers of epochs classified in each sleep stage for the reduced database are presented in the second line of Table 1.

Table 1 Description of the database used in this study. Number of epochs in the sleep stages

	wake	NREM I	NREM II	NREM III&IV	REM
Full database	*5 232*	*1 989*	*32 966*	*7 701*	*15 366*
Test database	*1 914*	*1 879*	*2 206*	*1 902*	*2 099*

The database used in this study consists then of 10,000 randomly selected epochs, each classified into one of the five sleep stages by both experts. This set S of 10,000 classified epochs is then split in ten subsets $S = \{S_1, S_2,..., S_{10}\}$, each subset S_k containing 1,000 epochs, equally distributed in the five classes.

2.2 Features extracted from the physiological signals

Each epoch stored in the database consists of a 20 seconds recording of six signals (four EEG, one EOG and one EMG). During these 20 seconds, the physiological signals are assumed to be stationary. Since the signals were sampled at 128 Hz, each one of the 6 recorded time series contains 2,560 samples. Features are extracted on each epoch using the 6 signals. Features are information describing different characteristics of the signal. They are extracted using different signal processing techniques. The PRANA software was used to visualize the polysomnographic recordings and to extract the features.

EEG features

- Five features express the relative power of the EEG signal in given frequency bands. They are calculated using Fourier transformation. Total spectral power (P_{tot}) is computed in frequency band [0.5 ; 32.5] Hz.

- $P_{rel}(EEG,\delta_{FT})$ with δ_{FT} = [0.5 ; 4.5] Hz;
- $P_{rel}(EEG,\theta_{FT})$ with θ_{FT} = [4.5 ; 8.5] Hz;
- $P_{rel}(EEG,\alpha_{FT})$ with α_{FT} = [8.5 ; 11.5] Hz;
- $P_{rel}(EEG,\sigma_{FT})$ with σ_{FT} = [11.5 ; 15.5] Hz;
- $P_{rel}(EEG,\beta_{FT})$ with β_{FT} = [15.5 ; 32.5] Hz.

- Five features characterize the wavelet coefficients generated by discrete wavelet transformation. In this study, a 4-level wavelet packet decomposition of the EEG signal has been used to construct a set of