Computer Aided System for Analysis of Heart Rate Data Oriented to Development of Automatic Diagnostics of Sleep

Antanas Zilinskas, Julius Zilinskas, Audrius Varoneckas

Abstract: *A computer aided system is created to aid development of new medical diagnostics of sleep. A toolbox of data mining methods including standard and advanced methods (multidimensional scaling, cross-spectrum analysis method, and optimization based classification) is made available to the users. Preliminary testing shows that knowledge extracted from heart rate data can be efficiently used for classification of sleep stages.*

Key words: *Computer Systems, Data Mining, Bioinformatics, Medical Applications.*

INTRODUCTION

A problem of feature extraction from heart rate data related to the diagnosis of sleep disorders and diseases is considered. An important factor for diagnostics is sleep structure. A basic method of evaluation of sleep structure is polysomnography where sleep stages are identified from the brain activity, eye movements, muscle activity, heartbeat, blood oxygen levels and respiration data recorded during overnight stay in a sleep laboratory [4]. However, this method is rather complicated, time-consuming, and expensive. On the other hand, the medical doctors-experts in sleep problems can evaluate the sleep structure heuristically from heart rate data. Since recording of interbeat interval (RR) and stroke volume (SV) sequences is much cheaper procedure than polysomnography, recognition of sleep stages from the features extracted from RR and SV sequences would be very advantageous. Therefore we have created a computer aided system to aid research leading to development of automatic classification of sleep stages from heart rate data. Although oriented to a specific problem, knowledge extraction on sleep disorders, the described system can be easily tuned to other diagnostic problems based on analysis of data represented by time series.

AVAILABLE DATA

The quality of sleep is defined by the structure of sleep which consists of Rapid Eye Movement (REM) sleep and four stages (1, 2, 3, 4) of NonREM sleep [4]. During normal sleep, stages regularly alternates and the amount of time spent in each sleep stage is quite stable for a given age range. Changed sleep structure may indicate disturbed health. Since recording RR and SV sequences is much cheaper procedure than polysomnography, recognition of sleep stages from the features extracted from RR and SV sequences would be very advantageous. The aim of our project is the development of a system for elicitation of information on sleep stages from RR/SV sequences. The extracted information is formalized as feature parameters vectors used for recognition of sleep stages. The example of a hypnogram, i.e. the graph of sleep changes in time course, obtained by means of polysomnography is shown in Figure 1. In the same figure the graphs of RR and SV sequences are shown. Hypothetically the automatic and reliable reconstruction of the hypnogram from RR and SV sequences is possible.

Automatic classification methods are based on several mathematical models of RR/SV sequences. The adequacy of mathematical models should be proved by means of comparison with real records. Besides of methods of testing statistical hypotheses the heuristic approval by medical experts is important. Therefore methods of generation of artificial RR/SV sequences and their visual presentation as well as possibility of storing them in the database have been implemented. For the feature extraction further used for recognition of the sleep stages different data models are used; most important models are stationary time series and nonlinear dynamical systems.

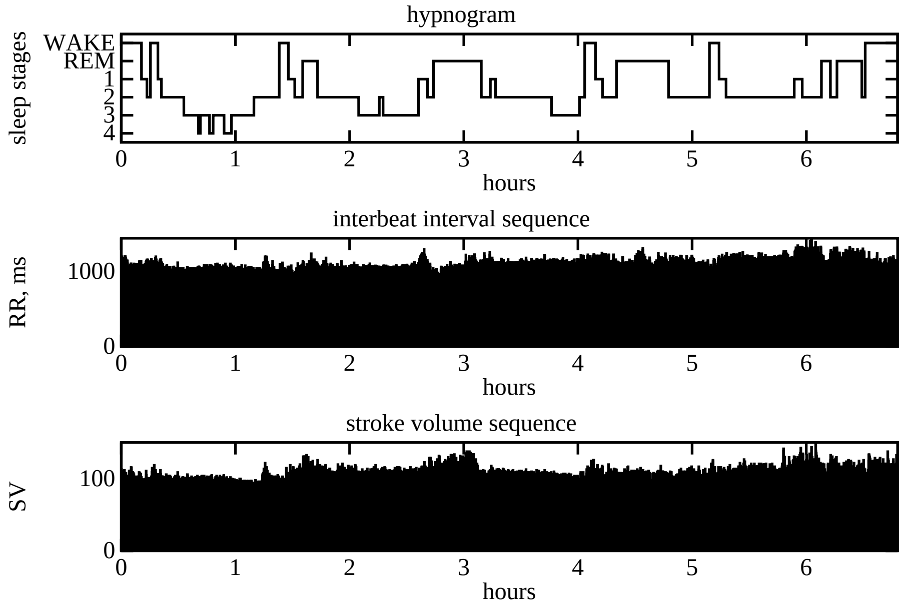

Figure 1. Example of hypnogram, interbeat interval and stroke volume sequences

SYSTEM ARCHITECTURE

Knowledge on sleep stages should be extracted from RR/SV vector time series. The standard methods for time series analysis are: statistical inference, hypotheses testing, correlation analysis, spectral analysis etc. The published results in the field of medical data mining concern one dimensional time series, e.g. RR time series. Our problem requires extension of the standard set of methods enabling analysis of vector time series, e.g. we need to estimate cross spectral density between RR and SV. Recently methods of nonlinear dynamics have appeared prospective in the analysis of medical time series [6,7]. We have included into our data mining tool frequently used in bioinformatics algorithms of estimating parameters of nonlinear dynamics as well as methods of empirical mode decomposition and progressive detrended fluctuation analysis.

One of the most important data mining tools is visualization of the available information, especially of multidimensional data. The visualization of several time series in one computer screen is implemented for the visual heuristic analysis of correspondence between estimated parameters and sleep stages. Here we use standard methods of 2d 3d graphics; for an example see the next section. Important tool of extension of heuristic abilities of medical experts to multidimensional space of estimated parameters is visualization of multidimensional data. Standard methods used in data mining are principal component analysis (PCA) and Kohonen' self organizing maps (SOM). However, PCA is a linear projection method not always well representing the structure of multidimensional data. SOM is not suitable to visualize large sets of multidimensional data, for example to visualize the data of subjective self assessment with respect to sleep quality of a group of 1500 patients. We propose to visualize large sets of multidimensional data by means of multidimensional scaling (MDS), a method implementing non linear projections of multidimensional space to two dimensions with minimal deformation of inter point distances [1]. Contemporary MDS algorithms are suitable for visualization of arrays of dimensionality up to several thousands. An example of visualization of the mentioned earlier data on subjective sleep quality is presented in the next section.

The other key data mining technique is classification. Well known classification methods include discriminant analysis, artificial neural networks (ANN), support vector machines (SVM). The choice of a proper classification method from our toolbox should be guaranteed, since the goal of the described data mining tool is to aid the development of a reliable automatic recognition method of sleep stages. Therefore we have included besides the standard methods also comparatively new classification methods based on linear programming; for an original idea we refer to [4]. It is worth to mention that our system includes also sophisticated methods absent in other systems and state-of the art reviews [3]; these methods are: analysis of vector time series, visualization by means of MDS, and linear programming based classification.

EXAMPLES

This section illustrates application and some advantages of the system described in previous sections.

An example of spectral analysis of RR and SV sequences is shown in Figure 2. Normalized very low frequency (up till 0.04Hz) power (VLF), normalized low frequency (0.04-0.15Hz) power (LF), normalized high frequency (0.15-0.4Hz) power (HF) and ratio of low and high frequency power (LF/HF) are some of the frequency domain measures of heart rate variability suggested in [7]. These measures normally are used to analyze RR sequences. We propose to take into account also the measures of SV power spectrum, and RR/SV cross power spectrum. It can be seen from the graphs of LF/HF during a night presented in Figure 2, that LF/HF graph of RR/SV cross power spectrum has peaks where have peaks either graph of RR or graph of SV coinciding in time with REM sleep. Therefore RR/SV cross power spectrum is more informative for detection of REM sleep than RR power spectrum normally used for this purpose.

The parameters estimated by means of statistical and spectral methods are used to evaluate the health status of a patient. Frequently it is important to compare the parameter vector of a patient with the set of parameter vectors of a population of patients in a region. An example of MDS based visualization of data on sleep quality of a population of 1500 patients is presented in Figure 3; the data of objective sleep quality is presented on the left, and the data of subjective one is on the right. The objective sleep quality is defined by 8 parameters estimated by standard statistical and spectral methods from RR and SV sequences of a patient. The subjective sleep quality is defined by 7 parameters estimated from the patient answers to the Pittsburg Sleep Quality Index (PSQI) questionnaire [2]. Four classes of objective sleep quality and five classes of subjective sleep quality are indicated. The data vector of a new patient can be immediately inserted showing in the figure its relative location with respect the data of other patients, and enabling a physician to make a preliminary diagnosis. Below a brief description of the indicated classes is presented.

The objective quality of sleep for class `I' is the best. The sleep effectiveness here is the largest, and the actual sleep time is the longest. The sleep structure is rather good. The wakefulness at nighttime is rather short. For the class `II' the sleep structure is good; although the wakefulness during night is rather long, but the sleep effectiveness is normal, and the total sleep time is satisfactory (< 6 hours). The sleep structure of the patients in class `III' is rather bad; the sleep effectiveness is about 75\%, the wakefulness is frequent and long, the time during the active sleep and during the stage 4 is too short. Class `IV' represents a very bad sleep structure; the sleep effectiveness is only 60\%, the wakefulness is increased to almost 40\% of total sleep period. The subjective quality of sleep can be expressed by patients with enormous variety of characteristics. To make these expressions comparable, the subjective assessments are unified using the PSQI [2]. Very many patients of the considered population evaluate their sleep very well. Therefore the class of good sleep is subdivided into two subclasses, and five classes on the whole are considered. The class `I' represents patients who evaluate their sleep quality excellent, and think that they do not have any sleep problems at all. PSQI of those patients was less than 5. The patients of the class `V' feel serious sleep problems; their PSQI exceeded 16.

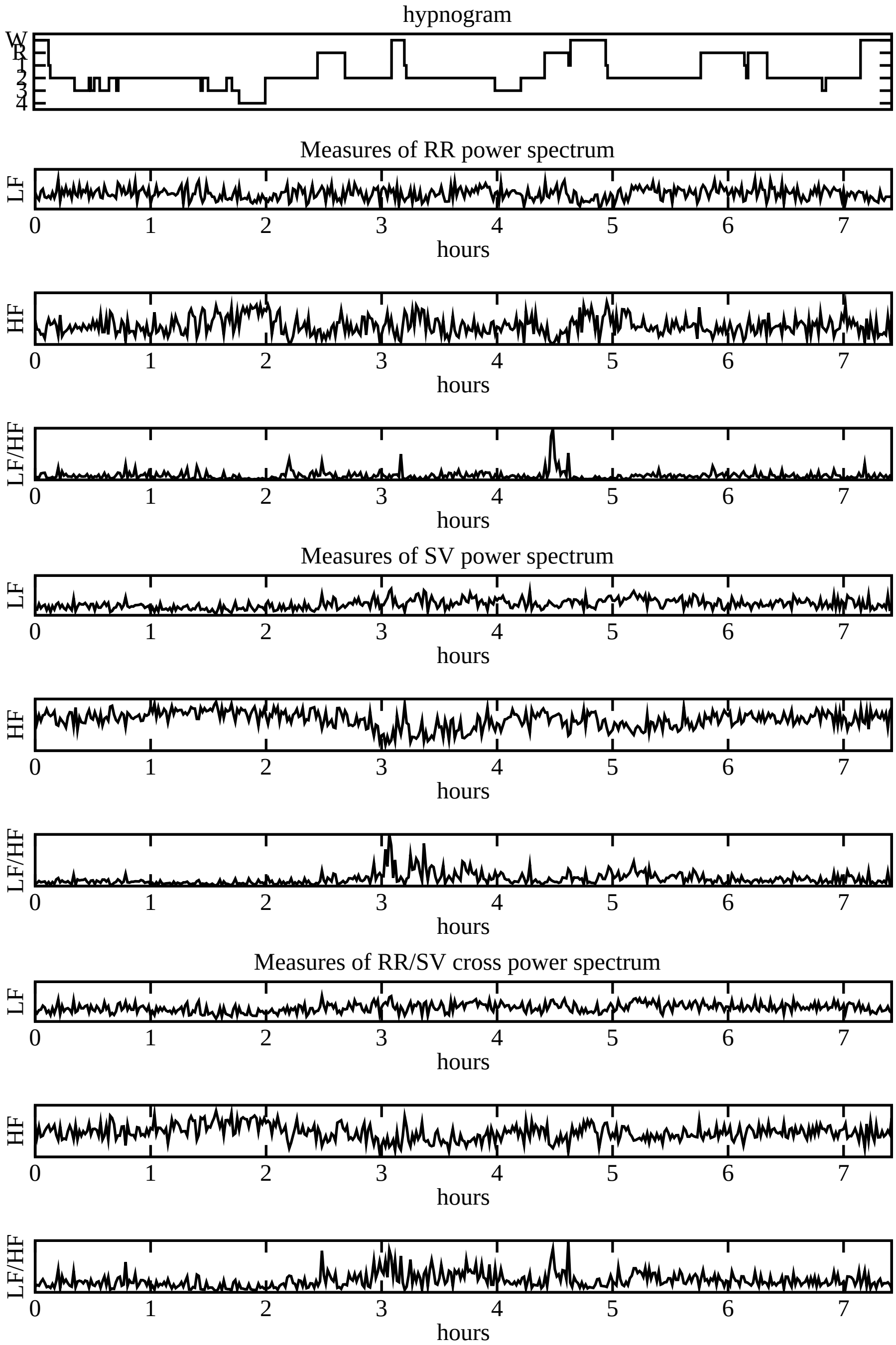

Figure 2. Hypnogram and graphs of spectral parameters of RR and SV sequences estimated during a night

The visualization of the data on objective and subjective sleep quality shows rather great discrepancy between subjective self assessment and real health status. Such a result of visualization is important in planning prophylactic and educational measures.

A system, especially oriented to aid development of new diagnostic methods, should include classifiers well suitable to the data relevant to the considered diagnostics. Since specific properties of the data in question at the initial stage of the project is not known, we have included standard classification methods, e.g. statistical discriminant analysis and ANN. SVM is an optimization based classifier as well as relatively new classifier based on linear programming. Although the latter is well known in operations research community, and its application for the breast cancer diagnostics [3] has been awarded the Lanchester Prize of the year 2000, it is still not frequently used in medical data mining. The optimization based classifier of [3] has been modified to separate two classes by a quadratic surface.

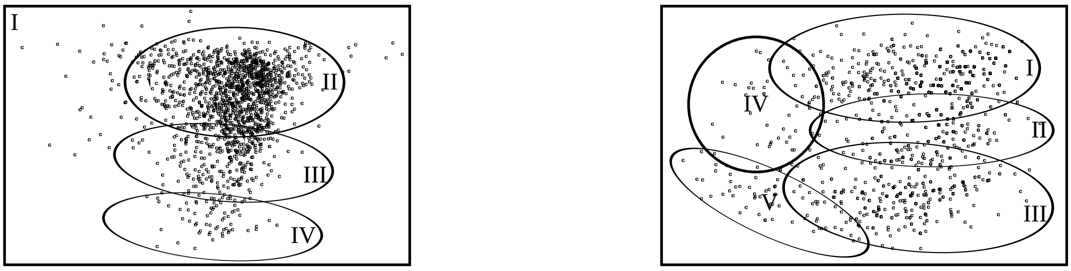

Figure 3. Visualization of data on sleep quality: the objective sleep quality data is on the left and the subjective one is on the right

The system includes a database which can be used as a test-bed of different data mining methods. For example, the population of 1500 patients is classified by medical doctors using polysomnography to 4 classes according to the objective sleep quality. The sizes of the classes are as follows: I class contains 36 patients, II does 1108, III does 287, and IV does 65. The data of these patients extracted from RR sequences consists of 8 dimensional vectors, and they are visualized in the Figure 3. The picture, however, does not unambiguously answer the question on separability of four classes in 8 dimensional space. To answer this question the data was attempted to separate by means of three classifiers: ANN, SVM and Quadratic Classification Function (QCF). The dichotomies were obtained by the classifiers with the following parameters tuned experimentally. There was used a two layers feed forward ANN with 10 neurons in the first layer, and 1 neuron in the second (output) layer. Transfer functions in both layers were hyperbolic tangent sigmoid. Gradient descent back propagation method was used for training the ANN. A nonlinear SVM with polynomial kernel of fifth degree was used. The results of separation of four classes of the (training) sample are presented in the Table 1.

Table 1. Failures of classification on training and exam samples

Training					Exam				
Class	Size	SVM	ANN	QCF	Class	Size	SVM	ANN	QCF
1	36	0	0	0	1	8	1	3	2
2	1109	4	15	20	2	44	13	1	6
3	288	1	16	0	3	78	34	62	29
4	67	0	6	0	4	20	0	8	1

The best separability is achieved by means of SVM. However, it is interesting to test also the generalization capability of all three classifiers. For the examination the data of a new group of 150 patients has been classified by the classifiers trained on the data presented above. The results are presented in the Table 1, and they can be interpreted in two ways. From the point of view of the generalization capability the optimization based classifiers outperform ANN. Such a conclusion is interesting for the experts in classification. Relatively large number of failures in recognition of the third class by SVM and ANN can be explained by the over-fitting of these classifiers. However, the quadratic separation surface of QCF can not be over-fitted. Therefore, the second interpretation of the results seems also possible, that the training data and the exam data can not be considered a unique sample. Such a conclusion is important for the medical experts implying further medical examination of all available patients' data in both groups.

CONCLUSIONS

A computer aided system for analysis of RR sequences is created including well known methods as well as some relatively new methods for knowledge elicitation. The applications of the created system to the problems related to the development of automatic diagnostic methods of sleep disorders and diseases are discussed. The usefulness of multidimensional scaling based visualization is demonstrated as well as advantages of optimization theory based classifiers. The created system can be applied without of major readjustment for other medical subjects where either RR or other time series contain essential information about a disease.

REFERENCES

[1] I. Borg, P.Groenen. Modern Multidimensional Scaling. New York: Springer, 1997.

[2] D. Buysse et al. The Pittsburg Sleep Quality index: A New Instrument for Psychiatric Practice and Research, Psychiatric Research, 28, 1989, 193/213.

[3] Cios, K., G. Moor. Medical Data Mining and Knowledge discovery: Overview of Key Issues, In: K.Cios (Ed.) Medical Data Mining and Knowledge Discovery. Heidelberg: Physica Verlag, 2001, 1/20.

[4] Mangasarian, O., W. Street, W. Wolberg. Breast Cancer Diagnosis and Prognosis via Linear Programming, Operations Research, 43, 1995, 570/577.

[5] Rechtschaffen, A., A. Kales. A Manual of Standardized Terminology, Techniques, and Scoring for Sleep Stages in Human Subjects, U.S. Government Printing Office, Washington D.C., 1968.

[6] Souza Neto, E.P., et al. Assessment of Cardiovascular Autonomic Control by Empirical Mode Decomposition,.4th International Workshop on Biosignal Interpretation, Como, Italy, June 24-26, 2002, Proceedings, 2002, 123/126.

[7] Task Force of the European Society of Cardiology and the North American Society of Pacing and Electrophysiology, Heart Rate Variability: Standards of Measurement, Physiological Interpretation, and Clinical Use, Circulation 93, 1996, 1043/1065

ABOUT THE AUTHOR

Antanas Zilinskas, Professor, Department of Informatics, Vytautas Magnus University, Vileikos 8, 44404 Kaunas, Lithuania, E-mail: antanasz@ktl.mii.lt.

Julius Zilinskas, Dr., Institute of Psychophysiology and Rehabilitation, Vyduno 4, 00135 Palanga, Lithuania, E-mail: julius.zilinskas@ktl.mii.lt

Audrius Varoneckas, Ph.D. student, Institute of Psychophysiology and Rehabilitation, and Vytautas Magnus University, Vileikos 8, 44404 Kaunas, Lithuania, E-mail: audrius_varoneckas@fc.vdu.lt