Sleep Stage Classification using Fuzzy Sets and Machine Learning Techniques

Pedro Piñero[1], Pavel Garcia[1], Leticia Arco[2], Alfredo Álvarez[3], M. Matilde García[2], Rolando Bonal[2]

[1] Grupo de bioinformática, Universidad de las Ciencias Informáticas
{ppp, cortina}@uci.cu
[2] Centro de Estudios Informáticos, Universidad Central "Marta Abreu" de Las Villas
{leticiaa, mmgarcia, bonal}@uclv.edu.cu
[3] Centro de Neurociencias de Cuba
alfredo@cneuro.edu.cu

Abstract. The hypnogram is determined after a study of electrophysiological records. In this paper we present the ISSSC (Intelligent System for Sleep Stages Classification). This system is divided into four different modules: the first processes the electrophysiological signals and determines its most relevant parameters; the second module establishes fuzzy rules that will be used during the classification process; the third module is an inference module, it implements a fuzzy model. Finally the system builds the patient's hypnogram and provides us different outputs. We present the classification results obtained from applying the systems to classify patients with different sleep disorders.

Keywords: sleep stages, fuzzy rules-based reasoning, fuzzy system, expert system, machine learning.

1 Introduction

Sleep stages automatic classification has become a key problem in neuroinformatics and generally speaking in neuroscience. It is used to describe the patient's transition through the sleep stages (hypnogram) taking into account the study of the electrophysiological records such as Electroencephalograms (EEG), Electro-oculograms (EOG) and Electromyograms (EMG).

The most popular technique to process the sleep stages based on rules were proposed by Rechtschaffen and Kales (R&K) in 1968 [12]. Today it is widely recognized that these rules have severe limitations which were not foreseen 30 years ago [7]. Nevertheless, these rules have survived all criticism raised in the past.

In this work we will assume that the sleep stage scoring is a kind of classification problem known as classification with learning or supervised classifications. In this problem the human sleep stages can be classified into one of 6 discrete stages according to R&K rules.

To solve our classification problem we use Machine Learning techniques and fuzzy logic theory.

Machine Learning is an interdisciplinary field with connections to artificial intelligence, information theory and statistics. Machine learning algorithms learn automatically from experience and use different forms to represent knowledge. Among the many ways to represent knowledge, if-then rules [4][2] have been used with other techniques to make inferences and classify some new cases.

On the other hand, the fuzzy rules, an example of if-then rules, are based on the use of fuzzy logic to represent knowledge. General fuzzy rules can be represented as: *If* x *is* A *then* y *is* B*;* where, B and A are fuzzy sets belonging to the Y and X linguistic variables respectively.

Fuzzy sets [6] are usually identified by membership functions. A fuzzy set μ of X is a mapping involving the set X up to the unit interval: $\mu: X \rightarrow [0, 1]$. Here $\mu(x)$ is known as the x-value membership degree.

In section 2 we provide details of the modules that constitute the Intelligent System for Sleep Stages Classification (ISSSC), and in section 3, the results of the use of this system are validated through the analysis of the system's implementation and the elements that affect the classification process. Finally, in the conclusions of this paper we evaluate the usefulness of the system .

2 Intelligent System for Sleep Stages Classification (ISSSC)

ISSSC version 1.0, consists of four different modules figure 1. The first module processes the patient's electrophysiological records that are stored in binary files.

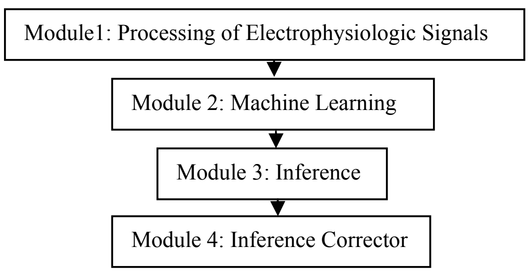

Fig. 1. Module diagram and cases of use

We carried out this research in the Clinical Neurophysiology Offices of the Neurosciences Center of Cuba. We recorded the EEG, EOG and EMG using Medicid -3E. In each case we work with silver-disk electrodes stuck with collodion to the skin or to the scalp.

The EEG derivations were: Fz, C3, C4, O1 and O2 of the 10/20 system. We established the frequency range from 0.5 to 30Hz and the gain of 10, 600 in the amplifiers. We used two monopole channels for recording the EOG.

We situated two electrodes under the patient's chin for recording the EMG. We established a frequency range of 10-100 Hz and the gain was of 10, 000. We studied the signals taken from some patients of the Neurosciences Center. The sample consisted of 4 patients (MRLL, ACGBCLA, AGSCLA and ARR) suffering from different illnesses: insomnia, depressive disorders etc. The manual sleep classification of the patients was made by different experts with an inter-rater reliability of 71 %.

The main purpose of this stage is to reduce data, trying to prevent missing important information; that is to reduce the number of variables without losing relevant information. The electrophysiological signals were captured every 30 s.

2.1 Module 1: Processing of Electrophysiological Signals and computation of relevant parameters

This module computes the necessary descriptors of patients that have been already studied, and forms a training set that could be used for learning and rules production. This module computes the descriptors of new patients, so that they can be classified.

The input files to this module are: MRK, CDC, INF and PLG. The files with the extension (INF) are text files that contain general information. The files with extension (.PLG) keep the registered signals. The file with the extension (.MRK) stores the sleep stages, marks and information inserted in the record once the specialists have analyzed it. Finally, the files with the extension .CDC store the calibration values and the DC level of every register channel.

Once the input information is analyzed, it has to be processed in order to form the training set. In this step, every epoch is considered as a case and the sleep stage the patient is in, is the class that the case belongs to.

If the sample for the analysis is from the record of a patient that has been already studied, the output in the first module will be the files (.lrn), (.data) and (.names). These files will be used in the learning process and fuzzy rules generation. If the sample is from the record of a new patient, then the output is a file (.cse). The file (.lrn) stores a training set and there is always a file (.names) where the description of this set is stored. Now, the file (.data) combines all the information in the first two files to supply two different ways to organize the output information of this module.

2.2 Module 2: Machine learning

Two algorithms that generate fuzzy rules ware created for this module. They use a training set with nominal and numeric data. The input files are (.lrn) and (.names) or (.data) files that were generated in the previous module. Two different ways of processing the parameters are identified according to the information in the input files : Discretization of parameters and clusterization of parameters.

Once the system has recognized the numeric intervals that correspond to each linguistic term, it automatically generates the membership functions and the fuzzy sets.

GENRUL5 and MLRUL are the algorithms developed for generating the fuzzy rules using numeric and nominal data [2]. In the first of these algorithms, the system goes over the whole training set generating the necessary rules. A case is covered if and only if there is an already generated fuzzy rule that has the same consequent or class of the objective feature of the case and the parameters values is equal to a linguistic term that precedes the rule. In the second algorithm (MLRUL), partitions are created by consecutive arrangements of the training set. The most relevant linguistic variable in the partition is taken as the comparison criteria. This relevance is determined by the use of methods that stimulates homogeneity among cases of the same class and the heterogeneity among cases of different classes. The following methods are proposed to determine the relevance of a linguistic variable: Mántaras Distance [9], descriptive statistics VCramer, computation of the entropy and the acquisition of information [4] and MLRelevance measure [2]

This module creates two files. The first is a .rul file that will be used in the inference process; and the second is a text file that describes the generated fuzzy rules and whose main objective is to explain the inference process.

2.3 Module 3: Inference

This module is used by the system's 2nd case of use and is related to the classification and study of new patients. The process of inference uses the components package FuzzyPack. This package is a platform that enables the generation of fuzzy systems that match the zero order Sugeno model.

A fuzzy system generated using FuzzyPack operates in the first place based on the fact that there is an input file .rul that stores all the information to determine the fuzzy rule system: description of the parameters domain, its relevance information, the membership functions of those parameters, their values and the description of all the system rules including their significance level.

Second, the inference using FuzzyPack is based on a connection mechanism among components. This mechanism anables the creation of a layers net that depends on the rules file (*.rul*) given as an input parameter to this module.

The first layer of this module is a parameters layer. It has as many nodes as different attributes. The second layer if the membership function layer. It has as many nodes as different membership functions. Every node of the first layer is connected to all membership functions linked to its linguistic term. There are as many nodes in the third layer as the amount of rules in the file (.rul). The connections between nodes of the second and third layers depend on the existance of the membership functions in every rule. In the fourth layer there is one node for every different consequent in the rules file. That is, there will be 6 consequents, one for every sleep stage (waking, REM, Stage1, Stage2, Stage3, Stage4). All nodes in the third layer that have the same consequent will be connected to the node in the fourth stage that corresponds to that consequent. There is in the fifth layer only one node and all the previous layer nodes are connected to it. Its function is to give a final answer and to return the explanatory files that help understand the inference of the system.

2.4 Module 4: Inference Corrector

Up to this point, the process of automatic classification takes place independently in each epoch. This means that the system determines the sleep stage in each 30-second interval. It does not take into account the relationship between the given stage and its adjacent stages. This may alter, in some cases, the logical sleep sequence because, even though the stages are distant from each other, they may have similar physiological characteristics. That is why it is necessary to implement corrector module that takes into account the relationship already mentioned.

This module's operation is based on a hard-rule system and a case-based reasoning system. The hard rules are supplied by experts [3] and the cases are those from the Polysomnogram of some patients that were classified and studied by experts.

3 Validation

Evaluation is a necessary and important process in the development and testing of any software. The main purpose of this process is to determine the system's degree of correct response according to the expectations.

Brender [5] proposes some metrics and other measures that express different quality parameters of medical knowledge. Among the metric measures used in the evaluation of the system we can mention total behavior, behavior conditioned to the class (expected and inferred values), kappa total behavior, kappa behavior conditioned to the class (inference values) and an error functional.

3.1 Elements that influence the classification process

Inter-expert reliability. The Inter-expert reliability moves from 67% to 91 % [1]. In 10 laboratories in Japan, it moved from 67% to 75.36% [1]. In the Center of Neuroscience of Cuba (CNC) [3], it was 71% for all cases and 72% in the healthy cases. In a similar study, B. Kemp had an inter-expert reliability of 75% in all cases and 77% in healthy cases [8].

Automatic classification of healthy and sick patients: Martin [10] had an inter-expert reliability of 80.8% in a study of 5 healthy subjects, Stanus [14] had an inter-expert reliability of 75% with healthy subjects and 70 % with depressive disorders and insomnia. In the CNC the agreement between the experts and the system was 69% with all cases and 70% with healthy subjects. B. Kemp in a similar comparison, reports 70% and 75%, respectively. This demonstrates that it is easier to classify healthy subjects than subjects suffering from sleep disease.

Analysis of each sleep stage: During the visual classification analysis of 9 healthy case studies, Smith and Karacan [13] stated a 83% of inter-expert reliability. Though during this process stages 1 and REM were analyzed as a whole. On the other hand, on [11], the net efficiency was only tested on stages 1, 2, REM and Waking having as a result a 77.6% of inter-expert reliability. If stages are disregarded while analyzing the case studies, the final results will not be the effective ones of the classification.

3.1 Test

The electrophysiological signals resulting from the 4 patients that were taken to create the bases of the case sets for each patient were: EMG (C3-A12), EEG (O1-A12), EOG (LOG-A12), EOG (ROG-A12) and EMG-25.

Each base of case sets was split, having a 70% on training and a 30 % on test.

Analyzing table 1, as a particular case of metrics and functional application for the GENRUL5 algorithm having the base of cases ACGBCLA, we can see satisfactory results.

Table 1: Metrics and functional using GenRul5

Patient (ACGBCLA)	
Metrics and funtionals	Value
Total behavior	0.8736
Total behavior of Kappa	0.7980
Functional of quality	0.8736
Functional of error	0.0081
Weight of error	0.0200

Table 2: Results from applying the corrector module with the MLRul algorithm

Bases of Cases	Without correction (%)	Corrected (%)
MRLL	70.41	79.48
ACGBCLA	81.01	87.37
AGSCLA	74.81	84.37
ARR	68.15	78.86

Table 3: Discretization and clusterization comparisons, using GenRul5 algorithm (test sets)

Bases of cases	Discretization (%)	Clusterization (%)
MRLL	62.98	67.97
ACGBCLA	61.43	71.90
AGSCLA	41.37	64.89
ARR	73.52	73.86

It is very important to analyze each stage within the context of the hypnogram, while classifying automatically the sleep stages of patients. The inference module determines each stage in an independent way but the inference corrector module uses the hypnogram, resulting from the inference module, following the stages transition rules. That is why, we can see on table 2 that corrected results surpass in approximately 10 % the first results of the classification.

Summarizing, we can see on table 3 that the process of clusterization has higher results than those from the process of discretization, due to the fact that intervals are not specified by the user but obtained automatically from data.

4 Conclusions

The ISSSC version 1.0 system is a helping system for medical diagnosis. This system allows us to create the hypnogram of patients taking into account the electrophysiological signals: EEG, EOG and EMG. The system is a useful tool for detecting sleep disorders, it also helps to spread the expert's expertise. The ISSSC system consist of four modules: signals preprocessing, machine learning, inference and inference corrector module. Each module works independently from the others. This characteristic of the system allows us to develop different applications using different module combinations. The result of the medical test run on this system were successful though a greater amount of patients must be tested with this system in order to prove the reliability of the results.

The heterogeneity of the patient's disorders used on the learning module may bring about false learning and thus affect the results inferred during the sleep analysis. For this reason, we propose a specialized learning process for each kind of sleep disorder aimed at improving the results obtained.

References

1. Agarwal, R., Gotman, J.: Computer-Assisted Sleep Staging. IEEE Transactions On Biomedical Engineering, Vol 48, No. 12. (2001)
2. Arco , L., Caballero, Y.: Aprendizaje automatizado para la generación de reglas borrosas. Trabajo de Diploma. Universidad Central "Marta Abreu" de Las Villas, Cuba (2001)
3. Barreiro , A.: Clasificación automática de etapas de sueño humano. Trabajo para optar por el título de Especialista de Primer Grado en Fisiología Normal y Patológica.Ciudad de La Habana, Cuba (1995)
4. Bender, E.: Mathematical Methods in Artificial Intelligence. IEEE Computer Society Press. Los Alamitos, California (1996)
5. Brender, J.: Measuring quality of medical knowledge. In Proceeding of the Twelfth International Congress of the European Federation for Medical Informatics (1994) 69-74
6. Cox E., Taber R. and O'Hagan M., The Fuzzy Systems Handbook, Second Edition AP Professional, Paperback, 2nd Bk&Cd edition, August 1998.
7. Himanen SL, Hasan J. Limitations of Rechtschaffen and Kales. Sleep Medicine Reviews 2000; p 149–167.
8. Kemp B.: A model-based monitor of human sleep stages. Biol. Cybern (1987)
9. Mántaras, R.: A Distance-Based Attribute Selection Measure for Decision Tree Induction. Machine Learning (1991)
10. Martin, W., Johnson, L.: Pattern recognition of EEG-EOG as a technique for all night sleep stage scoring. Electroencephalogr. Clin. Neurophisyol, Vol. 32 (1972) 417-427
11. Oropesa, E., Hans, C., Marc, J.: Sep Stage Classification using Wavelet Transform and Neural Network. International Computer Science Institute (1999)
12. Rechtschaffen, A., Kales, A.: A Manual of Standardized Terminology, Techniques, and Scoring System for Sleep Stages of Human Subjects. Los Angeles, CA: BIS/BRI, Univ. of California (1968)
13. Smith, J., Karacan, I.: EEG sleep stage scoring by an automatic hybrid system. Electroencephalogr. Clin. Neurophisyol, Vol. 31, (1971) 231-237
14. Stanus, E., Lacroix, B.: Automated slep scoring: A comparative reliability study of algorithms. Electroencephalogr. Clin. Neurophisyol, Vol. 66, (1987) 448-456