ESANN 2016 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Bruges (Belgium), 27-29 April 2016, i6doc.com publ., ISBN 978-287587027-8.
Available from http://www.i6doc.com/en/.

Automatic detection of EEG Arousals

Isaac Fernández-Varela[1], Elena Hernández-Pereira[1],
Diego Álvarez-Estévez[2] and Vicente Moret-Bonillo[1] *

1- Universidade da Coruña - Departamento de Computación
Facultade de Informática, Campus de Elviña, A Coruña - Spain

2- Sleep Center & Clinical Neurophysiology - MCH en Bronovo-Nebo
Lijnbaan 32 2512 VA, The Hague - Netherlands

Abstract. Fragmented sleep is commonly caused by arousals that can be detected with the observation of electroencephalographic (EEG) signals. As this is a time consuming task, automatization processes are required. A method using signal processing and machine learning models, for arousal detection, is presented. Relevant events are identified in the EEG signals and in the electromyography, during the signal processing phase. After discarding those events that do not meet the required characteristics, the resulting set is used to extract multiple parameters. Several machine learning models — Fisher's Linear Discriminant, Artificial Neural Networks and Support Vector Machines — are fed with these parameters. The final proposed model, a combination of the different individual models, was used to conduct experiments on 26 patients, reporting a sensitivity of 0.72 and a specificity of 0.89, while achieving an error of 0.13, in the arousal events detection.

1 Introduction

The American Academy of Sleep Medicine (AASM) defines the electroencephalographic arousal as an abrupt shift in electroencephalogram (EEG) frequency, including alpha, theta, and/or frequencies greater than 16 Hz, lasting at least 3 seconds and with at least 10 seconds of previous stable sleep [1]. Furthermore, during the rapid eye movement (REM) phase, a concurrent increase in the submental electromyography (EMG), lasting at least 1 second, is needed. As arousals cause fragmented sleep, sleep studies must identify these events. Sleep studies are performed with a polysomnography (PSG), recording a set of physiological signals from the patient — pneumological, electrophysiological and contextual information —. An expert physician can observe the EEG and EMG derivations to detect arousals. Since the recording of a PSG last a whole night, the amount of data is huge, making the detection of EEG arousals a very time-consuming task. Thus, automatic detection and analysis is desired.

In recent years, different approaches tried to solve this problem with different techniques, highlighting the proposal of Alvarez-Estevez *et al.* [2], as is based in the use of machine learning models after processing two EEG and one EMG derivations, and it is the approach followed in this work.

*This research was partially funded by the Xunta de Galicia (Grant code GRC2014/035) and by the Spanish Ministerio de Economía y Competitividad, MINECO, under research project TIN2013-40686P both partially supported by the European Union ERDF.

ESANN 2016 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Bruges (Belgium), 27-29 April 2016, i6doc.com publ., ISBN 978-287587027-8.
Available from http://www.i6doc.com/en/.

2 Proposed Method

The proposed method includes the following phases: signal conditioning, relevant intervals selection through a windowing processing, feature extraction after grouping selected intervals and, finally, the use of different machine learning models.

2.1 Signal Processing

Three signals are processed using different techniques: two EEG derivations (C3/A2 and C4/A1) and one submental EMG. Whereas the EEGs are studied in the frequency domain, the EMG is studied in time domain.

Signal Conditioning: It is well known that EEG and other related biosignals usually present artifacts that can mislead their interpretation [3]. After locating the QRS complexes on the ECG signal, we check if they are inducing an artifact. Because of the short peak duration, the solution proposed is to interpolate the affected segment, obtaining a near real signal as shown in Figure 1.

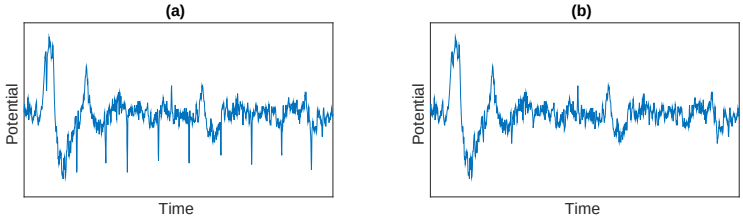

Fig. 1: EEG signal (a) before and (b) after signal conditioning.

Signal Windowing: After both EEG signals are conditioned, a windowing process takes place. The three available signals, the two EEG derivations and the submental EMG, are processed using a moving window.

The window used for the EEG derivation has a total duration of 3 seconds. Between two consecutive studied instants the skipped time was 0.2 seconds. All values were chosen empirically. According to the AASM, in order to score an arousal there must be an abrupt change in the alpha (α = 8-12 Hz), theta (θ = 4-7 Hz) and/or frequencies greater than 16 Hz. Transforming each window into frequency domain, using a hamming function and the Fourier's transform, power is obtained following the formula: $power = \frac{1}{n^2}\sum_{i=1}^{n}|X(n)|^2$ where n is the number of samples and X the bandpassed signal.

For each frequency, a baseline over the average of the previous 10 seconds is created. Intervals with abrupt changes are selected when the ratio between the power and the baseline is greater than 1.5 — selected empirically —.Intervals lasting less than 3 seconds are discarded.

The process carried with the EMG signal is similar but the windows displacement was augmented to 0.4 seconds and the parameter studied in the window

ESANN 2016 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Bruges (Belgium), 27-29 April 2016, i6doc.com publ., ISBN 978-287587027-8. Available from http://www.i6doc.com/en/.

is the peak to peak amplitude. A baseline over the average of the previous 30 seconds is created, but the threshold used to select high activity intervals is 2. Intervals lasting less than 3 seconds are discarded.

Intervals Grouping: To group the intervals of the different signals, an approach based on the *epochs* division is followed. Identifying the epoch of the interval using its middle point, a valid group is formed choosing one interval from each signal. If there is no interval for one signal, no group is formed, whereas if there is more than one, the one with the higher power on the frequencies band mentioned in Section 2.1 is selected.

2.2 Machine Learning Models

From the previous groups, features are extracted and used as input to several classifications models. We also investigate the combination of all of them.

Feature Extraction: Table 1 describes the complete set of features extracted from the signals. Regarding the EEG intervals, we include not only the powers already obtained (Section 2.1), but also delta (δ = 0.5-4Hz) and sigma (σ = 12-15Hz) powers. At last, Hjorth parameters are included, as they have been demonstrated to be a good characterization of the EEG [4]. These parameters are: $Activity = var(X(n))$, $Mobility = \sqrt{\frac{Activity(X'(n))}{Activity(X(n))}}$ and $Complexity = \frac{Mobility(X'(n))}{Mobility(X(n))}$, where X is the signal and X' the first derivative.

Within the EMG interval, the only features included were those obtained in Section 2.1.

Finally, we include two contextual features: the sleep stage, automatically obtained following the method in [5]; and the common time during which both EEG intervals appear simultaneously.

EEG features	EMG features	Contextual features
δ, θ, α, σ and $> 16Hz$ total power	Sum of amplitudes	Sleep stage
δ, θ, α, σ and $> 16Hz$ max. power	Max. amplitude	Common time
δ, θ, α, σ and $> 16Hz$ min. power	Min. amplitude	
Activity, Mobility and Complexity	Duration	
Duration		

Table 1: Intervals features description

Classification Models: As already mentioned, three classification models were considered for the arousal detection task.

Fisher's discriminant [6] uses a linear combination of the attributes at the input, splitting the space into classes. Between each pair of classes, a linear discriminant is defined. The goal is to maximize the distance between the means of the two classes while minimizing the variance within each class.

Support Vector Machines (SVM) [7] are a supervised classification technique that works by non linearly projecting the training data in the input space to a feature space of higher (infinite) dimension by the use of a kernel function. In this work, a RBF kernel was chosen because it maps the samples into a higher

ESANN 2016 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Bruges (Belgium), 27-29 April 2016, i6doc.com publ., ISBN 978-287587027-8.
Available from http://www.i6doc.com/en/.

dimensional space, being able to handle the case when the relation between class labels and attributes is nonlinear.

Finally, Artificial Neural Networks (ANN) are mathematical models biologically inspired on how the biological neurons work. Basically, they are composed of a set of interconnected layers of simple computing nodes that operate as nonlinear summing devices. Different models of ANNs are available throughout the literature depending on the architecture, on the process to adjust their weights, or in the propagation of the information from the inputs to the outputs [8]. In our case, a feedforward network with one hidden layer was used.

Combined Approach: The classification models were compared with an approach that combines the classifiers following Shortliffe and Buchanan (S&B) [9] certainty factors model. This model is based on the definition of certainty factors (CF). Given the hypothesis, i.e the presence of arousal, the CF can get a value between (-1, 1) where 1 asserts the hypothesis and -1 denies it — arousals are accepted with an empirical threshold of 0.7 —. The combination of two evidences — two individual classifiers outputs — referred to the same hypothesis is made as follows:

$$CF_{ij} = \begin{cases} CF_i + CF_j - CF_i \times CF_j & \text{if } CF_i > 0, CF_j > 0 \\ CF_i + CF_j + CF_i \times CF_j & \text{if } CF_i < 0, CF_j < 0 \\ (CF_i + CF_j)/(1 - min(|CF_i|, |CF_j|)) & \text{if } CF_i \times CF_j < 0 \end{cases}$$

2.3 Performance Measures

The performance of the method is evaluated in terms of following measures:

The classification error computed as the proportion of incorrectly classified positive and negative instances.

The sensitivity which quantifies the ability to correctly identify positive instances. It is the proportion of true positives that are correctly identified.

The specificity which quantifies the ability to correctly identify negative instances. It is the proportion of the true negatives that are correctly identified.

The AUC which compares simultaneously the sensitivity and specificity. In a two class problem is the average between both values.

3 Experimental Procedure

In order to develop the arousal detection method, and also to validate it, we use a set of PSG recordings from real patients. All the recordings used were taken from the Sleep Heart Health Study (SHHS) [10].

The training and validation set used contains a total of 2353 arousals in 18094 epochs. As this set is unbalanced, an undersampling technique was applied, to avoid biased classifiers. The testing set used contains a total of 688 arousals in 5878 epochs.

In order to configure the models before any experimentations, a ten-fold cross validation was followed trying different parameters. With this procedure, the selected values were $C = 2^{11}, \epsilon = 2^3$ for the SVM, and 40 neurons in the hidden layer for the ANN.

ESANN 2016 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Bruges (Belgium), 27-29 April 2016, i6doc.com publ., ISBN 978-287587027-8. Available from http://www.i6doc.com/en/.

4 Experimental Results

Four experiments were designed, varying the included parameters, with the results shown in Table 2. As ANN results are different with each execution, the values shown are the average values of 15 executions. Best results are highlighted in bold.

		Error	Sensitivity	Specificity	AUC		Error	Sensitivity	Specificity	AUC
Fisher's	All the features	0.196	0.762	0.810	0.786	Without Hjorth parameters	0.171	0.737	**0.843**	0.790
SVM		**0.134**	0.815	**0.874**	**0.847**		**0.161**	0.814	**0.843**	**0.828**
ANN		0.160	**0.855**	0.838	**0.847**		0.200	**0.860**	0.790	0.825
Fisher's	Without sleep stage	0.200	0.745	0.809	0.777	Without Hjorth and sleep stage	**0.171**	0.721	**0.845**	0.783
SVM		**0.163**	0.811	**0.841**	**0.826**		0.190	0.840	0.806	**0.823**
ANN		0.182	**0.835**	0.815	0.825		0.202	**0.851**	0.790	0.821

Table 2: Test results. Best classifier values marked in bold face.

With the complete set of features, we achieved the best results, both in terms of highest AUC and in terms of lower error. In the other experiments results were similar, with the SVM obtaining the highest AUC, but the ANN achieving the highest sensitivity. Focusing in the experiment with all the features, both SVM and ANN perform better than Fisher's discriminant. Although the AUC is equal for both nonlinear classifiers, SVM error is lower. While ANN gets higher sensitivity, SVM gets higher specificity. As the problem is unbalanced and dominated by the negative class, higher specificity means lower error.

To check if classification capabilities could be improved, we proposed a combined approach.

An independent set, containing 8416 arousals in 31080 epochs, is constructed to test the generalization capabilities with the proposed combination of the individual classifiers. Results are shown in Table 3.

	Error	Sensitivity	Specificity	AUC
Fisher's	0.181	0.784	0.815	0.799
SVM	0.160	0.816	0.834	**0.825**
ANN	0.230	**0.884**	0.734	0.809
S&B Combination	0.125	0.721	0.890	0.810

Table 3: Test results. Best classifier values marked in bold face.

As expected, the combination reduced the error achieved and the sensitivity, as with this approach, arousals are only scored when individual classifiers tend to agree. Thus, less arousals were scored but less errors were made. However, global performance in terms of AUC was similar to the best individual case. While the error reduction between the SVM and the S&B Combination was a 22 %, the AUC reduction was only a 2 %. Furthermore, this combination was better (in terms of AUC) than the other two individual classifiers.

ESANN 2016 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Bruges (Belgium), 27-29 April 2016, i6doc.com publ., ISBN 978-287587027-8. Available from http://www.i6doc.com/en/.

5 Conclusions

This paper proposes a new method for automatic arousals detection. Two phases summarize the method: signal processing and a machine learning approach. Signal processing includes the conditioning of the EEG to reduce the impact of artifacts; the analysis of EEG derivations searching for abrupt power changes; the analysis of EMG searching for high activity; and finally, the union of the relevant intervals found in the previous steps. The machine learning approach includes the extraction of features from the intervals selected before; the use of several models — one linear: Fisher's discriminant, and two non-linear: ANN and SVM — in four different experiments; and finally, the proposal of a combination method of the individual classifiers.

It has been demonstrated that using the complete set of features proposed the SVM achieves the best results, obtaining a sensitivity of 0.815 and a specificity of 0.874, with an error of 0.134. A combined approach was tried to demonstrate that individual classifiers performance was enhanced. In this case, the error is reduced in a 22 % while the AUC is maintained.

More general and adaptative artifacts removing methods are proposed as future work. The incorporation of features selection is also future work.

References

[1] American Academy of Sleep Medicine, R.B Berry, et al. *The AASM manual for he scoring of sleep and associated events: rules, terminology and technical specifications*. American Academy of Sleep Medicine, 2015.

[2] D Alvarez-Estevez and V Moret-Bonillo. Identification of electroencephalographic arousals in multichannel sleep recordings. *IEEE Transactions on Biomedical Engineering*, 58(1):54–63, 2011.

[3] P. Anderer, S. Roberts, A. Schlögl, et al. Artifact processing in computerized analysis of sleep EEG - A review. *Neuropsychobiology*, 1999.

[4] Bo Hjorth. Eeg analysis based on time domain properties. *Electroencephalography and clinical neurophysiology*, 29(3):306–310, 1970.

[5] D Álvarez Estévez, Fernández-Pastoriza, et al. A Method for the Automatic Analysis of the Sleep Macrostructure in Continuum. *Expert Systems with Applications*, 40(5):1796–1803, 2013.

[6] Ronald A Fisher. The use of multiple measurements in taxonomic problems. *Annals of eugenics*, 7(2):179–188, 1936.

[7] N Cristianini and J Shawe-Taylor. An introduction to support vector machines. 2000.

[8] J Principe, N Euliano, and C Lefebvre. Neural systems: Fundamentals through simulations, 2000.

[9] Edward H Shortliffe and Bruce G Buchanan. A model of inexact reasoning in medicine. *Mathematical biosciences*, 23(3):351–379, 1975.

[10] Stuart F Quan, Barbara V Howard, Conrad Iber, et al. The sleep heart health study: design, rationale, and methods. *Sleep*, (20):1077–85, 1998.