A Hybrid Approach for Artifact Detection in EEG Data

Jacqueline Fairley[1], George Georgoulas[2], Chrysostomos Stylios[2], and David Rye[1]

[1] Emory University, School of Medicine
Dept. of Neurology, Atlanta, Georgia, USA
{jafairl,drye}@emory.edu
[2] Laboratory of Knowledge and Intelligent Computing, Dept. of Informatics and Communications Technology, TEI of Epirus, Kostakioi, 47100, Artas Greece georgoul@gmail.com, stylios@teiep.gr

Abstract. This paper presents a hybrid approach for extreme artifact detection in electroencephalogram (EEG) data, recorded as part of the polysomnogram (psg). The approach is based on the selection of an "optimal" set of features guided by an evolutionary algorithm and a novelty detector based on Parzen window estimation, whose kernel parameter *h* is also selected by the evolutionary algorithm. The results here suggest that this approach could be very helpful in cases of absence of artifacts during the training process.

Keywords: Genetic Algorithms, Feature Selection, Parzen Novelty Detection, Artifact Detection.

1 Introduction

Valid automated computer-based sleep analysis system development relies upon the creation of efficient automated computer-based artifact processing methodologies [1]. This paper introduces a hybrid computational based method for the automated detection of a commonly observed electroencelphalogram (EEG) artifact within human psg data [2]. The artifact of interest is usually created by excessive patient movement (EPM), which is visually characterized by increased signal amplitude and variance values within the EEG. However, the characteristics of EPM make the underlying physiological EEG signal attributes visually unrecognizable and interfere with sleep technician and physician psg analysis [3]. During automated/computerized psg analysis these body movements may also be misinterpreted [4].

Due to the common occurrence of EPM artifacts within human psg recordings the signal analysis approaches obtained within this study are vital to the establishment of an efficient automated computer-based artifact processing methodology. Implementation of the latter will advance the development of a valid automated computer-based sleep analysis system, which will directly impact the diagnosis and treatment of people affected by sleep related illnesses. Further emphasis on the importance of this work is provided by the National Institutes of Health which states, "At least 40 million

K. Diamantaras, W. Duch, L.S. Iliadis (Eds.): ICANN 2010, Part I, LNCS 6352, pp. 436–441, 2010.

Americans each year suffer from chronic, long-term sleep disorders each year. These disorders account for an estimated $16 billion in medical costs each year" [5].

Presently, no standardized approaches for psg computer-based automated artifact removal and/or compensation are widely accepted within clinical practice of human sleep analysis. However, two main research approaches are used to address artifacts within human psg data and have been cited in the literature, which include psg artifact prevention and treatment [1], [2]. The second approach, artifact treatment focuses, on artifact removal by utilizing computational artifact data elimination and compensation techniques. The primary focus of this work is computational artifact data elimination.

The artifact problem can be formulated as a two class classification problem. However when one class is either under-sampled or not present at all, then the problem becomes more difficult and a different approach is needed. In the latter case a novelty detector can be used [6], which attempts to model, only, the known class.

2 Materials and Methods

As mentioned in the introductory section, the main idea behind this research work is to treat the artifact detection as a novelty (anomaly) detection or as a one class classification problem [6]. By this approach, we treat artifacts as anomalous situations and we focus on modeling the normal EEG behavior. Deviations from the "Standard EEG Model" are considered artifact.

As in the general two-class (or multi-class) classification problems, we usually have to move from the original space (the "raw data" space) to a feature space of a (much) smaller dimension through a feature extraction process. By doing so, we hope to condense the relevant information and get rid of potential "noise" and also alleviate the problem of the curse of dimensionality. Therefore we often tend to extract more features than are necessary based upon expert knowledge and intuition and then employ a feature selection stage to come up with a near optimum set of variables [7].

2.1 Data Description

The proposed work was tested using a psg record sampled at 200 Hz for a total duration of 7.25 hours provided in compliance with Emory University Institutional Review Board protocol by the Emory Clinic Sleep Disorders Center (ECSDC) located in Atlanta, Georgia, USA.

In order to extract the psg recording, surface electrodes from calibrated sleep monitoring equipment were attached to subjects by sleep technicians at the ECSDC. The electroencephalogram (EEG) data channel C3-A2 was extracted from the central electrode (C3) and referenced to anterior electrodes (A2), according to the international 10-20 electrode placement system [8].

A visual example of a 30 second epoch/segment of EPM artifact contamination in the EEG is displayed in the top panel of Figure 1. Increased EEG signal amplitude (C3-A2) is shown in the vertical axis displaying an amplitude value exceeding 700μV indicated by the arrow.

Twenty features (Table 1) were extracted from 1 sec time windows, EEG (C3-A2), after consultations with sleep physicians at ECSDC, an exhaustive review of prior bio-signal/psg data artifact detection methods, and a detailed visual analysis of the signal characteristics of EPM artifact and Non-artifact corrupted psg data sets [9].

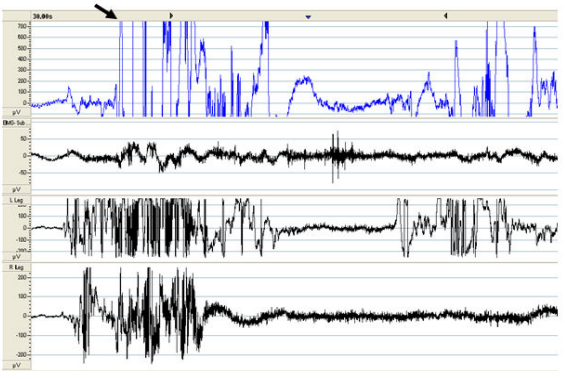

Fig. 1. Example of EPM artifact displayed in the top panel EEG channel

2.2 Genetic Algorithm Feature Selection Stage

Feature selection can be performed in more than one way utilizing different criteria [13]. In this work, a wrapper [10] approach was adopted to assist in selecting a set of features that maximizes a portion of the area under the ROC curve (AUC) [11].

More specifically a GA is utilized to select the features that are used by the novelty detector (which at the same time tries to optimize a design parameter of the detector). The GA population consists of binary chromosomes divided into five competing subpopulations [12]. Each of the subpopulations (containing different mutation rate ranges that provide varying degrees of search space exploration) competes for resources with the subpopulation having the best performance given a greater number of chromosomes/individuals [12].

The GA was implemented using the GEATBx toolbox [12] and it run for 500 "generations". Apart from twenty bits reserved to encode the selected set, we also genetically encoded the kernel parameter h of the Parzen detector using ten more bits.

Fitness Function. Due to the imbalanced nature of the data set, the selection of accuracy (overall classification rate) as a metric is not the best choice for this case. Thus, the classifiers are compared by using their corresponding receiver operating characteristic (ROC) curves. The ROC curve plots the True Positive (TP) rate (such that positive are the artifact free segments and negative are artifact segments) against the False Positive (FP) rate as an acceptance threshold is varied (Figure 2). AUC is a single scalar value that can be used for classifier comparison [11].

For this work because we were also looking for a high True Negative (TN) rate we substituted the AUC for a portion of the area that was between specific (small) values (0.005-0.05) of the FP rate which corresponded to high TN rates (Figure 2).

Table 1. Extracted Features

Feature Number	Feature Name	Symbol
1	Mean Absolute Amplitude	MAA
2	Curve Length	L
3	Mean Energy	MnE
4	Power Spectrum Sub-band, Delta Power	{Delta}
5	Power Spectrum Sub-band, Theta Power	{Theta}
6	Power Spectrum Sub-band, Alpha Power	{Alpha}
7	Power Spectrum Sub-band, Spindle Power	{Spindle
8	Power Spectrum Sub-band, Beta Power	{Beta}
9	Mean/Expected Value	E
10	Variance	Var
11	Standard Deviation	Std
12	Spectral Edge Frequency	SEF
13	Kurtosis	Kurt
14	Skewness	Skew
15	Mobility	Mobi
16	Complexity	CmP
17	Zero Crossing	C
18	Entropy	EnT
19	75^{th} Amplitude Percentile	75 Amp
20	Non-Linear Energy	NE

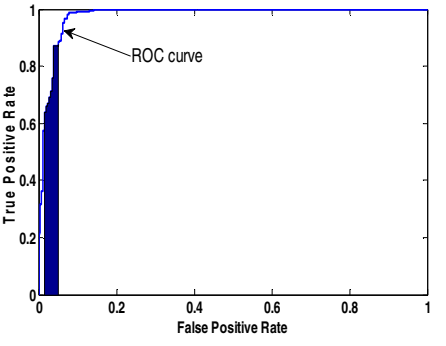

Fig. 2. Typical ROC curve for an imbalanced data set ("negative" class under-sampled)

2.3 Parzen Novelty Detector

In this work the Parzen window method is used to estimate the probability density function of the training data (the underlying stochastic model). It is a non-parametric kernel based method and the most widely used kernel is the Gaussian kernel which is controlled by a single parameter *h*. The latter is employed with *h* being automatically selected by the GA (10 bit encoding –(0.005-2.0 phenotype)).

The Parzen detector was implemented using the Data Description Toolbox [13]. The threshold to decide whether a sample comes from the underlying statistical model

is put such that a fraction (a user defined parameter) of the training objects is rejected (0.01 in our case).

2.4 Experimental Procedure

In order to test the proposed approach we used manually labeled EPM artifacts within the EEG data sets. Manual/expert EPM artifact labeling was based upon instruction from ECSDC physicians and technicians in visual artifact identification. A total of 26098 epochs (25882 not artifact and 216 artifact epochs) were involved in this work.

We employed the *k*-fold cross validation with *k* set equal to five. Due to the abundant number of non-artifact samples we used only part of the available non-artifact data in order to reduce processing time. After that, we divided the non-artifact training data into two sets and we used 70% of them to train the novelty detector (to build the statistical model and select the corresponding threshold) and the remaining 30% along with the corresponding artifact data to calculate the performance measure (the portion of the AUC as described above).

3 Results

The procedure described in section 2.4 was repeated 5 times and the results were averaged. The overall achieved performance was 98.44%±0.41 (mean±standard deviation) for the non artifact segments and 80.93%±11.33 for the artifacts. The TP rate is close to the expected value since a 0.01 rejection error was selected during the training process. The TN rate is worse but it can be improved on the expense however of the TP rate.

The GA always selected the curve length and the standard deviation in all but one repetition. Surprisingly it did not select the, variance which might be considered a viable feature based upon Figure1. It also frequently selected the mean energy and the nonlinear energy. The highest number of features selected were 8 with the most occurrences at 5 features being selected 7 times and the lease occurrences with 1 feature being selected zero times. Therefore, the GA was "biased" toward parsimonious solutions, selecting on average five (out of the original twenty) features.

4 Conclusions

In this paper, we proposed a novel hybrid approach to artifact detection based on a combination of a GA algorithm for feature selection and a novelty detector. The results indicate that this approach can be used as an alternative to the standard two class classification approach especially when the information about the artifact class is missing.

Even though the feature selection module was not directly dictated to favor solutions with a lower number of features it frequently used only ¼ of the original variables resulting in a more compact representation of the problem. It is important to mention that most of the features were selected in a concise manner revealing a certain inherent pattern of the problem.

On the other hand the results are not as good as in the case of a two class formulation [14]. Nevertheless they are promising and the proposed approach can be applied toward cases where minimal to zero information about the artifacts exist.

Acknowledgments. The authors would like to thank the National Institute of Neurological Disorders and Stroke along with the National Science Foundation sponsored program Facilitating Academic Careers in Engineering and Science (FACES) at the Georgia Institute of Technology and Emory University for providing research funding for this project. Authors would also like to thank the Operational Programme Education and Lifelong Learning of the Greek Ministry of Education, Lifelong Learning and Religious Affairs co-financed by the Greece and the European Union for its sponsoring. Gratitude is also extended to the technicians at ECSDC for data collection and technical assistance in psg interpretation.

References

1. Anderer, P., Roberts, S., Schlogl, A., Gruber, G., Klosch, G., Herrmann, W., Rappelsberger, P., Filz, O., Barbanoj, M.J., Dorffner, G., Saletu, B.: Artifact processing in computerized analysis of sleep EEG – a review. Neuropsychobiology 40, 150–157 (1999)
2. Klass, D.W.: The continuing challenge of artifacts in the EEG. American Journal of Eeg Technology 35(4), 239–269 (1995)
3. Butkov, N.: Atlas of Clinical Polysomnography, vol. I. Ashland, Synapse Media Inc. (1996)
4. Foundation, N.S.: Sleep in America Poll (2005), http://www.sleepfoundation.org
5. NINDS, N.: Brain Basics: Understanding Sleep in Brain Resources and Information Network (BRAIN) NINDS, Editor. National Institutes of Health, Bethesda (2007)
6. Tax, D.M.J.: One-class classification; concept-learning in the absence of counter-examples. Ph.D. thesis, Delft University of Technology (2001)
7. Sa, M.J.P.: Pattern recognition. Concepts, methods, and applications. Springer, Heidelberg (2001)
8. Bloch, K.E.: Polysomnography: a systematic review. Technol. Health Care 5(4), 285–305 (1997)
9. Fairley, J.: Statistical Modeling of the Human Sleep Process via Physiological Recordings, Ph.D. thesis, Electrical and Computer Engineering, p.167, Georgia Institute of Technology: Atlanta. (2008)
10. Guyon, I., Elisseeff, A.: An introduction to variable and feature selection. J. Machine Learning Research 3, 1157–1182 (2003)
11. Bradley, A.P.: The use of the area under the ROC curve in the evaluation of machine learning algorithms. Pattern Recognition 30(7), 1145–1159 (1997)
12. GEATbx The Genetic and Evolutionary Algorithm Toolbox for Matlab, http://www.geatbx.com/
13. The Data Description Toolbox, http://homepage.tudelft.nl/n9d04/dd_tools.html
14. Fairley, J., Johnson, A., Georgoulas, G., Vachtsevanos, G., Rye, D.: Multiple Intelligent Feature Selection vs. Single Feature Thresholding for Extreme Artifact Detection in EEG Data (2010) (in preparation)