Unsupervised Analysis of EEG signals for the discovery of ictal patterns

Evangelia Pippa, Iosif Mporas
Dept. of Computer Engineering and Informatics
University of Patras
Rion-Patras, Greece
pippa@ceid.upatras.gr, imporas@upatras.gr

***Abstract*— In this paper, patient-specific unsupervised analysis of electroencephalographic (EEG) signals for the discovery of unknown ictal patterns was performed. For the analysis we relied on the k-means clustering algorithm. The proposed methodology when applied to EEG recordings from an epileptic patient revealed patterns that appear with a significantly greater frequency during the epileptic seizures. The discovered patterns can be exploited as indicators of the ictal periods in an automated approach for the detection of seizures.**

Keywords: clustering; EEG; epileptic seizures; pattern discovery; machine learning

I. INTRODUCTION

Electroencephalography (EEG) is a technology developed to investigate the brain by measuring the electric fields generated by neurons in the cortex [13]. Electrodes, which act as the sensors to detect the electrical activity, are attached to the surface of the cerebral cortex and are connected by wires to a computer to capture the brain activities. EEG is non invasive and provides information which is retrievable, easily recorded and is obtainable using inexpensive technology compared with other brain activity diagnosis methods [12]. EEG has two main advantages; direct measure of neuronal activity and best temporal resolution by milliseconds [10]. In addition to the profound impact on our understanding of the brain [2,17], EEG signals are important to capture brain disorders. They are useful for analyzing the cognitive activity of the brain and diagnosing types of seizure and potential mental and neurological health problems, such as epilepsy [12].

Epilepsy is a chronic disorder of the central nervous system that predisposes individuals to experiencing recurrent seizures [14]. Approximately 1% of the world population suffers from seizures, while more than 20% of the epileptic patients suffer from seizures that are refractory to medication [3,15]. Epileptic seizures are brief episodes of abnormal excessive or synchronous neuronal activity in the brain of patients suffering from epilepsy [4]. During an epileptic seizure there are several specific changes recorded in the electroencephalogram (EEG) which is a sensitive and important test used to evaluate patients with suspected epilepsy. There are certain characteristic ictal neurophysiological patterns, already known by the clinical experts, that support the identification and detection of epileptic events and postictal and/or interictal abnormalities that can provide supplementary information.

The detection of such already well-known patterns related to epilepsy (e.g. spikes or spikes and waves) is a common problem that is frequently encountered in the literature [1,7,19]. Several powerful machine learning algorithms have been proposed for the detection of ictal EEG periods, as well. The most popular use supervised machine learning techniques including artificial neural networks [9, 11] and support vector machines [5,15,18]. However, the main goal of our work is not to use existing medical knowledge (clinical guidelines, etc.) to identify patterns but rather to contribute to medicine by discovering new knowledge, this way assisting in better understanding of epilepsy and related disorders as well as their manifestations.

Therefore, in this paper, we propose an unsupervised method for the discovery of potentially unknown electroencephalographic ictal patterns from epileptic patients. The proposed methodology reveals patterns which appear significantly more frequently during the epileptic seizures. The discovered patterns are strong indicators of the ictal class and thus, can be exploited as tools for the automatic detection of epileptic seizures.

The rest of this paper is organized as follows. In Section II the proposed methodology is presented. Section III provides details about the evaluation data and the experimental protocol followed and Section IV presents the achieved results. Finally in Section V we conclude this work.

II. METHODOLOGY

The proposed methodology for the discovery of potentially unknown patterns from the EEG data of a single channel consists of the following steps:

- Frame blocking
- Feature Extraction
- Clustering
- Post-processing

The block diagram of the overall architecture is illustrated in Fig. 1. The input to the illustrated architecture consists of

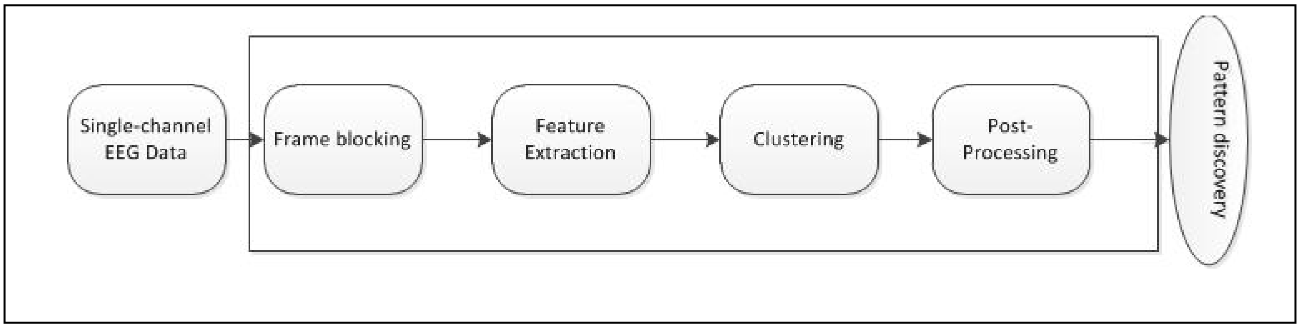

Figure 1. Proposed architecture for unsupervised pattern discovery from EEG

EEG signal samples from a single channel. As shown in Fig. 1, in a first step the EEG signal is preprocessed. Preprocessing consists of frame blocking of the incoming samples to epochs of constant length w with constant time-shift s. After preprocessing the extracted epochs are in parallel processed by a frequency-domain feature extraction algorithm. Here, the feature extraction algorithm computes the power spectral density of each epoch separately. Power spectral density (PSD) of a signal is a positive real function of a frequency variable associated with a stationary stochastic process. The PSD is defined as the discrete time Fourier transform (DTFT) of the autocorrelation sequence (ACS) [16].

$$\varphi(\omega) = \sum_{k=-\infty}^{\infty} r(k) e^{-i\omega k} \quad (1)$$

where the auto covariance sequence is defined as

$$r(k) = E\{y(t) y^{*}(t-k)\} \quad (2)$$

and $y(t)$ is the discrete time signal $\{y(t); t = 0, \pm 1, \pm 2, \dots\}$ assumed to be a sequence of random variables with zero mean.

After feature extraction, each epoch is represented by a single feature vector containing its power spectral density. The feature vectors derived from the whole set of epochs of the segmented signal are clustered using k-means clustering algorithm [8]. Given a set of observations, where each observation is a d-dimensional feature vector containing the power spectral density of each epoch, k-means clustering partitions the observations into k sets so as to minimize the within-cluster sum of squares using an iterative refinement technique which alternates between two steps:

- Assignment step: Assign each observation to the cluster with the "nearest" mean.
- Update step: Calculate the new means to be the centroids of the observations in the new clusters.

The algorithm converges to a (local) optimum when the assignments no longer change. There is no guarantee that the global optimum is found.

In a final post-processing step, the percentage of appearance of each cluster in the ictal and non-ictal epochs is calculated. Clusters with similar percentages of appearance in the two classes (ictal and non-ictal) are ignored since they do not carry any useful information for the discrimination of the two classes. On the other hand, clusters with significantly different percentages are captured since they indicate patterns related to the class in which they present the high percentage. Such patterns can be used as indicators of the class they appear in an automated method for the detection of epileptic seizures.

III. Experimental Setup

The previously described methodology for pattern discovery was evaluated on EEG recordings contained within the CHB-MIT database [14], which can be downloaded from the PhysioNet website [1]. This database, collected at the Children's Hospital Boston, consists of EEG recordings from pediatric subjects with intractable seizures. Subjects were monitored for up to several days following withdrawal of anti-seizure medication in order to characterize their seizures and assess their candidacy for surgical intervention. For each clinical seizure, an expert indicated the earliest EEG change associated with the seizure. The data was segmented into one hour long records. The data set consists of 916 hours of continuous scalp EEG sampled at 256 Hz. Records that do not contain a seizure are called non-seizure records and those that contain one or more seizures are called seizure records. All data were stored in EDF formatted files [6]. For the current analysis, one subject from the CHB-MIT database was chosen randomly and the selected EEG channel was CZ-PZ. Annotations for seven seizures were included in the recording.

During pre-processing the signal from the CZ-PZ channel was frame blocked to epochs of 1 second length, without time-overlap between successive epochs. The spectral power in 2Hz frequency bins from 1 to 30Hz was computed for each epoch. The computed feature vectors were used to cluster the whole set of epochs of the segmented signal using k=8, 16, 32 clusters. Finally, the percentage of the appearance of each cluster in the ictal and non-ictal epochs was calculated for each one of the clustering cases to reveal relevant patterns.

IV. Results

The unsupervised method presented in Section II was applied to the 10th subject of the CHB-MIT database following the experimental setup described in Section III. Fig. 2 shows distribution of the clusters among the two classes (ictal and non-ictal) for the first clustering case (k=8). As it can be seen, when eight clusters are used, the ictal period of the signal

[1] http://physionet.org/physiobank/database/chbmit/

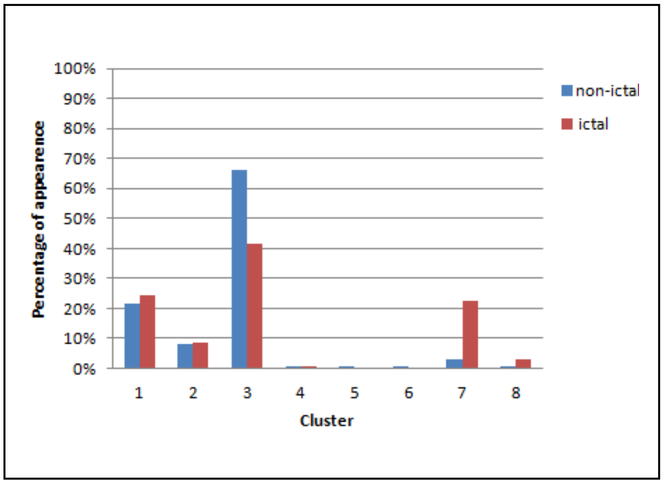

Figure 2. Appearence of each cluster in ictal and non-ictal epochs (k=8).

consists of cluster 7 by 22.69% while the non-ictal period contains only 3.19% the 7th cluster. As a result, for the specific patient, cluster 7 can be considered as a strong indicator of the presence of ictal state. Table I shows the percentage of appearance of each cluster in the set of ictal and non-ictal epochs when k=8 clusters were used with k-means.

Fig. 3 shows the distribution of the clusters among the two classes when k=16 clusters are used with k-means. Here, the ictal period of the signal consists of cluster 10 by 21.58% while the non-ictal period contains 10.54% the 10th cluster, a quite lower percentage. Furthermore, the non-ictal class is highly represented by cluster 1 (50.68%) when only the half of this percentage (24.23%) corresponds to ictal epochs represented by the first cluster. The distribution of the clusters among the two classes for this clustering case (k=16) are shown on Tables II and III.

Finally, Fig. 4 shows the distribution of the clusters among the two classes (ictal and non-ictal) for the last clustering case (k=32). Once again, patterns with discriminative power among the two classes such as cluster 4, cluster 19 and cluster 20 were revealed. The ictal period of the signal consists of cluster 4, cluster 19 and cluster 20 by 7.45%, 13.88% and 6.39% respectively while only 0.3%, 1.42% and 0.06% of non-ictal epochs correspond to the 4th, 19th and 20th clusters, respectively.

The discovered patterns indicate the existence of underlying information, related to seizure characteristics, within the electroencephalographic epochs clustered to groups with significantly different frequency of appearance within the ictal class. As a result the clusters with such a discriminative power among the two classes require further examination by clinical experts for the clinical interpretation of the results and the potential discovery of knowledge that remains unknown so far.

V. CONCLUSION

In this paper, we investigated the problem of the discovery of potentially unknown ictal patterns from single-channel EEG using clustering of spectral features. The experimental results of the proposed methodology on single channel data from one subject revealed patterns related to both ictal and non-ictal class. We deem the application of our methodology across

TABLE I. PERCENTAGE OF APPEARENCE OF EACH CLUSTER IN ICTAL AND NON-ICTAL EPOCHS (K=8)

k=8	**Clusters**							
	1	***2***	***3***	***4***	***5***	***6***	***7***	***8***
Non-Ictal (%)	21.5	8.1	66.1	0.1	0.01	0.2	3.2	0.8
Ictal (%)	24.5	8.4	41.1	0.2	0.00	0.0	22.7	2.9

TABLE II. PERCENTAGE OF APPEARENCE OF EACH CLUSTER IN ICTAL AND NON-ICTAL EPOCHS (K=16).

k=16	**Clusters**							
	1	***2***	***3***	***4***	***5***	***6***	***7***	***8***
Non-Ictal (%)	50.7	0.02	6.1	0.9	0.1	0.5	2.3	3.0
Ictal (%)	24.2	0.22	3.3	7.5	0.0	0.6	0.8	4.8

TABLE III. PERCENTAGE OF APPEARENCE OF EACH CLUSTER IN ICTAL AND NON-ICTAL EPOCHS (K=16)

k=16	**Clusters**							
	9	***10***	***11***	***12***	***13***	***14***	***15***	***16***
Non-Ictal (%)	1.4	10.5	0.01	14.9	0.1	0.04	0.2	9.2
Ictal (%)	12.6	21.6	0.00	9.0	6.6	0.22	0.0	8.4

subjects will reveal more generalized EEG patterns related to the epileptic seizures. Finally, we aim to use these patterns as tools to discriminate between ictal and non-ictal epochs as part of a wider framework for the automatic detection of epileptic seizures in the future.

ACKNOWLEDGMENT

This study is partially funded by the EC under the FP7/2007-2013 with grant ARMOR, Agreement Number 287720. This research has been co-financed by the European Union (European Social Fund – ESF) and Greek national funds through the Operational Program "Education and Lifelong Learning" of the NSRF - Research Funding Program: Thales. Investing in knowledge society through the European Social Fund.

REFERENCES

[1] M. Adjouadi, M. Cabrerizo, M. Ayala, D. Sanchez, I. Yaylali, P. Jayakar and A. Barreto, “Detection of Interictal Spikes and Artifactual Data Through Orthogonal Transformations, Journal of Clinical Neurophysiology, vol. 22, pp. 53-64, 2005.

[2] G. Buzsaki, “Rhythms of the brain,” Oxford University Press, 2006.

[3] J. Corsini, L. Shoker, S. Sanei and G. Alarcón, “Epileptic seizure predictability from scalp EEG incorporating constrained blind source separation,” IEEE Trans. Biomed Eng., vol. 53, pp. 790-9, 2006.

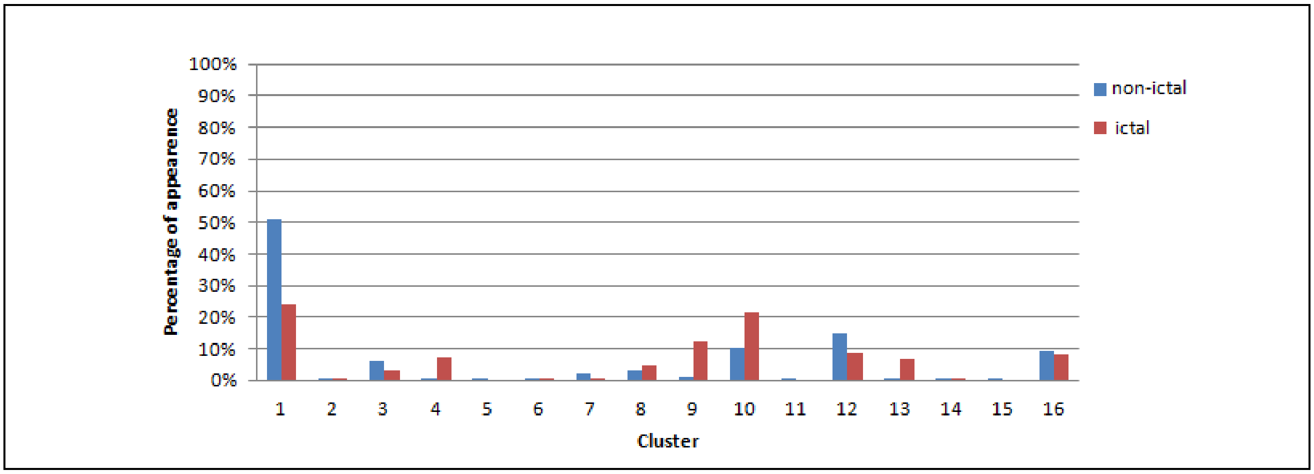

Figure 3. Appearence of each cluster in ictal and non-ictal epochs (k=16).

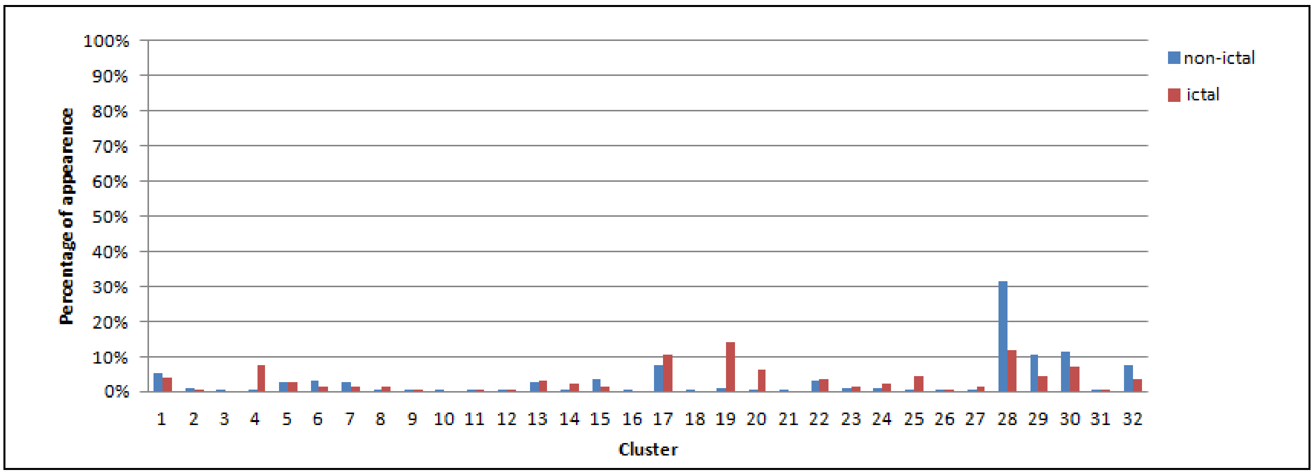

Figure 4. Appearence of each cluster in ictal and non-ictal epochs (k=32).

[4] R. S. Fisher, W. V. E. Boas, W. Blume, C. Elger, P. Genton, P. Lee and J. Engel, “Epileptic seizures and epilepsy: definitions proposed by the International League Against Epilepsy (ILAE) and the International Bureau for Epilepsy (IBE) ,” Epilepsia, vol. 46, pp. 470-472, 2005.

[5] B. Hunyadi, M. Signoretto, W. Van Paesschen, J.A. Suykens, S. Van Huffel and M. De Vos, “Incorporating structural information from the multichannel EEG improves patient-specific seizure detection,” Clinical Neurophysiology, vol. 123, pp. 2352-61, 2012.

[6] B. Kemp and J. Olivan, “European data format 'plus' (EDF+), an EDF alike standard format for the exchange of physiological data,” Clinical Neurophysiology, vol. 114, pp. 1755–61, 2003.

[7] Y. Kultu, M. Kuntalp and Kuntalp D, “Optimizing the performance of an MLP classifier for the automatic detection of epileptic spikes,” Expert Systems with Applications, vol. 36, pp. 7567-7575, 2009.

[8] J. B. MacQueen, “Some Methods for classification and Analysis of Multivariate Observations,” Proceedings of 5th Berkeley Symposium on Mathematical Statistics and Probability, University of California Press., pp. 281–297, 1967.

[9] H.R. Mohseni, A. Maghsoudi and M.B. Shamsollahi, “Seizure detection in EEG signals: a comparison of different approaches,” In Conf. Proc IEEE Eng Med Biol Soc., pp. 6724-7, 2006.

[10] S.A. Mousavi, M.R.H. Arshad, H.H. Mohamed and S.A. Alomari, “An efficient method for mining Event-Related Potential Patterns,” International Journal of Computer Science Issues, vol. 8, 2011.

[11] S. Nasehi and H. Pourghassem, “Seizure Detection Algorithms Based on Analysis of EEG and ECG Signals: a Survey,” Neurophysiology, vol.44, pp. 174-186, 2012.

[12] K.S. Ng, H.J. Yang and S.H. Kim, “Hidden pattern discovery on event related potential EEG signals, Biosystems,” vol. 97, pp. 15-27, 2009.

[13] E. Niedermeyer and F. L. da Silva, “Electroencephalography: Basic Principles, Clinical Applications, and Related Fields,” Lippincott Williams and Wilkins, 2005

[14] A.H. Shoeb, “Application of machine learning to epileptic seizure onset detection and treatment,” PhD Thesis, Harvard-MIT Division of Health Sciences and Technology, 2009.

[15] A. Shoeb, H. Edwards, J. Connolly, B. Bourgeois, S.T. Treves and J. Guttag, “Patient-specific seizure onset detection,” Epilepsy Behavior, vol. 5, pp.483-98, 2004.

[16] Stoica and R. L. Moses, “Spectral analysis of signals,” Pearson/ Prentice Hall Upper Saddle River, NJ, 2005.

[17] B. Swartz and E. Goldensohn, “Timeline of the history of EEG and associated fields,” Electroencephalography and clinical Neurophysiology, vol. 106 , pp.173–176, 1998.

[18] M. Valderrama, S. Nikolopoulos, C. Adam, V. Navarro, M. Le Van Quyen, “Patient-specific seizure prediction using a multi-feature and multi-modal EEG-ECG classification,” XII Med. Conf. on Medical and Biological Engineering and Computing, , vol. 29, pp.77-80, 2010.

[19] G. Xu, J. Wang, Q. Zhang, S. Zhang and J. Zhu, “A spike detection method in EEG based on improved morphological filter,” Computers in Biology and Medicine, vol. 37, pp. 1647-1652, 2007.