Biomedical Signal Processing and Control 40 (2018) 275–285

Contents lists available at ScienceDirect

Biomedical Signal Processing and Control

journal homepage: www.elsevier.com/locate/bspc

A robust unsupervised epileptic seizure detection methodology to accelerate large EEG database evaluation

Kostas M. Tsiouris [a,d], Sofia Markoula [b], Spiros Konitsiotis [c], Dimitrios. D. Koutsouris [a], Dimitrios I. Fotiadis [d,e,*]

[a] Biomedical Engineering Laboratory, School of Electrical and Computer Engineering, National Technical University of Athens, GR15773, Athens, Greece
[b] University Hospital of Ioannina, GR45110, Ioannina, Greece
[c] Department of Neurology, Medical School, University of Ioannina, GR45110, Ioannina, Greece
[d] Unit of Medical Technology and Intelligent Information Systems, Department of Materials Science and Engineering, University of Ioannina, GR45110, Ioannina, Greece
[e] Department of Biomedical Research, Institute of Molecular Biology and Biotechnology, FORTH, University of Ioannina, GR45110, Ioannina, Greece

ARTICLE INFO

Article history:
Received 31 March 2017
Received in revised form
15 September 2017
Accepted 30 September 2017

Keywords:
EEG
Epilepsy
Unsupervised seizure detection
Medical knowledge
Time-Frequency analysis

ABSTRACT

In this work an unsupervised methodology for the detection of epileptic seizures in long-term EEG recordings is presented. The design of the methodology exploits the available medical knowledge to tackle the lack of training data using a simple rule-based seizure detection logic, avoiding complex decision making systems, training and empirical thresholds. The Short-Time Fourier Transform is initially applied to extract the EEG signal energy distribution over the delta (<4 Hz), theta (4–7 Hz) and alpha (8–13 Hz) frequency bands. A set of four novel seizure detection conditions is proposed to isolate EEG segments with increased potential of containing ictal activity, by identifying segments where the EEG signal energy is intensively accumulated among the three fundamental frequency rhythms. A set of candidate seizure segments is extracted based on the intensity of the accumulated EEG activity per seizure detection condition. The clinician has to visually inspect only the extracted segments instead of the entire duration of the patient's EEG recordings to speed up the annotation process. The results from the evaluation with 24 cases of long-term EEG recordings, suggest that the proposed methodology can reach on average up to 89% of seizure detection sensitivity, by automatically rejecting 95% of the total patient's EEG recordings as non-ictal, without requiring any *apriori* data knowledge.

1. Introduction

Epilepsy is a neurological condition expressed with the occurrence of seizures that alter the normal electrical activity between the brain's neurons, leading to various clinical manifestations depending on the affected area of the brain. As much as 50 million people are estimated to suffer from epilepsy, 30% of whom do not respond to treatment with anti-epileptic medication (WHO, 2012). The electroencephalogram (EEG) is used to capture the electrical activity of the brain. It is considered the golden standard for epilepsy diagnosis and is the most cost effective means for long-term monitoring of patients with any suspected or diagnosed epileptic syndromes [1]. Nowadays, continuous recording of EEG signals, for up to several days, is a very common practice due to cost reduction in data storage. However, this practice yields extensive volumes of EEG data that have to be carefully inspected for signs of ictal activity in order to isolate seizure segments for further analysis. Visual inspection of large volumes of EEG signals has proven to be an extremely time consuming and tedious task even for domain experts. In the direction of relieving part of the annotation effort and automating the detection of epileptic seizures, machine learning techniques and unsupervised methodologies have been introduced in the analysis of EEG signals.

Over the past few years, many supervised methodologies have been proposed for the detection of epileptic seizures, highlighting the need for developing tools for faster annotation of large volumes of EEG data even in non-seizure detection applications [2,3]. The vast majority of the available machine learning techniques have been used in seizure detection, ranging from the more simplistic ones such as the k-nearest neighbors algorithm (k-NN) [4], to the very popular Support Vector Machines (SVMs) [5] and the more

* Corresponding author at: Unit of Medical Technology and Intelligent Information Systems, Department of Materials Science and Engineering, University of Ioannina, GR45110, Ioannina, Greece.
E-mail address: fotiadis@cc.uoi.gr (D.I. Fotiadis).

https://doi.org/10.1016/j.bspc.2017.09.029
1746-8094/

computationally demanding neural networks (NNs) [6,7]. The classification process was often complemented with the addition of clustering techniques, which exploit the EEG signal's variations to extract clusters of similar activity in an attempt to increase the classifier's seizure detection performance [8,9]. Besides clustering and classification, rule-based seizure detection methodologies have also been studied as an alternative solution. In these methodologies, a wide variety of EEG signal behavior criteria and/or threshold-crossing parameters have been combined to declare the occurrence of ictal activity when triggered [10–13]. A set of training EEG data was required to extract the seizure detection criteria and to fine tune the seizure detection thresholds, prior to the evaluation with the testing subset. Furthermore, seizure detection methodologies based on fuzzy rule-based decision-making systems have been tested as well [14], showing great potential in mimicking the EEG expert's reasoning in declaring ictal activity.

In an attempt to demonstrate the applicability of machine learning techniques in everyday clinical practice, researchers have also introduced external validation of their methodologies with out of sample EEG data. In most cases a classifier (i.e. SVM, k-NN, Bayesian and Neural Networks) was trained using a set of EEG recordings from a subgroup of the available patients and was then validated using the rest of the EEG dataset from the remaining patients [15–17]. The classifier's ability to adapt to new patients was evaluated by measuring its seizure detection performance on the latter. In more recent publications, researchers have also investigated the potential gains from using a much wider set of features when training their classifiers, in order to address their impact in classification performance [18,19]. In other such systems, instead of focusing on classification techniques, the researchers used the EEG recordings from the training group to extract seizure detection criteria and/or adjust feature-related detection thresholds, in an attempt to optimize their methodologies for better out of sample detection performance [20–22].

Compared to unsupervised methodologies, the main drawback of using machine learning techniques such as the above, is that they require pre-annotated EEG data or some kind of *apriori* information during the training process. On the other hand, the challenge in using unsupervised techniques is to find a universal methodology that would adapt to multiple patients. The possibility of developing an unsupervised seizure detection methodology that attributes variations in the EEG signals to the occurrence of epileptic seizures has been under investigation over the past few years. Researchers have initially shown that ictal EEG activity caused specific ictal EEG signal manifestations that were uncommon during the interictal state or further away from the epileptogenic zone [23,24]. These works, however, examined small segments of EEG recordings (usually a seizure and a few minutes of EEG before and after) showing a proof of concept rather than a complete evaluation. Recently, Birjandtalab et al. [25] presented an unsupervised methodology for the detection of epileptic EEG activity using features extracted from spectral analysis and clustering, further enhancing the semi-automated framework set by Smart and Chen [26]. These methodologies were evaluated using the same EEG database as in this work, but only a few of the available cases and only seizure adjacent EEG segments. However, despite the limited dataset, the results provided significant evidence for the effectiveness of EEG spectral analysis in identifying ictal activity in an unsupervised manner.

In this work, a methodology is proposed to detect and isolate segments with epileptic EEG activity in a completely unsupervised manner, based on four novel seizure detection conditions (Section 2.3). Using the proposed conditions, the EEG segments that are more likely to contain epileptic activity are automatically isolated for visual inspection and validation without requiring any user intervention, *apriori* information about the patient or any EEG data for training. The advantage of the proposed conditions is that they have been formulated combining the available medical knowledge and understanding regarding the different expressions of ictal activity in EEG signals and previous literature findings of EEG signal processing techniques in the automated seizure detection field. Our primary goal is to provide a system that offers high seizure detection rates while drastically reducing the time required for the annotation of long-term EEG recordings. Furthermore, to the best of our knowledge, this is the first time that an unsupervised seizure detection methodology has been evaluated using such an extensive volume of EEG data from a public database, consisting of about 983 h of scalp EEG recordings and 198 seizures [27].

2. Materials and methods

2.1. EEG dataset

The CHB-MIT EEG database contains scalp EEG signals from 23 subjects with intractable epileptic seizures (5 males, aged 3–22 years old; 17 females, aged 1.5-19 years old; 1 missing gender/age data), recorded at the Children's Hospital Boston in cooperation with the Massachusetts Institute of Technology [28]. The subjects were monitored for up to several days, following withdrawal of their anti-epileptic medication in order to isolate the epileptogenic area and evaluate the possibility of surgical intervention. In total, there are more than 980 h of raw EEG signals available, collected following the International 10–20 standard system of EEG electrode positions. The recordings were organized in 24 cases, each containing EEG data from a single subject, except cases 1 and 21 which were obtained from the same subject with a 1.5 year interval. All signals were sampled at 256 samples per second with 16-bit resolution. As changes in EEG montage were frequent within each case, the number of the available EEG signals varies, with most of the files consisting of 23 signals, while 24 and 26 signals were included in others. During the annotation phase, 198 events were identified as epileptic seizures by clinical experts who marked their electroencephalographic onset and offset after visual inspection. The contents of the CHB-MIT database are presented in full detail in Table 1. The database is available at PhysioNet.org/pn6/chbmit [27,29].

2.2. Medical knowledge background

According to the International League Against Epilepsy, ILAE, an epileptic seizure can be defined as a transient occurrence of signs and/or symptoms due to abnormal excessive or synchronous neuronal activity in the brain [30]. The EEG patterns of epileptic seizures contain one of the well-established spike, wave and spike-wave combinations and usually include dominant bursts of rhythmical activity that begin suddenly and, depending on the type of seizure, may evolve to other regions of the brain during the seizure [31]. That is why the spectral contents of the EEG signals are important for the detection of epileptic seizures. These ictal patterns are most often represented with rhythmical activity within the delta (<4 Hz), theta (4–7 Hz), alpha (8–13 Hz) and gamma (13–30 Hz) waves, or the slow component of the spike-wave complex, in seizures where such EEG activity is expressed.

The medical background is further supported by the findings of previous EEG signal processing methodologies targeting epileptic seizure detection, where the EEG activity within the 1–30 Hz frequency band, in particular, was found to be highly seizure-related [20,32–36]. Furthermore, ictal rhythmical activity is proven to be a solid indication of seizure occurrence and a more reliable way to detect such events compared to spike and/or wave symiological expressions. This is greatly depicted by the prevalence of EEG spec-