A System for Epileptic Seizure Focus Detection Based on EEG Analysis*

Maria J. Santofimia[1], Xavier del Toro[1], Jesus Barba[1], Julio Dondo[1], Francisca Romero[2], Patricia Navas[2], Ana Rubio[2], and Juan C. Lopez[1]

[1] Computer Architecture and Network Group, School of Computing Science, University of Castilla-La Mancha, Spain
[2] Neuroscience Institute, Hospital Regional Universitario Carlos Haya, Malaga, Spain
{MariaJose.Santofimia,Xavier.delToro,Jesus.Barba,Julio.Dondo, JuanCarlos.Lopez}@uclm.es

Abstract. This work presents a recognition system for epileptiform abnormalities based on electroencephalogram (EEG) analysis. The proposed system combines a Support Vector Machine classifier automatically trained by an implementation of machine learning approach known as Bag of Words.

Keywords: Epileptic seizure, epileptic focus, SVM classified, bag of words

1 Introduction

Epilepsy is a disease that, according to [1], approximately affects 1% of the world's population. However, despite the great impact that epilepsy has on society, there are few computational systems and tools supporting doctors during the process diagnosing, treatment, and the identification of those cases that can greatly benefit from surgery interventions.

In fact, the bottleneck is found at the begging of the process, at the diagnosis stage that becomes very time-consuming due to its complexity. According to the neurologist office of the University's Regional Hospital Carlos Haya, approximately 50% of people attended in their offices, regardless of presenting epilepsy symptoms, are afterwards diagnosed as non-epileptic.

The bottleneck is therefore in the analysis of the EEG, that requires from a neurophysiologist going through the EEG logs that should ideally record as much cerebral activity as possible to increase the probability of recording seizure occurrences. This is a tedious and resource consuming activity, which requires very advanced training and experience.

Additionally, as stated in [1], more than 20% of the patients can not be treated with medication. In this cases, the identification of the epileptogenic

* This research was supported by the Spanish Ministry of Science and Innovation under the project DREAMS (TEC2011-28666-C04-03).

focus is essential to determine whether surgery can be an option to overcome medication problems.

One of the first and main attempts to achieve an automatic system to recognize epileptic seizures is found at [2]. This work attempted to characterize, for first time, the problem of analyzing an EEG in seeking for epileptic seizure activity. To date, the main steps of the methodology proposed in that paper remain the same, which basically are: pre-filtering, characterization of wave epochs, and detection. Moreover, the works in [3] and [4] present a state of the art revision of the most relevant methods and advances found in the literature.

Following this new line of research, many different approaches can be found in the literature. Some methods resort to neural networks to train the classifier in detecting the seizures. For example, the work in [5] uses a single-layer neural network for classifying EEG signals that have been characterized by introducing a novel nonlinear feature, the fractal intercept. Some other works either overlook or generally refer to expert training without giving detailed information about how the classifier has been trained. For example, the work in [6] presents an approach for seizure detection in neonatal patients. The work uses a Support Vector Machine (SVM) classifier which has been manually trained by an expert neurophysiologist. The work in [7] also presents a classifier-based system. In this case, a classifier based on a Gaussian mixture model (GMM) is used to distinguish between normal and epilepticform activity. However, that work simply refers to a *trained algorithm* as the mechanism enabling the classification.

The use of machine-learning algorithms in the field of epilepsy has been already proposed for different purposes. The work in [8] proposes the use of a novel machine-learning algorithm for detecting seizure termination. However, the training of the proposed classifier is carried out by means of training vectors rather than being automatically accomplished as proposed here.

From the computational research field point of view, efforts are addressed at devising mechanisms that speed up the time the process of seizure detection and recognition. In this sense, the work in [9] proposes the use of FPGA to implement a Soft Decision Tree algorithm intended to classify the risk of epilepsy.

The reality found at hospitals, however, is quite different. Theoretical systems presented in journals and conferences never reach the doctor's office, probably because most of them are just focusing in detecting epileptic seizures. Interviews with experts (neurologist and neurophysiologist) bring into light that normal activity filtering would be a very valuable feature for computational system supporting their analysis of EEGs.

In this sense, one of the main challenge faced by systems performing automatic analysis of EGGs is the detection of *artifacts*. An artifact is an anomalous waveform generated by certain movements of the patient such as blinking, that does not have and epileptic origin, but introduces complications on the analysis of the EEG. Experts claim that a first filtering of artifacts and the waveforms associated to normal activity would ease the analysis of EEGs and improve the attention received by patients.

Moreover, the detection of an epileptic seizure brings into light a second challenge, as it is the detection of the focus or epileptogenic area. This aspect is essential to determine whether the epileptic patient can go through surgery to eradicate the disease [10].

The majority of the works found in the literature are targeted at identifying epileptic seizures. In order to do so, different signal processing approaches are implemented in combination with classification mechanisms. Having analyzed the requirements and challenges faced by experts in this field, this work proposes a two stage system. A primary artifact filtering stage is followed by a wave classification of EEG activity. In order to do so, a machine learning algorithm, known as Bag of Words is used to train a Support Vector Machine (SVM) classifier. The training and testing data, from anonymous patient, have been provided by the University's Regional Hospital Carlos Haya. The provided dataset consists of EEGs from epileptic and non-epileptic patients. Those not showing any epileptic activity, however, present artifacts, which could be erroneously interpreted as abnormal or epileptic activity.

The rest of the paper is organized as follow. Section 2 presents some relevant background information to better understand the contributions of this work. Section 3 describes the overall system, organized in two stages: training and testing. Section 4 discusses how artifact detection could be improved by combining video sequences with the recordings of the EEG equipment. Section 5 highlights some of the most important conclusions withdrawn from the proposed work.

2 Background on classification methodologies

The main contribution of this work lays on the proposed classifier and automatic training system. Inspired in the work presented in [11], aimed at recognizing human actions from video sequences, this paper proposes a similar methodology applied to EEG analysis and epilepticform wave recognition.

The revision of the state of the art for automatic EEG analysis brings into light that most of the approaches found on literature propose the use of different classification systems. Basically, they all follow the same methodology. First, each channel of the EEG is split into time windows, which are processed and characterized in order to be computed. Second, the system is trained to identify seizure and normal states, discarding those disturbances that have not been caused by brain activity but, on the contrary, result from movements of the human body (i.e., eye blinking). Finally, as a result of the training phase, a model for each state is computed and provided to a classifier. The classifier is therefore responsible for classifying each time window into any of the two or three considered states (if artifacts and normal waves are considered as separate states).

Despite the fact that the majority of the studies in this field are targeted at the classification stage, few have been to the automation of the training stage. As mentioned in the previous section, most of the works found in the literature either overlook or manually undertake the training phase. This imposes a major drawback in deploying those systems in real environments.

The work in [12] claims that non-specific patient training yields poor results in detecting seizures onsets due to the high variability in wavelet forms. Recent studies address the development of patient-specific classifiers, such as [13]. However, this work is mainly concerned about how to distinguish artifacts from epileptic activity, independently of patient-dependent characteristics, in an autonomous and automatic fashion. Only when epileptic activity has been detected, the proposed system is engaged in detecting the epileptogenic area, therefore helping doctors in their diagnosis of those epileptic types that are suitable for surgery treatments.

In this sense, this paper proposes an approach, which has already been validated in the fields of computer vision and natural language processing. A machine learning algorithm, know as Bag of Words (BoW) is in charge of computing the models for the different events that are going to be identified. Then, the output models are provided to a Support Vector Machine (SVM) classifier, which ultimately performs the element identification.

As it has been already mentioned, the BoW algorithm was originally proposed for natural language processing in [14]. Then, this approach has been gaining importance in the computer vision field and, recently, in the field of video action recognition [11]. This work proposes the use of BoW in the field of EEG analysis for epilepticform waves and epileptogenic focus detection.

Following the methodology described for BoW in natural language and computer vision, the first step consists in extracting the most relevant features that characterize an EEG signal. Then, these features should be quantified so that they can be computationally managed. The most appropriate mean to do so is by compiling them into feature vectors and then, cluster them to compute the model. The clustering criterion is based on the different events provided to the system. Some works adopt a linear approach, in which only normal and seizure events are considered. This work considers three different types of events: normal, artifact, and seizure.

The clustering phase is known, under the BoW terminology, as the *code book generation* phase, since it was originally targeted at text classification. This study therefore proposes the use of a three-word language, in which the equivalent for a word is simply one of the three considered events.

3 System Description

This work has a threefold aim: first, filtering out artifacts from the EEG dataset; second, analyzing EEG in seeking for those epilepticform waves; and, finally, when epileptic activity has been detected, identifying the epileptonogenic area. Figure 1 depicts the different phases involved in the training and testing stages. The filtering and feature extraction phases are common for both stages, the training and testing. The feature extraction phase is required as a mean to transform a continuous signal into a discrete set of values fed to the computational system proposed here.

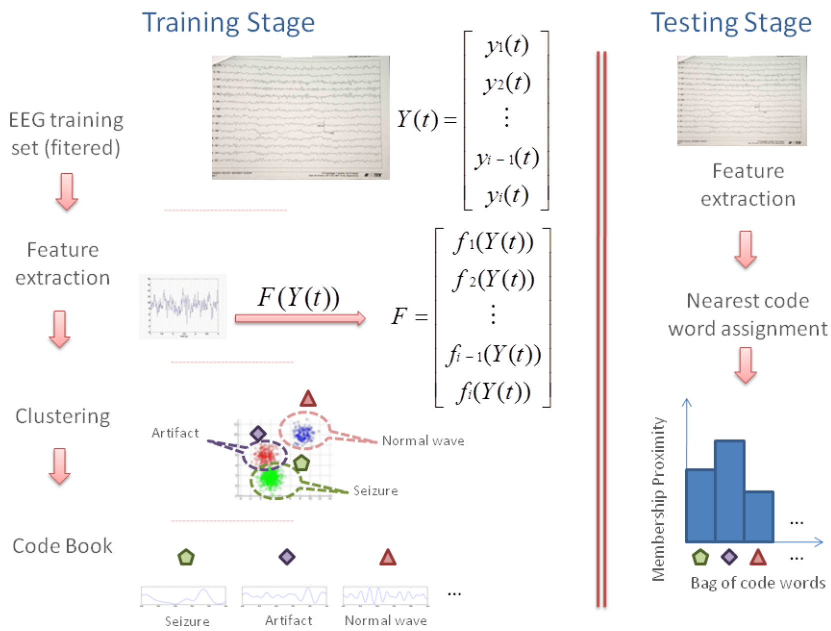

Fig. 1. Stage overview

In order to fulfill these three main aims, this paper presents a two-stage process system. First, a machine learning algorithm is trained to identify artifacts, normal waveforms, and seizures. Then, the training process outputs a set of models that are provided to a SVM classifier, in charge of performing the classification. Finally, whenever epileptic activity is identified from the EEG analysis, an identification of the epileptogenic area is undertaken.

The dataset employed in this paper, provided by the University's Regional Hospital Carlos Haya, consists of EEG results recorded using a *XLTEK Neuroworks* equipment.

3.1 Signal Pre-processing

The original EEG signals, recorded by the *XLTEK Neuroworks* equipment is sampled at 512Hz. Those signals are then band-pass filtered (2 - 200Hz). Signals are also split into 12 seconds epochs, implementing a sliding window with 50% overlap between epochs. Each epoch represents the range normally observed by the neurophysiologist when analyzing an EEG.

3.2 Signal Feature Extraction

The feature extraction of the EEG datasets is common to the training and testing stages. During this procedure, several calculations in the frequency and time domain are performed on the vectors corresponding to the time windows of the various EEG channels (y1, y2,..., yi). In this work the following signal features are calculated: average value, variance, maximum value, minimum value, RMS value, peak frequency, and time derivative and integral.

A matrix of features F is obtained by computing each feature for each channel in a given dataset. The resulting matrix of features if then provided to the BoW or the SVM classifier, depending on whether the training or the testing phase are considered. It should be noted that the computational cost of this procedure depends on the computational complexity of the selected features, the number of channels considered and the size of the time windows. Moreover, due to the repetitive nature of the computations to be performed, it is apparent that solutions based on parallel architectures can considerably reduce the lattency of the procedure calculation.

3.3 Training Stage

The first step in the training stage consists in transforming the training dataset, composed of one signal for each of the 16 channels, into 12 seconds epochs suitable for computation purposes. Using the terminology proposed by the BoW algorithm, each feature is now consider a word and channels are assimilated to documents. However, channels are continuous signals that should be converted into discrete values so that they can be computationally treated. Channel signals are therefore divided into 12 seconds epochs, that are going to be represented by a vector of attributes consisting in sets of *attributes-values*.

Signal epochs are split into epochs and for each one of them, the feature vector is computed, calculating the corresponding value for each of the composing features.

The SVM training has been accomplished by using the Matlab implementation of the algorithm. A model is computed by providing the SVM with three different labeled datasets: one for normal activity, one for artifacts, and one for epileptic seizures.

3.4 Testing Stage

The SVM classifier is tested by using the model computed in the previous stage, the training stage. The SVM implements a non-linear kernel in order to classify input epochs into one of the three sets: normal, artifact, or seizure.

Once the training model has been computed using Matlab, the testing can be carried out by using the SVMpython library[3]. In order to maximize the use of the available dataset, training and testing can use the same data, by implementing

[3] http://tfinley.net/software/svmpython1/

the *leave one out* cross validation. Basically, it consists in repeatedly taking one observation of the dataset for testing purposes and using the remainder as training observations. Iteratively, each single EEG patient composing the dataset is left out of the training data and used for testing purposes.

4 Combining Video for a Better Identification of Artifacts

Identifying artifacts is not always a simple task. There are certain situations in which, not even experts, can positively discern between artifacts and seizures if only EEG data are considered. However, if EEG data analysis is enhanced with annotations whenever patients blink, activate their facial muscles, or move artifacts could be more easily identified.

In fact, EEG equipment do normally incorporate a video recording devices to support doctors in their analysis of the EEG data. The same process could be automatically implemented so that whenever the classifier identifies a potential artifact, video images recorded at that exact time instant can be analyzed to support the recognition system decision.

However, video analysis is a quite sensitive task, prone to significative variations in the results, due to changes, for example, in the lighting of the room the capture is taking place or in the position of the camera (the calibration problem). These are open issues in the video analysis knowledge area, which nowadays can only be minimized in a controlled environment as it is the case of clinical premises.

OpenCV[4] (Open Source Computer Vision) library can be used to develop the set of artifact recognition algorithms (ARA). OpenCV is free, multiplatform, and contains hundreds of functions that span several areas in artificial vision such as face recognition, object recognition, camera calibration, etc.

5 Conclusions

A system for epileptiform abnormalities based on electroencephalogram (EEG) analysis has been presented in this paper. The proposed system combines an SVM classifier and training method based on an implementation of the Bag of Words approach, which has been successfully applied to language and image processing.

The proposed system is intended to alleviate the bottleneck in analyzing the EEG results of patients that, in half percent of the cases, only normal activity has been recorded. Moreover, whenever epileptic activity is detected by the system, the system can also be used to point out the channel in which the seizure was originated. Ultimately, the detection of the epileptogenic area plays an essential role in determining whether that type of epilepsy can be operated or not. At the moment, the neurophysiologist analyzes the EEG and tries to identify the

[4] OpenCV web site. http://opencv.org/. Last visit on 29/07/2012

channel in which the seizure was originated. However, seizures can spread so quickly to different parts of the brain that it is often very difficult to determine, by a simple ocular exploration of the EEG data, in which part of the brain the seizure was originated. Signal analysis at a high sampling rate can enhance the possibilities of detecting the focus where the seizure was originated.

References

1. Shoeb, A., Edwards, H., Connolly, J., Bourgeois, B., Treves, S.T., Guttag, J.: Patient-specific seizure onset detection. Epilepsy and Behavior **5**(4) (2004) 483–498
2. Gotman, J.: Automatic recognition of epileptic seizures in the EEG. Electroencephalography and clinical neurophysiology **54**(5) (1982) 530–540
3. Carney, P.R., Myers, S., Geyer, J.D.: Seizure prediction: Methods. Epilepsy and Behavior **22**(4) (2011) 94–101
4. Winterhalder, M., Maiwald, T., Voss, H.U., Aschenbrenner-Scheibe, R., Schulze-Bonhage, A., Timmer, J.: Quantitative neuroscience. Kluwer Academic Publishers, Norwell, MA, USA (2004) 103–116
5. Yuan, Q., Zhou, W., Liu, Y., Wang, J.: Epileptic seizure detection with linear and nonlinear features. Epilepsy and Behavior **24**(4) (2012) 415 – 421
6. Runarsson, T.P., Sigurdsson, S.: On-line detection of patient specific neonatal seizures using support vector machines and half-wave attribute histograms. Computational Intelligence for Modelling, Control and Automation, International Conference on **2** (2005) 673–677
7. Thomas, E.M., Temko, A., Lightbody, G., Marnane, W.P., Boylan, G.B.: A Gaussian mixture model based statistical classification system for neonatal seizure detection. In: IEEE Workshop on Machine Learning for Signal Processing. (2009)
8. Shoeb, A., Kharbouch, A., Soegaard, J., Schachter, S., Guttag, J.: A machine-learning algorithm for detecting seizure termination in scalp eeg. Epilepsy and Behavior **22**(4) (2011) 36–43
9. Harikumar, R., Balasubramani, M.: Fpga synthesis of soft decision tree (sdt) for classification of epilepsy risk levels from fuzzy based classifier using eeg signals. International Journal of Soft Computing and Engineering **1**(4) (2011) 206–211
10. Luders, H., Bongaman, W., Najm, I.M.: Textbook Of Epilepsy Surgery. Informa Healthcare (2000)
11. Nebel, J.C., Lewandowski, M., Thévenon, J., Martínez, F., Velastin, S.: Are current monocular computer vision systems for human action recognition suitable for visual surveillance applications? In: Proceedings of the 7th international conference on Advances in visual computing - Volume Part II. ISVC'11, Berlin, Heidelberg, Springer-Verlag (2011) 290–299
12. Wilson, S.B., Scheuer, M.L., Emerson, R.G., Gabor, A.J.: Seizure detection: evaluation of the Reveal algorithm. Clinical neurophysiology **115**(10) (2004) 2280–2291
13. Shoeb, A.H., Guttag, J.V.: Application of machine learning to epileptic seizure detection. In: ICML'10. (2010) 975–982
14. Joachims, T.: Text categorization with support vector machines: learning with many relevant features. In Nédellec, C., Rouveirol, C., eds.: Proceedings of ECML-98, 10th European Conference on Machine Learning, Heidelberg et al., Springer (1998) 137–142