+ MODEL

Available online at www.sciencedirect.com

ScienceDirect

CAAI Transactions on Intelligence Technology xx (2016) 1—13
http://www.journals.elsevier.com/caai-transactions-on-intelligence-technology/

Original Article

Classification of epilepsy using computational intelligence techniques

Khurram I. Qazi [a,*], H.K. Lam [a], Bo Xiao [a], Gaoxiang Ouyang [b], Xunhe Yin [c]

[a] *Department of Informatics, King's College London, Strand, London, WC2R 2LS, United Kingdom*
[b] *State Key Laboratory of Cognitive Neuroscience and Learning, School of Brain and Cognitive Sciences, Beijing Normal University, No.19, XinJieKoWai St., HaiDian District, Beijing, 100875, PR China*
[c] *School of Electronic and Information Engineering, Beijing Jiaotong University, Beijing, PR China*

Available online ■ ■ ■

Abstract

This paper deals with a real-life application of epilepsy classification, where three phases of absence seizure, namely pre-seizure, seizure and seizure-free, are classified using real clinical data. Artificial neural network (ANN) and support vector machines (SVMs) combined with supervised learning algorithms, and *k*-means clustering (*k*-MC) combined with unsupervised techniques are employed to classify the three seizure phases. Different techniques to combine binary SVMs, namely One Vs One (OvO), One Vs All (OvA) and Binary Decision Tree (BDT), are employed for multiclass classification. Comparisons are performed with two traditional classification methods, namely, *k*-Nearest Neighbour (*k*-NN) and Naive Bayes classifier. It is concluded that SVM-based classifiers outperform the traditional ones in terms of recognition accuracy and robustness property when the original clinical data is distorted with noise. Furthermore, SVM-based classifier with OvO provides the highest recognition accuracy, whereas ANN-based classifier overtakes by demonstrating maximum accuracy in the presence of noise.

Keywords: Absence seizure; Discrete wavelet transform; Epilepsy classification; Feature extraction; *k*-means clustering; *k*-nearest neighbours; Naive Bayes; Neural networks; Support vector machines

1. Introduction

Epilepsy is a neurological condition such that it affects brain and the nervous system. It is a very commonly known neurological disorder and approximately 1% of general population is affected [1]. Only in the UK, around 1 in 100, more than half a million people suffer from epilepsy. There can be many causes of epilepsy and sometimes it is not possible to identify them. In the domain of epilepsy, seizure is referred to as an epileptic seizure and brain is the source. During an epileptic seizure normal functioning of the brain is disturbed for that certain time period, causing disruption on signalling mechanism between brain and other parts of the body. These seizures can put epilepsy patients at higher risk for injuries including fractures, falls, burns and submersion injuries, which are very common in children [2]. These injuries happen because seizure can happen anytime and anywhere without prior warning and the sufferer would continue his or her activity with an unconscious mind. If a system can effectively predict the pre-seizure phase (the transition time of the brain towards developing seizure), it could then generate an early warning alarm so that precautions can be taken by the sufferer.

Absence seizure is one from many forms of generalized epileptic seizures in which larger part of the brain is disturbed. These seizures are very short and sometime may go un-notice. The patient seems confused and may not remember the seizure afterwards. The complex spike-and-wave patterns generated by the brain during these seizures can be recorded on the electroencephalogram (EEG) and a neurologist can identify the three absence seizure phases namely seizure-free, pre-seizure and seizure [3,4]. To automate this process, EEG data is converted into a digital format and fed into a computerized

* Corresponding author.
E-mail addresses: khurram_ishtiaq.qazi@kcl.ac.uk (K.I. Qazi), hak-keung.lam@kcl.ac.uk (H.K. Lam), bo.xiao@kcl.ac.uk (B. Xiao), ouyang@bnu.edu.cn (G. Ouyang), xhyin@bjtu.edu.cn (X. Yin).
Peer review under responsibility of Chongqing University of Technology.

http://dx.doi.org/10.1016/j.trit.2016.08.001

seizure detection (classification) system, which can automatically recognize the input pattern. The two core modules of this classification system are: feature extraction and design of a classifier using these features. A feature extraction method extracts the most discriminative information from the EEG recordings, which means an ideal feature can have the property of differentiating among three phases of absence seizure.

In literature, there is a long list of methods that can be used to extract features from the EEG signals. These methods include Fourier Transforms (FT); good for analysing stationary signals, and Time Frequency Distribution (TFD); provides a tool for examining continuous segments of EEG signals. However, EEG signals are non-stationary in nature and conventional frequency analysis methods may not capture the full details of the brain signals. Lyapunov exponents; discussed the detection and prediction of epileptic seizure in Refs. [5], analysis of correlation structure [6], and high order spectral analysis (HOS) [7] are the examples of non-linear methods for EEG signal analysis in the domain of epilepsy. Advances in wavelet theory has also made it a very suitable for bio-medical signal processing. It has a built-in advantage of capturing repeated and irregular patterns. It can also deal with the analysis of transient and sudden signal changes [8,9]. This is possible because this technique provides variable window size, narrow at high and wide at low frequency levels. Furthermore [8], has discussed several different methods for EEG signal analysis and concluded that Wavelet Transform (WT) has more advantages over other methods. The next step in designing the classification system is to combine these extracted features with an appropriate learning method to design a classifier. These methods can be divided into traditional classification methods and modern learning algorithms also known as computationally intelligent algorithms or machine learning algorithms. Bayesian methods based on statistical theory, *k*-nearest neighbour (*k*-NN) and decision trees based on logical branching, are considered to be in the category of traditional classifiers. On the other hand, Support Vector Machine (SVM), Artificial Neural Network (ANN), *k*-Means Clustering (*k*-MC) and Self-organising Maps (SOMs) exhibit intelligent behaviour by learning the complex and non-linear patterns hidden inside the input feature vector, are considered to be computationally intelligent approaches. Below is the brief introduction about all the methods used in this research.

Classification systems based on Bayesian statistical theory have been widely and successfully used commercially [10]. This method gives a way to represent sensory evidence; features extracted from the raw data, and prior information about the problem in hand that is collected from domain knowledge. Considering the equal prior probabilities for all classes and hence, ignoring the hassle of obtaining the domain knowledge, the analysis becomes very straight forward. Although, it is a powerful and simple rule to handle and implement, yet estimating posterior probabilities from the data is a non-trivial task [11] and the distribution of data may not be uniform. *k*-NN rule as the name implies, classify an observation by giving it a label after probing the labels on all the *k*-NNs and making decision based on majority voting. Usually, Euclidean distance is used to measure the distances between the neighbouring instances. This algorithm mostly provides an acceptable performance in many applications [10] such as visual category recognition [12] and is also very easy to implement. However, *k*-NN algorithm suffers due to large memory requirements and also there is no logical way to choose the best *k* value, which would affect the classification problem and may not yield very good results.

Inspired by the human brain functioning and architecture, McCulloch and Pitts in 1940s presented a logical threshold unit (LTU) [13]. This basic idea of LTU has been generalized in many ways since then and it is the building block for modern ANNs. In 1962 a single layer perceptron neural network was introduced by Rosenblatt [14] by extending the idea of LTU along with a trainable network with adaptive elements [15]. Despite the huge success of this neural network model, it was only limited to solve linearly separable problems. A traditional feed-forward neural network [13] has three types of layers, i.e., input, hidden and output layers. Each layer consists of nodes connected in a layer-to-layer manner. The feed-forward neural networks can learn any smooth non-linear functions in a compact domain to any degree of accuracy and are considered to be universal approximators. They have many applications such as classification, prediction, clustering and approximation [13]. The drawback of this technique is that it requires a lot of parameters to be tuned for training the neural network. Also, there is no set of rules for finding the number of hidden layers and neurons in these layers. Furthermore, ordinary neural networks suffer from the "over-fitting" and "local optimization" problems [16].

The SVM method maps the non-linear and inseparable data from an input space into a higher dimensional feature space where the data would then be linearly separable [17]. This task is accomplished by utilising the concept of separating hyper planes [17]. Instead of computing a mapping function, the use of kernel function saves the computational demand especially for feature mapping function of higher dimensional space. The SVM algorithm aims to maximize the margin (the region separating the support vectors on either side of the hyperplane) and tries to find an optimal hyperplane. Hence, also called the maximal margin classifier. Although, the training parameters for SVM technique are very few, it can still be a computationally time consuming and highly complex. Nonetheless, SVM has a good generalization ability, solution to the over-fitting problem and also performs well in a high dimensional feature space.

k-MC is an unsupervized learning technique that clusters samples based on similar properties. Whereas, *k* represents the number of clusters or classes. Within a cluster the samples are similar but different from the samples grouped in other clusters. This is an iterative process in which samples are grouped together according to their closet mean and cluster positions are adjusted until these positions do not change for some iterations. The advantage of this technique is that it does not require labels with the input examples. The convergence is also faster if *k* is small. However, if clusters are of different