HYBRID SUPPORT VECTOR MACHINE FOR CLASSIFICATION OF EEG SIGNALS

Mohammad Zaini and Ali Omar
MARA University of Technology, Malaysia
E-Mail: mohamad.zainni@gmail.com

ABSTRACT

Reading EEG signals manually is a very difficult and time-consuming task. In many situations, we like to get the results in a very short amount of time (e.g. monitoring seizure patients). In other cases, we like to study huge amount of data. In both cases, reading EEG manually is not practical and therefore automatic approach is preferred. In this paper, we propose a simple system that can achieve the state of the art results for IED classification (accuracy of 82%) while using a relatively simple algorithm. The advantage of using a simple algorithm is to make it possible to implement this system on cheap consumer devices like phones.

Keywords: EEG, EEG signal processing, EEG classification, hybrid SVM.

1. INTRODUCTION

Electroencephalograms (EEGs) are referred to signals collected using electrodes placed on the scalp. EEGs are one the main means by which neurologists diagnose brain-related diseases such as epilepsy and seizures 0. Commonly diagnosed abnormalities include epileptic events, seizures and strokes [2]. Neurologists manually read and interpret EEGs, which is very time-consuming process due to complex nature of these signals.

Figure-1 shows a typical clinical EEG [3]. As we can see from this picture, each EEG consists of several channels and as a result an EEG is multi-dimensional signal. Diagnosing diseases using an EEG recording requires a detailed knowledge of the patient's medical history, mental and physical state during the recording process so that abnormal variations in the signal are properly interpreted.

EEG machine records electrical potential between two electrodes using surface electrodes and that potential can be recorded using surface electrode. Electrode impedances should be maintained between 100 and 5000 ohms. The International Federation of Societies for Electroencephalography and Clinical Neurophysiology has

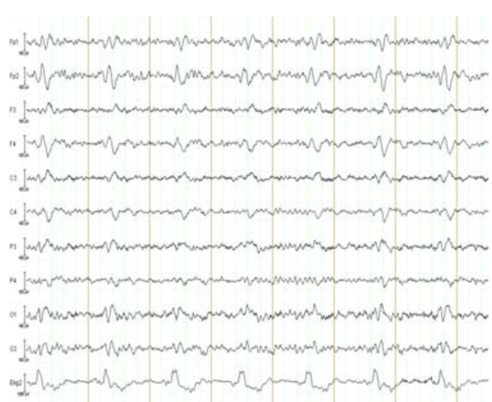

Figure-1. Example of EEG signal [3].

recommended the conventional electrode setting (also called 10–20) for 21 electrodes (excluding the earlobe electrodes) [4] [5] [6]. Electrode placement has been standardized by international 10-20 system that uses anatomical landmarks on the skull including a heart pulse electrode that is known by EKG. The designations; F_p (fronopolar), F(frontal), T(temporal), O(occipital), C(central), P(parietal) are utilized in the 10-20 system. Numbers combined following the letters for location reflect either the left (odd numbers) or right (even numbers) hemisphere of electrode placement. The "z" designation reflects midline placement (i.e. C_z is central midline). While electrodes conduct electrical potentials from the patient scalp to an electrode box (or jack box), a montage selector allows physicians to use either *bipolar* montage or *referential* montage.

Figure-2 shows this montage.

EEGs composed of 1 dimensional signals gathered from different parts of the scalp. Together these 1 dimensional signals create a high dimension signal that can be used by trained physician to diagnose different conditions such as abnormal activities for a patient. One of the major challenges for many doctors and hospitals is the fact that reading EEGs is very specialized expertise and can only be performed by licensed technician or doctors. Moreover, in case of seizure patients, we often want to know about the abnormal activity in real time (so we can help the patient). If

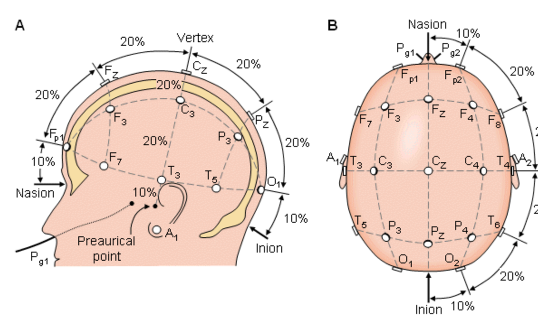

Figure-2. 10-20 Montage used in this research [6].

ISSN 1819-6608
ARPN Journal of Engineering and Applied Sciences

we only rely on human experts to read and interpret EEGs we cannot solve these issues.

The ability to automatically predict critical events from EEG has been actively researched for the past 40 years. Unfortunately, clinical use of such systems is limited due to bad performance. EEG events are defined as critical points in a signal, such as a spike or interictal epileptiform discharges (IED) that correlate with the presence of different diseases. A classification error rate of 5% for these EEG events would be acceptable clinically [7]. However, current systems do not operate at this level of accuracy due to a lack of adequate algorithms.

We have developed a system, that automatically interprets EEGs, and delivers good results on clinical data. An overview of the system is shown in

Figure-3. It incorporates a traditional support vector machine (SVM) based system. First an N channel EEG is transferred into N independent signals and each signal is processed using a moving window. The output of each window in then feed into each SVM machines. Our system is a hybrid SVM machine that include two SVMs that classifies its input into one of the 3 classes: 1-background 2- eye-blink 3- IED. We will show our system can compete with other state of the art systems despite the fact we are using a very simple machine learning algorithm.

The Paper is organized as follows: In the next two sections, we will introduce SVM and hybrid SVM machines. Next experiments will be presented. We show our algorithm can compete with the state of the art while enjoying much simpler architecture.

2. SUPPORT VECTOR MACHINES

Classifiers are typically optimized based on some form of risk minimization. Empirical risk minimization is one of the most commonly used techniques where the goal is to find a parameter setting that minimizes the risk function [8]:

$$R(\alpha) = \frac{1}{2N}\sum_{i=1}^{N}\left|y_i - f\left(x_i,\alpha\right)\right|, \tag{1}$$

where α is the set of adjustable parameters and x_i, y_i are the input and output, respectively. However, minimizing (1) does not necessarily imply the best classifier possible. For example,

Figure-4 (adapted from [8]) shows a two-class problem and the corresponding decision boundaries in the form of hyper-planes. All the hyper-planes can achieve

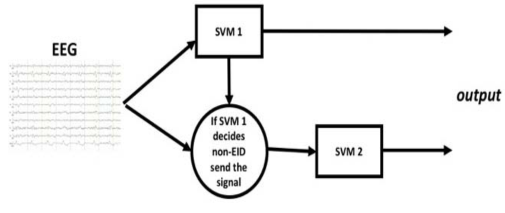

Figure-3. Block diagram of the proposed system.

perfect classification and, hence, zero empirical risk. However, C_0 is the optimal hyper-plane because it maximizes the distance between the margins, therefore offering better generalization. This form of learning is an instance of Structural Risk Minimization (SRM) where the goal is to learn a classifier that minimizes a bound on the expected risk, rather than the empirical risk [9] [10]. SVM learning is based on this SRM principle.

SVMs can transform data into a high dimensional space where the data can be separated using a linear hyper-plane. The optimization process for SVM learning therefore begins with the definition of a functional that needs to be optimized in terms of the parameters of a hyper-plane. The function is defined in such a way that guarantees good classification (if not perfect classification) on the training data and also maximizes the margin. The points that lie on the hyper-plane follow [8]:

$$w.x + b = 0\ . \tag{2}$$

The goal of the optimization process should be to maximize the margin. Posing this as a quadratic optimization problem has several advantages and the functional can be compactly written as [8]:

$$L_p = \frac{1}{2}\|w\|^2 - \sum_{i=1}^{N}\alpha_i y_i\left(x_i w + b\right) + \sum_{i=1}^{N}\alpha_i\ . \tag{3}$$

Only a few training instances have an impact on the function and the optimal decision surface. This comes from the fact that, at the end of the optimization, only a small percent of the training examples have non-zero multipliers. These instances are called Support Vectors. Note that we have assumed that the data are perfectly separable. This is not the case in most real data. This problem is handled by introducing slack variables into (3):

$$y_i\left(x_i w + b\right) - 1 + \zeta_i \geq 0 \qquad \forall i. \tag{4}$$

Notice that the linearity in the SVM is manifested in the dot products. Suppose we transform the data into a higher dimension space where the data is linearly separable. The theory of Kernel functions is used to avoid dealing directly with the high dimensional space and the excessive computations that result from such transformations [8] [11]. For example, RBF Kernel used in this work defined as:

$$K(x,y) = \exp\left\{-\gamma\left|x-y\right|^2\right\} \tag{5}$$

The final classifier takes the form:

$$f(x) = \sum_{i=1}^{N}\alpha_i y_i K(x,x_i) + b \tag{6}$$

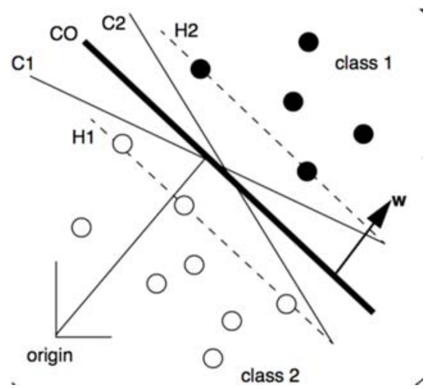

Figure-4. Example of Linear classification [8].

where N is the number of support vectors. This is definition of a binary classifier and sign of *f* decides about the class that SVM predicts.

3. HYBRID SVM MACHINE

SVMs are very powerful binary classifiers, e.g. they can classify the input into two classes. In our problem, we are interested to classify the input into three classes: 1- IEDs 2- eye-blinks 3-background. We proposed architecture (

Figure-1) that makes this goal possible. The first step is to classify any input signal into IED and non-IED classes. The next step is to classify non-IEDs into eye-blink and background. This is an example of hierarchical classification [12]. During training of the first SVM, all eye-blink and backgrounds are lumped together with one label (non-IED). For training the second SVM all IEDs are removed from the training set and we only train them on non-IEDs. One can think of the second SVM a fine-grain classifier.

During evaluation, we first use the first SVM to find all IEDs. If a signal is classified as non-IED then it should pass through the second SVM that classify it into either background or eye-blink. The final result is a three-way classification of the input signal.

4. EXPERIMENTS

In this section, we will present the experimental results for a classifier that classify IEDs, eye-blinks and background events. We have used a publicly available data [7] to conduct

Table-1. Number of different classes in our data.

Event	Number in annotated subsection
IED	5128
Eye-Blink	331
Background	201

these experiments. The set was not annotated and therefore we have annotated a portion of it by hand and divided the data to test and training subsets. We have trained two SVM machines that each classifies the events into two classes. The first machine classifies between IEDs and non-IEDs and the second one classifies between eye-blinks and backgrounds. In other words, after we decided something is not IED then we decide if it is background or eye-blink. Each EEG file consists of multiple channels. We have defined events as one second of one channel; e.g. each event is a 1 dimensional signal. Following [13] [14] [15] we call these second-channel. **Error! Reference source not found.** shows the number of each class in our dataset.

Feature extraction goal is to represent a raw EEG signal by small number of attributes or features which contains all the relevant information for a given task [16]. It is important to notice that the size of feature vector relative to raw EEG is very small. The reason is to avoid the curse of dimensionality (e.g. when the dimensionality of the data is high learning all parameters become prohibitive). In this work, we have used short term Fourier transform (STFT) to extract features. These features represent energy in different frequency bands [17]. In STFT, the signal is divided into small segments with overlapping data and fast Fourier transform (FFT) applied to each one. The output of successive STFTs can provide a time-frequency representation of the signal. To accomplish this, the signal truncated by multiplying it by a window so that the signal is zero outside the window. The STFT is defined as [17]:

$$S_{STFT}[m,n] = \sum_{k} x[k]w[k-m]e^{-\frac{2\pi nk}{N}} \tag{7}$$

where x[k] denotes a signal and w[k] denotes an L-point window function.

The STFT is applied to one second EEG signal segmented into a 256 point segments with 50% overlapping between each successive segments. Each segment is multiplied by a 256 point Triangle window, then the FFT algorithm is applied to each segment.

After feature extraction, we can train our model using SVM machines described in the previous section. To train SVM machines we present them with features for 1 second of signal for each channel (second-channel) and the

Table-2. comparing proposed system with other state of the arts systems.

Algorithm	Accuracy (%)
RBF Neural Network [17]	75%
Neural Network [19]	80%
Associate System [20]	87%
HMM [21]	84%
This work	82%

label for that second-channel. The labels are annotated manually with help of professional EEG technicians. We have used 3200 number of data points to train a binary classifier. First SVM classifies the input into none IED and IED. The second one classifies only non-IEDs into background and eye-blink. For evaluation, we have used the trained SVMs and feed them with features extracted from test section of data and our hybrid SVM predicts a label for every second-channel. It is important to notice those training and evaluation subsets are mutually exclusive. Table-2 shows the result of our approach along with some state of the art results reported in the literature for similar problems (but on different data). As we can see, our results are close to other state of the art results despite the fact that we have used a much simpler model.

5. CONCLUSIONS

In this paper, we have introduced a hybrid-SVM algorithm that can classify the input EEG signal into three class: 1- IEDs 2- Eye-blink 3- Background. In the context of this paper background refers to everything other than IEDs and eye-blink. We have shown that our proposed algorithm can achieve state of the art results for this task while using a relatively simple system. Next step for this research is to use other type of machine learning like neural network, deep-learning, Bayesian and KNN in place of SVM machines.

REFRENCES

[1] Teplan M. 12002. Fundamentals of EEG measurements. Measmt Sci. Rev. 2(2).

[2] Bickford R. D. 1987. Electroencephalography. in Encyclopedia of Neuroscience, Ed. G. Adelman, Birkhauser, Cambridge (USA). 371-373.

[3] Stern John M. 2005. Atlas of EEG patterns. Lippincott Williams and Wilkins.

[4] On F. R., R. Jailani, H. Norhazman and N. Mohamad Zaini. 2013. Binaural beat effect on brainwaves based on EEG. In Signal Processing and its Applications (CSPA), 2013 IEEE 9th International Colloquium on, pp. 339-343. IEEE.

[5] Norhazman, H., N. Mohamad Zaini, M. N. Taib, Hassan Omar, R. Jailani, S. Lias, Lucyantie Mazalan and Maizura Mohd Sani. 2012. Behaviour of EEG alpha asymmetry when stress is induced and binaural beat is applied. In Computer Applications and Industrial Electronics (ISCAIE), 2012 IEEE Symposium on. pp. 297-301. IEEE.

[6] Malmivuo Jaakko and Robert Plonsey. 1995. Bioelectromagnetism: principles and applications of bioelectric and biomagnetic fields. Oxford University Press, USA.

[7] Harati A., Soon-Mi Choi, Mehriar Tabrizi, Iyad Obeid, Joseph Picone and M. P. Jacobson. 2013. The Temple University Hospital EEG Corpus. In Global Conference on Signal and Information Processing (GlobalSIP), IEEE. pp. 29-32.

[8] Ganapathiraju Aravind, et al. 2000. Hybrid SVM/HMM architectures for speech recognition. INTERSPEECH.

[9] Cortes Corinna and Vladimir Vapnik. 1995. Support vector machine. Machine learning. 20.3, 273-297.

[10] Suykens Johan AK and Joos Vandewalle. 1999. Least squares support vector machine classifiers. Neural processing letters. 9.3, 293-300.

[11] Burges, Christopher JC. 1998. A tutorial on support vector machines for pattern recognition. Data mining and knowledge discovery. 2.2, 121-167.

[12] Gordon Allan D. 987. A review of hierarchical classification. Journal of the Royal Statistical Society. Series A (General). 119-137.

[13] Shoeb, Ali H., and John V. Guttag. 2010. Application of machine learning to epileptic seizure detection. In Proceedings of the 27th International Conference on Machine Learning (ICML-10). pp. 975-982.

[14] Shoeb Ali, Alaa Kharbouch, Jacqueline Soegaard, Steven Schachter and John Guttag. 2011. A machine-learning algorithm for detecting seizure termination in scalp EEG. Epilepsy and Behavior 22, S36-S43.

[15] Ahammad Nabeel, Thasneem Fathima and Paul Joseph. 2014. Detection of epileptic seizure event and onset using EEG. BioMed research international.

[16] Suleiman, Abdul-Barry Raouf, and Toka Abdul Hameed Fatehi. 2011. Features Extraction Techniques of EEG Signals for BCI Application.

[17] Allen J. 1977. Short-term spectral analysis and modification by discrete Fourier transform. IEEE Transactions on Acoustics Speech and Signal Processing. 235-238.

[18] A. Saastamoinen, T. Pietila, A. Va̎rri, M. Lehtokangas and J. Saarinen. 1998. Waveform detection with RBF network-Application to automated EEG analysis. Neurocomputing. 20: 1-13.

[19] Exarchos Themis P., Alexandros T. Tzallas, Dimitrios I. Fotiadis, Spiros Konitsiotis, and Sotirios Giannopoulos. 2006. EEG transient event detection and classification using association rules. Information Technology in Biomedicine, IEEE Transactions on. 10(3): 451-457.

[20] C. Castellaro, G. Favaro, A. Castellaro, A. Casagrande, S. Castellaro, D. V. Puthenparampil, and C. F. Salimbeni. 2002. An artificial intelligence approach to classify and analyse EEG traces. Clin. Neurophysiol. 32: 193-214.

[21] Harati, A., Golmohammadi, M., Lopez, S., Obeid, I. and Picone J. 2015. Improved EEG Event Classification Using Differential Energy. 2015 IEEE Signal Processing in Medicine and Biology Symposium (SPMB), USA.