Feature Extraction and Classification of EEG Signals using Wavelet Transform, SVM and Artificial Neural Networks for Brain Computer Interfaces

M. R. Nazari Kousarrizi, A. Asadi Ghanbari
Electrical Engineering Department
K. N. Toosi University of Technology
Computer department of Islamic Azad University
Science and Research branch
Tehran, Iran
mr.nazari@ee.kntu.ac.ir, asadi.abdolreza@gmail.com

M. Teshnehlab, M. Aliyari, A. Gharaviri
Electrical Engineering Department
K. N. Toosi University of Technology
Tehran, Iran
Teshnehlab@eetd.kntu.ac.ir, m_aliyari@eetd.kntu.ac.ir, gharaviri@ieee.org

***Abstract*—Brain Computer Interface one of hopeful interface technologies between humans and machines. Electroencephalogram-based Brain Computer Interfaces have become a hot spot in the research of neural engineering, rehabilitation, and brain science. The artifacts are disturbance that can occur during the signal acquisition and that can alter the analysis of the signals themselves. Detecting artifacts produced in electroencephalography data by muscle activity, eye blinks and electrical noise is a common and important problem in electroencephalography research. In this research, we used five different methods for detecting trials containing artifacts. Finally we used two different neural networks, and support vector machine to classify features that are extracted by wavelet transform.**

Keywords-component; Brain computer interface; Independent component analysis; Artifact; support vector machine (SVM) and Artificial neural networks (ANN).

I. INTRODUCTION

Brain computer interface (BCI) technology has shown remarkable promise for providing individuals with severe motor disabilities a means to communicate via computers and to have command and control over their environment [1]. Previous published work has shown that EEG related to both movement and imagination of movement can be used as control signals for BCI [2].

There are a few non-invasive methods for obtaining these brain signals to be utilized in a BCI design. The common methods could be grouped into 3 types. The first type uses EEG signals recorded at the scalp during some mental tasks [3]. The second uses single-trial visual evoked potential (VEP) signals where the subjects gaze at a screen of alphabets or menus [4]. The third uses synchronization and resynchronization of μ-rhythm extracted during sensory motor tasks. Reviews of some of these technologies and developments in this area are given by Vaughan *et al* [5] and Wolpaw *et al* [6].

In this research, we show how we can convert EEG activity into cursor movement by a BCI using an appropriate feature extraction scheme. The proposed automated method for the classification of EEG activity is based on signal preprocessing, feature extraction and classification. For signal preprocessing, we use different methods for rejecting artifacts.

For feature extraction we use the power spectrum, variance and mean of the Haar mother wavelet. Finally, we implemented a feed-forward multi-layer perceptron (MLP) with a single hidden layer with five neurons, a probabilistic neural network (PNN).and support vector machine (SVM) classifier with Gaussian RBF kernel.

II. MATERIALS AND METHODS

A. Database

In this research, we used an EEG signal as the basic signal for classification. The EEG data is from an open EEG database of University of Tuebingen. Two versions of the EEG database are employed as [7].

1) Dataset I: The datasets were taken from a healthy subject. The subject was asked to move a cursor up and down on a computer screen, while his cortical potentials were taken. During the recording, the subject received visual feedback of his slow cortical potentials (Cz-Mastoids). Each trial lasted 6s. During every trial, the task was visually presented by a highlighted goal at either the top or bottom of the screen to indicate negativity or positivity from second 0.5 until the end of the trial. The visual feedback was presented from second 2 to second 5.5. Only this 3.5 second interval of every trial is provided for training and testing. The sampling rate of 256 Hz and the recording length of 3.5s results in 896 samples per channel for every trial. This dataset contain 266 trials that 70% of this datasete is considered as train dataset and the rest are considered as test.

2) Dataset II: The datasets were taken from an artificially respirated ALS patient. The subject was asked to move a cursor up and down on a computer screen, while his cortical potentials were taken. During the recording, the subject received auditory and visual feedback of his slow cortical potentials (Cz-Mastoids). Each trial lasted 8s. During every trial, the task was visually and auditorily presented by a highlighed goal at the top or bottom of the

978-0-7695-3739-9/09 $25.00 © 2009 IEEE
DOI 10.1109/IJCBS.2009.100

screen from second 0.5 until second 7.5 of every trial. In addition, the task ("up" or "down") was vocalised at second 0.5. The visual feedback was presented from second 2 to second 6.5. Only this 4.5 second interval of every trial is provided for training and testing. The sampling rate of 256 Hz and the recording length of 4.5s results in 1152 samples per channel for every trial. This dataset contain 200 trials that 70% of this datasete is considered as train dataset and the rest are considered as test.

Artifacts rejected from the dataset. In the second step, features are extracted from the EEG signals using the Wavelet (WT), which demonstrated to be the most promising feature extraction method in other studies. Finally, support vector machine (SVM*)* and two different neural network types (MLP, PNN) are employed as classifiers to classify moving a cursor up and down on a computer screen.

III. ARTIFACT REJECTION

Artifact rejection plays a key role in many signal processing applications. The artifacts are disturbance that can occur during the signal acquisition and that can alter the analysis of the signals themselves. Our aim is to automatically remove the artifacts, in particular from the EEG recordings. The artifact rejection is based on the independent component analysis (ICA). Five different methods used for detecting trials containing artifacts.

1) Extreme values: First, we used standard thresholding of potential values. Here, data trials were labeled as artifactual if the absolute value of any data point in the trial exceeded a fixed threshold.

2) Linear trends: Marked linear trends at one electrode typically indicate transient recording-induced current drifts. To detect such events, we measured the goodness of fit of EEG activity to an oblique straight line within a sliding time window. We then either marked or not the data trial depending on the minimum slope of this straight line and its goodness to fit (in terms of r^2).

3) Data Improbability: Most artifacts have “unusual” time courses, e.g., they appear as transient, ‘odd’, or unexpected events, and may be so identified by the outlying values of their statistics relative to normal brain activity. We tested the use of the joint-probability of the observed distribution of data values and the kurtosis of the data value distribution for detecting such artifacts. The joint probability was computed for every data trial at each electrode or independent component.

4) Kurtosis: We used a second measure to detect unusually ‘peaked’ distributions of potential values the kurtoss (K) of the activity values in each trial:

$$K = m_4 - 3m_2^2 \tag{1}$$

$$m_n = E\{(x - m_1)^n\} \tag{2}$$

Where m_n is the n-order central moment of the variable, m_1 is the mean and E an expectation function (here, the average). If all activity values are similar, or the values alternate between two or more extremes, the kurtosis will be highly negative. Again, this type of activity is not typical of brain EEG signals. Strong negative kurtosis values usually reflect AC (alternating current) or DC (direct current) artifacts, for example those induced by screen currents, strong induced line noise from electrical machinery, lighting fixtures, or loose electrode contacts. If the kurtosis is highly positive, the activity distribution is highly peaked (usually around zero) with infrequent appearance of extreme values, and the identified data is likely to contain an artifact. Eye blink artifacts, as for example extracted from scalp EEG data by ICA, also have relatively high kurtosis.

5) Spectral pattern: Finally, some EEG artifacts have specific activity and scalp topographies that are more easily identifiable in the frequency domain. For instance, temporal muscle activations typically induce relatively strong *20-60 Hz* activity at temporal electrodes, while saccadic eye blinks produce unusually strong *(1-3 Hz)* low frequency activity at frontal electrodes. Software routines for performing the artifact detection methods described above are available within the EEGLAB toolbox [8].

For each trial we divided the trials to six epochs and do ICA for them. Then employed five methods described above for rejecting artifacts and mark the illegal epoch, if the percent of the unmarked epochs in each trial less than 30% that trial mark as a corrupt trial and delete from the training or testing set.

IV. FEATURE EXTRACTION

For features extraction from the raw EEG data many methods such as time domain, frequency domain, and time–frequency domain are used. Since the EEG is non-stationary in general [9], it is most appropriate to use time–frequency domain methods like wavelet transform (WT) as a mean for feature extraction. The WT provides a more flexible way of time–frequency representation of a signal by allowing the use of variable sized windows. In WT long time windows are used to get a finer low-frequency resolution and short time windows are used to get high-frequency information. Thus, WT gives precise frequency information at low frequencies and precise time information at high frequencies. This makes the WT suitable for the analysis of irregular data patterns, such as impulses occurring at various time instances. The EEG recordings were decomposed into various frequency bands through fourth-level wavelet packet decomposition (WPD). The decomposition filters are usually constructed from the Haar or other sharp mother wavelets, when the data has discontinuities. In this research, based on the analysis of the data, Haar mother wavelet was used in the decomposition. The power spectrum, variance and mean of the signal (each channel) are extracted as features. So the feature set for each subject in each trial consisted of 3*number of channels. As a result, the feature matrix was

266*18 and 200*21 for subject A and B respectively. Finally the feature matrix is normalized.

V. CLASSIFICATION APPROCHES

A. Neural Networks

An artificial neural network (ANN) is an interconnected group of artificial neurons simulating the thinking process of human brain. One can consider an ANN as a “magical” black box trained to achieve expected intelligent process, against the input and output information stream. ANN are useful in application areas such as pattern recognition, classification etc.

1) Multilayered Perceptron Neural Networks: The decision making process of the ANN is holistic, based on the features of input patterns, and is suitable for classification of biomedical data. Typically, multilayer feed forward neural networks can be trained as non-linear classifiers using the generalized backpropagation (BP) algorithm.

Our network has one hidden layer with five neurons and output layer with one neuron. We use generalized BP algorithm with momentum for training procedure. Momentum is a standard training technique which is used to speed up convergence and maintain generalization performance [10]. For hidden and output layers, we used bipolar and unipolar sigmoid functions respectively as decision function on the other hand we normalized weights and inputs. With these methods we achieved a NN classifier that is the most suitable classifier for the task at hand. We determined the most effective set as well as the optimum vector length for high accuracy classification. This NN classifier was trained and tested by using the feature sets described above.

By means of minimizing error, we optimized the number of neurons in hidden layer to five. Weights and sigmoid function slope were trained at the next step. With 500 epochs and learning rate equal to 0.05, we achieved classification accuracy equal to 88.75% and 80.87% for subject A and B respectively.

2) Probabilistic Neural Network: The probabilistic approach to neural networks has been developed in the framework of statistical pattern recognition. Probabilistic neural network (PNN) is derived from radial basis function (RBF) network which is an ANN using RBF. RBF is a bell shape function that scales the variable nonlinearly. PNN is adopted for it has many advantages [11]. Its training speed is many times faster than a BP network. PNN can approach a Bayes optimal result under certain easily met conditions. Additionally, it is robust to noise examples. We choose it also for its simple structure and training manner. The most important advantage of PNN is that training is easy and instantaneous. Weights are not “trained” but assigned. Existing weights will never be alternated but only new vectors are inserted into weight matrices when training. So it can be used in real-time. Since the training and running procedure can be implemented by matrix manipulation, the speed of PNN is very fast.

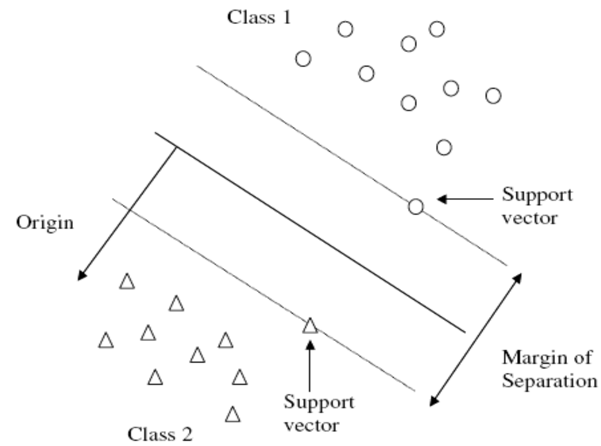

Figure 1. SVM classification with a hyperplane that maximizes the separating margin between the two classes (indicated by data points marked by “Δ”s and “**O**”s). Support vectors are elements of the training set that lie on the boundary hyperplanes of the two classes.

B. Support Vector Machine

The SVM is a relatively new classification technique developed by Vapnik [12] which has shown to perform strongly in a number of real-world problems, including BCI.

The invention of SVM was driven by underlying statistical learning theory, i.e., following the principle of structural risk minimization that is rooted in *VC* dimension theory, which makes its derivation even more profound. The SVMs have been a topic of extensive research with wide applications in machine learning and engineering. The output of a binary SVM classifier can be computed by the following expression:

$$y = \operatorname{sgn}\left(\sum_{i=1}^{N} \alpha_i y_i k(x_i, x) + b.\right) \tag{3}$$

where $\{x_i, y_i\}_{i=1}^{N}$ are training samples with input vectors $x_i \in R^d$, and class labels $y_i = \{-1,1\}, \alpha_i \geq 0$, are Lagrangian multipliers obtained by solving a quadratic optimization problem, b is the bias, and $k(x_i, x_j)$ is called kernel function in SVM. The most commonly used kernel function is the Gaussian RBF as:

$$k(x_i, x_j) = \exp\left(\frac{-\|x_i - x_j\|^2}{2\sigma^2}.\right) \tag{4}$$

The SVM for the linearly separable case find optimal separating hyper plane, as shown in “Fig. 1” [13].

TABLE I. RESULTS OF THE DATASET I

Accuracy of classifiers	Training	Test
MLP	99.56%	88.75%
PNN	100%	83.75%
SVM	100%	82.25%

TABLE II. RESULTS OF THE DATASET II

Accuracy of classifiers	Training	Test
MLP	99.57%	80.87%
PNN	100%	68.35%
SVM	100%	68.33%

VI. SIMULATION RESULTS

To classify cursor movements two versions of the EEG database are used. We used 70% each of datasets for training and the rest for test classifiers. Generally, the classification accuracy over files, which were included in training, is higher than the accuracy for the testing set. Tables I and II indicate the results of classification accuracy during training and test stages for both dataset. In comparison with the neural network classifier, SVM classifier has a better training accuracy rate but test accuracy of neural network classifier is better than SVM, but because of the nature of SVM classifier, this classifier is more general than neural network and this specification is very important in the use of classifiers.

VII. CONCLUSION

We implemented a multilayer perceptron with one hidden layer, a PNN, and SVM classifiers for classification. In our neural network structure, the output layer unit has sigmoid function, which makes our network capable of nonlinearly mapping and capturing dynamics of signals. This performance results from a rigors artifact rejection methodology and procedure.

In our SVM classifier we examine different values for σ which is a very essential parameter in designing a SVM classifier with Gaussian RBF kernel and then the one which obtained the best result was selected.

REFERENCES

[1] J.R. Wolpaw, N. Birbaumer, D.J. McFarland, G. Pfurtscheller, and T.M. Vaughan "Brain-computer interfaces for communication and control," Clin. Neurophysiol, 2002, pp. 767-791.

[2] J. Den, J. Yao and J.P. Dewald "Classification of the intention to generate a shoulder versus elbow torque by means of a time-frequency synthesized spatial patterns BCI algorithm," J.Neural Eng. 2005, 2, pp. 13 1-8.

[3] R. Palaniappan, P. Paramesran, S. Nishida, and nN. Saiwaki, "A New Brain-Computer Interface Design Using Fuzzy ARTMAP," IEEE Transactions on Neural System and Rehabilitation Engineering, vol. 10 pp. 140-148, Sept. 2002.

[4] E. Donchin, K.M. Spencer, and R. Wijesinghe, "The mental prosthesis: assessing the speed of a P300-based brain-computer interface," IEEE Transactions on Rehabilitation Engineering, vol. 8 no. 2, pp. 174-179, June 2000.

[5] T.M. Vaughan, J.R. Wolpaw, and E. Donchin, "EEG based communications: Prospects and Problems," IEEE Transactions on Rehabilitation Engineering, vol. 4, no. 4, December 1996.

[6] J.R. Wolpaw, et. al., "Brain-Computer Interface Technology: A Review of the First International Meeting," IEEE Transactions on Rehabilitation Engineering, vol. 8 no. 2, pp. 164-173, June 2000.

[7] BCI Competition 2003. http://ida.first.fraunhofer.de/projects/bci/competition

[8] A. Delorme, and S. Makeig, "EEGLAB: an open source toolbox for analysis of single-trial EEG dynamics including independent component analysis," J Neurosci Methods *134*, 9-21, 2004.

[9] H. Ocak, "Optimal classification of epileptic seizures in EEG using wavelet analysis and genetic algorithm," Signal Processing 88, 2008, pp. 1858–1867.

[10] M. Hagan, H. Demuth, and M. Beale, "Neural Network Design," PWS Publishing Company 1995, ISBN: 0-534-94332-2.

[11] T. Master, "Practical Neural Network Recipes," New York: John Wiley, 1993.

[12] V. N. Vapnik, "Statistical Learning Theory," New York: Wiley, 1998.

[13] S. Chandaka, A. Chatterjee, S. Munshi, "Cross-correlation aided support vector machine classifier for classification of EEG signals," Expert Systems with Applications,2008.