EEG Signal Classification using Linear Predictive Cepstral Coefficient Features

S. Pazhanirajan
Annamalai University
Assistant Professor
Department of Computer Science and Engineering

P. Dhanalakshmi,Ph.D
Annamalai University
Associate Professor
Department of Computer Science and Engineering

ABSTRACT

An electroencephalogram (EEG) is a procedure that records brain wave patterns, which are used to identify abnormalities related to the electrical activities of the brain. In this study an effective algorithm is proposed to automatically classify EEG clips into two different classes: normal and abnormal. For categorizing the EEG data, feature extraction techniques such as linear predictive coefficients (LPC) and linear predictive cepstral coefficients (LPCC) are used. Support vector machines (SVM) is used to classify the EEG clip into their respective classes by learning from training data.

General Terms

Feature Extraction, Classification

Keywords

EEG, EEG classification, EDF format, Feature extraction, Linear Prediction Cepstral Coefficients (LPCC), Support Vector Machines (SVM).

1. INTRODUCTION

The electroencephalogram is a measure of brain waves. It is a readily available test that provides evidence of how the brain functions over time. EEG is used in the evaluation of brain disorders [8]. Most commonly it is used to show the type and location of the activity in the brain during a seizure. It is also used to evaluate people who are having problems associated with brain function [7]. These problems might include confusion, coma, and tumors, long-term difficulties with thinking or memory, or weakening of specific parts of the body. An EEG is also used to determine brain death.

      The EEG data is converted into WAV format using audacity software. In this work, a WAV format EEG audio sound is reduced to a small set of parameters using various feature extraction techniques. The terms linear predictive coefficients, linear predictive cepstral coefficients refers to the features extracted from audio data. In the second step, Support vector machines are applied to classify audio into their respective classes by learning from training data. A pth order linear prediction (LP) analysis is used to capture the properties of the signal spectrum. In the LP analysis of audio, each sample is predicted as linear weighted sum of the past p samples, where p represents the order of prediction. In LPC, the cepstral is a common transform used to gain information from an audio signal.

2. RELATED WORK

In [1] different features were extracted from electro encephalogram (EEG) signals which were recorded from 32 subjects during hypnosis suggestion to classify hypnotizability levels. Based on the obtained result, a method was suggested to estimate the hypnosis susceptibility level from hypnosis EEG signals instead of using traditional clinical subjective tests. In order to characterize the non-Gaussian information contained within the EEG signals, a new feature extraction method based on bispectrum was proposed in [2] and applied to the classification of right and left motor imagery for developing EEG-based brain–computer interface systems. The proposed system showed satisfactory results. In [3] a system has been proposed to use the Covariance Matrix Adaptation Evolution Strategy (CMA-ES) for automatically tuning the filters. Spatial and frequency-selection filters were evolved to minimize both classification error and the number of frequency bands used. A signal-to-noise ratio (SNR) saliency measure was employed in [4] to determine saliency of input features of probabilistic neural networks (PNNs) used in classification of EEG signals. This method increases the classification accuracy by reducing noise in the classification model. In [5] an expert system is proposed which detects variations in EEG signals. The expert systems mixture of experts (ME) and modified mixture of experts (MME) are provided with diverse or composite features. The result demonstrates that the MME trained on diverse features achieved accuracy rates which were higher than that of the ME. In [6] a novel detector is proposed based on a support vector machine assembly classifier (SVMA). The SVMA consists of a group of SVMs each trained with a different set of weights between the seizure and non-seizure data and the users can selectively control the output of the SVMA classifier. The proposed algorithm improves the detection performance compared to that of traditional methods.

3. OUTLINE OF THE WORK

      In this work, automatic EEG signal feature extraction and classification approaches are presented. In order to discriminate the two categories of EEG namely normal and abnormal, two features namely LPC, LPCC are extracted to characterize the EEG content. Support vector machine (SVM) is applied to obtain the optimal class boundary between the classes by learning from training data. The experimental results show that the classification accuracy of SVM with linear predictive cepstral features can provide a better result.

The work is organized as follows. Feature extraction for EEG is presented in Section 4. Modeling techniques for EEG classification is described in Section 5. Experimental results using SVM is reported in Section 6. Finally, conclusions and future work are given in Section 7.

4. FEATURE EXTRACTION FOR EEG

EEG features representing the information can be extracted from the EEG signal at the segmental level. The segmental features are the features extracted from short (3 seconds) segments of the EEG signal. These features represent the short-time spectrum of the EEG signal. The spectral information of the same data uttered by different person may differ depend on the person brain signal, and the standard method recognized worldwide as International Digital EEG system. The selected features include LPC and LPCC.

4.1 Linear Prediction Analysis

EEG features representing the characteristics of the EEG signal can be extracted from the signal for classifying EEG. In this work, LPCC are used as EEG features. These features represent the short- time spectrum of the EEG signal. For EEG feature extraction, the differenced EEG signal is divided into frames of 20 ms, with a shift of 10 ms. A pth order LP analysis is used to capture the properties of the signal spectrum.

In the LP analysis of EEG, each sample is predicted as linear weighted sum of the past p samples, where p represents the order of prediction. If s(n) is the present sample, then it is predicted by the past p samples as

$$\hat{s}(n) = -\sum_{k=1}^{p} a_k\, s(n-k) \qquad (1)$$

The LPC are obtained using Levinson-Durbin recursive algorithm. This is known as LPC analysis. The difference between the actual and the predicted sample value is termed as the prediction error or residual, and is given by

$$(n) = s(n) - \hat{s}(n)$$

$$= s(n) + \sum_{k=1}^{p} a_k s(n-k) \qquad (2)$$

$$= \sum_{k=0}^{p} a_k s(n-k), a_0 = 1 \qquad (3)$$

Optimal predictor coefficients will minimize this mean square error. At minimum value of E,

$$\frac{\partial E}{\partial a_k} = 0, \qquad k = 1,2,\dots p \qquad (4)$$

Differentiating and equating to zero we get,

$$Ra = r \qquad (5)$$

Where, $a = \begin{bmatrix} a_1 & a_2 & \dots & a_p \end{bmatrix}^r$,

$$r = [r(1) r(2) \dots r(p)]^r,$$

and **R** is a Toeplitz symmetric autocorrelation matrix given by,

$$\boldsymbol{R} = \begin{bmatrix} r(0) & r(1) & \cdots & r(p-1) \\ r(1) & r(0) & \cdots & r(p-2) \\ \vdots & \vdots & \ddots & \vdots \\ r(p-1) & \cdots & \cdots & r(0) \end{bmatrix} \qquad (6)$$

Eq.(5) can be solved for predictor coefficients using Durbin's algorithm as follows:

$$E^{(0)} = r[0] \qquad (7)$$

$$k_i = \frac{r[i] - \sum_{j=1}^{L-1} \alpha_j^{i-1} \cdot r[|i-j|]}{E^{(i-1)}} \qquad (8)$$

$$1 \le i \le p$$

$$\alpha_j^i = k_i \qquad (9)$$

$$\alpha_j^i = \alpha_j^{(i-1)} - k_i \cdot \alpha_{i-j}^{(i-1)} \qquad (10)$$

$$E^i = \left(1 - k_i^2\right) . E^{i-1} \qquad (11)$$

The above set of equations is solved recursively for i = 1, 2 . . . , p. The final solution is given by

$$a_{m\,=\alpha_m^{(p)}} \qquad 1 \le m \le p \qquad (12)$$

where, $a_m\,'s$ are linear prediction coefficients (LPCs). In this work, 14th order LP coefficients are extracted.

4.2 Linear Prediction Cepstral Coefficients

The cepstral is a common transform used to gain information from an EEG signal. It can be used to separate the excitation signal (which contains the words and the pitch) and the transfer function (which contains the voice quality). The cepstrum can be seen as information about rate of change in the different spectrum bands. The recursive relation between the predictor coefficients and cepstral coefficients is used to convert the LP coefficients (LPC) into LP cepstral coefficients c_k.

$$c_0 = \ln\sigma^2$$

$$c_m = a_m + \sum_{k=1}^{m-1} \left(\frac{k}{m}\right) c_k a_{m-k}\,; \quad 1 \le m \le p$$

$$c_m = \sum_{k=1}^{m-1} \left(\frac{k}{m}\right) c_k a_{m-k}; \quad p < m \le d \qquad (13)$$

where σ^2 is the gain term in the LP analysis and d is the number of LP cepstral coefficients. A 19 dimensional weighted linear prediction cepstral coefficient (LPCC) for each frame is used as a feature vector.

5. SUPPORT VECTOR MACHINE

Support vector machine (SVM) is based on the principle of structural risk minimization. SVM learns an optimal separating hyper plane from a given set of positive and negative examples. It minimizes the structural risk, that is, the probability of misclassifying yet to be seen patterns for a fixed but unknown probability distribution of the data. This is in contrast to traditional pattern recognition techniques of minimizing the empirical risk, which optimizes the performance on the training data. SVM can be used for pattern classification. For linearly separable data SVM finds a separating hyper plane which separates the data with the largest margin. For linearly inseparable data, it maps the data in the input space into a high dimension space

$$x \in R^I \rightarrow \Phi(x) \in R^H$$

with kernel function $\Phi(x)$, to find the separating hyper plane. SVMs are evaluated as popular tools for learning from the given data. The reason is that SVMs are more effective than the traditional pattern recognition approaches based on the combination of a feature selection procedure and a conventional classifier.

SVM is a statistic machine learning technique that has been successfully applied in the pattern recognition area. SVM constructs a linear model to estimate the decision function using non-linear class boundaries based on support vectors. If the data are linearly separated, SVM trains linear machines for an optimal hyper plane that separates the data without error and into the maximum distance between the hyper plane and the closest training points. The training points that are closest to the optimal separating hyper plane are called support vectors.

Fig.1 shows the architecture of the SVM. SVM maps the input patterns into a higher dimensional feature space through some nonlinear mapping chosen a priori. A linear decision surface is then constructed in this high dimensional feature space. Thus, SVM is a linear classifier in the parameter space, but it becomes a non-linear classifier as a result of the non-linear mapping of the space of the input patterns into the high dimensional feature space.
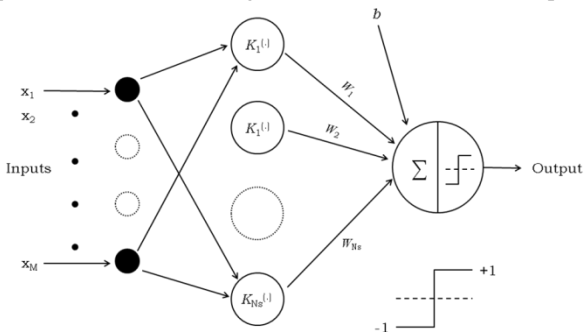

Fig 1: Architecture of the SVM (Ns is the number of support vectors).

An example for SVM kernel function Φ (x) maps 2-Dimensional input space to higher 3-Dimensional feature space as shown in Fig. 2. SVM was originally developed for two class classification problems. The N class classification problem can be solved using N SVMs. Each SVM separates a single class from all the remaining classes.

SVM generally applies to linear boundaries. In the case where a linear boundary is inappropriate SVM can map the input vector into a high dimensional feature space. By choosing a non-linear mapping, the SVM constructs an optimal separating hyper plane in this higher dimensional space, as shown in Fig. 2. The function K is defined as the kernel function for generating the inner products to construct machines with different types of non-linear decision surfaces in the input space.
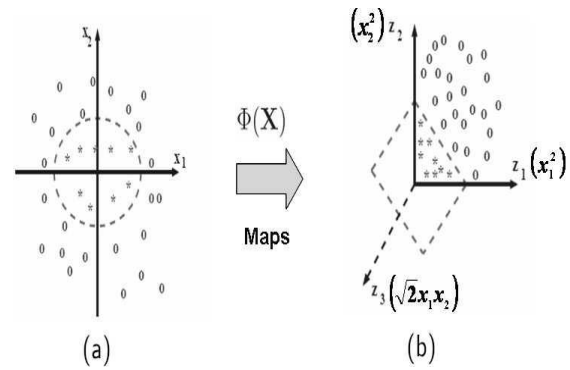
$$K (x,x_i) = \Phi(x) .\Phi(x_i)$$

Fig 2: An example for SVM kernel function $\Phi(x)$ maps 2-dimensional input space to higher 3-dimensional feature space. (a) Nonlinear problem. (b) Linear problem.

The kernel function may be any of the symmetric functions that satisfy the Mercer's conditions (Courant and Hilbert, 1953). There are several SVM kernel functions as given in Table 1.

Types of Kernels	Inner Product of Kernel K(x^t,x_i)	Details
Polynomial	$(x^t x_i + 1)^p$	Where x is input patterns, x_i is support vectors, p is degree of the polynomial. $\sigma 2$ is variance, $1 \le i \le Ns$, Ns is number of support vectors. β_0, β_1 are constant values.
Gaussian	$Exp[-\|x^t - x_i\|^2 / 2 \sigma 2]$	
Sigmoidal	$Tanh(\beta_0(x^t x_i) + \beta_1)$	

Table 1. Types of SVM inner product kernels.

The dimension of the feature space vector $\Phi(x)$ for the polynomial kernel of degree p and for the input pattern dimension of d is given by

$$\frac{(p + d)!}{p! \, d!}$$

For sigmoidal kernel and Gaussian kernel, the dimension of feature space vectors is shown to be infinite. Finding a suitable kernel for a given task is an open research problem. Given a set of EEG data corresponding to N categories for training, N SVMs are trained. Each SVM is trained to distinguish between one category and all other categories in the training set. During testing, the class label l of an EEG x can be determined

$$l = \begin{cases} n, if\ d_n(x) + t > 0 \\ 0, if\ d_n(x) + t \leq 0 \end{cases}$$

where $d_n(x) = \max\{d_i(x)\}_{i=1}^n$ and di(x) is the distance from x to the SVM hyper plane corresponding to category i. The classification threshold is t, and the class label $l = 0$ stands for unknown.

6. EXPERIMENTAL RESULTS
LPC and LPCC features are extracted from EEG data. The extracted features are classified using the SVM classifier into two classes. The feature extraction techniques use 2 hours of EEG data obtained from Raja Muthiah Medical College and Hospital, Annamalai University. The data is collected from 20 number of patients, in which 10 are male and 10 are female. The data is sampled at 8khz for optimized features extraction. The features are separated into 3 seconds of training and testing group, which is given to the SVM classifier.

The SVM learning is used to calculate the classification parameters. The EEG frames are classified into the respective classes by the training process. For each frame the LPC and LPCC features are obtained based on the calculated features. The SVM learning algorithm is applied to generate the classification parameters. The EEG data is classified using the generated classification parameters. The classification results for the different features are shown Fig.3. From the result it is observed that the overall classification accuracy is 84% using LPCC as features.
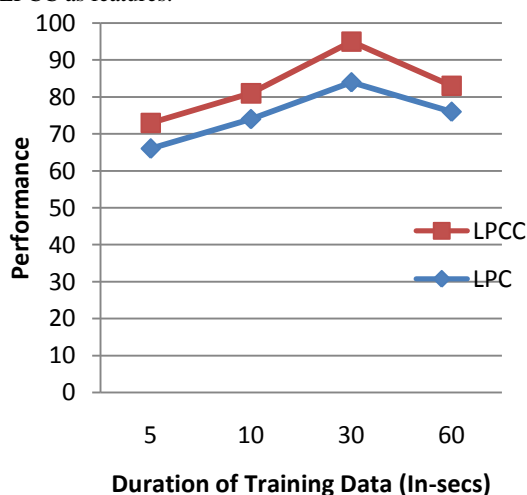

Fig 3: Performance of SVM for EEG classification

7. CONCLUSION
In this work, we proposed an automatic EEG classification system using SVM. Linear Prediction Coefficients (LPC) and Linear Prediction Cepstral Coefficients (LPCC) are calculated as features to characterize EEG content. A non linear support vector machine learning algorithm is applied to obtain the optimal class boundary between the various classes, normal and abnormal by learning from training data. Experimental results show that the proposed EEG classification scheme is very effective and accuracy. In future, extracting different features and developing better classification algorithms and integration of classifiers to reduce the classification errors will be implemented.

8. REFERENCES
[1] Golnaz Baghdadi, Ali Motie Nasrabadi (2012). Comparison of different EEG Features in Estimation of Hypnosis Susceptibility Level. Elsevier(pp.590-597).

[2] Shang-Ming Zhou, John Q. Gan, Francisco Sepulveda (2007). Classifying mental tasks based on features of higher-orderstatistics from EEG signals in brain–computer interface. Elsevier(pp.1629-1640).

[3] Ricardo Aler, Inés M. Galván, José M. Valls (2012). Applying evolution strategies to preprocessing EEG signals for brain–computer interfaces. Elsevier(pp.53-66).

[4] Elif Derya Ubeyli, Inan Culer (2009). Measuring saliency of features representing EEG signals using signal-to-noise ratios. Elsevier(pp.501-509).

[5] Elif Derya Ubeyli, Inan Culer (2006). Features Extracted by Eigenvector Methods for Detecting Variability of EEG Signals. Elsevier(pp.592-603).

[6] Y. Tang, D.M. Durand (2012). A Tunable Support Vector Machine Assembly Classifier for Epileptic Seizure Detection. Elsevier(pp.3925-3938).

[7] Z. Khalili1, M. H. Moradi (2008). Emotion Detection Using Brain and Peripheral Signals. Biomedical engineering faculty. IEEE Transactions on Proceedings of the CIBEC.

[8] Ashish Panat, Anita Patil (2012). Analysis of Emotion Disorders based on EEG signals of Human Brain. International Journal of Computer Science, Engineering and Applications (IJCSEA) Vol.2, No.4.

[9] Abdulhamit Subasi (2007). EEG Signal Classification Using Wavelet Feature Extraction and a Mixture of Expert Model. Elsevier(pp.1084 -1093).

[10] Duda, R. O., Hart, P. E., & Strok, D. G. (2001). Pattern Classification. New York: John Wiley-Interscience.

[11] P. Dhanalakshmi, S. Palanivel, V. Ramalingam (2009). Classification of Audio Signals Using SVM and RBFNN. Elsevier(pp. 6069-6075).

[12] R. Courant, H. Hilbert-methods of Mathematical Physics-vol.1- Interscience Publishers Inc.1953.