

Wavelet packet
feature extraction and
support vector
machine psychology
essay



\n[[toc title="Table of Contents"](#)]\n

\n \t

1. [1. Introduction](#) \n \t
2. [2. DATA SELECTION](#) \n \t
3. [3. methods](#) \n \t
4. [3. 1 Analysis using DWT and WPT](#) \n \t
5. [3. 2 Principal component analysis](#) \n \t
6. [3. 3 Linear discriminant analysis](#) \n \t
7. [3. 4 SVM classifier](#) \n \t
8. [4. results and discussion](#) \n \t
9. [5. conclusion](#) \n

\n[/toc]\n \n

ABSTRACT- The aim of this work is an automatic classification of the electroencephalogram (EEG) signals by using statistical features extraction and support vector machine. From a real database, two sets of EEG signals are used: EEG recorded from a healthy person and from an epileptic person during epileptic seizures. Three important statistical features are computed at different sub-bands discrete wavelet and wavelet packet decomposition of EEG recordings. In this study, to select the best wavelet for our application, five wavelet basis functions are considered for processing EEG signals. After reducing the dimension of the obtained data by linear discriminant analysis and principal component analysis, feature vectors are used to model and to train the efficient support vector machine classifier. In order to show the efficiency of this approach, the statistical classification performances are

evaluated, and a rate of 100% for the best classification accuracy is obtained and is compared with those obtained in other studies for the same data set.

Keywords- EEG; Discrete Wavelet Transform, Wavelet Packet Transform, Support Vector Machine, Statistical analysis, classification.

1. Introduction

In neurology, the electroencephalogram (EEG) is a non-invasive test of brain function that is mostly used for the diagnosis and classification of epilepsy. The epilepsy episodes are a result of excessive electrical discharges in a group of brain cells. Epilepsy is a chronic neurological disorder of the brain that affects over 50 million people worldwide and in developing countries, three fourths of people with epilepsy may not receive the treatment they need [1]. In clinical decisions, the EEG is related to initiation of therapy to improve quality of epileptic patient's life. However, EEG signals occupy a huge volume and the scoring of long-term EEG recordings by visual inspection, in order to classify epilepsy, is usually a time consuming task. Therefore, many researchers have addressed the problem of automatic detection and classification of epileptic EEG signals [2, 3]. Different studies have shown that EEG signal is a non-stationary process and non-linear features are extracted from brain activity recordings in order to specific signal characteristics [2, 4, 5, 6]. Then these features are used as input of classifiers [11]. Subasi in [7] used the discrete wavelet transform (DWT) coefficient of normal and epileptic EEG segments in a modular neural network called mixture of expert. For the same EEG data set, Polat and Günes [8] used the feature reduction methods including DWT,

autoregressive and discrete Fourier transform. In Subasi and Gursoy [9], the dimensionality of

the DWT features was reduced using principal component analysis (PCA), independent component analysis (ICA) and linear discriminant analysis (LDA). The resultant features were used to classify normal and epilepsy EEG signals using support vector machine. Jahankhani, Kodogiannis and Revett [10] have obtained feature vectors from EEG signals by DWT and performed the classification by multilayer perceptron (MLP) and radial basis function network. Wavelet packet transform (WPT) appears as one of most promising methods as shown by a great number of works in the literature [11] particularly for ECG signals and relatively fewer, for EEG signals. In [12], Wang, Miao and Xie used wavelet packet entropy method to extract features and K-nearest neighbor (K-NN) classifier. In this work, both DWT and WPT split non stationary EEG signals into frequency sub-bands. Then a set of statistical features such as standard deviation, energy and entropy from real database EEG recordings were computed from each decomposition level to represent time-frequency distribution of wavelet coefficients. LDA and PCA are applied to these various parameters allowing a data reduction. These features were used as an input to efficient SVM classifier with two discrete outputs: normal person and epileptic subject. A measure of the performances of these methods is presented. The remaining of this paper is organized as follows: Section 2 describes the data set of EEG signals used in our work. In Section 3, preliminaries are presented for immediate reference. This is followed by the step up of our experiments and the results in section 4. Finally, some concluding remarks are given in Section 5.

2. DATA SELECTION

We have used the EEG data taken from the artifact free EEG time series database available at the Department of Epileptology, University of Bonn [23]. The complete dataset consists of five sets (denoted A-B-C-D-E). Each set contains 100 single-channel EEG signals of 23, 6s. The normal EEG data was obtained from five healthy volunteers who were in the relaxed awake state with their eyes open (set A). These signals were obtained from extracranially surface EEG recordings in accordance with a standardized electrode placement. Set E contains seizure activity, selected from all recording sites exhibiting ictal activity. All EEG signals were recorded with the same 128 channel amplifier system and digitized at 173.61 Hz sampling. 12 bit analog-to-digital conversion and band-pass (0.5-40 Hz) filter settings were used. For a more detailed description, the reader can refer to [13]. In our study, we used set A and set E from the complete dataset.

Raw EEG signal

Feature extraction: Energy, Entropy and Standard deviation from DWT and WPT decomposition coefficients

Dimensionality reduction by LDA and PCA

Classification and

Performance measure

Healthy

Epileptic

<https://assignbuster.com/wavelet-packet-feature-extraction-and-support-vector-machine-psychology-essay/>

Figure 1 The flow chart of the proposed system

3. methods

The proposed method consists of three main parts: (i) statistical feature extraction from DWT and from WPT decomposition coefficients, (ii) dimensionality reduction using PCA and LDA, and (iii) EEG classification using SVM. The flow chart of the proposed method is given in figure 1. Details of the pre-processing and classification steps are examined in the following subsections.

3. 1 Analysis using DWT and WPT

Since the EEG is a highly non-stationary signal, it has been recently recommended the use of time-frequency domain methods [14]. Wavelet transform can be used to decompose a signal into sub-bands with low frequency (approximate coefficients) and sub-bands with high frequency (detailed coefficients) [15, 16, 17]. Under discrete wavelet transform (DWT), only approximation coefficients are decomposed iteratively by two filters and then down-sampled by 2. The first filter $h[.]$ is a high-pass filter which is the mirror of the second low pass filter $l[.]$. DWT gives a left recursive binary tree structure. We processed 16 DWT coefficients. Wavelet packet transform (WPT) is an extension of DWT that gives a more informative signal analysis. By using WPT, the lower, as well as the higher frequency bands are decomposed giving a balanced tree structure. The wavelet packet transform generates a full decomposition tree, as shown in figure 2. In this work, we performed five-level wavelet packet decomposition.

The two wavelet packet orthogonal bases at a parent node (i, p) are obtained from the following recursive relationships Eq. (1) and (2),

where $l[n]$ and $h[n]$ are low (scale) and high (wavelet) pass filter, respectively; i is the index of a subspace's depth and p is the number of subspaces [15]. The wavelet packet coefficients corresponding to the signal $x(t)$ can be obtained from Eq. (3),

l

$(3, 0) (3, 1) \dots (3, 6) (3, 7)$

h

$l h l h$

$l h$

$h l h$

$l h$

l

SIGNAL

$(0, 0)$

$(1, 0)$

$(1, 1)$

(2, 0)

(2, 1)

(2, 2)

(2, 3)

Figure 2 Third level wavelet packet decomposition of EEG signal

Table 1 gives the frequency bands for each level of WPT decomposition.

Figures 3 and 4 show the fifth level wavelet packet decomposition of EEG segments, according to figure 2. We processed 32 WPT coefficients.

Therefore, in this study, three statistical parameters: energy feature (En), the measure of Shannon entropy (Ent) and standard deviation (Std) are computed,

(4)

(5)

(6)

3. 2 Principal component analysis

To make a classifier system more effective, we use principal component analysis (PCA) for dimensionality reduction. The purpose of its implementation is to derive a small number of uncorrelated principal components from a larger set of zero-mean variables, retaining the maximum possible amount of information from the original data. Formally,

the most common derivation of PCA is in terms of standardized linear projection, which maximizes the variance in the projected space [18, 19]. For a given p -dimensional data set X , the m principal axes W_1, \dots, W_m where $1 \leq m \leq p$, are orthogonal axes onto which the retained variance is maximum in the projected space. Generally, W_1, \dots, W_m can be given by the m leading eigenvectors of the sample

Table1 Frequency band of each wavelet decomposition level.

Decomposition

level

Frequency band (Hz)

1

2

3

4

5

0-86. 8; 86. 8-173. 6

0-43. 5; 43. 5-86. 8; 86. 3-130. 2 ; 130. 2-173. 6

0-21. 75; 21. 75-43. 5; 43. 5-54. 375; 54. 375-86. 3; 86. 3-108. 05; 108. 05-130. 2; 130. 2 130. 2-151. 95; 151. 95-173. 6;

0-10. 875; 10. 875-21. 75; 21. 75-32. 625; 32. 625-43. 5; 43. 5-54. 375; 54. 375-65. 25; 65. 25-76. 125; 76. 125-87; 87-97. 875; 97. 875-108. 75; 108. 75-119. 625; 119. 625-130. 5; 130. 5-141. 375; 141. 375-152. 25; 152. 25-163. 125; 163. 125-173. 6

0-5. 44; 5. 44-10. 875; 10. 875-16. 31; 16. 31-21. 75; 21. 75-27. 19; 27. 19-32. 625; 32. 625-38. 06; 38. 06-43. 5; 43. 5-48. 94; 48. 94-54. 375; 54. 375-59. 81; 59. 81-65. 25; 65. 25-70. 69; 70. 69-76. 125; 76. 125-81. 56; 81. 56-87; 87-92. 44; 92. 44-97. 87; 97. 87-103. 3; 103. 3-108. 75; 108. 75-114. 19; 114. 19-119. 625; 119. 625-125. 06; 125. 06-130. 5; 130. 5-135. 94; 135. 94-141. 38; 141. 38-146. 81; 146. 81-152. 25; 152. 25-157. 69; 157. 69-163. 125; 163. 125-168. 56; 168. 56-173. 6

covariance matrix where \bar{x} is the sample mean and N is the number of samples, so that $SW_i = \lambda_i W_i$, where λ_i is the i th largest eigenvalue of S . The m principal components of a given observation vector x_i are given by the reduced feature vector .

3. 3 Linear discriminant analysis

Linear discriminant analysis (LDA) projects high-dimensional data onto a low-dimensional space where the data can achieve maximum class separability [19]. The aim of LDA is to create a new variable that is a combination of the original predictors, i. e. the derived features in LDA are linear combinations of the original variables, where the coefficients are from the transformation matrix i. e. LDA utilizes a transformation matrix W , which can maximize the ratio of the between-class scatter matrix S_B to the within-class scatter matrix S_W , to transform the original feature vectors into lower dimensional feature

space by linear transformation. The linear function $y = W^T x$ maximizes the Fisher criterion $J(W)$ [19],

where $x_j(i)$ represents the j th sample of the i th of total c classes. k is the dimension of the feature space, and μ_i is the

Figure 3 Fifth level wavelet packet decomposition of healthy EEG signal (set A).

Figure 4 Fifth level wavelet packet decomposition of epileptic EEG signal (set E).

mean of the i th class. M_i is the number of samples within classes i in total number of classes.

where μ is the mean of the entire data set.

As a dimensionality reduction method, LDA has also been adopted in this work.

3. 4 SVM classifier

In this work, SVM [20] has been employed as a learning algorithm due to its superior classification ability. Let n examples $S = \{x_i, y_i\}_{i=1}^n$, $y_i \in \{-1, +1\}$, where x_i represent the input vectors, y_i is the class label. The decision hyperplane of SVM can be defined as (w, b) ; where w is a weight vector and b a bias. The optimal hyperplane can be written as,

where w_0 and b_0 denote the optimal values of the weight vector and bias.

Then, after training, test vector is classified by decision function,

To find the optimum values of w and b , it is required to solve the following optimization problem:

subject to

where ξ_i is the slack variable, C is the user-specified penalty parameter of the error term ($C > 0$), and ϕ the kernel function [21]. A radial basis function (RBF) kernel defined as,

was used, where σ is kernel parameter defined by the user.

4. results and discussion

Before we give the experimental results and discuss our observations, we present three performance measures used to evaluate the proposed classification method. (i) Sensitivity, represented by the true positive ratio (TPR), is defined as

(ii) Specificity, represented by the true negative ratio (TNR), is given by,

(iii) and average classification accuracy is defined as,

(16)

where FP and FN represent false positive and false negative, respectively.

All the experiments in this work were undertaken over 100 segments EEG time series of 4096 samples for each class set A and set E. There were two diagnosis classes: Normal person and epileptic patient. To estimate the reliability of the proposed model, we utilize ten-fold cross validation method.

The data is split into ten parts such that each part contains approximately

<https://assignbuster.com/wavelet-packet-feature-extraction-and-support-vector-machine-psychology-essay/>

the same proportion of class samples as in the classification dataset. Nine parts (i. e. 90%) are used for training the classifier, and the remaining part (i. e. 10%) for testing. This procedure is repeated ten times using a different part for testing in each case. As illustrated in Fig. 3 and 4, feature vectors were computed from coefficient of EEG signals. Taking energy as feature vector, figure 5 shows that the features of both normal and epileptic EEG signals are mixed. The proposed analysis using wavelets was carried out using MATLAB R2011b.

In literature, there is no common suggestion to select a particular wavelet. Therefore, a very important step before classifying EEG signals is to select an appropriate wavelet for our application. Then, five wavelet functions namely Daubechies, Coiflets, Biorthogonal, Symlets and Discrete Meyer wavelets are examined and compared, in order to evaluate the performance of various types of wavelets. Figure 6 shows accuracy, sensitivity and specificity from different wavelets. We see that the best wavelet giving good correct rate is the Db2, Db4, coif3 and Bior1. 1. The choice of the mother wavelet is focused on daubechies where the length of the filter is $2N$, while coifflet wavelet filter is $6N$ and biorthogonal wavelet $(2N + 2)$. After EEG signal Db2 wavelet decomposition and dimensionality reduction, results of correct rate classification are showed in Table 2. The classification accuracy varies from the optimum value (100%) to a lowest value (87%). The results using standard deviation are the best results obtained and using entropy is better than using energy in EEG signals classification. In this study, experimental results show that linear discriminant analysis based on wavelet packet decomposition improves classification and the optimum SVM results are

obtained by using standard deviation feature computed from wavelet packet coefficient and LDA reduction method. For this proposed scheme, the accuracy of the classification is 100%. This method presents a novel contribution and has not yet been presented in the literature. Figure 7 shows the average rate of classification (accuracy, sensitivity, specificity) obtained with different methods of decomposition (DWT or WPT), two reduction methods (LDA or PCA) and three characteristic features (standard deviation, energy, entropy) using the four best wavelet (Db2, Db4, coif3 and Bior1. 1). We see that the combination of LDA with standard deviation have an optimum average accuracy rate of 99.90% and combination of standard deviation with PCA reaches 99.50%. Table 3 gives a summary of the accuracy results obtained by other studies from the same dataset (set A and set E) using extraction of features from EEG signal and their classification.

5. conclusion

In this paper, EEG signals were decomposed into time-frequency representations using discrete wavelet transform, wavelet packet transform and statistical features were

Figure 5 Energy feature vector coefficient D3 versus D2 (adapted from [22]).

Table 3 Epilepsy classification accuracies evaluation obtained in literature from the same data sets

Authors

Method

Accuracy (%)

[7] Subasi

DWT + Mixture of Expert

94. 50

[8] Polat and Günes

DWT+DFT+ Auto-regres-sive model + Decision Tree

99. 32

[9] Subasi and Gursoy

DWT+PCA+ LDA+ICA +SVM

98. 75(PCA)

100(LDA)

99. 5(ICA)

[12] Wang, Miao and Xie

WPT+ Entropy-hierarchical K-NN classification

99, 44

[14] Übeyli

Burg autoregressive + LS-SVM

99.56

Our method

WPT + Standard deviation+

LDA + SVM

100

computed to represent their distribution. The most suitable mother wavelets for feature extraction and classification were found. The selection of the suitable mother wavelet and using reduction methods lead to the improvement of performance of EEG signal classification. It has been shown by experiments that for the SVM and the combination of the standard deviation with LDA have the highest correct classification rate of 100% in comparison with other techniques. The interest in expert systems for detection and classification of epileptic EEG signal is expected to grow more and more in order to assist and strengthen the neurologist in numerous tasks, especially, to reduce the number of selection for classification performance.

These promising results encourage us to continue with more depth our study and to apply it to other databases recorded with other diseases.