Dear Author,

Please check your proof carefully and mark all corrections at the appropriate place in the proof (e.g., by using on-screen annotation in the PDF file) or compile them in a separate list.

For correction or revision of any artwork, please consult http://www.elsevier.com/artworkinstructions.

Any queries or remarks that have arisen during the processing of your manuscript are listed below and highlighted by flags in the proof. Click on the Q link to go to the location in the proof.

<table>
<thead>
<tr>
<th>Location in article</th>
<th>Query / Remark: click on the Q link to go please insert your reply or correction at the corresponding line in the proof</th>
</tr>
</thead>
<tbody>
<tr>
<td>Q1</td>
<td>Please check affiliation &quot;a&quot;, and correct if necessary.</td>
</tr>
<tr>
<td>Q2</td>
<td>Please note that the reference style has been changed from Name-Date style to Numbered style as per the journal specifications.</td>
</tr>
<tr>
<td>Q3</td>
<td>Please provide vol. no. and also check the journal title in Ref. [16].</td>
</tr>
<tr>
<td>Q4</td>
<td>Please provide vol. no. in Ref. [30].</td>
</tr>
</tbody>
</table>

Thank you for your assistance.
A comparative study of wavelet families for EEG signal classification

Tapan Gandhi a, Bijay Ketan Panigrahi b, Sneh Anand a

a Brain Dynamics and Cognitive Engineering Laboratory, Centre for Biomedical Engineering, Indian Institute of Technology, All India Institute of Medical Sciences, New Delhi, India
b Machine Intelligence Research Laboratory, Department of Electrical Engineering, Indian Institute of Technology, Delhi, India

- The most useful wavelet function was found out for EEG signal analysis. Coiflets 1 is the suitable candidate for accurate classification of the EEG signal. Higher efficiency and accuracy were obtained in less computational time.
A comparative study of wavelet families for EEG signal classification

Tapan Gandhi, Bijay Ketan Panigrahi, Sneh Anand

ARTICLE INFO

Article history:
Received 9 March 2010
Received in revised form 4 April 2011
Accepted 13 April 2011
Communicated by V. Palade

Abstract

Over the past two decades, wavelet theory has been used for the processing of biomedical signals for feature extraction, compression and de-noising applications. However the question as to which wavelet family is the most suitable for analysis of non-stationary bio-signals is still prevalent among researchers. This paper attempts to find the most useful wavelet function among the existing members of the wavelet families for electroencephalogram signal (EEG) analysis. The EEGs considered for this study belong to both normal as well as abnormal signals like epileptic EEG. Important features such as energy, entropy and standard deviation at different sub-bands were computed using the wavelet functions—Haar, Daubechies (orders 2–10), Coiflets (orders 1–10), and Biorthogonal (orders 1, 1.4, 2.4, 3.5, and 4.4). Feature vectors were used to model and train the Probabilistic Neural Network (PNN) and the classification accuracies were evaluated for each case. The results obtained from PNN classifier were compared with Support Vector Machine (SVM) classifier. From the statistical analysis, it was found that Coiflets 1 is the most suitable candidate among the wavelet families considered in this study for accurate classification of the EEG signals. In this work, we have attempted to improve the computing efficiency as it selects the most suitable wavelet function that can be used for EEG signal processing efficiently and accurately with lesser computational time.

Keywords:
Wavelet transform
Feature extraction
Epileptic EEG
Probabilistic Neural Network (PNN)
Support Vector Machine (SVM)

1. Introduction

Processing and analysis of bio-signals using software techniques have come into play since the early 1960s providing physicians with fast and accurate means toward more precise diagnosis [1]. Feature extraction and classification of the signal (required for diagnostic purposes), however, have always been the two most critical problems encountered in time domain analysis [2]. The sole purpose of feature extraction, is to extract salient characteristics from digitized data collected from the data acquisition phase [3] followed by classification based on the extracted features [4–6]. The feature of the signal is derived from its linear expansion coefficients where the most common linear expansion method used is Fourier transform [9]. Since the early days of digital processing, Fourier transform has been most commonly applied for signal representation. However, bio-signals frequently characterized by a non-stationary time behavior if processed with Fourier transform, would not yield the best result. Hence, for such transient signals, a time–frequency representation is highly desirable, with an aim to derive meaningful features [10]. From the variety of approaches available [2,3,7–9,11,12], the Wavelet transform was found to be an effective time–frequency analysis tool for analyzing the transient signals, as this method unifies different tools that have been developed for processing application till now. The feature extraction and representation properties can be used to evaluate various transient events in biological signals [13]. Several wavelet families are available for signal characterization and selection of appropriate wavelet is very important for the analysis of signals. Depending on the type of bio-signal to be analyzed, the mother wavelet is chosen according to the convenience and the requirement of the experimenter. The research work done till date for bio-signal classification using wavelet technique has been carried out mostly using the Daubechies family of order mostly 2 or 4 [3,14–16]. Moreover, the automated diagnostic system designed for detection purposes gives an accuracy of 70–90% (depending on the bio-signal classified) [14–16]. The research presented in this paper deals with the selection of the most suitable wavelet function for signal analysis of EEG signature in particular. Here, the wavelet techniques were used to decompose the epileptic/normal EEGs for feature extraction followed by classification of the signals using SVM and PNN with an impressive diagnostic accuracy of about 99.3% [7,8]. The reason for selection of epileptic EEG signal for this work is its variation in morphologies like sharp, spikes, and slow waves, each of which is in a different frequency range [17].

* Corresponding author.
E-mail addresses: gandhitk@gmail.com (T. Gandhi), sneha@cbme.iitd.ac.in (S. Anand), bkpanigrahi@ee.iitd.ac.in (S. Anand).

© 2011 Elsevier B.V. All rights reserved.
The property of wavelet theory can be best explored in such type of transient signal as it decomposes the epochs into different frequency bands (spike—13.5–50 Hz; sharp—5–12.5 Hz, and slow—1–2.5 Hz) which can then be analyzed depending upon their scanning function [18].

Till date, several soft-computing methods have been proposed in the literature for the diagnosis of the epileptic activity in EEG signal [19–26]. They include template matching [27] time domain, frequency domain [27,28] and time–frequency domain [27–29] which are very few. There is no standard method for selecting the best wavelet for processing EEG signals [11,14,30]. The choice of wavelet has significant impact on the quality of results with regard to the classifier, which takes the wavelet coefficients as input features. Using efficient classification tool, precise learning ability and processing capacity of neural network can be found out to analyze EEGs efficiently in minimal time for reliable diagnosis.

The goal of this present work is to find out the most suitable wavelet function which can be used to extract features from EEG signals for various applications like brain machine interfacing or to design expert systems for diagnosis of epileptic activity efficiently.

2. Data collection and analysis

Two categories of data were selected for the present study. The first category of EEG data used for this study was recorded on a Grado Telefactor EEG Twin3 machine with sampling frequency 400 Hz available at Sir Ganga Ram Hospital, New Delhi. EEG recordings of twenty one subjects were collected from the hospital’s epileptic seizure database. Supracranial data was acquired using 16 gold electrodes on the scalp according to international 10–20 system of electrode placement. Two sets of data were selected for the study. The first set of data was epileptic signals having 300 epochs of 5 s duration. The second set of data was background activity of the same person when signals were seizure-free (300 epochs of 5 s duration). The data was selected under supervision of experienced neurologists from the large epileptic database of the hospital.

The second category of EEG data used for this study is publicly available, described in Ref. [27]. The detailed description of the signals can be obtained from reference as mentioned. The complete data set consists of five sets (denoted A–E), each containing 100 single channel EEG signals of 23.6 s duration. Sets A and B have been taken from surface (extracranial) EEG recordings of five healthy volunteers with eye open and closed, respectively. Signals from sets D and C have been measured in seizure-free intervals from five subjects in the epileptic zone and from the hippocampal formation of the opposite hemisphere of the brain. Set E comprises of epileptic signals recorded during seizure (ictal) from all recording sites. Sets C–E have been recorded intracranially.

2.1. Wavelet based feature extraction and parameter estimation

Performance of the expert system depends on the signal analysis, feature selection and classification methods used. Wavelet decomposition was employed and features were extracted from EEG (normal/abnormal) signals. However, the output of wavelet transform can be significantly affected by the choice of the mother wavelet (the basic wave shape) with which the signal is analyzed [31–33]. A wavelet family $\psi_{a,b}(t)$ is a set of elementary function generated by dilations and translations of a unique admissible mother wavelet $\psi(t)$

$$\psi_{a,b}(t) = \frac{1}{\sqrt{|a|}} \psi \left( \frac{t-b}{a} \right)$$ (1)

where $a, b \in \mathbb{R}$, $a \neq 0$, $a, b$ are the scaling (dilation) and translation parameters, respectively, and ‘t’ is the time. The scale parameter will decide the oscillatory frequency and the length of the wavelet, the translation parameter will decide its shifting position.

The WT can be implemented with a specially designed pair of FIR filters called quadrature mirror filter (QMFs) pairs. QMFs are distinctive because the frequency responses of the two FIR filters separate the high- and low-frequency components of the input signal. The dividing point is usually half-way between 0 Hz and half of the data sampling rate (the Nyquist frequency). The outputs of the QMF filter pair are decimated (or de-sampled) by a factor of two. The low-frequency (low-pass) filter output is fed into another identical QMF filter pair. This operation can be repeated recursively as a tree or pyramid algorithm, yielding a group of signals that divides the spectrum of the original signal into octave bands with successively coarser measurements in time as the width of each spectral band narrows and decreases in frequency. The tree or pyramid algorithm can be applied to the WT by using the wavelet coefficients as the filter coefficients of the QMF filter pairs. In WT multi-resolution algorithm (MRA), same wavelet coefficients are used in both low-pass (LP) and high-pass (HP) filters. The LP filter coefficients are associated with the scaling function, and the HP filter is associated with the wavelet function. Fig. 1 shows the tree algorithm of a multi-resolution WT for a discrete EEG signal sampled at 400 Hz.

The outputs of the LP filters are called the approximations (A), and the outputs of the HP filters are called the details (D). In MRA, any time series can be completely decomposed in terms of the approximation and detail coefficients based on the level of decomposition as shown in Fig. 1. Application of DWT on raw signal produces a multi-resolution analysis (MRA) of various statistical and non-statistical parameters across time and frequency. The subsets of the wavelet coefficients of the decomposition tree were selected as input vectors to the classifier.

2.3. Parameters for feature extraction

Feature sets were constructed using MRA analysis shown in Fig. 1 and coefficients were stored for further processing.

In order to reduce the feature dimension, few statistical and non-statistical parameters were considered instead of directly training the classifier with detail ($D_1(t)$) and approximation ($A_1(t)$) coefficients. For this purpose, we have selected Energy (EDA), Entropy (ENT), and standard deviation (SD) as parameters based on which the PNN was trained for the classification of signals.

The energy at each decomposition level was calculated using the following equations:

$$ED_i = \sum_{j=1}^{N} |D_j|^2, \quad i = 1, 2, \ldots, l$$

$$EA_i = \sum_{j=1}^{N} |A_j|^2$$ (2)

The entropy at each decomposition level was calculated using the following equation.

$$\text{Entropy} = -\sum_{j=1}^{N} |A_j| \log |A_j|$$

Entropy:

\[ EN_i = - \sum_{j=1}^{N} D_j \log(D_j), \quad i = 1, 2, \ldots, l \]  

where \( i = 1, 2, \ldots, l \) is the wavelet decomposition level from levels 1 to \( l \). \( N \) is the number of coefficients of detail or approximate at each decomposition level.

The standard deviation at each decomposition level was calculated using the following equation:

\[ \text{STD}_{\sigma_i} = \left( \frac{1}{N-1} \sum_{j=1}^{N} (D_{ij} - \mu_i)^2 \right)^{1/2}, \quad i = 1, 2, \ldots, l \]

where \( \mu_i \) is the mean and is given by

\[ \mu_i = \frac{1}{N} \sum_{j=1}^{N} D_{ij}, \quad i = 1, 2, \ldots, l \]

In this manner \((l+1)\) dimensional feature vector was obtained from \( l \) level decomposition as shown in Fig. 3.

Thus for an \( l \) level decomposition, the feature vector of any parameter can be represented as Feature=[\( XD_1 XD_2 \ldots XD_lXA_l \)] where \( X \) is the parameter, \( D \) is the detail coefficient, \( A \) is the approximation coefficient, and \( l \) is the level of decomposition.

However, as we need to optimize the technique and hence we choose one of the suitable parameters out of three described above by statistical means.

3. Probabilistic Neural Network model and classification of signals

An AI-based classifier is essentially a mapping \( f: \mathbb{R}^m \rightarrow \mathbb{Z}^n \) from the feature space to the discrete class space. The Artificial Neural Network (ANN) implements such a mapping by using a group of interconnected artificial neurons simulating the human brain. An ANN can be trained to achieve expected classification results against the input and output information stream. The Probabilistic Neural Network (PNN) is a kind of distance-based supervised learning ANNs, using a bell shape activation function. This technique makes decision boundaries nonlinear and hence it can approach the Bayesian optimal [28]. Compared with traditional back-propagation (BP) neural network, PNN is considered more suitable to medical application since it uses Bayesian strategy; a process familiar to medical decision makers [4,24,25,35]. The real-time property of PNN is also crucial for the research. There is no need to train the network over the entire data set again. Therefore one can quickly update the network as more and more patient data becomes available. The difference between the inference vector and the target vector are not used to modify the weights of the network. Moreover there is no need to set the initial weights of the network and also there is no relationship between the learning process and recalling processes. The learning speed of the PNN model is very fast, making it suitable for signal analysis and classification problems in real time.

3.1. K-fold cross-validation

Cross-validation, sometimes called rotation estimation [36–38], is the statistical practice of partitioning a sample of data into subsets such that the analysis is initially performed on a single subset, while the other subset(s) are retained for subsequent use in confirming and validating the initial analysis. The initial subset of data is called the training set; the other subset(s) are called validation or testing sets.

In K-fold cross-validation, the original sample is partitioned into \( K \) sub samples. Of the \( K \) sub samples, a single sub sample is retained as the validation data for testing the model, and the remaining \( K-1 \) sub samples are used as training data. The cross-validation process is then repeated \( K \) times (the folds), with each of the \( K \) sub samples used exactly once as the validation data. The \( K \) results from the folds then can be averaged (or otherwise combined) to produce a single estimation. The advantage of this method over repeated random sub-sampling is that all observations are used for both training and validation, and each observation is used for validation exactly once. The variance of the resulting estimate is reduced as \( k \) is increased.

The disadvantage of this method is that the training algorithm has to be run again from scratch \( k \) times, which means it takes \( k \) times as much computation to make an evaluation [39]. In this study, we have used 10-fold scheme to achieve best time as well as performance accuracies for the Indian data and after getting the optimized net, we have tested our classifier on international available database mentioned...
4. Results and discussion

Feature vectors were created from the extracted nodes of decomposed wavelet coefficients of EEG signals. It was shown in Fig. 2 that the features of both normal and abnormal EEG signals are highly mixed (taking energy as a feature vector).

Hence, our approach is to find out a suitable mother wavelet as well as features so that the classification of the signals can be achieved with higher accuracy in lesser time. All mother wavelets have different filter lengths. Obviously, the longer the wavelet filters’ length, higher the computational cost [34]. In this part, we investigated the influence of different kinds of wavelet to achieve satisfactory classification accuracy. In order to entangle this issue, four commonly used wavelets; Daubechies, Haar, Coiflets, and Biorthogonal were taken into account. The characteristics of these wavelets were shown in Table 1. The compact support and support width are used to measure the domain of wavelet function with non-zero values on it. Compact support means these non-zero values are only for a finite duration and support width indicates length of the non-zero duration. These parameters influence the frequency characteristics of the wavelet transform. A wavelet with small support width is fast to compute, but the narrowness in time-domain implies a large width. In frequency domain, conversely, wavelets with large compact support are smoother and have finer frequency resolution.

Decomposition levels also play an important role in optimization process. In 'l' level decomposition, (l+1) dimensional feature vector was created as shown in Fig. 1. In order to reduce computational cost, care should be taken in choosing the level of decomposition corresponding to classification performance. In the present study, although we have tested the classification performance based on selecting level from 1 to 10, only few wavelets were cited according to suitability and applicability. With training and testing in 7:3 ratios to PNN, we found that when decomposition levels were relatively as small as l < 4, the classification accuracy was about 90%.

When decomposition level was more than 6, accuracy of more than 99.33% was achieved with our data and nearly 100% accuracy was achieved with the data available from public database [27] used for comparative study.

Moreover, it was observed that there is no significant change in classification accuracy after 6th level of decomposition. In order to evaluate the performance of various types of wavelets, decomposition level at ‘6’ was fixed based on the accuracy based on Figs. 3 and 4. The tradeoff between the classification accuracy, time and spread were maintained to get the best classifier. In this work, 100 epochs of each category of seven subjects of our data and the publicly available data were chosen to test our hypothesis. To test the difference between these wavelet families under the spread range 0.5–2, the extracted features were used in PNN and the simulated testing accuracies were fed into SPSS-15 data view window for desired statistical analysis. Classification accuracy of Indian data were computed when publicly available data were used for training corresponding to various wavelets, spreads and parameters. For convenience, classification accuracy and computational time for few wavelet families with a single spread were shown in Table 2. From the descriptive analysis, we rejected the null hypothesis which means there is significant difference between the mean values. Moreover the statistical test was conducted within Daubechies wavelet family. The results showed no statistical difference, but based on the filter length, db2/db3 was found to be the best mother wavelet for signal decomposition.

In order to validate our hypothesis, SVM (Support Vector Machine) was also employed to classify the EEG signals. Though

![Fig. 2. Feature set representing energy of coefficients D2 and D3.](image)

<table>
<thead>
<tr>
<th>Wavelet</th>
<th>Biorthogonal</th>
<th>Orthogonality</th>
<th>Symmetry</th>
<th>Compact support</th>
<th>Support width</th>
<th>Filter length</th>
</tr>
</thead>
<tbody>
<tr>
<td>Db</td>
<td>No</td>
<td>Yes</td>
<td>Far from</td>
<td>Yes</td>
<td>2N−1</td>
<td>2N</td>
</tr>
<tr>
<td>Haar</td>
<td>No</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>2N−1</td>
<td>2N</td>
</tr>
<tr>
<td>Coif</td>
<td>No</td>
<td>Yes</td>
<td>Near from</td>
<td>Yes</td>
<td>6N−1</td>
<td>6N</td>
</tr>
<tr>
<td>Blor</td>
<td>Yes</td>
<td>No</td>
<td>Yes</td>
<td>Yes</td>
<td>2N+1 for recon</td>
<td>Max(2N, 2N0)+2</td>
</tr>
</tbody>
</table>

the accuracy of classification was less and computational time was more compared to PNN classifier, Coif1 and Db2 were still found to be the best for EEG signal classification, as shown in Tables 2 and 3.

Table 4 shows the classification accuracy of publicly available data, when Indian data was used for training.

So far, from the above analysis we found Coif1 with spread 2 and Db2/Db3 with spread 1 st
EEG classification when different wavelets were chosen. This means, the above mentioned wavelet has robustness in classification performances.

Again to find out the best parameter for feature extraction out of standard deviation, energy and entropy and to validate the accuracy of PNN, descriptive statistical analysis was conducted using SPSS-15. From the descriptive statistics of homogeneity it was found that both energy and standard deviation were equally significant at \( P < 0.001 \). Hence for EEG signal processing and feature extraction, both energy and standard deviation are equally suitable. On the basis of these results, we can choose Coif1 with energy/standard deviation as a feature to achieve higher classification accuracy as well as lesser computational cost compared to Db2/Db3.

5. Conclusion

In this work, the most suitable mother wavelet for feature extraction and classification of EEG signals was found based on the accuracy of expert system. Eighteen features were extracted from the sub-bands obtained after the decomposition by each of the wavelet family and then PNN was employed to classify the cases. Based on the classification accuracy and computational time, it was found that Coiflet of order 1 (Coif1) is the best wavelet family for analysis of EEG signal as the support width of the mother wavelet function resembles that of the EEG signal and also has a compact filter length, thus reducing the processing time. Moreover the selection of the suitable mother wavelet through parameterization leads to the improvement of performance of EEG signal analysis in comparison with random selection of the mother wavelet.

Acknowledgment

The authors are thankful to the authority of Sir Ganga Ram Hospital, New Delhi for giving access to their epileptic database and R.G. Andrzejaket et al. for their publicly available database, described in Ref. [38].

References


