Fuzzy entropy relevance analysis in DWT and EMD for BCI motor imagery applications

Análisis de relevancia con entropía difusa en aplicaciones BCI con imaginación motora mediante descomposiciones DWT y EMD

Abstract

Rhythm analysis in advanced signal processing methods has long of interest in application areas such as diagnosis of brain disorders, epilepsy, sleep or anesthesia analysis, and more recently in brain computer interfaces. In this paper the Discrete Wavelet Transform (DWT) and Empirical Mode Decomposition (EMD) techniques are applied to extract the brain rhythms from electroencephalographic (EEG) signals in motor imagination tasks, of left-and right hand, using public dataset BCI Competition 2003. Then the brain rhythms are characterized by statistical features. Additionally, fuzzy entropy algorithm was used to perform the relevance analysis to determine the most important features in the training set. Classification stage was performed using K-NN classifiers and SVM, obtaining classification accuracy up to 100 % with EMD. Classification results allow us to infer that the techniques used are appropriate to generate solutions in BCI applications for recognizing motor imagination in people with motor disabilities.

Key words: fuzzy entropy, wavelet, EMD, BCI.

Resumen

El análisis de ritmos en métodos avanzados de procesamiento de señal es de interés en áreas de aplicación, tales como diagnóstico de trastornos cerebrales, epilepsia, análisis del sueño o anestesia, y más recientemente en las interfaces cerebro-computador. En este trabajo se aplican la transformada wavelet discreta (DWT) y descomposición por modos empíricos (EMD) para extraer los ritmos cerebrales de señales electroencefalográficas (EEG) en tareas de imaginación de motora, de mano izquierda y derecha, utilizando la bases de datos pública BCI Competition 2003. Los ritmos cerebrales se caracterizan mediante funciones estadísticas; además, se utilizó el algoritmo de entropía difusa para realizar el análisis de relevancia y determinar las características más importantes en el conjunto de entrenamiento. La etapa de clasificación se realizó utilizando clasificadores K-NN y SVM, de la que se obtuvieron porcentajes de precisión de hasta el 100 % de clasificación. Los resultados de la clasificación permiten inferir que las técnicas utilizadas son adecuadas para generar soluciones en aplicaciones BCI para el reconocimiento de la imaginación motora en las personas con discapacidad motora.

Palabras claves: entropía difusa, transformada wavelet, EMD, BCI.
1. Introduction

The left-right hand motor imagination can alter neuronal activity in major sensorimotor areas, leading to changes in mu rhythm (noted $\mu$, with frequencies $8 - 12 \text{ Hz}$) and beta rhythm (noted $\beta$, with frequencies $> 13 \text{ Hz}$). The BCI systems require efficient online processing methods to classify these EEG signals in order to build a system that allows communication with its environment to patients with physical disabilities [1].

Several feature extraction techniques have been used for this purpose, including autoregressive models [2],[3], Common Spatial Patterns [4] and Wavelet Transform [5],[6]. Two methods for characterizing EEG signals in motor imagination task are used in this work. First, the Discrete Wavelet Transform (DWT), and second, the Empirical Mode Decomposition (EMD), which have shown promise in the treatment of EEG signal [6],[7] are used to decompose the EEG signal in its brain rhythms. Then, the brain rhythms are characterized with statistical functions as relative energy and standard deviation. Additionally, fuzzy entropy algorithm is used to perform a relevance analysis and determinate the most important features in the training set. Finally, different classifiers are used to recognize patterns of motor imagination. The application of this method was used to analyze the data set provided by Graz BCI Competition 2003.

2. Methods

2.1. Discrete Wavelet Transform (DWT)

Specifically, DWT analyzes the signal in different scale bands (frequency bands) $\Delta F_n$, with different resolutions by decomposing the signal into approach and detail components. DWT employs two sets of functions called scaling functions and wavelet functions, which are associated with low-pass filters and high-pass filters, respectively. The decomposition of the signal into different frequency bands $\Delta F_n$, is performed iteratively, producing a set of approximation and detail coefficients at different levels or scales [8]. For $x(t)$ signal, DWT decomposition equation is given by (1):

$$x(t) = \sum_{k=-\infty}^{k=+\infty} C_{n,k} \phi (2^{-n} t - k) + \sum_{j=1}^{n} \sum_{k=-\infty}^{k=+\infty} d_{j,k} 2^{-j/2} \psi (2^{-j} t - k) \tag{1}$$

Where $x(t)$ represents a time-domain signal, $C_{n,k}$ represents the approximation coefficients to the level $n$ and traslation factor $k$, while $d_{j,k}$ $(j = 1 \text{ to } n)$ represents the detail coefficients at level $j$, $\psi(t)$ is the wavelet function and $\phi(t)$ is a scale function. The mother wavelet selection, determines the signal representation. Daubechies, Coiflet and Symlet are generally used as mother wavelet families. The coefficients derived from wavelet decomposition are too long to be used as features for classification; in this paper, two features were used to characterize the wavelet coefficients in each sub-levels of decomposition.
2.2. Empirical mode decomposition (EMD)

Through EMD, used for the analysis of nonstationary and nonlinear signals, the different functions are obtained by intrinsic modes (IMF) that compose it. The general procedure used to decompose the EEG signals into IMF is proposed in [9], described in Table I, where \( i \) represents the total number of IMF produced.

<table>
<thead>
<tr>
<th>Decomposition of EEG signals into IMF</th>
</tr>
</thead>
<tbody>
<tr>
<td>Step 1: Read EEG signal, ( x_i )</td>
</tr>
<tr>
<td>Step 2: ( c_0 = x_i, k = 0 )</td>
</tr>
<tr>
<td>Step 3: Identification of maximum and minimum of ( x_k, k = k + 1 )</td>
</tr>
<tr>
<td>Step 4: Set upper envelopes, ( u_k ), and lower envelopes, ( l_k ).</td>
</tr>
<tr>
<td>Step 5: Calculate the mean of the envelopes ( m_k = [u_k + l_k]/2 )</td>
</tr>
<tr>
<td>Step 6: Calculate ( c_k = c_k - 1 - m_k )</td>
</tr>
<tr>
<td>Step 7: if ( c_k ) is an IMF: ( i = i + 1, d_i = c_k, x_i = x_i - 1 - d_i ) go to step 2, otherwise go to step 3</td>
</tr>
<tr>
<td>End If</td>
</tr>
<tr>
<td>Step 8: If more function cannot be extracted, Residue ( r_i = x_i )</td>
</tr>
<tr>
<td>End</td>
</tr>
</tbody>
</table>

The original data are the sum of the IMF \( c_j(t) \) components plus the final residue \( r_n(t) \), expressed as in (2).

\[
x(t) = \sum_{j=1}^{n} c_j(t) + r_n(t)
\]  

(2)

In EMD analysis, the first IMF has the highest frequency content, this frequency content decreases with increasing in the IMF component, until n-IMF component, which is almost a linear function of time. Compared with wavelet decomposition, the components of the IMF have lower frequency content, are useful for analyzing low-frequency oscillations, and can be used in the Hilbert spectral analysis to obtain instantaneous frequency [10]. Hilbert transform is mainly used to construct the analytical signal \( z(t) = x(t) + iy(t) \). The analytical signal can be separated into its component signal amplitude \( \alpha(t) \) and instantaneous phase \( \theta(t) \), whose real part is equal to the original signal, as shown in (3), where \( w_j \) is instantaneous frequency.

\[
x(t) = \text{Real} \sum_{j=1}^{n} \alpha(t) e^{i \int w_j(t) dt}
\]  

(3)

2.3. Fuzzy entropy

Information theory founded by Shannon, has been formally engaged with communication systems in its infancy, but has also been applied to other areas such as clustering, fuzzy sys-
tems and decision making. The theory of information relates to the quantification of the information. The quantity of information is defined as the amount of information conveyed in an event and depends on the probability of the event [11]. The definition of the amount of information means the logarithm of the probability of an event $x$, $P(x)$. The average amount of information on all events, based on the classical information theory, called entropy, and is given by (4).

In the probabilistic approach, Shannon entropy is a well-known measure of uncertainty and is extensively covered in the literature. An extension of the Shannon entropy is the fuzzy entropy concept, in which the fuzzy sets used to assist estimation of entropy. Fuzzy entropy differs from Shannon classical entropy from the point at which fuzzy entropy contains fuzzy uncertainties (possibilistic) while Shannon entropy contains randomness uncertainties (probabilistic).

Fuzzy entropy, as Shannon entropy satisfies the four axioms of De Luca-Termini [12]. Shannon entropy is defined from a discrete random variable $x$ with probability function $p(x_i)$, given by (4), where $i$ denotes the number of elements and $p(x_i)$ denotes the occurring probability of the element $x_i$.

$$H(X) = -\sum_i p(x_i) \log_2 p(x_i) \tag{4}$$

From Shannon entropy, in [13] Khushaba defines fuzzy entropy of the elements of the class $i$ along each specific feature function $f$ (fuzzy set), as in (5), where $P(f, c_i)$ can be interpreted as the degree to which the sample is pre-defined to belong to class $i$ really contributes to that specific class.

$$H(f, c_i) = -P(f, c_i) \log P(f, c_i) \tag{5}$$

Complete entropy diffuse along the C-class is given by (6):

$$H(f, C) = \sum_i H(f, c_i) \tag{6}$$

Fuzzy equivalent for the joint probability of the training patterns belonging to class $i$, is given by (7).

$$P(f, c_i) = \frac{\sum_{k \in A_i} \mu_{ik}}{NP} \tag{7}$$

Where $A_i$ is the set of indices of the training patterns belonging to class $i$, $NP$ is the total number of patterns, and $\mu_{ik}$ represents the $k$–th fuzzy membership value belonging to class $i$, given by (8), with fuzziness parameter $m$, for the $k$–th measurement vector $x_k$ and the $i$–th centroid vector $v_i$.

$$u_{ik} = \frac{1}{\sum_{j=1}^{c} \left( \frac{|x_k - v_i|}{|x_k - v_j|} \right)^{\frac{m-1}{m}}} \tag{8}$$

The fuzzy entropy measure was used to evaluate the significance of the different characteristics of the signal, so a high value represents fuzzy entropy function which does not contribute
greatly to the deviation between classes, and an entropy value under presents more informative features. The most informative features translate into groups with minimum overlap and maximum class separability [14].

3. Experimental Set-up

In this paper, we used data collected from online BCI experiments, and analyzed as described in this section.

3.1. Data Bases

Database Graz BCI Competition 2003 – III was provided by the Department of Medical Informatics, Institute of Biomedical Engineering, Graz University of Technology [15]. This dataset was recorded from a normal subject (female, 25y) during a feedback session. The subject sat in a relaxing chair with armrests. The task was to control a feedback bar by means of imagery left or right hand movements. The order of left and right cues was random. The positions of the electrodes and the distribution system are shown in Figure 1.

The experiment consists of 7 runs of 40 trials each. All of them were carried out the same day. In total there were 280 trials 9 s length. During the first two seconds, \( t = 2 \) s an acoustic stimulus indicates the beginning of the test, trigger channel (# 4) was located in a low to high

Figura 1. Schematic of time and position of electrodes in the experiment.
and a cross “+” is displayed for 1 second, then $t = 3\, s$, an arrow (left or right) is shown for reference. At the same time, the subject was asked to move one bar in the direction of the signal. The recording was performed using an amplifier g.tec and Ag/AgCl.

Three bipolar channels (anterior EEG “+” , later “-“) were measured at C3,Cz and C4. EEG samples were taken with 128 Hz, filtered between 0.5 and 30 Hz.

Database Graz BCI Competition 2003 – IV was provided by Fraunhofer -FIRST, Intelligent Data Analysis Group (Klaus- Robert Müller), and Department of Neurology at Freie Universität Berlin [16]. This data set was recorded from a normal subject during a session without feedback. The subject was seated in a normal chair, relaxed arms resting on the table, fingers in the standard position to write on the computer keyboard. The task was to press with the index and little fingers the corresponding keys on your order when your own pace. The experiment consisted of 3 sessions of 6 minutes each. All sessions were held on the same day with few breaks in between. Pressing was performed at an average rate of 1 key per second.

3.2. Experiments

Because the channel Cz is independent in motor imagination [6], only C3 and C4 channels of each database are used for feature extraction. First, the average power spectral density (PSD) in C3 and C4 channels was calculated, in order to provide an effective reference in wavelet decomposition stage.

3.2.1. Wavelet features

According to the EEG signals of motor imagination and power spectrum in C3 and C4 channels, were chosen five levels of decomposition and the mother wavelet Daubechies10. Table II shows the frequency bands obtained for each level of wavelet decomposition, with 128Hz and 100Hz sampling frequency for the respective datasets. The extracted wavelet coefficients provide the distribution of motor imagination signal in time and frequency.

<table>
<thead>
<tr>
<th>Decomposition level</th>
<th>Frequency BCI 2003- III</th>
<th>Frequency BCI 2003- IV</th>
</tr>
</thead>
<tbody>
<tr>
<td>D1</td>
<td>64 a 128 Hz</td>
<td>50-100 Hz</td>
</tr>
<tr>
<td>D2</td>
<td>32 a 64 Hz</td>
<td>25-50 Hz</td>
</tr>
<tr>
<td>D3</td>
<td>16 a 32 Hz</td>
<td>12.5-25 Hz</td>
</tr>
<tr>
<td>D4</td>
<td>8 a 16 Hz</td>
<td>6.25-12.5 Hz</td>
</tr>
<tr>
<td>D5</td>
<td>4 a 8 Hz</td>
<td>3.12-6.25 Hz</td>
</tr>
<tr>
<td>A5</td>
<td>0 a 4 Hz</td>
<td>0-3.12 Hz</td>
</tr>
</tbody>
</table>

Table II. Wavelet Decomposition Levels.
3.2.2. EMD features

First, we apply EMD to each EEG signal. Subsequently, we determined by PSD, which IMFs are on the same frequency band of the rhythms in order to group it.

By grouping the IMF signals at different levels that show similar spectral behavior, the Hilbert transform is applied to each level shown in Table III. In this case, after applying Hilbert transform given by (3) to the signals formed in each level, we obtain the real component of the amplitude of the analytic signal.

To characterize the brain rhythms and represent the distribution of the frequency and amount of changes in it, were calculated two statistical functions on the dataset obtained from the wavelet decomposition and datasets extracted through EMD for each frequency level. These features are Standard Deviation and Relative Energy. Each statistical function is applied to the respective sub-levels of decomposition (6 DWT coefficients and 4 EMD levels), for both C3 and C4 channels. Thus, 24 and 16 statistical features of DWT and EMD were respectively obtained.

<table>
<thead>
<tr>
<th>Level</th>
<th>IMFs Cluster</th>
<th>Frequency range</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>IMF1 + IMF2</td>
<td>&gt; 12 Hz</td>
</tr>
<tr>
<td>2</td>
<td>IMF3</td>
<td>8 Hz a 12 Hz</td>
</tr>
<tr>
<td>3</td>
<td>IMF4</td>
<td>4 Hz a 8 Hz</td>
</tr>
<tr>
<td>4</td>
<td>IMF5 + IMFn</td>
<td>&lt; 4 Hz</td>
</tr>
</tbody>
</table>

3.2.3. Results and discussion

We use the function fuzzy c-means of MATLAB®, to construct the fuzzy membership values for the features obtained of the two databases; each of these reflects the degree of membership of the sample for each of the two classes. Combining the membership values obtained $\mu_{ik}$ and equations (5), (6) and (7), were calculated the fuzzy entropy on specific features space. Table IV details the number of features, in order, with the lowest value of fuzzy entropy that identifies patterns with greater information content. Figure 2 shows the box plots of the average evaluation of fuzzy entropy algorithm applied to different numbers of selected features. In this way, to reduce the space dimensions, the ten features with less fuzzy entropy value were trained and classified by cross-validation (10 fold cross-validation) for $k−nn$ classifiers and support vector machines (SVM).
Table IV. Wavelet decomposition levels.

<table>
<thead>
<tr>
<th>Features selected</th>
<th>DWT-Fuzzy Entropy</th>
<th>EMD-Fuzzy Entropy</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dataset III</td>
<td>Dataset IV</td>
<td>Dataset III</td>
</tr>
<tr>
<td>1. Std C4 D3</td>
<td>Std C4 D3</td>
<td>Std C3 L4</td>
</tr>
<tr>
<td>2. Std C4 D4</td>
<td>Std C4 D5</td>
<td>Std C3 L3</td>
</tr>
<tr>
<td>3. EnR C4 D1</td>
<td>Std C4 D4</td>
<td>Std C3 L2</td>
</tr>
<tr>
<td>4. Std C4 D1</td>
<td>Std C3 A5</td>
<td>Std C3 L1</td>
</tr>
<tr>
<td>5. EnR C3 D2</td>
<td>Std C4 A5</td>
<td>EnR C3 L3</td>
</tr>
<tr>
<td>6. EnR C4 D2</td>
<td>Std C4 D2</td>
<td>EnR C3 L2</td>
</tr>
<tr>
<td>7. EnR C4 D4</td>
<td>Std C3 D4</td>
<td>EnR C4 L3</td>
</tr>
<tr>
<td>8. Std C3 D2</td>
<td>Std C4 D1</td>
<td>EnR C3 L1</td>
</tr>
<tr>
<td>9. Std C3 D3</td>
<td>Std C3 D3</td>
<td>EnR C4 L1</td>
</tr>
<tr>
<td>10. ERW C3 D1</td>
<td>Std C3 D2</td>
<td>Std C4 L1</td>
</tr>
</tbody>
</table>

Figura 2. Accuracy Fuzzy entropy applied to different number of features.

Classification results of the features extracted by the proposed method are shown in Table V for \( k - nn \) \( (k = 3) \) and SVM classifier with linear kernel function to map the training points to a higher-dimensional space.

The results of classification based on the characterization by wavelet transform and fuzzy entropy widely evident that the SVM classifier has better results than k-nn classifier in most occasions when validation is performed. However, the feature extraction based on the EMD and fuzzy entropy, both classifiers tend to show the same results in accuracy, which makes them generally have the yield potential for use in pattern recognition tasks of motor imagination.

Generally, the method of fuzzy entropy maintains or improves the accuracy of the feature of EMD and DWT where it was evaluated for the database Competition BCI 2003. The results,
using fuzzy entropy, show greater impact when evaluated using k-nn classifier, increasing by up to 3.36 % of accuracy using less features.

Figure 3 shows the computational cost of the training in EMD and DWT in both databases. It is worth to notice that DWT takes less than 70 % time to obtain the results with similar accuracy. Determining the computational cost is achieved by running each of the proposed algorithms using Matlab on a laptop with Intel second generation Core i5 processor and 4GB RAM. The time of highest execution was considered 100 %, and by rule of three, the remaining percentages were made.

Other feature methods applied to the databases used in this work show lower results, as in [17] with accuracy 90 % using wavelet coefficients and autoregressive parameters; also, Qiao [18] obtained accuracies of 90 % applying entropy coefficients Wavelet Packet (WEP). Fang et al. [19] used Approximate Entropy (ApEn) and linear classifier LDA with accuracy results of 91.4 %, and Zou et al. [6] obtained 3.94 % misclassification using DWT and SVM classifier.

Table V. Accuracy of the proposed method using SVM classifier and $K-NN$.

<table>
<thead>
<tr>
<th>Dataset</th>
<th>Classifier</th>
<th>DWT</th>
<th>DWT-Fuzzy Entropy</th>
<th>EMD</th>
<th>EMD-Fuzzy Entropy</th>
</tr>
</thead>
<tbody>
<tr>
<td>BCI III</td>
<td>SVM</td>
<td>97.94±1.29</td>
<td>99.09±0.82</td>
<td>99.48±0.76</td>
<td>99.27±0.89</td>
</tr>
<tr>
<td></td>
<td>K-NN</td>
<td>93.97±2.19</td>
<td>97.33±1.73</td>
<td>99.17±0.98</td>
<td>99.39±0.86</td>
</tr>
<tr>
<td>BCI IV</td>
<td>SVM</td>
<td>100±0.00</td>
<td>99.87±0.27</td>
<td>100±0.00</td>
<td>100±0.0</td>
</tr>
<tr>
<td></td>
<td>K-NN</td>
<td>97.72±0.87</td>
<td>98.70±1.00</td>
<td>96.25±1.17</td>
<td>99.01±0.63</td>
</tr>
</tbody>
</table>

Figura 3. Computational cost EMD vs. DWT.
4. Conclusions

In this paper we proposed and compared two methods for feature extraction of EEG signal for motor imagery of right and left hand, based on DWT and EMD. Also, the fuzzy entropy algorithm was used to select the most relevant features. The methodology described in this paper enable the automated recognition of motor imagery in BCI systems. Obtained results are very close in EMD and DWT decompositions; however the computational cost of DWT is less than 70% of EMD.

We have analyzed the signals electroencephalography contained in datasets Competition 2003 dataset III and dataset IV, according to the classification results obtained effectively check the signals extracted through electrodes positioned on the C3 and C4 channels contain information interest concerning rhythms engines left and right hand.

The classification of patterns extracted through the wavelet decomposition methods and EMD, experienced two different data sets, has yielded good results accuracy. Also, by applying the fuzzy entropy algorithm, the classification accuracy is maintained using a lower dimensionality space. The fact of using different databases and different methods of characterization of EEG signals in addition to optimal classification results, makes the fuzzy entropy algorithm, a technique that could be extrapolated to other datasets and characterization techniques for used in tasks of motor imagination through BCI systems.

5. Acknowledgment

This work is supported by project P13259 from Instituto Tecnológico Metropolitano de Medellín-Colombia, and also Master thesis project “Diseño de una metodología para la extracción de características de electroencefalografía en sistemas con interfaz cerebro-computador”.

References


