14th European Signal Processing Conference (EUSIPCO 2006), Florence, Italy, September 4-8, 2006, copyright by EURASIP

RANKING FEATURES OF WAVELET-DECOMPOSED EEG BASED ON SIGNIFICANCE IN EPILEPTIC SEIZURE PREDICTION

Pedram Ataee, Alireza Nasiri Avanaki, Hadi Fatemi Shariatpanahi, Seyed Mohammadreza Khoee

Control and Intelligent Processing Center of Excellence,
School of Electrical and Computer Engineering, University of Tehran
P.O. Box 14395-515, Tehran, Iran
phone: +98-912-348-2130, fax: +98-21-8877-8690, email: ataee@ieee.org

ABSTRACT

A method for ranking features of wavelet-decomposed EEG in order of importance in prediction of epileptic seizures is introduced. Using this method, the four most important features (extracted from each level of wavelet decomposition) are selected from ten features. The proposed set of features is then used to recognize "pre-seizure" signal, thus predicting a seizure. Our feature set outperforms previously used sets by achieving higher class separability index and correct classification rate.

1. INTRODUCTION

Epilepsy is known to be the second most frequent brain dysfunction, from which about 50 million people suffer world wide. Epileptic seizures cannot be controlled with medication in a third of this population [1]. The rest of this population has to put up with the side effects of the seizure-control drugs. A pre-seizure warning can help the patient to prepare for the seizure (e.g., pull over if driving). Knowing when a seizure would occur, the medication can be used just in time to reduce the unwanted side effects.

The patient's EEG signal is known to contain much information about the occurrence of epileptic seizures [2]. It is also shown (see Section 2) that the EEG signal conveys enough information to *predict* an epileptic seizure.

The information content of a signal can be summarized in a few features computed for that signal. Several features are suggested to represent information of an EEG signal with regards to (imminent) occurrence of an epileptic seizure [3, 4]. One can calculate a number of such features for each consequent segment of the EEG waveform and feed them to a classifier, which recognizes its input set of features to be either in the "pre-seizure" class (meaning, *most probably* a seizure happens and the patient should be warned), or in the "normal" or the "seizure" classes[1].

By increasing the number of features extracted from a signal, it is less likely to miss any information relevant in classification. With a large number of features, however, it is also more difficult to design and/or train a classifier to effectively recognize the desired pattern in the signal (a.k.a. "the curse of dimensionality"). Therefore, it is important to select only the significant features that carry the relevant information of the signal.

In this paper, we introduce a method of measuring the significance of each feature in correct recognition of the "pre-seizure" class (i.e., measuring utility of a feature in seizure warning). We also apply our method to determine the significance of ten features that are either used in previous work or are believed to contain EEG information useful to our purpose.

The rest of the paper is organized as follows. We first review the related previous work in Section 2. The way of extracting features from the EEG signal, the database on which the method is applied, as well as the method for selection of the most important features are described in Section 3. The selected significant features, and other set of features are used for classification of the EEG signals in Section 4. We wrap up the paper with a few concluding remarks and a number of directions for future work in Section 5.

2. PREVIOUS WORK

To the best of our knowledge, there is no method for ranking the individual features of (decomposed) EEG waveform based on their importance in seizure prediction. [5] gives a method of identifying the significant EEG channels for a certain purpose. Such a method must be used as the first step of customizing an EEG-based seizure warning system for every patient to determine the minimum number and the locations of the electrodes.

Our work addresses the problem of effective feature extraction from EEG to achieve a high ratio of true to false seizure warning. Several features are considered to represent EEG information: [6] suggests zero crossings and extremes information of the wavelet decomposed signal. [7, 4] use spectral content of the signal. Mean of absolute value [3], average power [3], standard deviation [3, 8], mean [8], maximum [8], minimum [8], various informational measures (such as Shannon's entropy) [9] of each wavelet decomposition level (sub-band), and the ratios of absolute mean values of the adjacent wavelet decomposition levels [3] are also used. The most promising of these features along with a few we suggested are compared in their utility in seizure prediction.

[1] It is a good idea that the system requests medical assistance in cases of "pre-seizure" or "seizure" via patient's cellular phone, for example.

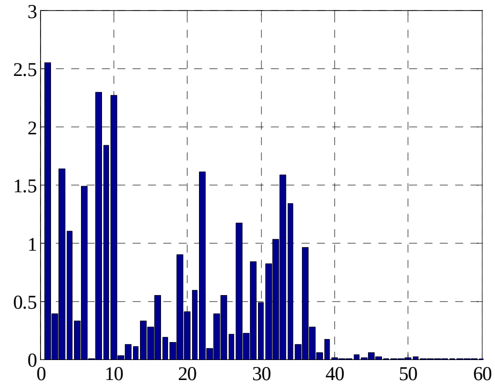

Figure 1 – Effect on separability for each of the ten tested features extracted for each wavelet decomposition level.

3. FEATURE SELECTION

3.1 Analysis specifications

Due to the non-stationary nature of EEG signal, it should be analyzed both in time and in frequency. Therefore, the wavelet transform, as a time-frequency analysis tool, is a suitable choice [3, 4, 10, 11, and 12]. The wavelet (basis) function should be similar to the usual EEG pattern, to capture as much information as possible. That is why Daubechies function family are proved to be a successful choice as the wavelet basis for our problem [2, 9]. We also show that the fourth-order Daubechies function is a good choice for our problem as well (see Section 4). Since the useful frequency content of the EEG signal is up to 30 Hz (i.e., higher frequency components are mainly noise or artifacts), the wavelet analysis is performed in 4 to 6 levels only [2, 13].

3.2 Database

We use the data provided in [14] and used in [8, 15], where a detailed description is also given. The data consists of five groups, each comprised of 100 single-channels, 23.6-second EEG. The first two groups are recorded from five healthy subjects: with open (**g1**) and closed eyes (**g2**). The third and fourth groups are recorded prior to a seizure from part of the brain with the syndrome (**g3**) and from the opposite (healthy) hemisphere of the brain (**g4**). The fifth group (**g5**) is recorded from part of the brain with the syndrome during the seizure.

3.3 Method

The following features were extracted from each level of wavelet decomposition for each waveform in the database.

1. Number of zero crossings.
2. Number of extremes.
3. Time of reaching the maximum point.
4. Maximum value.
5. Peak to peak value.
6. Mean value.
7. Energy.
8. Standard deviation.
9. Third-order moment.
10. Shannon's entropy (given by (1)).

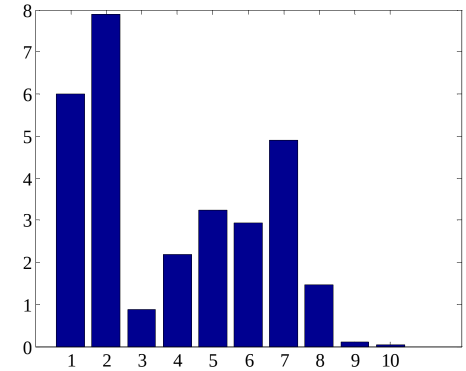

Figure 2 – Effect on separability of **g3** and **g5**, for each feature summed over all decomposition levels.

$$\begin{aligned} E_j &= \sum_k |C_j(k)|^2 \\ E_{tot} &= \sum_{j<0} E_j \\ p_j &= \frac{E_j}{E_{tot}} \\ S_s &= -\sum_{j<0} p_j \ln[p_j] \end{aligned} \tag{1}$$

These features are either used in [3, 6, 8, and 9] or we believed to contain useful information of EEG signal.
The feature set, X, is distributed in L classes with probabilities of p_i . S_w, S_B, and S denote in-class scatter, between-class scatter, and separability matrices, respectively. The traces of S_W and S_B express the average within-class scatter and the distance between two classes respectively [17]. For a given set of features, a measure of separability of the classes can be considered as the following.

$$J_1 = \frac{trace(S_B)}{trace(S_W)}$$

A separability index invariant to linear transformations of features is given by the following formula.

$$J_2 = trace(S_W^{-1} S_B)$$

$Disc_i$, the effect of the *i*-th feature in separability, is given by

$$S = S_w^{-1}.S_B = VDV^T$$

$$Disc_i = \sum_{j=1}^{n} \hat{e}_i^{\,T} V^{(j)} \lambda_j$$

$$\hat{e}_i = (P_1, P_2, ..., P_n)^T \qquad P_k = \begin{cases} 1 & k = i \\ 0 & k \neq i \end{cases}$$

For a specific set of features, one measure for utility for that set is given by the trace of its separability matrix [16]. Also, $Disc_i$ gives the effect of the *i*-th feature of the set.

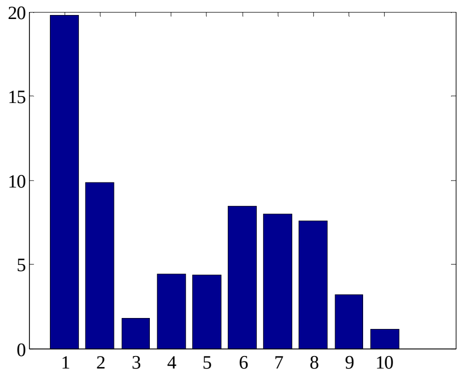

Figure 3 – Effect on separability of all five classes, for each feature summed over all decomposition levels.

3.4 Ranking features based on their significance

Figure 1 illustrates the amount of $Disc_i$ for each level of wavelet decomposition (a column of height $Disc_i$ for decomposition level k is depicted at horizontal location $i+10(k\text{-}1)$). In this experiment, all ten features listed above are used, and all L=5 classes are assumed equiprobable ($p_i = 0.2$). It is observed that the three most effective features in this scenario are computed from the first decomposition level (corresponding to the lower frequency components) and are the number of zero crossings, the standard deviation, and Shannon's entropy. Another interesting observation is that the number of zero crossings has little effect in separability when computed for the second decomposition level.

In Figure 2, features are compared based on their effect on separability of classes **g3** and **g5**, summed over all decomposition levels. One can see that the number of extremes, the number of zero crossings, and energy of the signal are the most useful features, considering overall.

In Figure 3, features are compared based on their effect on separability of all classes.

4. CLASSIFICATION RESULTS WITH THE MOST SIGNIFICANT FEATURES

In Table 1, our proposed set of features (the last row) is compared to those suggested by [3, 6, 8, and 9] (the first three rows) in their ability to separate and classify all five classes. Our proposed set of features, listed in the last row (the four features with highest $Disc_i$ in Figure 3), gives the best performance. The data is classified with a multi-layered perceptron (MLP), which is a simple neural network classifier with one hidden layer and is commonly used for small feature numbers [17]. MLP is trained by a back-propagation method, with 50% of the data (the rest is used for test). The second column lists the separability index ($Disc_i$). The third column gives the average correct classification rate. Our set of features achieves the highest separability index, and it is only 1% below the second set in correct classification rate with MLP. Note that a higher separability means that our proposed set can result in higher correct classification rate by using a more sophisticated classifier.

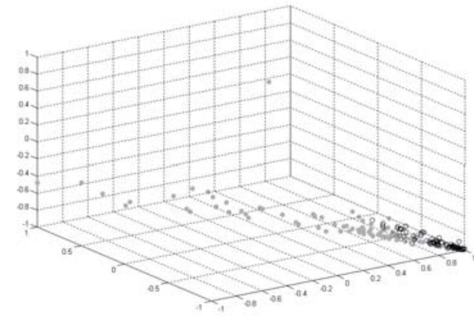

Figure 4 – Successful separation of g3 and g5 by the three most significant features.

In Table 2, utility of five feature sets are compared in classification of two equiprobable data groups **g3** and **g5**. In addition to the correct classification rates, the confusion matrices are also listed. For example, with the feature set given in the third row, the 50 examples of **g3** are recognized correctly in 43 cases, and are mistakenly classified as **g5** in 7 cases. The average correct classification rate is 89%. The highest separability is again achieved by our proposed feature set (those with highest $Disc_i$ in Figure 2), listed in the last row. Note that this set of features (optimal for 2-class operation) differs in one feature with the feature set optimal for 5-class case.

In Table 3, three basis functions for wavelet decomposition are compared in a setup similar to the previous experiment. The fourth order Daubechies function provides the highest separability and correct classification rate, and therefore is used for analysis. This observation can be justified by the fact that Daubechies functions of lower orders lack the details required to represent the relevant information in EEG signal.

5. CONCLUSION AND FUTURE WORK

In this work, we introduced a number of features for EEG signal to recognize the signal as a sample of "normal", "pre-seizure", or "seizure" classes. Our main contributions are measuring the effect (significance) of each feature in classification and using the four most significant ones. We showed a high separability can be achieved among the classes using the proposed set. Thus, even with a classifier as simple as MLP, we reached high correct classification rates that are comparable to previous work that used sophisticated classifiers on other feature sets.

We plan to investigate utility of other features (such as higher order statistics) for classification of EEG signal. We will also try using more sophisticated classifiers over a larger set of features.

Other bodily signals that do not seem directly related to epilepsy (temperature of certain points on body, and ECG for example) may have significant information about occurrence of seizures. Using the method of significance measurement presented in this work, we plan to check the utility of such signals in seizure prediction as well.

Feature set	**Separability**	**Correct Classification Rate**
min, max, mean, variance	3.63	54.0 %
Zero crossing, mean	6.10	73.0 %
entropy, energy	2.75	63.6 %
zero crossing, number of extremes, mean, energy	7.69	72.0 %

Table 1– Comparison of various features set in their ability to discriminate between all five classes. The last row indicates our selected most significant features.

Feature set	Separability	Correct Classification rate	Confusion Matrix
min, max, mean, variance	2.17	94.0 %	45 1 5 49
zero crossing, number of extremes, mean, energy	3.54	94.0 %	47 3 3 47
zero crossing, entropy, variance,3rd order moment	2.98	89.0 %	43 7 4 46
zero crossing, peak to peak, entropy, energy	3.37	90.0 %	44 4 6 46
zero crossing, number of extremes, peak to peak, energy	3.92	94.0 %	47 3 3 47

Table 2 – Comparison of various feature sets in their ability to discriminate between **g3** and **g5**. The last row indicates our selected most significant features and the other rows were proposed in previous works.

Basis Function	**Separability**	**Correct Classification rate**
db4	3.92	96.0 %
db2	3.27	95.3 %
db1 (Haar)	3.33	93.5 %

Table 3 – Finding the best order of Daubechies function to be used as our wavelet basis function.

REFERENCES

[1] L. D. Iasemidis, “Epileptic Seizure Prediction and Control,” *IEEE Transaction on Biomedical Engineering*, vol. 50, No. 5, pp. 549-558 , May 2003.

[2] H. Adeli, Z. Zhou, and N. Dadmehr, “Analysis of EEG records in an epileptic patient using wavelet transform,” *Journal of Neuroscience Methods*, vol. 123, pp. 69-87, 2003.

[3] A. Subasi, “Automatic recognition of alertness level from EEG by using neural network and wavelet coefficients,” Expert *Systems with Applications*, vol. 28, pp. 701-711, 2005.

[4] J. Muthuswamy and N. V. Thakor, “Spectral analysis methods for neurological signals,” *Journal of Neuroscience Methods*, vol. 83, pp. 1-14, August 1998.

[5] T. N. Lal, *et al*, “Support vector machines in BCI,” *IEEE Trans on Biomedical Engineering, special issue on brain computer interface*, Vol. 51, No. 6, pp. 1077-1080, 2 004.

[6] P. Zarjam and M. Mesbah, "Discrete Wavelet Transform Based Seizure Detection In Newborns EEG Signals," in *Proc. of the Seventh International Symposium on Signal Processing and Its Applications*, pp.459-462, July 2003.

[7] B. Clemens, G. Szigeti, and Z. Barta, “EEG frequency profiles of idiopathic generalised epilepsy syndromes,” *Epilepsy Research*, vol. 42, pp. 105-115, 2000.

[8] I. Guler and E. D. Ubeyli, “Adaptive neuro-fuzzy inference system for classification of EEG signals using wavelet coefficients,” *Journal of Neuroscience Methods*, vol. 148, pp. 113-121, 2005.

[9] O. A. Rosso, M.T. Martin, and A. Plastino, “Brain electrical activity analysis using wavelet-based informational tools,” *Physical A*, vol. 313, pp. 587-608, 2002.

[10] C. Yamaguchi, "Fourier and Wavelet Analysis of Normal and Epileptic Electroencephalogram (EEG)," in *Proc. the 1st International IEEE EMBS Conference on Neural Engineering*, Capri Island, Italy, March 20-22, 2003, pp. 406-409.

[11] R. Saab, M. J. McKeown, L. J. Myers, and R. Abu-Gharbieh, "A Wavelet Based Approach for the Detection of Coupling in EEG Signals," in *Proc. of the 2nd International IEEE EMBS Conference on Neural Engineering*, Arlington, Virginia, March 16-19, 2005, pp. 616-620.

[12] M. K. Kiymik, I. Guler, A. Dizibuyuk, and M. Akin, "Comparison of STFT and wavelet transform methods in determining epileptic seizure activity in EEG signals for real-time application," *Computers in Biology and Medicine*, vol. 35, pp. 603-616, 2005.

[13] L. Szilagyi, Z. Benyo, and S. M. Szilagyi, “A New Method for Epileptic Waveform Recognition Using Wavelet Decomposition and Artificial Neural Network,” in *Proc. of the Second Joint EMBS/BMES Conference*, Houston, TX, USA, October 23-26, 2002, pp. 23-26.

[14] www.meb.uni-bonn.de/epileptologie/science/physik/eegdata.html

[15] R. G. Andrzejak, K. Lehnertz, F. Mormann, C. Rieke, P. David, and C. E. Elger, "Indications of nonlinear deterministic and finite-dimensional structures in time series of brain electrical activity: Dependence on recording region and brain state," *PHYSICAL REVIEW E*, vol. 64, 061907, 2001.

[16] R. O. Duda, P. E. Hart, and D. G. Stork, *Pattern Classification.* Address: 2nd Ed, John Wiley & Sons, 2000.

[17] C. M. Bishop, *Neural Networks for Pattern Recognition.* Address: 2nd Ed, Clarendon Press Oxford, 1995.