

“© 2007 IEEE. Personal use of this material is permitted. Permission from IEEE must be obtained for all other uses, in any current or future media, including reprinting/republishing this material for advertising or promotional purposes, creating new collective works, for resale or redistribution to servers or lists, or reuse of any copyrighted component of this work in other works.”

Brain-Computer Interface Analysis using Continuous Wavelet Transform and Adaptive Neuro-Fuzzy Classifier

Sam Darvishi and Ahmed Al-Ani

Abstract—The purpose of this paper is to analyze the electroencephalogram (EEG) signals of imaginary left and right hand movements, an application of Brain-Computer Interface (BCI). We propose here to use an Adaptive Neuron-Fuzzy Inference System (ANFIS) as the classification algorithm. ANFIS has an advantage over many classification algorithms in that it provides a set of parameters and linguistic rules that can be useful in interpreting the relationship between extracted features. The continuous wavelet transform will be used to extract highly representative features from selected scales. The performance of ANFIS will be compared with the well-known support vector machine classifier.

I. INTRODUCTION

EEG signals are the most popular way of interpreting the brain activities in the realm of non-invasive BCI [1]. Among the different types of EEG signals used for BCI communication, (P300, visual evoked potentials, slow cortical Potential, and Mu rhythms) – the last one, which is also known as motor imagery signal, issued from the central motor cortex, is the most suited for paralyzed patients and asynchronous BCIs [2]. There are a number of challenges that face the implementation of BCI systems. Two of the most important are the classification accuracy and speed of data transfer. This study focuses on the first task through the use of an Adaptive Neuron-Fuzzy Inference System (ANFIS) classifier. Data set III of the BCI competition 2003, which was provided by Graz University of Technology, Austria, is used here.

The four basic steps in a traditional classification system are: pre-processing, feature extraction, classification and post processing [3]. We will concentrate upon the 2nd and 3rd steps, as the EEG data has already been filtered between 0.5 and 30 Hz. Actually, a previous study has shown that the most prominent changes for motor imagery data take place in the Alpha [8-13 Hz] and Beta [18-25 Hz] frequency bands [4]. The feature extraction method that we adopted is based on the Continuous Wavelet Transform (CWT), where the EEG data is transformed using CWT with correspondent scales of alpha and beta ranges. The student's t-test is exploited to extract the features, which are then classified using an ANFIS classifier. Finally, the obtained results are compared to those obtained using a Support Vector Machine (SVM) classifier.

The paper is organized as follows: the next section

describes the EEG data representation. Section III describes classification of EEG trials using both ANFIS and SVM. The results are presented in section IV, and a conclusion is given in section V.

II. EEG DATA REPRESENTATION

The EEG dataset was recorded using three channels (C3, Cz and C4) with 140 training trials and 140 testing trials. The recording length of each trial was set to 9 seconds, with the first three seconds for preparation purpose, while the last 6 seconds represent the data during and after a stimulus is shown on a screen placed in front of the subject. We decided not to use the first 3 seconds, as they barely contain any useful information. Moreover, we used channels C3 and C4 only, where a previous study has shown that they contain most of the useful information for this particular application [5].

A number of feature extraction methods have been proposed in the literature to represent EEG signals, such as wavelet transform [6], power spectra [7] and adaptive autoregressive (AAR) [8]. We have decided to use wavelet transform, as it has been found to provide a good way to visualize and decompose EEG signals into measurable component events [9]. Both the Discrete Wavelet Transform (DWT) and the Continuous Wavelet Transform (CWT) have been used in EEG analysis. DWT is more computationally efficient than CWT, where CWT seems very redundant. However, if processed properly, CWT can clarify subtle information that DWT cannot extract [10]. The Morlet mother wavelet is used here, as it proved to be useful in analyzing EEG signals [9].

The relationship between CWT scales and frequency is expressed as follows: low scale corresponds to high frequency and vice versa. The Wavelet Toolbox of MATLAB® uses the following formula to map between a scale and a pseudo-frequency:

$$F_a = \frac{F_c}{a \cdot \Delta} \quad (1)$$

where a is a CWT scale, Δ is the sampling period (1/128 s for the used dataset), F_c is the center frequency of the wavelet function (0.8125 Hz for Morlet) and F_a is the pseudo-frequency corresponding to scale a . Hence, $F_a = 104/a$. Accordingly, the corresponding scale ranges for the Alpha and Beta bands become [8-13] and [4.16-5.78] respectively.

We have adopted the feature extraction method described

in [10], and introduced some changes to suit this particular application. The training dataset that consists of 140 trials was split according to the output class labels, i.e., left and right hand movements, each with 70 trials of the two EEG channels. The Morlet mother wavelet was then convolved with each trial and produced a 4-dimensional matrix for each channel for both the Alpha and Beta bands. The matrix is basically the wavelet coefficients that present the level of correspondence between the Morlet mother wavelet and the EEG signal.

To increase computation efficiency as well as extracting the most discriminant features from EEG signals, we calculated the averaged energy of wavelet coefficients through a number of timing windows. After extensive experiments, we found that two timing windows, each with 3 s duration, represent the optimal choice for the ANFIS classifier. Eight timing windows were found to give better results for the SVM classifier, as will be shown in section IV. Hence, for each trial of a given channel, there are two (or eight for SVM) averaged energy coefficients for a particular scale.

Afterwards, the student's t-test was used as an optimum tool for extracting the maximal points of discrepancy between averaged wavelet energy coefficients of the two classes, i.e., to find the scales and timing windows that achieve the maximum discrimination between the left and right hand movements. To accomplish this task, the mean and standard deviation of coefficients in each class and the pool variance is used [10].

Finally, the optimum number of features was set to two per frequency band per channel for the ANFIS classifier (eight for the SVM classifier). Hence, for a given trial, the number of extracted features for the ANFIS and SVM classifiers was 8 and 48 respectively. A detailed description of the feature extraction process is shown in Fig. 1.

III. EEG CLASSIFICATION

There have been a number of attempts to implement BCI using various types of classifiers. Nevertheless, to the best of our knowledge, there have been no previous attempts to implement BCI using an ANFIS classifier. Thus, in this study we managed to classify the extracted features using ANFIS and compare its results with the well accomplished SVM classifier.

A. The Adaptive Neuro-Fuzzy Inference System Classifier

This classifier combines the properties of fuzzy logic and neural networks to produce a set of parameters and linguistic rules, which are expected to be beneficial in interpreting the relationship between the extracted features.

ANFIS is normally a multi-layer feed-forward neural network with its weight and bias parameters estimated by fuzzy rules and fuzzy membership functions [11].

The key Challenge for classifying data with ANFIS is its

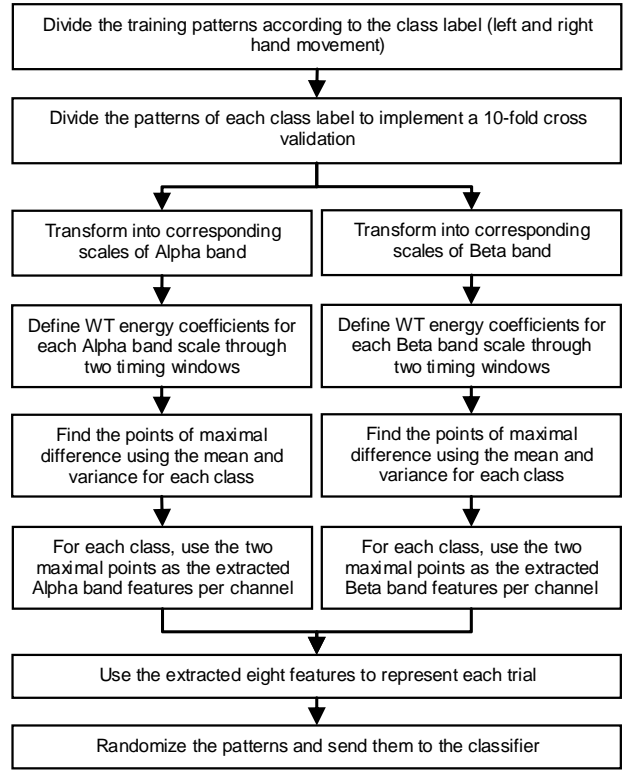


Fig. 1. Flow chart of the feature extraction process

ability to generalize when presented with small number of training data (like the 140 trials used here). A typically generated fuzzy inference system makes many fuzzy rules that in turn, lead to a large number of ANFIS parameters that need to be adjusted. These parameters will not be properly adjusted if using limited number of training data. For example, in our case as we have 8 features for every trial, if we define three fuzzy membership functions for each input feature, then the total number of possible rules will be 6561, which can not be trained using our small number of training patterns. To overcome this problem we used subtractive clustering to generate the Fuzzy Inference System (FIS), which can generate limited number of rules. The parameters of the obtained FIS would then be trained using neural networks.

The fuzzy subtractive clustering aims to identify useful patterns in data by finding the optimal data points to define cluster centers. The obtained clusters are used to extract a set of fuzzy rules. The generated fuzzy inference system would be trained by identifying the membership function parameters of the inputs and output.

B. Support vector Machine (SVM):

SVM proved to be a powerful classification algorithm, which can be useful when dealing with large number of features and limited number of training patterns. The main objective of SVM is to find hyper planes that maximize the separation between classes. Both linear and non-linear SVMs have been developed. The non-linear boundary decision usually gives better results and is implemented by

transforming the inputs space to another space, called the feature space. The main reason behind this transformation is that linear operation in the feature space is equivalent to non-linear operation in the input space. [12]. However, because of our limited data, a linear SVM is used here, as experiments have shown that a non-linear SVM would have a poor generalization on unseen data.

IV. EXPERIMENTAL RESULTS:

A. Classification using ANFIS

As explained above, 8 features have been used for the ANFIS classifier. The first 4 feature are related to the Alpha band and the rest to Beta. These features represent the average value of wavelet energy coefficients through two timing windows. The clusters that are formed are found to be roughly split between the two classes. For instance, Table 1 shows the centre of each cluster and the corresponding class when using five clusters obtained after normalizing the input features between 0 and 1.

TABLE 1.
CENTERS OF CLUSTERS WITH RESPECT TO FEATURES AND CLASSES

Cluster	F1	F2	F3	F4	F5	F6	F7	F8	Class
1	0.15	0.11	0.31	0.05	0.70	0.22	0.17	0.15	1
2	0.57	0.23	0.16	0.11	0.30	0.04	0.69	0.23	2
3	0.15	0.15	0.58	0.23	0.09	0.08	0.44	0.01	2
4	0.63	0.17	0.08	0.17	0.45	0.22	0.14	0.10	1
5	0.31	0.15	0.34	0.15	0.10	0.19	0.40	0.17	2

The identified clusters will lead to the formation of fuzzy rules (five rules for the above example). The `Genfis2.m` function in MATLAB® has been used to generate the rules. This function extracts rules by first using the outcome of the subtractive clustering to determine the number of rules and antecedent membership functions and then uses linear least squares estimation to determine each rule's consequent equations. Giving the 8 input features, each rule will consist of 8 antecedents and one consequent. The radius of each cluster specifies the range of influence of the cluster center. Specifying a smaller cluster radius will usually yield more, smaller clusters in the data, and hence more rules.

The interpretation of rules can provide useful information about the interaction between features. For example, the following rule describes the range and relationship between the Alpha band features for both channel 1 (F1 and F2) and channel 2 (F3 and F4), and the Beta band features (F5-F8) in selecting class 1.

If (F1 is C1F1) and (F2 is C1F2) and (F3 is C1F3) and (F4 is C1F4) and (F5 is C1F5) and (F6 is C1F6) and (F7 is C1F7) and (F8 is C1F8) then (out is class1)

In order to choose the best values of cluster radius and number of rules, we carried out two experiments. In the first experiment, the radius value was varied between [0.18, 1.2]. The classification accuracy of a 10-fold cross validation for both the training and test sets are shown in Fig. 2. As mentioned above, the smaller the radius, the larger the

number of rules. However, a given radius value does not guarantee the same number of rules for all 10 folds of the cross-validation. For instance, for a radius of 0.8, the generated number of rules varied in the 10 folds between 3 and 5. The figure indicates that setting the radius to a small value gave very high classification accuracy for the training set, but a low accuracy for the test set. This is expected, as this would lead to the generation of high number of rules and hence many parameters to adjust, which can not be done appropriately using a small number of training patterns, and thus, will lead to poor generalization. On the other hand, better generalization can be achieved when using a larger radius, with a best classification accuracy of 80.71%.

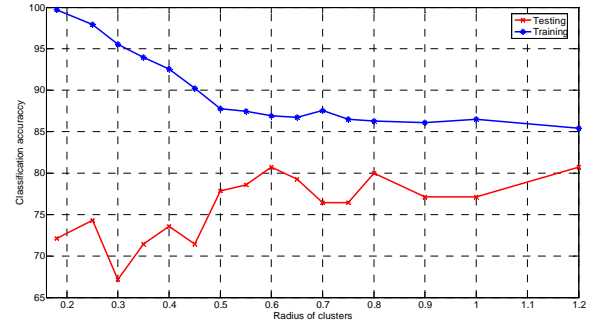


Fig. 2. Radius of clusters versus classification accuracy

The second experiment dealt with evaluating the effect of number of rules on classification performance. The number of rules was varied between [2, 100]. The same number of rules was applied to all 10 folds, which sometimes required using different cluster radius. The classification results are shown in Fig. 3. Similar to the previous experiment, the performance of the test set got worse with the increase of number of rules. The best result was achieved when using three rules with a classification accuracy of 82.14%, which is slightly better than fixing the cluster radius.

B. Classification using SVM

In order to classify EEG trials with SVM, we first used the same features of the ANFIS classifier, i.e., CWT energy coefficient calculated using two timing windows. We then used eight timing windows for the sake of comparison. The classification accuracy of a 10-fold cross-validation is shown Table 2.

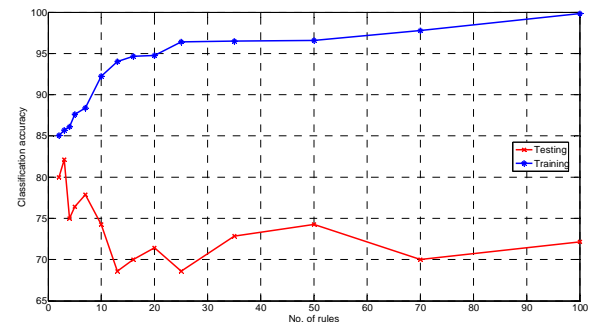


Fig. 3. No. of rules versus classification accuracy

TABLE 2
CLASSIFICATION ACCURACY OF SVM

	Class. Acc. Training	Class. Acc. Testing
2 timing windows	81.90	77.14
8 timing windows	86.19	81.43

The above table indicates that using 8 timing windows, which make 48 features per trial, provided better results than using 2 timing windows. The obtained classification accuracy is close to that of ANFIS, with ANFIS being slight better.

V. CONCLUSION

In this paper, we presented the incorporation of Continuous Wavelet Transform (CWT) based feature extraction and Adaptive Neuron Fuzzy Inference System (ANFIS) classifier. An ANFIS classifier can provide useful information about the interaction between input features and their relationship with the class labels. The ANFIS classifier that is implemented using fuzzy subtractive clustering and trained to adjust the membership parameters of the inputs and output has been carefully examined. Various values of cluster radius and number of rules have been tested to maximize the classification accuracy. The obtained results have found to be slightly better than those obtained using a linear support vector machine.

ACKNOWLEDGEMENT

The Authors would like to thank the institute of Human-Computer Interfaces, University of Technology, Graz, Austria, for providing the EEG data.

REFERENCES

- [1] T. Ebrahimi, J.M. Vesin and G. Garcia. Brain-computer interface in multimedia communication. *IEEE Signal Processing Magazine*, vol. 20, pp 14-24, 2003.
- [2] G.S. Dharwarkar. "Using Temporal Evidence and Fusion of Time-Frequency Features for Brain-Computer Interfacing". Master Thesis, University of Waterloo, 2005.
- [3] R.O. Duda, P.E. Hart and D.G. Stork. *Pattern Classification*. Wiley, 2nd ed. 2001.
- [4] L. Shoker, S. Sanei and A. Sumich. Distinguishing between left and right finger movement from EEG using SVM. *IEEE Engineering in Medicine and Biology Society conference (EMBC 2005)*, 5420-5423, 2005.
- [5] S. Lemm, C. Schafer and G. Curio. BCI competition 2003–dataset III: probabilistic modeling of sensorimotor Mu rhythms for Classification of imaginary hand movements. *IEEE Trans. on Biomedical Engineering*, vol. 51, pp 1077-1080, 2004.
- [6] M. Kawada. Analysis on synchronous time frequency components of human movement related cortical potentials using wavelet transform. *IEEE Engineering in Medicine and Biology Society conference (EMBC 2004)*, 333-336, 2004.
- [7] L.J. Trejo, R. Rosipal and B. Matthews. Brain-computer interface for 1-D and 2-D cursor control designs using volitional control of the EEG spectrum or steady-state visual evoked potentials. *IEEE Trans. on Neural Systems and Rehabilitation Engineering*. Vol. 14, pp. 225-229, 2006.
- [8] C. Vidaurre, A. Schlogl, R. Cabeza, R. Scherer, G. Pfurtscheller. A fully on-line adaptive BCI. *IEEE Trans. on Biomedical Engineering*, vol. 53, pp. 1214-1219, 2006.
- [9] V.J. Samar, A. Bopardikar, R. Rao and K. Swartz. Wavelet analysis of neuroelectric waveforms: a conceptual tutorial. *Brain Language*. vol. 66, pp. 7-60, 1999.
- [10] V. Bostanov. BCI competition 2003–data sets Ib and IIb: feature extraction from event-related brain potentials with the continuous wavelet transform and the t-value scalogram. *IEEE Trans. on Biomedical Engineering*, vol. 51, pp 1057-1061, 2004.
- [11] J-S. R. Jang, C-T. Sun and E. Mizutani. *Neuro-fuzzy and soft computing: a computational approach to learning and machine intelligence*. Prentice Hall, 1997.
- [12] C.W. Hsu, C-C. Chang and C-J. Lin. "A practical guide to Support vector classification". Technical report, National Taiwan University, 2003.