USING A MULTIPLE CLASSIFIER SYSTEM FOR IMPROVING THE PERFORMANCE OF ASYNCHRONOUS BRAIN INTERFACE SYSTEMS

Mehrdad Fatourechi [1], Gary E. Birch [1,2], and Rabab K. Ward [1,3]

[1]Department of Electrical and Computer Engineering, University of British Columbia, Vancouver, BC, Canada V6T 1Z4
[2] Neil Squire Foundation, Burnaby, BC, Canada V5M 3Z3
[3] Institute for Computing, Information and Cognitive Systems, Vancouver, BC, Canada
Email: { mehrdadf, rababw} @ece.ubc.ca, garyb@neilsquire.ca

ABSTRACT

To improve the performance of asynchronous brain interface (ABI) systems, a new classifier design is proposed. The spatial information of multiple EEG channels data is first used to create independent classifiers for different channels. A subset of these classifiers is then selected by a genetic algorithm to form a multiple classifier system (MCS) to decide whether a trial is an intended control or a no control signal. The analysis of the data from 4 subjects shows the effectiveness of the proposed method in improving the performance of an ABI system compared to the results obtained using only the best performing channel.

1. INTRODUCTION

A brain interface (BI) provides an alternative communication channel between a user's brain and a device. A successful BI design enables its users to control their environment (e.g., light switch or a wheelchair), a neural prosthesis or a computer by thinking of it only. This is done by measuring specific features of a person's brain signal that relate to his/her intent to affect control. These features are then translated into signals that are used to control/actuate devices. For a review of the field, see [1].

An Asynchronous Brain Interface (ABI) is a BI system that can always accept a control command from the user once the system is on. Such a system should not only be able to detect a specific pattern related to control , but it should also be robust enough so as not to be activated when the user does not intend to control the device [2] . Hence, it is very important to keep the FP rate as low as possible in order to prevent the user frustration. Unfortunately, the FP rates of current ABI systems are too high for such purposes.

In this paper, a new scheme that decreases the FP rate and increases the TP rate of ABI systems is presented. In this scheme, every EEG channel is regarded as a separate source of information. Then a separate ABI transducer or an "expert", consisting of a feature extractor followed by a feature classifier, is designed for each channel. If the experts are all accurate (they have performances better than chance) and independent (they make errors on different patterns) , it is then expected from the theory of multiple classifier systems (MCS) that improvements in the performance of the system can be achieved by combining the outputs of these expert systems. In [3], it is shown that by using a MCS (consisting of neural networks where each network processes a separate EEG channel), improvements in the performance of a synchronous BI system is achieved.

Since not every channel necessarily has a positive contribution to the performance of the system, a channel selection method should be incorporated in order to select the best channel combination. In this paper, this is carried out by using a genetic algorithm. The analysis of the data of 4 subjects shows that the performance of the new scheme is significantly better compared to the performance of the best performing single channel classifier.

The organization of this paper is as follows: In Section 2, ABI systems are discussed briefly. In Section 3, the proposed scheme for improving the performance of an ABI system is described. In Section 4, the results of analysis for 4 able-bodied subjects are presented. Section 5 is dedicated to discussion and conclusions.

2. ASYNCHRONOUS BRAIN INTERFACE SYSTEMS

To date, most BI researchers have focused their attention on synchronous control applications. In these applications, a user can initiate a command only during certain periods specified by the system. It is assumed that a user intends some control action during these specified times. In contrast, many real-life applications, such as the control of a wheelchair, require user- initiated control at any time. These types of applications (called asynchronous control applications) are characterized by allowing the system to be always in the "*Ready*" state, once it is switched on. These asynchronous systems are characterized by two states: Intentional Control (IC) and No Control (NC). During a NC period, a user of an ABI system may perform tasks other than control. It is not necessary that the user remains in a complete idle state (i.e., a user may be idle, thinking about a problem, or performing some action other than control). During an IC period, a user intends to issue a specific control command. Thus users consciously control their state only when they desire to control a device [4].

The performance of an ABI system is usually evaluated through two metrics: a true positive (TP) rate and a false positive (FP) rate. A FP rate is the percentage of misclassifying a NC trial as an IC trial, and a TP rate is the percentage of correctly classifying an IC trial. The FP rates of current ABI systems are still very high for practical applications. The main reason is the very noisy nature of the input signals of a BI system (especially the ones using EEG signals). This makes the correct detection of patterns difficult. Nevertheless, it is crucial to keep the FP rate as low as possible in order to prevent user frustration [2]. In the next section, a scheme is proposed which uses the spatial information of multiple EEG channels to improve the performance of an ABI system.

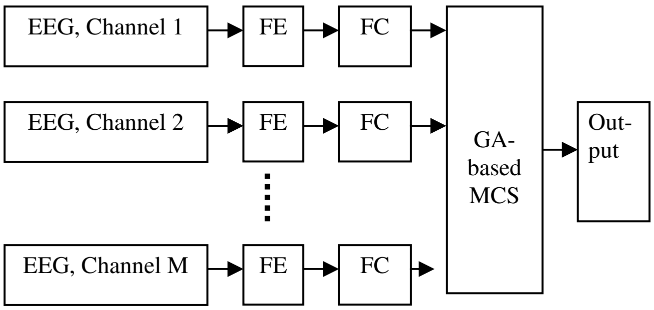

Fig.1. the overall structure of the proposed scheme

3. THE METHOD

The current main approach in the design of many ABI systems is to extract a number of features from different channels, form a feature vector (FV) and then classify this FV as an IC or NC state. The main challenge in such approach is that of finding a tradeoff between exploring more features and handling the complexity of the algorithm due to the size of the FV. Hence, prior to classification, feature selection may be applied in order to reduce the size of the FV. There are two main approaches for feature selection: filtering and wrapper. Filtering methods use a criterion other than classification accuracy to filter the irrelevant coefficients prior to classification and then the remaining features are used for classification. The main issue with such approach is that it is not guaranteed that once classified, the selected features work well together. The second approach (wrapper) selects a subset of features which yield the best classification accuracy. The main issue with this approach is the computation complexity.

There are, however, other designs for an ABI transducer that are based on a single channel data. In ([5] and [6]), ABI systems based on the data of a single ECOG channel are developed. Although the results based on single channels are promising, in some cases the FP rates remain too high for a practical system. In this study, we explore the effects of considering the spatial information in improving the performance of such systems.

In [3], the authors used a multiple classifier system (MCS) for channel combination in order to classify three types of movements in a synchronous BI system. The successful application of a MCS in [3] further motivated us to develop a similar scheme for ABI systems, where it is crucial to decrease the FP rates as much as possible. The theory of MCS provides an excellent means to achieve this goal.

The main idea behind the proposed scheme is to effectively use the spatial information of the EEG channels to improve the performance of the system. Fig.1 shows the overall structure of the proposed scheme. A separate transducer or an *expert* system (a feature extractor, FE, followed by a feature classifier, FC) is designed for each channel. In this paper, matched filtering is used for feature extraction. The input is cross-correlated with a template and a simple thresholding scheme is used for feature classification. A template represents MRP pattern related to the intended right index finger movement. Since the EEG signals vary from one channel to another, it is expected that a specific movement pattern results in different templates in different EEG channels. This is crucial in a MCS: having independent classifiers. If the designed expert systems have performances better than chance, it is then expected that by combining them, improvements in the performance of the system are achieved. In this section, the proposed scheme will be discussed in details.

3.1 Data Recording

The off-line data used in this paper, were collected from 4 able-bodied subjects (all male and right-handed). Subjects pressed a finger switch by their right index fingers after a cue appeared on the monitor screen. The EEG signals were recorded from 13 monopolar channels at F1, Fz, F2, FC3, FC1, FCz, FC2, FC4, C3, C1, Cz , C2 and C4 locations. Electro-oculographic (EOG) activity was measured as the potential difference between two electrodes, placed below and at the corner of the right eye. All signals were sampled at 128 Hz and referenced to the ear electrodes. Then the bipolar EEG data were generated by calculating the difference between the adjacent monopolar channels resulting in 18 bipolar EEG channels. The details of experiments carried out for recording the data were reported in [7].

3.2. Datasets

All the EEG signals were first low pass filtered to frequencies below 4 Hz using a linear phase FIR filter. This frequency band was selected because in our previous studies, it was found to be suitable for the study of Movement Related Potentials (the neuromechanism that is used for driving our BI system).

The IC trials data were collected during the 2.5 seconds period spanning 1.5 seconds before to 1 second after the switch activation. After rejecting trials contaminated with EOG artifacts (see [7] for details), an average of 300 IC trials were collected for each subject.

For the NC trials, the data was collected by sliding a window of width 2.5 seconds over the EEG signals recorded during the NC sessions. The sliding (shifting) step was 1/8 seconds. The 1/8 seconds shift was determined based on the maximum frequency of the neuromechanism of interest (4 Hz). After rejecting artifact-contaminated trials, approximately 5000 NC trial data were collected for each subject.

Both IC and NC datasets were then randomized and divided in training, validation and test sets. The training set is used to generate the templates. The validation set is used to identify the best configuration of the MCS. The configuration which yields the least error on the validation set, is selected. The performance of the system on the test set is then calculated. In order to reduce the bias in the results, we used a 5 fold nested cross validation for the analysis. For each outer validation set, 20% of the data are used for testing and the rest are used for training and model selection. In order to select models, the datasets in each inner cross-validation are further divided into 5-fold themselves. For each fold, 80% of the data are used for generating a template the rest are used for model selection.

3.3 Creating a template

For each subject, the IC data in each training set were averaged and low-pass filtered to frequencies below 4 Hz, using the same filter described in 3.2, to create the templates. This second stage of filtering was done in order to remove any potential high-frequency component, resulting from averaging.

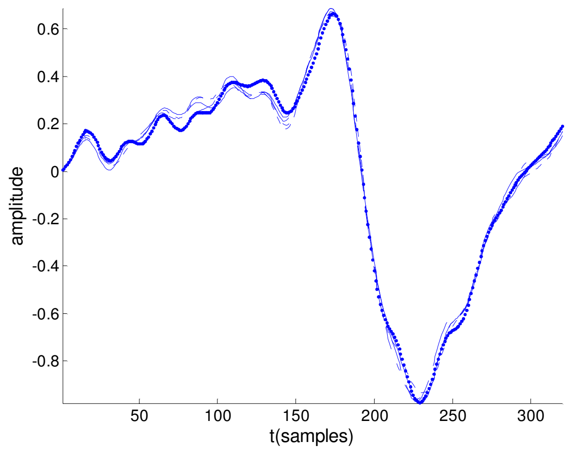

Fig 2. Sample templates for channel FCz-Cz, Subject AB1.

Fig. 2 shows sample templates of the MRP pattern in the bipolar channel FCz-Cz of Subject AB1. The activation time of the switch is at sample number 192 (Note that 128 samples equal one second). The templates are plotted for each of the five folds of the training sets (roughly 180 trials for this subject). As seen in this figure, the general shapes of the averages are similar over different training sets. These averages will be used later as templates for cross correlation with the input signals.

3.4 Cross Correlation

Cross correlation is an effective, yet easy to implement, method for waveform detection in the presence of noise. Since it is expected that the shapes of the single-trial IC data in the cross-correlogram, have some resemblance to the template, we postulated that a peak should be found around 320 samples (2.5 seconds). Fig.3 shows the histograms of the time of peaks of the cross correlogram of an IC validation set (channel FCz-Cz of Subject AB1). As seen in this figure, the timing of the peaks of the cross correlogram, occurs around the 320 sample (2.5 seconds) after the start of the sequence. However, this is not the case for NC trials and the timing of the peak of the cross-correlogram, has relatively a flat distribution. This is also to be expected, because of the random nature of the NC trials. Based on these observations, which were consistent among the subjects and over different channels, we decided to extract the maximum of the cross-correlogram in the range of [WindowSize-8, WindowSize+7] samples as the feature from each trial (WindowSize =320 samples). The size of this observation window (16 samples or $1/8^{th}$ of second) was chosen such that none of the features corresponding to NC trials is missed in the process of generating NC features (note that the NC trials are generated by extracting windows with the length of WindowSize and shifting this window by 16 samples over the EEG signals).

3.5. Threshold classifier

A simple threshold classifier was used for classification. If the value of an IC feature exceeded the value of a threshold, a TP is detected and if an NC feature exceeded a threshold, a FP is detected. A total of 101 threshold values from 0 to 1 (with the

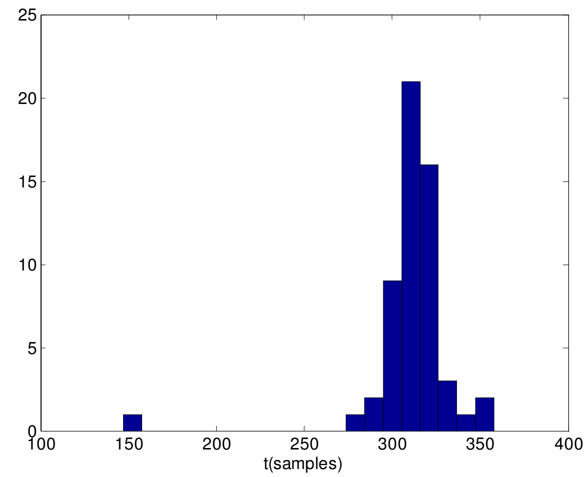

Fig 3. The time histogram of the maximum of cross-correlogram over a typical validation set (channel FCz-Cz, Subject AB1).

increments of 0.01) were tested and the values were stored.

For participation in a MCS, a classifier should not only be diverse compared to the other classifiers, but it should also be accurate enough (i.e. classify better than chance). As a result, only thresholds, which yielded TP > 50% and FP<50% on average on inner cross-validation sets were considered acceptable. For each channel, out of all the acceptable thresholds, the one which yielded the maximum TP/FP rate over the inner cross validation sets was selected as the final thresholding value for that particular channel.

3.6 Multiple classifier systems (MCSs)

The basic rationale behind using a MCS in our analysis is that the design of a single classifier with high performance for an ABI system is not a straightforward task at all. This is mainly due to the very noisy nature of the EEG signals, which makes the detection of the MRPs from the background EEG very difficult. The general hypothesis is that designing a strong classifier by fusing multiple weak classifiers is simpler than the design of a single high performance classifier.

One important issue in the design of a MCS is to create independent (diverse) classifiers. There are many methods of creating diverse classifiers, including boosting, bagging, using different features, etc. [8]. In this paper, we used the spatial information of channels for creating a diverse set of classifiers. Majority voting [8] is used for combining the outputs of classifiers.

3.7 Multi-objective GAs for channel combination

We applied a Genetic Algorithm (GA) [9] for the selection of classifiers with the highest classification accuracy yielded on the validation sets. With this approach, we were seeking to minimize the dependency on using diversity measures for classifier selection.

Each individual in the population has M bits, where bit i specifies whether or not the classifier # i is present in a particular MCS scheme. At least 3 objective functions should be considered in the fitness function: the TP rate, the FP rate and the number of classifiers. In order to simplify the fitness function, we combined the TP and the FP rates in a simple measure : TP/FP . This

Table 1. Comparison of the results obtained from the best channel and the proposed MCS method.

Subject	TP best	FP best	TP MCS	FP MCS
AB1	59.25	4.60	64.29	1.28
AB2	61.50	3.47	66.37	0.59
AB3	63.63	11.08	63.57	5.71
AB4	56.10	11.41	61.28	9.05

measure is maximized by maximizing the TP rate and minimizing the FP rate. Then a lexicographic approach [9] was used for multi-objective optimization of the structure of the system. Very briefly, the objectives are ranked according to their priorities before optimization. The objective with the highest priority (TP/FP) is used first when comparing the members of the population and the individuals are ranked in a single-objective fashion. Any ties are resolved by comparing the relevant individuals again with respect to the number of classifiers. The parameters and operators of the GA include, tournament-based selection (tournament size =3), uniform crossover, uniform mutation, the size of the initial population: 40 (random initialization), the size of the population 20. The number of evaluations was set to 1500. Also if for more than 10 consecutive generations, the amount of improvement in the best solution found so far is less than 1%, the algorithm is terminated. No tuning is performed on the GA parameters.

4. RESULTS

The results obtained for 4 subjects are shown in Table 1. As seen in this table, the proposed method yields a considerable improvement in the performance of the system (see columns 4 and 5) compared to the single best performing channel (see columns 2 and 3). The number of selected channels varied from 5 to 11 for different subjects.

The results show that for the first two subjects, by using the MCS scheme, the FP rates approach an acceptable level such that the system can be used in a practical application. For the other two subjects, although the performance of the system is improved (in terms of increasing the TP rate and/or decreasing the FP rate), yet for a practical system further improvement is desired.

It should be mentioned that the quality of the templates generated for the first two subjects, were visually much better than the templates derived for the last two subjects (perhaps the latter subjects were not as engaged as the first two subjects during the experiments). The amount of artifacts was also a factor which resulted in the reduction of the available IC trials for the last two subjects (especially for Subject AB4).

5. CONCLUSION

In this paper, a multiple classifier system (MCS) is used for improving the performance of an ABI system. In this scheme, the spatial information of multiple EEG channels is used for building independent classifiers. A genetic algorithm was then used to select the best MCS configuration.

Since this scheme deals with each channel separately, one of its advantages is its modularity. The other advantage of this scheme is that it is not necessary to design a complicated classifier based on a high-dimension feature vector generated by the data of all channels. It is only sufficient to build classifiers with moderate performance. The reduction in the dimension of the feature vector and the simplicity of the design are among the other advantages of using the proposed. The main advantage of using this scheme over the filtering selection approaches is to avoid the uncertainty of having features which may not work well together (in our scheme, all channels can potentially contribute to the classification performance of the system). The main advantage over the wrapper feature selection methods is that there is no need to train the classifiers for every feature subset. Once the classifiers are trained on each channel separately, they are treated as black boxes and their outputs are then combined based on majority voting scheme. Finally, the results in Table.1 shows great improvements in the performance of an ABI system when evaluated over the datasets for 4 able-bodied subjects compared to the performance of best channel. This suggests that by using a MCS method, the results reported in [5] and [6] may be improved further.

Future works include exploring more complicated feature extraction and feature classification methods in order to build up more powerful expert systems for each channel. We should also explore different cost functions for summarizing the confusion matrix. Using a GA for selection of the threshold of classifiers is also an area that is worth exploring. Finally, it would be interesting to test the performance of the system over continuous EEG data and ultimately during an online ABI experiment.

6. REFERENCES

[1] J.R. Wolpaw, N. Birbaumer, D.J. McFarland, G. Pfurtscheller, and T.M. Vaughan, “Brain-computer interfaces for communication and control”, *Clin. Neuro.*, vol.113, no. 6, pp. 767-791, 2002.

[2] S.G. Mason, and G. E. Birch, “Temporal Control Paradigms for Direct Brain Interfaces-Rethinking the Definition of Asynchronous and Synchronous”, Proc. of HCI Int. Conf., Las Vegas, 2005.

[3] B.O. Peters G. Pfurtscheller and H. Flyvbjerg, “Prompt recognition of brain states by their EEG signals”, *Theory Biosci*, vol. 116, pp.290–301, 1997.

[4] S.G. Mason, and G. E. Birch , “A brain-controlled switch for asynchronous control applications”, *IEEE Trans. Biomed. Eng.*, vol.47, no.10 , pp.1297-1307, 2000.

[5] B. Graimann, et.al, “Towards a direct brain interface based on human subdural recordings and wavelet packet analysis," *IEEE Trans Biomed. Eng.*, vol. 51, no.6, pp.954-962, 2004.

[6] J.E. Huggins, et.al, "Detection of Event-Related Potentials for Development of a Direct Brain Interface." *Journal of Clinical Neuro.*, vol. 16, no.5, p. 448-455, 1999.

[7] G. E. Birch, Z. Bozorgzadeh, and S. G. Mason, “Initial on-line evaluations of the LF-ASD brain-computer interface with able-bodied and spinal-cord subjects using imagined voluntary motor potentials”, *IEEE Trans. Neural Syst. Rehab. Eng.*, vol.10, pp.219-224, Dec.2002.

[8] A. K. Jain, R. P. W. Duin, and J. Mao, “Statistical pattern recognition: a review”, *IEEE Trans. On pattern analysis and machine intelligence*, vol.27, no.1, pp.4-37, Jan. 2000.

[9] T. Back, D. B. Fogel and T. Michalewicz, *Evolutionary Computation*, Institute of Physics Publishing, Bristol and Philadelphia, 2000.