BIOINFO Computer Engineering
Volume 1, Issue 1, 2011, pp-05-09
Available online at: http://www.bioinfo.in/contents.php?id=322

EEG Based Brain-Computer Interface for Classification Algorithms on Asynchronous Interface

[1]M. Senthilmurugan [2]N. Malmurugan [3]P.Jayapal
[1] Professor, Department of Information Technology (PG), A.V.C College of Engineering, Mayiladuthurai, Mannampandal – 609305, Nagapattinam District, Tamil Nadu, S.India
Email: smsenthil@hotmail.com, Mobile No: 9842091911
[2]Principal, Oxford Engineering College, Trichy, India, Email: n_malmurugan@yahoo.com
[3]Lecturer, Department of Information Technology (PG), A.V.C College of Engineering, Mayiladuthurai, Mannampandal – 609305, Nagapattinam District, Tamil Nadu, S.India
Email: jayapal.pal@rediffmail.com

Abstract- When we talk about the interfacing with a computer we typically mean typing at a keyboard or using a mouse. The EEG, or electroencephalogram, is electrical activity recorded from the scalp and produced by neurons in the brain. The growth of a Brain Computer Interface, or in our case, an EEG-based communication device, requires the raw EEG signal to be converted into a new output channel through which the brain can communicate and control its environment. This includes a discussion of an EEG-based user interface, covering all aspects of this topic ranging from the EEG input, over the processing stage, all the way to the corresponding output signals There is a growing awareness that for BCI's to be most useful for people with severe motor disabilities they must support self-paced (or "asynchronous") operation.
Keywords: *Brain Computer Interface (BCI), Electroencephalography (EEG), Artificial Neural Networks (ANN), Event-related potentials (ERP), Event-related desynchronization (ERD), Event-related synchronization (ERS).*

1. Introduction

Hans Berger first measured human brainwaves in 1924. Today, the EEG has become one of the most useful tools in the diagnosis of epilepsy and other neurological disorders. An electroencephalographic (EEG) pattern is an important and challenging biomedical signal processing problem. Such classification can be utilized to enable a patient to communicate without any overt physical movement. Developments of faster digital computers and better EEG devices have motivated many researchers to work on BCI systems [1] [2]. The EEG classification is one important part of the brain computer interface (BCI) - user interface which allows to work with computer and thus to communicate even for the disabled person (like those with the spinal cord injury, etc.). The EEG classification verifying physiology hypotheses about the brain can be also found in the field of physiology.

Different research groups have examined and used different methods to achieve this. Almost all of them are based on electroencephalography (EEG) recorded from the scalp. The EEG is measured and sampled while the user imagines different things (for example, moving the left or the right hand). Depending on the BCI, particular preprocessing and feature extraction methods are applied to the EEG sample of certain length. It is then possible to detect the task-specific EEG signals or patterns from the EEG samples with a certain level of accuracy.

2. Electroencephalography (EEG)

Electroencephalography (EEG) is a method used in measuring the electrical activity of the brain. This activity is generated by billions of nerve cells, called neurons. Each neuron is connected to thousands of other neurons. Some of the connections are excitatory while others are inhibitory. The signals from other neurons sum up in the receiving neuron. When this sum exceeds a certain potential level called a threshold, the neuron fires nerve impulse. The electrical activity of a single neuron cannot be measured with scalp EEG. However, EEG can measure the combined electrical activity of millions of neurons.

The temporal resolution of EEG is very good: millisecond or even better. However, the spatial resolution is poor. It depends on the number of electrodes, but the maximum resolution is in centimeter range whereas, for example, in MEG, PET or fMRI it is in millimeter range. The ongoing EEG is characterized by amplitude and frequency. The amplitudes of the EEG signals typically vary between 10 and 100 μV (in adults more commonly between 10 and 50 μV).

The electrical activity goes on continuously in every living human's brain. We may sleep one third of our life times, but the brain never rests. Even when one is unconscious the brain remains active. Much of the time, the brain waves are irregular and no general pattern can be observed [3].

However, there exist various properties in EEG, which can be used as a basis for a BCI:

1. Rhythmic brain activity
2. Event-related potentials (ERPs)
3. Event-related desynchronization (ERD) and event-related synchronization (ERS).

Table –1 Common EEG frequency ranges

Band	**Frequency Range**
Alpha (α)	8 – 13 Hz
Beta (β)	14-30 Hz
Delta (δ)	0.5 – 3 HZ
Theta (θ)	4 – 7 Hz
Mu (γ)	22 – 40 Hz

3. Two different BCI approaches

An ideal BCI could detect the user's wishes and commands directly. However, this is not possible with today's technology. Therefore, BCI researches have used the knowledge they have had of the human brain and the EEG in order to design a BCI. There are basically two different approaches that have been used. The first one called a pattern recognition approach is based on cognitive mental tasks. The second one called an operant conditioning approach is based on the self-regulation of the EEG response.

4. BCI components

A typical BCI device consists of several components. These include electrode cap, EEG amplifiers, computer and subject's screen. A critical issue is how the user's commands, i.e., the changes in the EEG, are converted to actions on the feedback screen or the application. This process can be divided into different stages.

4.1 Measurement of EEG

This is done by using the electrodes. Many BCIs use a special electrode cap, in which the electrodes are already in the right places, typically according to the international 10-20 system. It saves time because the electrodes do not have to be attached one by one. Typically, less than 10 electrodes are used in online BCIs with sampling rates of 100-400 Hz.

4.2 Preprocessing

This includes amplification, initial filtering of EEG signal and possible artifact removal. Also A/D conversion is made, i.e. the analog EEG signal is digitized.

4.3 Feature extraction

In this stage, certain features are extracted from the preprocessed and digitized EEG signal. In the simplest form a certain frequency range is selected and the amplitude relative to some reference level measured [6]. Typically the features are certain frequency bands of a power spectrum. The power spectrum (which describes the frequency content of the EEG signal) can be calculated using, for example, Fast Fourier Transform (FFT), the transfer function of an autoregressive (AR) model or wavelet transform. No matter what features are used, the goal is to form distinct set of features for each mental task. If the feature sets representing mental tasks overlap each other too much, it is very difficult to classify mental tasks, no matter how good a classifier is used. On the other hand, if the feature sets are distinct enough, any classifier can classify them.

4.4 Classification

The features extracted in the previous stage are the input for the classifier. Different BCIs can classify different number of classes, typically 2 to 5 classes. The classifier can be anything from a simple linear model to a complex nonlinear neural network that can be trained to recognize different mental tasks. With the exception of simple threshold detection, the classifier can calculate the probabilities for the input belonging to each class. Usually the class with the highest probability is chosen. However, in some BCI protocols none of the classes may be chosen, if the classification probability does not exceed some predefined level. This kind of classification result can be called “nothing” or “reject”.

Some critical properties of features need to be taken into consideration to select an algorithm [8]:
- Raw signals have a very low signal-to-noise ratio.
- Feature Vectors are often of high dimensionality.
- BCI features are non-stationary, may vary over time and particularly over sessions, which may imply doing training in each session.

Learning sets are usually small compared to the number of features, because training is time consuming for the subject, and the features often change over time. Among the five main categories of classifiers for BCI defined in [8] (linear classifiers, neural networks (NN), non linear Bayesian classifiers, nearest neighbours, and combinations of classifiers), we focus our attention on Neural Networks and Linear Classifiers, which are the most readily available and widely used.

Neural Networks, along with linear classifiers are widely used in BCI research, especially Multilayer Perceptrons (MLPs), which consist of several layers of neurons, each neuron being connected with the outputs of the previous layer: the first layer is connected with the input (ie: the vector of features) and the output of the last layer gives the label. NN are very flexible classifiers which have been used in many different BCI problems (binary, multiclass, synchronous, asynchronous...). However, since they can approximate any continuous function, they are sensitive to overtraining especially with noisy and non-stationary data.

Linear discriminant analysis (LDA) is a statistical method used to investigate differences among multivariate classes, to determine which attributes discriminate between the classes, and to determine an optimal way to distinguish among classes in the linear sense. The classic method of linear discrimination is the so-called Fisher’s linear discriminant. It seeks a linear combination of the signal features that maximizes the linear class separability. Mathematically, the discriminant function can be described as

$$\mathbf{d(x) = w^T x + w_0.}$$

The first part, $w^T x$, is the inner product of a feature vector x and a weight vector w, and, in general, it transforms a feature vector from a high-dimensional feature space to a 1-dimensional feature space. This transformation maximizes the distance between the means of the two classes and simultaneously minimizes the variance within each class. The second part, w0, is the threshold or bias. In a two-class discrimination problem, it can be used to assign features to classes. For instance, if the transformed feature vector is greater than the threshold, the feature is assigned to class 1, otherwise it is assigned to class 0. Furthermore, the distance between the value obtained from the linear combination of the feature vector with the weight vector and the threshold can be employed as a measure of classification performance. This time varying distance can be used to provide continuous feedback in a BCI system.

LDA can be used as a supervised linear statistical classifier. As a linear method, it gives the advantage that overfitting to the training data is unlikely, which is in contrast to nonlinear classifiers like neural networks that have a tendency to fit the training data very well, but have poor generalization abilities. Another advantage of LDA is its simplicity and computational efficiency. For that reason, LDA has been used in a number of online and offline BCI studies. So, for example, LDA was used to classify AAR parameters in a two-class discrimination task (left- and right-hand imaginary) with continuous feedback. Preprocessed data by the CSP method, from which band power features were derived, were classified with LDA. Other examples where LDA was used as classifier in brain-computer communication can be found elsewhere.

Support Vector Machines (SVM) uses a discriminant hyperplane that maximizes the margins, which is known to allow better generalization. SVM also permit non linear decision boundaries by introducing a kernel, for example Gaussian or Radial Basis Functions (RBF). SVM have several advantages: they have good generalization properties and are not too sensitive to the curse of dimensionality. SVM, which are stable and have a low variance, are efficient with noisy data that often contain outliers. In a review article, Lotte et al. [8] noted that a Gaussian SVM applied to a correlative time-frequency representation had 86% accuracy. Non-linear SVM have also outperformed an MLP in experiments. For these reasons, we chose to classify our data with a Gaussian SVM. We used a soft margin SVM [9] and had to optimize the margin constraint in order to prevent overfitting.

Table-2 Rate of correct classification for S1, S2 and S3

		CB	C	HM	FM
CB	S1		0.91	1	0.81
	S2		0.52	0.47	0.05
	S3		0.7	0.85	0.21
C	S1	0.96		0.88	1
	S2	1		0.84	0.90
	S3	0.58		0.66	0.52

HM	S1	0.03	0.48		0.38
	S2	0.75	0.54		0.45
	S3	0.47	0.79		0.31
FM	S1	0.52	0.90	0.92	
	S2	1	0.53	0.83	
	S3	0.90	0.59	0.82	

Table-3 Error rate for S1, S2 and S3 (Section 5.4.1)

		CB	C	HM	FM
CB	S1		0.03	0.48	0.37
	S2		0	0.34	0
	S3		0.37	0.39	0.34
C	S1	0.09		0.38	0.1
	S2	0.31		0.34	0.34
	S3	0.33		0.25	0.41
HM	S1	0	0.18		0.18
	S2	0.4	0.24		0.25
	S3	0.22	0.30		0.39
FM	S1	0.26	0	0.4	
	S2	0.50	0.12	0.38	
	S3	0.46	0.46	0.47	

Table-4 Comparison of Classification results for S1, S2 and S3, with the performance measure defined in 5.4.1

	CB/C	CB/HM	C/HM	CB/FM	C/FM	HM/FM
S1	0.94	0.62	0.70	0.66	0.94	0.67
S2	0.80	0.64	0.71	0.65	0.76	0.66
S3	0.66	0.68	0.72	0.58	0.56	0.57

4.4.1 Statistical analysis

In task X vs task Y classification, the test data were composed of 2n feature vectors, from which n belonged to label X and n belonged to label Y. Consider that the SVM was able to correctly label p X tasks out of the set of n X tasks and q Y tasks out of n Y tasks in the testing part. P/n is the rate of correct classification (or recognition rate) in the set of X tasks and q/n is the rate of correct classification in the set of Y tasks. The corresponding numbers are reported in Table 1 : p/n is given at (line X, column Y) and q/n is given at (line Y, column X). For example, 0.92 (line 1, column 2) is the proportion of correctly classified CB task in the CB vs C testing set, meaning that the algorithm was able to correctly classify 92% of the CB tasks in the set, whereas 97% of the C tasks (line 2, column 1) were correctly classified in the same set.

The error rate is the probability of misclassification of a task, assuming it was predicted by the SVM. Consider that the SVM classifies r tasks with label X and s tasks with label Y (r+s=2n). Out of the r (resp. s) tasks, only p (resp q) truly belongs to class X (resp. Y). In Table 2, the rate (1-p/r) is reported at (line X, column Y) and (1-q/s) is given at (line Y, column X). For example, 0.03 (line1, column 2) is the error rate the algorithm obtained when predicting label CB in C vs CB classification. The lower the rate, the more accurate the prediction of a given mental task is.

The best classification results were obtained for high recognition rates and low error rates. To assess global performances, we can weigh the rates of correct classification and the error rate as below:

$$\mathbf{Performance=0.5 * \lambda * (p/n+q/n) + 0.5 * \mu * (p/r+q/s)}$$

Where $\lambda+\mu=1$

We calculated for each set of two tasks the performances of the SVM using $\lambda=\mu=0.5$, meaning that error rates and rates of correct classification had the same weight. The maximum performance is 1 whereas chance level is 0.5. We report in Table 3 the results for all three subjects.

5. Results

Out of the original set of four mental tasks, we performed an SVM classification on each set of two mental tasks (a total of 6 classifications) for each subject. We calculated two statistics for each mental task, as explained in section 2.3 : the rate of correct classification and the error rate. We also calculated the performance of the algorithm on each pair of mental tasks as explained in section 2.3. For S1, the best recognition results (Table 1) were observed for task C vs task FM (97% - 92%), and task FM vs task C (100% - 89%). Generally speaking, C was the best recognized task (97% 89% 100%). However both task CB in CB vs HM (100%) and task FM in FM vs HM (92%) achieved better recognition rates than task C in C vs HM (89%). Some tasks were not recognized better than chance level (HM in HM vs C: 49%). For S2, the best recognition rates were obtained for C in C vs CB and FM in FM vs CB classification, where 100% was achieved. However, CB was poorly recognized in both cases (51% in C vs CB and 5% in CB vs FM), For S3, the best recognition rates were obtained for FM in FM vs CB (89%) and CB in CB vs HM (86%).

For S1, the lowest error rate (Table 2) was achieved on FM in FM vs C and HM in HM vs CB classification. Since the corresponding rate of correct classification was very high (89%) for FM in FM vs C, the algorithm was able to classify FM vs C in a very accurate way. However, for HM in HM vs CB, the corresponding percentage of correctly labeled HM task (Table 1) was very low (3 %). For S2, the error rate was also 0% for CB in CB vs C and CB vs FM classification. Generally speaking, the error rates achieved by S3 were higher than those achieved by S1 and S2. For S3, the lowest error rate was 23% for HM in HM vs CB classification. When assessing the global performances for each set of two mental tasks (Table 3), the best result was achieved on CB vs C and C vs F for S1 (0.95). In general, the algorithmic performances were lower for S2 than for S1, the highest score achieved by S2 being also for CB vs C. The global results were also poorer for S3 than for S1 and S2, although S3 outperformed S1 and S2 in two classifications (CB vs HM and C vs HM)

6. Conclusion

We observed that the classification performances of SVM strongly depended on the type of mental tasks performed by the subject. The reasons why such differences are observed still need to be explored. Do some mental tasks imply steadier brain states than others? Do some subjects have a stronger power of concentration on some tasks than on others? For example, two subjects (S1 and S2) chose a Mathematics or applied Mathematics major at University and obtained very good classification results on task C classification (especially on C vs CB which are both abstract cognitive tasks, one involving visuospatial abilities and the other involving calculus). A further classification with a multiclass SVM showed that task C was the best classified task for S1 (data not shown). To generalize our assumptions, we intend to apply our method to asynchronous BCI and assess whether the best labelled mental tasks are the same in both cases to ensure the learning process we applied here is effective in more realistic applications.

References

[1] C. Guger, et al. How many people are able to operate an EEG-based brain-computer interface (BCI). IEEE Trans. Rehab. Engng,, vol. 11, 145-147, 2003.
[2] S. Lemm, B. Blankertz, G. Curio, and K.-R. Muller., "Spatio-spectral filters for improved classification of single trial EEG.," IEEE Tranactions on. Biomedical Enineering., vol. 52, pp. 1541-1548, 2005.
[3] W. Wu, Y. Gao, E. Bienenstock, J. P. Donoghue, and M. J. Black, "Bayesian population decoding of motor cortical activity using a kalman filter," Neural Computing, vol. 18, pp. 80-118, 2006.
[4] Coyle, S., Ward, T., & Markham, C. (2003). Brain-computer interfaces: A review. Interdisciplinary Science Reviews, 28(2), 112-118.
[5] Mason, S.G., & Birch, G.E. (2003). A general framework for brain-computer interface design. IEEE Transactions on Neural Systems and Rehabilitation Engineering, 11(1), 70-85.
[6] Hjorth, B. (1970), "EEG analysis based on time series properties", Electroencephalography and Clinical Neurophysiology, 29:306-310.
[7] Electroencephalogram processing using neural networks, Clinical Neurophysiology, Vol. 113, No. 5, 2002, pp. 694-701.
[8] Brainmaster. http://www.brainmaster.com
[9] EEG Spectrum International Inc. Homepage. http://www.eegspectrum.com/, 2002.
[10] Andrea K″ubler et al. Brain-computer communication: Unlocking the locked in. *Psychological Bulletin*, 127(3):358–373, 2001.
[11] Ernst Nieder Meyer and Fernando Lopes da Silva, editors. *Electroencephalography*, chapter 7, pages 123–141. Lippincott Williams&Wilkins, 1999.
[12] Bio-Medical Instrumentation and Measurements by Leslie Cromwell, Fred J.Weibell and Erich A.Pfeiffer Perason Education Second Edition.
[13] Medical Instrumentation Application and Design, Third Edition by John G.Webster, John Wiley & Sons Inc, Newyork.
[14] Jasper, H. (1958). The Ten Twenty Electrode System of the International Federation. Electroencephalography and Clinical Neurophysiology 10, 371–375.
[15] Erfanian, Real-time eye blink suppression using Neural adaptive filters for EEG-based brain computer interface, 24th Ann Int. Conf.IEEE/EMBS, 2002.