Adaptive classification of mental states for asynchronous brain computer interfaces

Jean Le Pavec, Maureen Clerc

ADAPTIVE CLASSIFICATION OF MENTAL STATES FOR ASYNCHRONOUS BRAIN COMPUTER INTERFACES

Jean Le Pavec[1] , Maureen Clerc[1 2]
[1] Equipe-Projet Odyssée
INRIA/ENS
2004, route des Lucioles
Sophia-Antipolis, France
[2] Universite Paris-Est, Ecole des Ponts ParisTech, CERTIS
email: jlepavec@sophia.inria.fr, Maureen.Clerc@sophia.inria.fr

ABSTRACT

Brain Computers Interfaces (BCI) are emerging as a new means of communication, aiming to make a direct link between the brain and an external device, bypassing conventional motor outputs, such as peripheral nerves and muscles. A BCI extracts features from a brain signal and classifies them in order to interpret them in terms of the user's volition. For communication to be effective, the computer has to provide feedback to the user allowing him/her to judge how the brain activity is being classified and interpreted. Similarly, the user must produce patterns of brain activity which can easily be learned and recognized by the computer. Here, we describe a method for selecting mental tasks that are best classified by a subject using support vector machines (SVM).

KEYWORDS
brain computer interfaces, electroencephalography, learning, artificial intelligence, support vector machines.

1. Introduction

Brain Computer Interfaces, which aim to make a direct link between the brain and an external device, may offer the hope to provide the severely disabled with a new functional means to communicate with the outer world.

BCI was defined at the first international meeting for BCI technology in 1999 as a system that “must not rely on normal output pathways of peripheral nerves and muscles”. It measures signals coming directly from the brain, either invasively or non-invasively. Invasive recordings have a high signal to noise ratio (SNR) and are specific but have a small sensitivity volume. Non-invasive measurements such as electroencephalography (EEG), magnetoencephalography (MEG) or functional MRI have wider fields of view but are more difficult to analyze statistically. This paper focuses on EEG because of its simpler use in everyday life for people without normal motor abilities.
Because of the difficulty of reliably extracting information, up to now, successful BCI systems mostly function in a “synchronous” mode, i.e. where the system controls the timings at which the user's commands can be interpreted. This is for example the case of BCI spellers which use the evoked potentials elicited by a visual stimulation to detect which letter a subject is attending to [1].
It would seem more natural for the user to decide the pace at which commands can be issued to the system. BCI systems which comply with the user's timings are said to operate in an “asynchronous” mode [2]. Since asynchronous BCI are not triggered by external stimuli, they rely instead on the analysis of the individual characteristics of the subject's spontaneous activity. Among these, Slow Cortical Potentials (SCPs) and cortical oscillations are of particular interest.
SCPs correspond to low frequency (0.5 - 3Hz) EEG signals. Negative SCPs are often associated with movement and high cortical activity whereas positive SCPs reflect reduced cortical activation. SCPs can be modulated to control the movement of an object on a computer screen or to select a letter by a succession of two choices [3].
Primary sensory and motor cortical areas often display 8-12Hz activity when they are not engaged in processing sensory inputs or producing motor outputs. *Mu rhythms* are mostly localized in the somatosensory and motor cortices whereas *alpha rhythms* are more prominent in the visual cortex. *Beta rhythms* (18-26 Hz) are often associated with mu rhythms [4]. These oscillatory activities are involved in a variety of mental tasks – for instance, mu and beta rhythms decrease during the preparation of movements and increase after movements and during relaxation. By producing mu and beta rhythms through mental states such as imaginary hand movements, whole body activity or relaxation, it is possible to control a cursor in one or two directions [5].
Mental task classification relies on state-of-the-art classification algorithms. For a given classification algorithm, the learning stage is subject- and session-dependent. It is unlikely that all types of mental tasks should perform equally well for all subjects with respect to their classification [6]. It is therefore interesting to evaluate the mental tasks that are best classified by the algorithm for a given subject, and this paper provides a method to guide the selection of mental states which are best distinguishable by an SVM classifier.
In section 2, we describe our BCI system, and experimental protocol. Section 3 focuses on the classification algorithm, and section 4 presents our preliminary results.

2. Material and Methods

Although our long-term objective is asynchronous BCI, this study is concerned with a learning stage in a synchronous mode. During this stage, the subject is required to perform a set of mental tasks, one after another, with a timing provided by visual prompts. After the learning, the performances of the algorithm were compared on tasks, taken two by two. The method enabled us to select the labels that were best classified by the algorithm and therefore that were likely to give the best results in asynchronous BCI, assuming that the mental states and concentration of the subject do not differ too much between synchronous and asynchronous BCI.

2.1 Experimental protocol

2.1.1 Mental states

In this paper, we report on results obtained with three volunteers, two 23 year-old (S1, S2) and one 24 year-old (S3) male students, who were asked to perform a set of four mental tasks:

- cube rotation (CB)
- calculus: iterative subtraction (C)
- imagination of hand movement (HM)
- imagination of foot movement (FM)

Each mental task lasted ten seconds. There was a break of three seconds between two successive mental tasks. The computer indicated both the starting and ending times, together with the nature of the tasks to perform, which was selected randomly and displayed on the screen, using eevoke (Advanced Neuro Technology, Enschede, The Netherlands). Each task was performed four to six times depending on the subject:

S1:

- CB : 4 times
- C : 6 times
- HM : 4 times
- FM : 5 times

S2 and S3:

- CB : 6 times
- C : 6 times
- HM : 6 times
- FM : 6 times

2.1.2 Signal acquisition and preprocessing

EEG potentials were recorded at 64 standard positions over the scalp, with a commercial cap and integrated electrodes. The data acquisition was performed using ASA-lab (Advanced Neuro Technology, Enschede, The Netherlands) at a 256 Hz sampling rate for S1. For S2 and S3, the data were recorded at 512 Hz and we resampled them at 256 Hz using eeglab. Apart from a 50Hz filtering, we chose not to perform further preprocessing because our purpose is to use BCI for real-time applications and preprocessing would slow down the process of data analysis. Moreover, data recorded outside the laboratory are likely to be noisier than those recorded inside. So we assumed that processing noisier data would have better generalization properties.

2.1.3 Feature extraction

The features we used were based on a short-time Fourier transform. Although other features can be used (parameters of autoregressive models or wavelets...), their benefits are not self-evident [7]. To estimate the power spectrum of each channel over one second, we used a Welch periodogram. Specifically, we averaged the FFT of three segments of 0.5 second with 50% overlap, which yields a frequency resolution of 1Hz. We divided the frequency spectrum in five bands (4-7 Hz, 8-12Hz, 13-17Hz, 18-26Hz and 27-35 Hz), averaged the powers in each frequency band and normalized them, following the M1 method described in [6] to obtain 5 power values, for each channel. Thus an EEG sample had 320 features (64 channels times 5 components each). A new feature vector was obtained every 0.25 second.

2.2 Learning and classification

Formally, classification consists of finding the label of a feature vector x, using a mapping f, where f is learnt from a training set T. The purpose of the learning stage is to provide the algorithm with preclassified labelled data (here, vectors of 320 features), from which the algorithm builds the mapping in order to predict the labels of new data.

2.2.1 Methodology

To perform the learning, we divided our data into two parts: the testing part containing the last epoch (10 seconds) for each mental task and the learning part, with the rest of the data. For S1, we removed the last FM epoch because S1 reported having made a mistake in performing the last FM task. Each 10 second epoch was composed of 37 feature vectors. Thus the learning set was composed of 518 feature vectors for S1 and 740 feature vectors for S2 and S3 (for four tasks). For all three subjects, the testing set was composed of 148 feature vectors. The results obtained with the testing part were used to assess the performance of the algorithm on the mental tasks.

2.2.2 Choice of algorithm

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 Disciminant Analysis (LDA) assumes that all classes have an equal variance and calculates the hyperplane that minimizes interclass variance while maximizing the distance between the classes' means. They introduce a regularization parameter that allows or penalizes classification errors and thus can accommodate outliers, which are common in BCI signals. However, because of its linearity, LDA often gives poor results on complex nonlinear EEG data.
- Support Vector Machines (SVM) use 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.

		CB	C	HM	FM
	S1		0.92	1	0,.81
	S2		0.51	0.49	0.05
CB	S3		0.7	0.86	0.21
	S1	0.97		0.89	1
	S2	1		0.84	0.91
C	S3	0.59		0.65	0.51
	S1	0.03	0.49		0.38
	S2	0.76	0.54		0.46
HM	S3	0.46	0.78		0.3
	S1	0.51	0.89	0.92	
	S2	1	0.54	0.84	
FM	S3	0.89	0.59	0.81	

Table 1 Rate of correct classification for S1, S2 and S3 (section 2.3)

		CB	C	HM	FM
	S1		0.03	0.49	0.38
	S2		0	0.33	0
CB	S3		0.37	0.38	0.33
	S1	0.08		0.37	0.1
	S2	0.32		0.35	0.33
C	S3	0.33		0.25	0.41
	S1	0	0.18		0.18
	S2	0.4	0.23		0.26
HM	S3	0.23	0.31		0.39
	S1	0.27	0	0.4	
	S2	0.49	0.13	0.39	
FM	S3	0.47	0.45	0.46	

Table 2 Error rate for S1, S2 and S3 (section 2.3)

	CB/C	CB/HM	C/HM	CB/FM	C/FM	HM/FM
S1	0.95	0.63	0.71	0.67	0.95	0.68
S2	0.8	0.63	0.7	0.64	0.75	0.66
S3	0.65	0.68	0.72	0.58	0.56	0.56

Table 3 Comparison of classification results for S1, S2 and S3, with the performance measure defined in 2.3

2.3 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 belong 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 (line 1, 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:

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

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.

3. 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 labelled 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)

3. Discussion

Although raw data are known to be noisy and variable, the algorithm achieved very good results on some tasks, depending on the subjects, for example a global performance of 0.95 on CB vs C and C vs FM was achieved by S1 and 0.8 on CB vs C, by S2. S3 obtained his best result on C vs HM (0.72).

- The global performances of BCI depended both on the subjects and tasks. Generally speaking, S1 achieved better results than S2 and S3, although S3 slightly outperformed S1 and S2 in two cases (CB vs HM and C vs HM). No better results were achieved than 0.95 on CB vs C and C vs FM for S1. In all other cases, S2 achieved better results than S3, which implies that the ability to use BCI, ie: to produce stable patterns of brain activity, specifically to mental tasks, depends on subjects.
- The performances of the algorithm on each pair of tasks also depended on the subject. CB vs C led to very good classification performances for S1 (0.95) and S2 (0.80), poorer for S3 (0.65). C vs HM ranked first for S3, third for S1 and S2. This result encourages to perform this kind of study for each subject to choose the pairs of mental tasks that lead to the best classification performances.
- A mental task is not well classified in general but rather vs another mental task. For example, task CB was well classified vs C (0.95) but poorly vs HM (0.63) and FM (0.67) for S1. However, some tasks seem to lead to poorer classification results than others, such as HM for S1 and FM for S3.
- Although Table 3 synthesizes the results for each set of two mental tasks, both rate of correct classification and error rate should be taken into account, when assessing the performance of BCI, depending on the needs for communication. Thus, high λ and low μ tend to reduce the rate of false negatives whereas high μ and low λ tend to reduce the rate of false positives. For example, in CB vs C classification for S2, the percentage of correctly labelled CB tasks was low (51%), slightly higher than chance. However, the algorithm was never mistaken when predicting C. Such a low error rate could be successfully used for mental tasks that should absolutely be labelled correctly even if the recognition rate is low. Other tasks achieved both low error rate and low recognition rate (3%), such as HM in HM vs CB classification for S1. Their use in BCI should not prove as efficient.
- Although HM vs FM classification has already been extensively performed in BCI research [4], it ranked fourth for S1 and S2 and fifth for S3 in our study. Both tasks, which are body movements, encompass close areas in the brain, typically in the motor and somatosensory cortices with similar frequency bands, such as mu and beta. The modulation of power spectra in those bands and areas in the brain for movement tasks did not seem to be the most efficient patterns of brain activity to classify mental tasks for all three subjects.
- When asked to make predictions about the performances of the algorithm (data not shown), most subject considered that a pair of movement tasks (HM vs FM) or a pair of cognitive tasks (CB vs C) would lead to poorer algorithmic performances than a pair of cognitive vs movement tasks. Consequently, the relatively poor performances of HM vs FM did not surprised them much. However, the subjects' prediction were negated for the results obtained in HM vs FM classification were better than those obtained in CB vs HM and CB vs FM for both S1 and S2, although CB is a cognitive task and both FM and HM are movement tasks. In the same vein, the huge difference in classification results between FM vs C and HM vs C for S1 and S3 seemed quite surprising for both HM and FM are body movements. Those results suggest that the subjective distance between mental tasks is not sufficient to predict the performance of BCI. The subject's emotional or arousal state, together with the type of movement performed may also affect the performances of the algorithm, hence the need for such a method as to select the most classifiable mental tasks objectively.

4. 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 visuo-spatial 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] E E Sutter, The brain response interface: communication through visually-induced electrical brain responses. *J. Microcomput Appl* 1992:15:31—45.

[2] M Besserve, J Martinerie, L Garnero, Non-invasive classification of cortical activities for brain computer interface: a variable selection approach. *Proceedings of ISBI conference* 2008.

[3] N Birbaumer, Slow cortical potentials: their origin, meaning and clinical use. In: van Boxtel, Böcker, editors. *Brain and behavior: past, present and future*, Tilburg University Press, 1997.

[4] J R Wolpaw, N Birbaumer, D McFarland, G Pfurstcheller, T M Waughan, Brain-computer interfaces for communication and control *Clinical Neurophysiology 113* (2002) 767-791

[5] J R Wolpaw and D McFarland, Control of a two-dimensional movement signal by a noninvasive brain-computer interface in humans. *PNAS* , 101(51), 2004.

[6] J Le Pavec, Adaptive learning for asynchronous brain-computer interfaces. Master's thesis (Cogmaster, Paris, 2008)

[7] J del R Millan, F Renkens, J Mourino, W Gerstner, Brain-actuated interaction *Artificial Intelligence 159* (2004) 241-259

[8] F Lotte, M Congedo, A Lécuyer, F Lamarche and B. Arnaldi, A Review of Classification Algorithms for EEG based Brain-Computer Interfaces, *Journal of Neural Engineering* (2007) 4R1-R13

[9] “SVM and Kernel Methods Matlab Toolbox” was developed by S. Canu and Y. Grandvalet and V. Guigue and A. Rakotomamonjy,"Perception Systemes et Information, INSA de Rouen, Rouen, France" (2005)