Ensemble of SVMs for improving Brain Computer Interface P300 speller performances

A. Rakotomamonjy V. Guigue G. Mallet V. Alvarado

P.S.I CNRS FRE 2645 INSA de Rouen,
Avenue de l'université
76801 Saint Etienne du Rouvray France
alain.rakotomamonjy@insa-rouen.fr

Abstract. This paper addresses the problem of signal responses variability within a single subject in P300 speller Brain-Computer Interfaces. We propose here a method to cope with these variabilities by considering a single learner for each acquisition session. Each learner consists of a channel selection procedure and a classifier. Our algorithm has been benchmarked with the data and the results of the BCI 2003 competition dataset and we clearly show that our approach yields to state-of-the art results.

1 Introduction

Some people who suffer some neurological diseases can be highly paralyzed due to the fact that they do not have anymore control on their muscles. Therefore, their only way to communicate is by using their electroencephalogram signals. Brain-Computer interfaces (BCI) research aim at developing systems that help those disabled people communicating with machines.

Research on BCI is a fast growing field and several EEG-based techniques have been proposed for realizing BCI. The BCI framework that is of interest for us is based on Event Related Potentials (ERP) which appear in response to some specific stimuli. Hence, this BCI produces an appropriate stimuli for which the patient is expected to respond. The core principle underlying such system is then based on the recognition of ERP which corresponds to the stimuli. In other words, this BCI is essentially based on classification of a EEG signals which is a difficult problem due to the low signal-to-noise ratio and the variability of the ERP responses within a single subject.

We propose some solutions for improving the performance of such BCI by addressing this variability problem through an ensemble approach based on linear SVMs [6].

The paper is structured as follows : section 2 describes the BCI classification problem with more details and the methodology that have been used for channel selection and for building the multiple classifier systems. Section 3 presents the

[1] This work was supported by grants from the IST programme of the European Community under the PASCAL Network of excellence, IST-2002-506778.

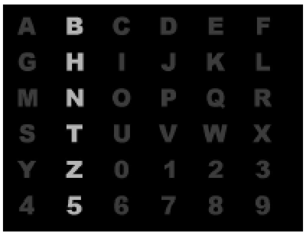

Fig. 1. P300 speller matrix with an highlighted column.

results that have been achieved whereas section 4 concludes the paper with comments and perspectives on the work.

2 Methods

The BCI problem we are addressing in this paper concerns the P300 speller. This speller has been introduced by Farwell and Donchin [3] who developed a protocol whereby a subject is presented a 6×6 characters matrix in which a row or column is randomly intensified. Then large P300 evoked potentials can be recorded in response to the intensification of a desired character. Hence the objective of the problem is to classify these potentials whether they correspond or not to the desired character . The data that we used in this study comes from the BCI 2003 competition dataset [1] and they have been provided by the Wadsworth institute [5]. In the following, we give a short description of the datasets. The data corresponds to the three separate spelling sessions of respectively 5,6, 8 words by a single subject. These recordings correspond to 64 channels which have been digitized at 240 Hz. However, in our case, we are only interested in the part of the signal that follows the intensification of a row or column. The experimental protocol, which have already been described in [3] is the following. For selecting a given character, each row and column of the matrix is highlighted in a random sequence (hence, the desired character appears on 2 out of 12 intensifications) and this procedure is repeated 15 times for the same character. Then after a short pause, the user has to focus on another character. In our case, the objective is to predict a word correctly by means of the fewer sequence repetitions as possible.

Brain Computer Interfaces classification problems are challenging essentially due to (i) the low signal-to-noise ratio of the signal, (ii) the variability of the EEG signal for a given user, and (iii) the variability between different users. Then, in order to achieve interesting results, it is expected that a classification strategy addresses all these points.

The problem we face is thus the following. We have $\{x_i, y_i\}_{i=1,\cdots,\ell}$ examples where each $x_i \in \mathbb{R}^d$ and $y_i \in \{-1, 1\}$. In our case, $d = 64 \times 240 \times t_d$ where t_d corresponds to the duration of the signal of interest after intensification. Since we have fixed t_d to $0.667s$, the number of features is equal to 10240 which suggests that a variable selection procedure would be helpful for a least, reducing the processing time.

According to the competition rules, the first two spelling sessions have been used for training our system while the last one has been kept for test. Hence, our training data is composed of 7560 stimuli responses which 1260 of them contain a true P300 ERP responses.

Signals from channels are. Here, we have filtered each signal using a 8-order bandpass filter with cut-off frequencies of 0.1 Hz and 20 Hz and then decimated each signal according to this latter frequency. Since we are interested only in a time window of 667 ms after intensification, at this stage the data dimension is 896. At this point, one can directly use the signals from preselected or user-defined channels as inputs of a classifier. This approach has been investigated by Meinicke et al. After appropriate scaling of the inputs, they trained a SVM with equal number of positive and negative examples. In this work, we propose an approach that tries to take into account the variability of the signals during different sessions or even during different words spelling in the same session.

The idea is to train a complete recognition stage for each word in the training set and then to combine the output of all the resulting classifiers for producing a single score for each signal.

Hence for training a single classifier, we have performed the following steps for each word spelling session signals:

- signals of each channel are normalized to zero mean and unit variance
- a channel selection is performed in order to retain only the most relevant channels. Our channel selection algorithm is based on a recursive channel elimination. The algorithm starts with all channels and then according to a user defined performance criterion C, the criterion $C^{(-j)}$ (the criterion when channel j is removed) is evaluated on all the remaining words of the training session. In our case, our criterion is defined as $C = \frac{tp}{tp+fp+fn}$ where tp, fp, fn are respectively the number of true positive, false positive and false negative. The channel that is suppressed is thus the one which removal maximizes the criterion C. Then we continue this procedure until all channels have been removed.
  Hence, this channel selection algorithm allows us to rank all the channels according to the criterion C and only the most relevant channels are subsequently used for classification.
- an linear SVM is then trained using all available examples described by the selected channels for this single word.

In this BCI problem, one should distinguish an evoked potential response classification (as described above) and a character recognition. Remember that the character that is spelled is characterized by a row and column of the matrix, and thus a character recognition is achieved by recognizing as positive a given row and column. Hence, for a given sequence of all rows and columns matrix illumination (which corresponds to 12 intensifications), if $f(x)$ is the score of a given row or columns rc given by our classifier then $S_{rc}(k) = S_{rc}(k-1) + f(x)$

where x is the input signal associated to the highlighted row or columns rc, $f(x)$ is the score (SVMs output) associated to x and $S_{rc}(k)$ is overall score of rc at sequence k. Then at a given sequence k, the character that should be recognized is the one with maximal row and column score. In our approach, we are using an ensemble approach since we have a classifier for each word of the training session, the score $f(x)$ is actually : $f(x) = \sum_{i=1}^{n} f_i(x)$ where $f_i(x)$ is the score given by each classifier to x.

3 Results

Several experiments have been carried out in order to analyze the benefits of our approach.

All SVM classifiers that we used are linear SVM classifier. We justify this choice by stating that what makes this problem (and many other biosignal classification problems) difficult is the variability of the datasets. Thus, we believe that dealing with a linear classifier will prevent from overfitting the training data.

Our results are compared to those obtained by BCI competition winner. We have reported in the table (1), the results obtained by Bostanov [2] and a result from a similar algorithm than the one described by Meinicke et al. [4]. In this experiment, we have used the 10 channel proposed by the authors and a linear SVM (instead of a gaussian kernel SVM) for which the C regularization parameter is 0.1. We have also evaluated the gain of using a multiple classifiers system in which each classifier is dedicated to a single word learning. As expected, taking into account the variability of the P300 within subject by multiplying the number of classifiers greatly enhances the recognition rate performances. From table (1), one can see that using all channels achieves a state of the art result since all words are correctly recognized with only four sequences. This is very interesting since it confirms our hypothesis that variability of responses play a very important role in the P300 recognition, and it is necessary to cope with this variability. Again, if we compare performance of a single SVMs and a mixture of SVMs approach with the 10 channels used by Meinicke et al., we can see that our approach only slightly improves performances. This latter point highlights the need of an appropriate channel selection which, again should take into account the problem variability.

We have analyzed the channel selection procedure. What we expect from the channel selection is two-fold : a reduced number of explicative channels and an increased processing speed, which will be useful in a real-time application context. First of all, we have analyzed the variability of the channel selection procedure for a single subject using the P300 speller within different sessions and runs. Table (2) shows the 10 top ranked channels for a given session according to the above described criterion $tpr/(tpr + fpr + fnr)$. The first remark that can be drawn from this table is that only few channels (55, 56, 58, 60 which are also known as P8 ,Po7 ,POz ,Po8) are considered as equally relevant for almost all the sessions. This point focuses again the variability of the data and justifies the

	Nb. of sequences							
Algorithms	1	2	3	4	5	6	7	10
Bostanov [2]	11	5	2	1	1	0	0	0
10 preselected channels and single SVM	14	6	6	0	1	0	0	0
all channels and single SVM	14	10	9	5	5	5	1	0
10 preselected channels and 1 SVM per word	13	8	3	1	2	0	0	0
all channels and 1 SVM per word	7	4	3	0	0	0	0	0
4 relevant channels and 1 SVM per word	8	7	4	0	1	0	0	0
10 relevant channels and 1 SVM per word	8	5	5	1	0	1	0	0
26 relevant channels and 1 SVM per word	4	2	0	0	0	0	0	0
30 relevant channels and 1 SVM per word	5	3	0	0	0	0	0	0
optimal relevant channels and 1 SVM per word	4	2	1	0	0	0	0	0

Table 1. Number of mispellings in the test words with respects to the number of sequences and the algorithm.

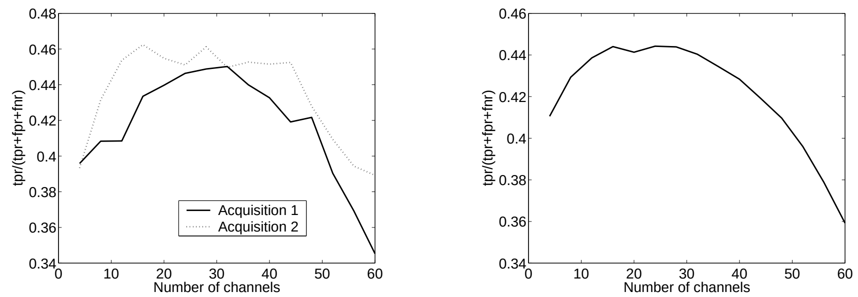

Fig. 2. Examples of channel selection criterion variation with respect to the number of channels. On the left, we have two examples for each training word and on the right we have plotted the averaged criterion value over all the training words.

need of taking into account separately the different word spelling acquisition. If we compare the channels top-ranked by our channel selection algorithm to those used by the other competitors of on this datasets [4], we can see that only few channels are in common. Figure (2) describes the variation of the channel selection criterion for different words spelling sessions. Again, we can see that the optimal number of channels differs considerably from a session to another and on average, the optimal number of channels is between 15 and 30.

Regarding the spelling performance given in table (1), two interesting points have to be highlighted. First of all, one can see that the if the number of relevant channels to be used is chosen appropriately, then all the words in the test set can be recognized correctly with only 3 sequences. However, our fully automated procedure (last line of the table) need 4 sequences for achieving equivalent performance. The second interesting point is that within 15 and 30 used channels, the overall performance is rather stable with a number of misspellings and number of needed sequences varying respectively from 6 to 10 and 2 to 3.

Sessions	10 Top Ranked Channels
1	9 15 18 36 40 55 56 59 63 64
2	18 39 53 55 56 58 59 60 61 64
3	9 18 40 48 53 55 56 58 61 64
4	10 18 33 42 46 55 56 58 60 64
5	16 22 39 40 50 56 57 60 61 62
6	2 10 36 42 48 50 55 56 58 60
7	10 17 21 25 31 43 46 51 55 56
8	10 32 41 44 49 52 55 56 60 61
9	10 23 42 48 55 56 58 60 62 63
10	4 10 17 41 42 49 55 56 58 64
11	13 34 41 48 55 56 58 60 62 64

Table 2. 10 Top Ranked channels for the differents word spelling sessions

4 Conclusions and perpectives

We have described in this work a methodology for classifying event related potentials for a Brain-Computer Interface. The strength of our approach is based on an ensemble SVMs which allows us to deal with the variability of EEG responses. In this context, we have trained a SVM classifier and selected the most relevant channels associated to each word spelling sessions signal. This method yields to state-of-the-art results since with a fully automated procedure, we were able to correctly recognized all words with only 4 sessions. Our future works now will deal with the variability of EEG responses within different subjects.

References

1. B. Blankertz, K-R Mller, G. Curio, T. Vaughan, G. Schalk, J. Wolpaw, A. Schlgl, C. Neuper, G. Pfurtscheller, T. Hinterberger, M. Schrder, and N. Birbaumer. the BCI competition 2003: Progress and perspectives in detection and discrimination of eeg single trials. *IEEE Trans. Biomed. Eng*, 51(6):1044–1051, 2004.
2. 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 Transactions on Biomedical Engineering*, 51(6):1057–1061, 2004.
3. L. Farwell and E. Donchin. Talking off the top of your head : toward a mental prosthesis utilizing event-related brain potentials. *Electroencephalography and Clinical Neurophysiology*, 70(6):510–523, 1988.
4. P. Meinicke, M. Kaper, F. Hoppe, M. Heumann, and H. Ritter. Improving transfer rates in brain computer interfacing : A case study. In S. Thrun In S. Becker and editors K. Obermayer, editors, *Advances in Neural Information Processing Systems 15*, volume 15, pages 1107–1114, 2003.
5. G. Schalk, D. McFarland, T. Hinterberger, N. Birbaumer, and J. Wolpaw. BCI2000 : a general-purpose brain-computer interface (BCI) system. *IEEE Transactions on Biomedical Engineering*, 51(6):1034–1043, 2004.
6. B. Scholkopf and A. Smola. *Leaning with Kernels*. MIT Press, 2001.