A method for predicting the success of a BCI training session based on the classification of the CSP filters itself

C. Grozea[1], S. Fazli[1], G. Nolte[1], M.Danóczy[1], F. Popescu[1]

[1]Fraunhofer FIRST (IDA), Berlin, Germany

cristian.grozea@first.fraunhofer.de

Abstract

We present an offline analysis of a large set of BCI experiments, focusing on common spatial filters and patterns (CSP). First, we show that it is possible to infer from the CSP filters whether the cross-validation error of LDA-classified EEG data preprocessed by this CSP will be high or low and predict thus the future performance of the feedback sessions following the calibration. Our test is 7 to 10 times faster to compute than the cross-validation. Second, from the CSP patterns, we calculate the corresponding source localization of the activations on the cortex. We explore the possibility of applying our method towards the improvement of calibration procedure quality and thus reduce the phenomenon of BCI illiteracy.

1 Introduction

Common Spatial Pattern (CSP) is an established method of processing raw EEG signals in order to obtain a suitable signal projection for doing BCI in a two-class (e.g left/right hand movement imagination) setup [1]. It has benefitted from many enhancements over the last decade, some of which are described in the context of the Berlin Brain Computer Interface (BBCI) in [2].

CSP is a supervised learning algorithm for two classes, which assumes that the signal measured by EEG sensors is a linear spatial mixture of (unknown) original sources. The rows of the unknown mixing matrix are called patterns, whereas the columns of the demixing matrix, which is the solution of the inverse problem, are called filters. The goal of CSP is to find spatial projections in sensor space that optimally demix the measured signal by maximizing the variance in one class while minimizing the variance in the other class, thereby achieving optimal discriminability for later classification. The filters are obtained by solving a generalized eigenvalue problem to simultaneously diagonalize the covariances of both classes.

A researcher experienced with CSPs is able to decide if a given CSP filter is good or not, by visual inspection – see Figure 1. By 'good' we mean that the subject can perform BCI with reasonably high accuracy (80% or higher). However, the difference is not always as clear as in this illustrative example. Moreover, it would be useful to understand, both from a machine learning perspective as well as a physiological perspective, what makes a subject - and his CSP - "bad".

In this paper we develop algorithms which can decide whether a given CSP filter is good, and predict from the first session if the subject will be able to perform BCI well in future sessions, albeit with a lower accuracy than the prediction of the quality of the CSPs.

By employing source localisation techniques we can further explain why it should be possible to detect in the CSP filters and patterns how discriminable the mental imagination of the subject was during the calibration phase.

To automatically learn the mapping between CSP filters and the cross-validation error on the training set, we used recorded data from a large corpus of BCI experiments, computed the CSP on

[0]This work was supported by the European Commission's Marie Curie Excellence Team grant MEXT-CT-2004-014194 "Brain2Robot"

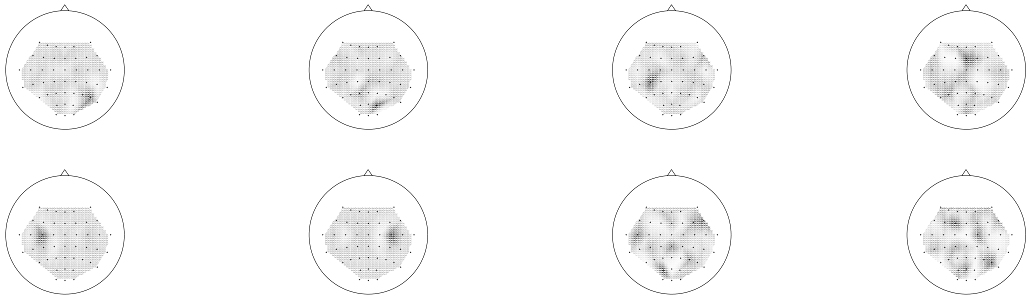


Figure 1: The 4 scalp plots on the left show a good CSP filter - cross-validation error 1%. The 4 scalp plots on the right show a bad CSP filter - cross-validation error 50%

the biggest common subset of channels and took this as the input of the mapping to be learned. As output, we took the cross-validation error, computed using information from all the channels that were recorded in the experiment.

We use feature selection methods based on the Markov blanket of the target. In Bayesian networks, the Markov blanket of a node is the set of all nodes that are needed to explain that target node. It contains all parents (the direct causes) of that node, all children (direct effects) and the other parents of its children as well.[3]

2 Methods

2.1 Data and preprocessing

The dataset used contains 148 experiments performed at the IDA group between 2001 and 2005 with 25 subjects. The paradigm was either LR (left/right) - 49 times, or LF (left/foot) - 53 times or RF (right/foot) - 46 times. Figure 2 shows descriptive statistics of the dataset.

The data has been filtered in the frequency domain by applying a wide-band band-pass filter from 5 to 30 Hz.

We processed these data using the Condor HTC system on a computing cluster. The processing we performed used the BBCI toolbox functions to first evaluate the cross-validation error on all channels available in the experiment. Then, only the channels common throughout the whole dataset were retained and the CSP was computed for each experiment. This was used as the initial input to the predictor. As output (binary valued), we took the membership or exclusion from the class of "good" experiments (i.e. less than 20% cross-validation error on the trials recorded). Here are the 45 channels available in all experiments considered: 'F5' 'F3' 'F1' 'Fz' 'F2' 'F6' 'FC5' 'FC3' 'FC1' 'FCz' 'FC2' 'FC4' 'FC6' 'T7' 'C5' 'C3' 'C1' 'Cz' 'C2' 'C4' 'C6' 'T8' 'TP7' 'CP5' 'CP3' 'CP1' 'CPz' 'CP2' 'CP4' 'CP6' 'TP8' 'P5' 'P3' 'P1' 'Pz' 'P2' 'P4' 'P6' 'P8' 'PO3' 'POz' 'PO4' 'O1' 'Oz' 'O2'.

The dataset for the learning problem we thus obtained had 148 samples each with 180 (45 channels multiplied by 4 filters) continuously valued features and a binary target.

2.2 Algorithms

Having more features than samples is always a problem, thus feature reduction and sparsification are to be considered. The best approach to feature reduction we found on this dataset was the "causal explorer" [4] able to provide us with Markov blanket estimations for a target feature. Out of all algorithms available in that toolkit, we used HITON, described in [5], very well suited for feature selection.

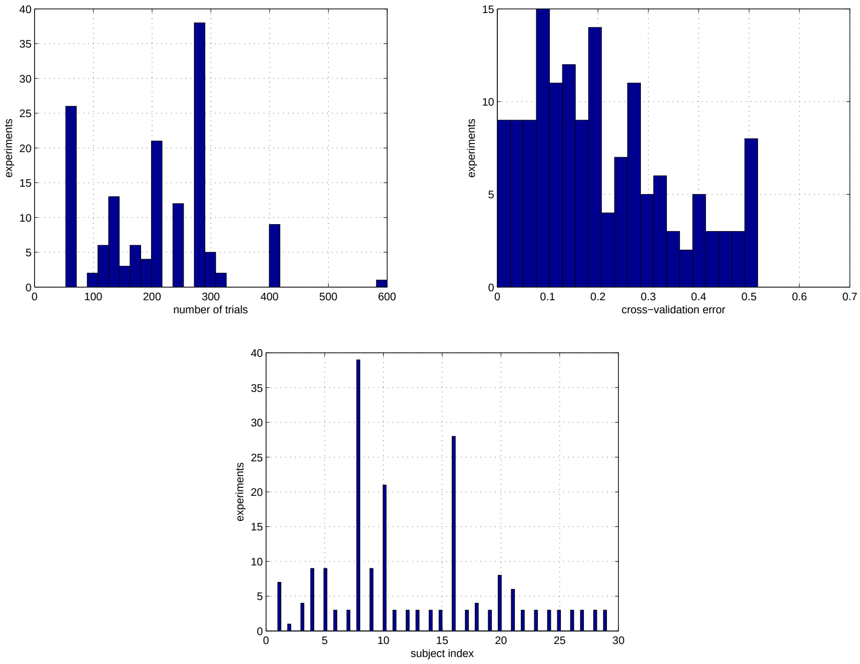


Figure 2: Left: the distribution of the number of trials. Right: The distribution of the cross-validation error. Bottom: The number of experiments per subject.

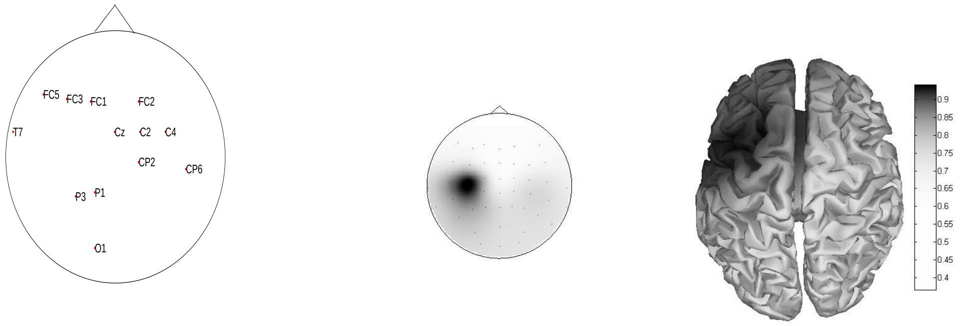
With the features thus selected, we performed a sparsifying linear norm-1 SVM training, by dividing the currently available training set into two equal sets, use one subset for training the SVM and one for testing the effect of the SVM parameters.

In order to validate the process, everything that has been described so far is wrapped into a leave-one-out cross-validation procedure, that iteratively leaves an experiment out and trains on the data derived from the retained experiments, and then tests on the left out experiment (CSP), after keeping only the features inferred as important on the data used for training.

A typical such feature set contains the following channels: 'FC5' 'FC3' 'FC1' 'FC2' 'T7' 'Cz' 'C2' 'C4' 'CP2' 'CP6' 'P3' 'P1' 'O1'. We remind the reader that we have in the dataset both experiments where the classes correspond to imaginary movements of the left and right hands and experiments where one of the classes corresponds to imaginary movements of one foot. In Figure 3(a) the approximate placement of these channels on the scalp can be seen.

ALGORITHM 1.

initialize the number of errors with 0
foreach experiment, i
    hold out the experiment i
    use HITON_MB to find the Markov blanket estimation of the target
    keep only the selected features, discarding all others
    split the set of remaining experiments in half
    foreach value of the SVM hyperparameter C in a predefined set of 20 values,
        train a norm 1 SVM on the first half
        test it on the second half
    retain the hyperparameter value that gave the best result and the corresponding linear model
    keep only the selected features in the hold-out experiment, discarding all others
    apply the linear model on the hold-out experiment
    if classification fails to give the correct class, increase the number of errors



(a) Approximate placement on the scalp of the channels most relevant to predict CSP-BCI performance.

(b) Imagined movement of a limb

Figure 3: (a) The channels selected by the causal feature selection. (b) Imagined movement of a limb. On the left, CSP pattern. On the right, corresponding source localization on the cortex obtained using the MUSIC model.

2.3 Implementation

The method has been implemented in Matlab. `CVX`, a package for specifying and solving convex programs [6, 7] has been used to implement and solve various flavors of norm-1 SVMs, seen as convex programming instances. Causal Explorer has been used to compute the Markov blanket. Processing of the dataset (on all prefixes of each experiment) took 300 cpu-hours. The cross-validation of our method took about two hours.

2.4 Inverse methods

To evaluate whether the CSP-patterns correspond to focal brain sources, which we expect to be the case for useful patterns, we apply an inverse method for each pattern. We chose the well-known MUSIC approach [8] which scans a predefined grid for dipolar sources and returns for each voxel the goodness-of-fit of the best dipole placed at that voxel. The respective scan shown over the grid, which in our case was confined to be on the cortical surface, provides a qualitative picture of areas which are most likely involved in the generation of the respective CSP pattern. We emphasize that the results are too blurred to represent true brain sources and can only be understood as a rough indication of the source origin.

The calculations were done for a three-shell realistically shaped volume conductor using a semi-analytic expansion of the electric lead field [9]. The volume conductor itself was chosen to be a publically available standard head [10], and electrode locations were adjusted to this head model.

Typical appropriate locations of the sources are obtained for the good calibration sessions – Figure 3(b), and typical mistakes for the failed calibration sessions are obtained and illustrated in Figure 4. Please note that this source localization analysis of the CSP patterns was purely qualitative, as opposed to the quantitative analysis that we did on the CSP filters.

3 Results

The cross-validation process produced 27% error. Thus we expect the method to be able to identify the experiments leading to less than 20% cross-validation error with 73% accuracy. Note that we used only 45 channels the are common to all BCI sessions in our dataset. On the other hand, the performance to be predicted corresponds to the classifier using all electrodes for which there is recorded data.

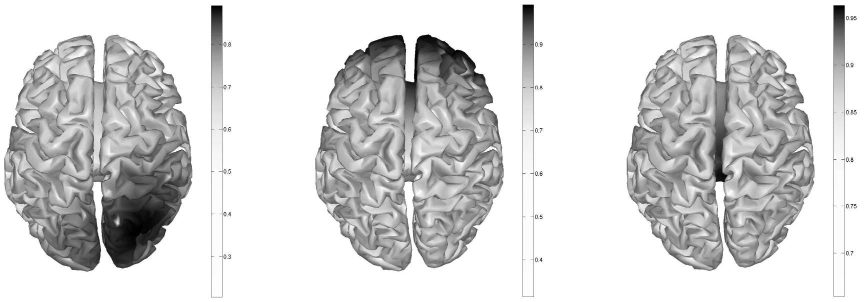


Figure 4: Source localization on the cortex obtained from the 'bad' CSP patterns that are most likely related to non-movement imagination activity, using the MUSIC model. On left: seeing/eye related. On middle: abstract cognitive processing/frontal activity. On right: no (or widely distributed) cortical activity.

The results are good, given the ambitious task of predicting with the less informed CSP computed on only 45 channels the performance of the (not yet computed) classifier on data processed with all electrods for which there is recorded data. As a further advantage, the classification algorithm presented here is on our data 7 to 10 times faster to compute than the 8-fold cross-validation – in both cases the CSP calculation was included.

4 Discussion

About half of the experiments were under the considered threshold of 20% used to label the subject performance as "good" or "bad". Therefore, for the learning problem, the dataset was fairly balanced, which makes the error measure used appropriate.

The use of the causal feature selection techniques to BCI sensitivity analysis is new to our knowledge and has produced set of channels that are relevant either for left hand, right hand, foot/feet movement imagination and for general alpha power level. This sensible choice of the channels further validates the use of this technique.

We have also run a different analysis where the input for the learning problem was the same set of CSP filters as explained before, but the output was 1 if the minimum cross-validation error amongst all known experiments of the same subject was below 20%, and 0 otherwise. In other words, we tried to predict from the CSP filter of one experiment the best performance of all, future and past, experiments of the same subject. The precision we obtained in predicting whether the subject will "ever" have a good training was lower, with a cross-validation error of 35%. What came out interesting out of this was that the set of channels usually selected was slightly different. Here is an example: 'F5' 'F3' 'Fz' 'F2' 'FC5' 'FC1' 'FC2' 'FC4' 'FC6' 'Cz' 'C4' 'T8' 'Pz' 'POz' 'O1'. The difference seems to be the higher occurrence of centro-parietal channels.

While looking at the localized sources for the CSP patterns one may easily identify the activations of cortical regions. For the low performance sessions, this enables the experimenter to pinpoint possible causes of the lack of performance in BCI for a particular subject, since he can more accurately determine the origin of activation and thus instruct the subject on how to improve his mental task performance.

5 Conclusion

A method to predict the success of a training session in which a subject's EEG is recorded on at least 45 channels while the subject performs imaginary limb movements in the Berlin BCI setup has been presented. By employing a causal feature selection technique, based on the Markov blanket of the target, we have been able to greatly reduce the number of features (channels) in the input CSP filters, and in this case proved critical to the success of the algorithm which mapped the CSP filters to sessions accuracy. As a result, we have been able to predict whether a BCI training was successful (low cross-validation error, i.e. below 20%), with 73% accuracy.

Furthermore, source localization has been employed to qualitatively inspect the CSP patterns and explain individual performances of subjects. Whereas good CSPs correspond to expected cortical sources, 'bad' ones may be due to a variety of mental task performance 'errors' which are explainable.

This justifies the claim of the experienced BCI lab researchers of being able to see the success of a training session from the initial CSP filters. Also, this opens the perspective – to be confirmed with further online studies – of being able to reduce the BCI illiteracy by instructing properly (e.g. "try to imagine a concrete movement." or "are you visualizing the scene?") the subjects who, initially, do not have a very good performance.

References

[1] J. Müller-Gerking, G. Pfurtscheller, and H. Flyvbjerg. Designing optimal spatial filters for single-trial EEG classification in a movement task. *Clinical Neurophysiology*, 110(5):787–798, 1999.

[2] Benjamin Blankertz, Ryota Tomioka, Steven Lemm, Motoaki Kawanabe, and Klaus-Robert Müller. Optimizing spatial filters for robust eeg single-trial analysis. *IEEE Signal Proc Magazine*, 25(1):41–56, 2008.

[3] J. Pearl. *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference.* Morgan Kaufmann Publishers, 1988.

[4] C.F. Aliferis, I. Tsamardinos, A. Statnikov, and L.E. Brown. Causal Explorer: A Causal Probabilistic Network Learning Toolkit for Biomedical Discovery. *International Conference on Mathematics and Engineering Techniques in Medicine and Biological Sciences (METMBS'03)*, pages 371–376, 2003.

[5] CF Aliferis, I. Tsamardinos, and A. Statnikov. HITON: A Novel Markov Blanket Algorithm for Optimal Variable Selection. *AMIA... Annual Symposium proceedings [electronic resource]*, 2003:21, 2003.

[6] M. Grant and S. Boyd. CVX: Matlab software for disciplined convex programming (web page and software), 2008.

[7] M.C. Grant and S.P. Boyd. Graph Implementations for Nonsmooth Convex Programs. *Recent Advances in Learning and Control, V. Blondel, S. Boyd, and H. Kimura, Eds. Springer*, 2008.

[8] JC Mosher, PS Lewis, RM Leahy, and R.B. TRW. Multiple dipole modeling and localization from spatio-temporal MEGdata. *Biomedical Engineering, IEEE Transactions on*, 39(6):541–557, 1992.

[9] G. Nolte and G. Dassios. Analytic expansion of the EEG lead field for realistic volume conductors. *Physics in Medicine and Biology*, 50(16):3807–3823, 2005.

[10] AC Evans, M. Kamber, DL Collins, and D. MacDonald. An MRI-based probabilistic atlas of neuroanatomy. *Magnetic Resonance Scanning and Epilepsy*, 264:263–274, 1994.