Regularised CSP for Sensor Selection in BCI

ARTICLE · JANUARY 2006

Source: OAI

REGULARISED CSP FOR SENSOR SELECTION IN BCI

J. Farquhar, N. J. Hill, T. N. Lal, B. Schölkopf

Max Planck Institute for Biological Cybernetics, Tübingen, Germany
e-mail: `jdrf@tuebingen.mpg.de`

SUMMARY: The Common Spatial Pattern (CSP) algorithm is a highly successful method for efficiently calculating spatial filters for brain signal classification. Spatial filtering can improve classification performance considerably, but demands that a large number of electrodes be mounted, which is inconvenient in day-to-day BCI usage. The CSP algorithm is also known for its tendency to overfit, i.e. to learn the noise in the training set rather than the signal. Both problems motivate an approach in which spatial filters are sparsified. We briefly sketch a reformulation of the problem which allows us to do this, using 1-norm regularisation. Focusing on the electrode selection issue, we present preliminary results on EEG data sets that suggest that effective spatial filters may be computed with as few as 10–20 electrodes, hence offering the potential to simplify the practical realisation of BCI systems significantly.

INTRODUCTION

BCI data sets typically consist of multiple time-series that are highly correlated, particularly so when measured by EEG, since EEG signals suffer from a high degree of spatial blurring. When transduction is based on a nonlinear transformation of the time-series, such as one that extracts band-power for the detection of Event-Related Desynchronisation (ERD), a *spatial filtering* preprocessing stage that performs *source separation* before nonlinear feature extraction will often improve results (see for example [1]). This can be done by Independent Component Analysis, or in some cases by the computationally much cheaper Common Spatial Pattern (CSP) method [2] and related algorithms [3, 4, 5].

One practical problem with spatial filtering is that it typically requires a large number of electrodes to be applied, whereas in everyday clinical application it is desirable to have to apply only a few. An additional problem associated with the supervised CSP algorithm in particular is its tendency to *overfit*, leading to poor generalisation (for illustration and discussion of this effect see [1, 4, 5]). This is a particular problem when the number of electrodes is large, and when the number of available trials is small.

Both problems argue for an approach which can *sparsify* the spatial filters that one computes, i.e. to force them to be based on a small number of electrodes, and to trade this characteristic off against performance on the training data. The goal is twofold: firstly to identify (based on an initial setting with a full EEG cap) which electrodes should be attached in future sessions and which can be omitted; secondly to *regularise* the computation of spatial filters, leading to improved generalisation in cases where overfitting is a problem. Regularisation by sparsification is a common approach in machine learning, and was described in the context of a CSP-like algorithm by Dornhege et al. [5]. The latter authors apply regularisation in the domain of the temporal FIR filters used in their algorithm. Here we apply the same principle to the spatial filters themselves, focusing on the question: what is the tradeoff between number of electrodes and performance, within the CSP framework?

THE RCSP ALGORITHM

CSP operates on the covariance matrix Σ_T between the d channels, computed using all trials, and the class-covariance matrix Σ_c which is computed using only trials from a given class c. Each filter is a vector $\mathbf{w}$ of length d, found by maximising the variance in one class whilst simultaneously minimising the variance in the other class(es). Equivalently, CSP can be seen as maximising the *Rayleigh quotient* which is the ratio of the variance of the filtered signal in class c to its variance overall. In addition to this criterion, we add a regularisation term incorporating a cost hyperparameter C. As is common in regularisation-by-sparsification approaches, our C is a penalty term on the L1-norm (i.e. the sum of the absolute values of the elements) of $\mathbf{w}$. Our $\mathbf{w}$ is therefore found by solving the following unconstrained optimisation problem:

$$\operatorname*{argmax}_{\mathbf{w}} \; \frac{\mathbf{w}^\top \Sigma_c \mathbf{w}}{\mathbf{w}^\top \Sigma_T \mathbf{w}} - \frac{C|\mathbf{w}|_1}{\sqrt{d}|\mathbf{w}|_2}. \tag{1}$$

The first term is the Rayleigh quotient: optimising this alone (i.e. setting $C = 0$) can be shown to be equivalent to solving the generalised eigenvalue problem $\Sigma_c \mathbf{w} = \lambda \Sigma_T \mathbf{w}$, which gives the ordinary CSP solution. We obtain a solution to (1) using the conjugate gradient method (see [6]). Once each filter is found, subsequent filters are found by *deflating* Σ_c as follows:

$$\Sigma_c \leftarrow \Sigma_c \left(I - \frac{\mathbf{w}^\top \mathbf{w} \Sigma_T}{\mathbf{w}^\top \Sigma_T \mathbf{w}} \right), \tag{2}$$

and then iterating the procedure. If C is set to 0, (1) and (2) together recover the ordinary CSP decomposition in full. With $C > 0$, we call the algorithm regularised CSP or rCSP, and its solutions are sparser, i.e. the resulting **w** vectors have fewer non-zero entries, meaning that fewer electrodes are used.

EXPERIMENTS

We tested the effect of varying C on the data from a number of two-class motor imagery experiments without feedback. 39-channel EEG was recorded from each subject as they performed 400 trials of imagined left- or right-hand movement. Regularised CSP was applied using the 7–30 Hz band, and a linear Support Vector Machine was used to classify the resulting variances of the spatially filtered signals. Offline performance was estimated using 2 repeats (with different random seeds) of 10-fold cross-validation, and the SVM's own regularisation parameter was optimised using 10-fold cross-validation nested within that (i.e. within the training subset of each outer fold).

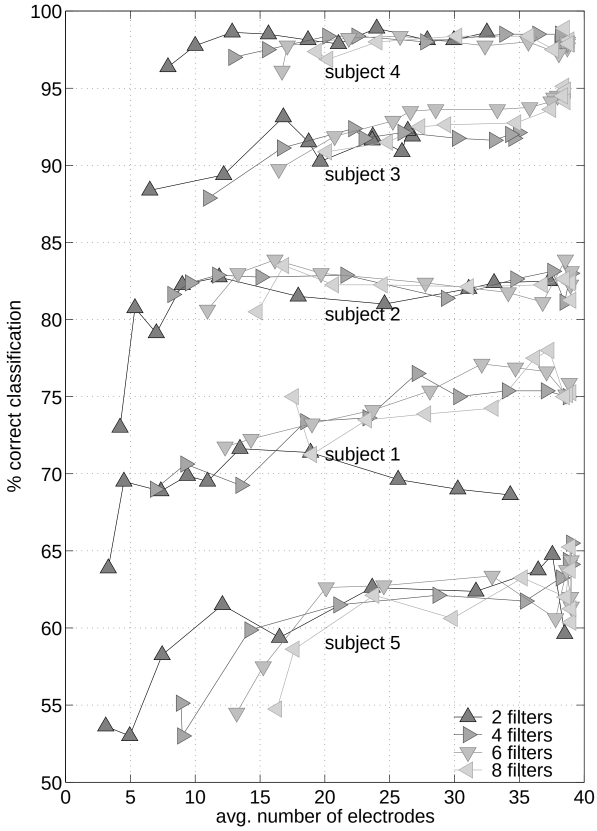

FIGURE 1: Classification accuracy for 5 subjects, as a function of number of electrodes required.

We varied the number of filters we wished to extract, $n \in \{2, 4, 6, 8\}$, and the cost parameter $C \in \{0, 0.01, 0.02, 0.05, 0.1, 0.2, 0.5, 1, 2, 5\}$. For each setting, we plot classification accuracy (averaged across the 20 outer folds) against the number of electrodes required in total to implement the n computed filters (also averaged across outer folds). We show results for 5 of the 6 subjects—the sixth subject showed similar trends, but we omit his results for readability since the curves overlap those of subjects 1 and 5.

Figure 1 gives a quantitative impression of the effect of the number of electrodes needed. For some subjects (for example, subjects 2 and 4) the curves are surprisingly flat: using only two spatial filters, one can reduce the number of electrodes to around 10 without any appreciable drop in classification accuracy. For the others, best performance was achieved with the maximum available number of electrodes, although close-to-optimal performance may still be achieved with around 20. In practice, the optimal choice of C and n should, as in most CSP implementations, be found for each subject by cross-validation.

Note that these are only preliminary results—our subjects started with a relatively small number of electrodes, 39, which meant they were widely spaced relative to those, say, a 128-electrode cap. It is possible that sparser electrode montages are effective if the candidate electrodes are more closely spaced.

CONCLUSIONS

Formulating the CSP problem as a Rayleigh quotient optimisation allows us to modify the formulation easily, with potential applications in both spatial and spatio-spectral filtering. The current modification, rCSP, allows automatic selection of a subset of electrodes during the optimisation of the spatial filter, showing that in some cases the number of electrodes can be reduced to 20 or fewer with little loss in performance.

REFERENCES

[1] Hill N. J, Lal T. N, Schröder M, Hinterberger T, Widman G, Elger C. E, Schölkopf B, and Birbaumer N. Classifying event-related desynchronization in EEG, ECoG and MEG signals. In: Dornhege G, del R. Millan J, Hinterberger T, McFarland D. J, and Müller K.-R (Eds.), *Towards Brain-Computer Interfacing.* MIT Press, Cambridge, MA (2006). In press.

[2] Koles Z. J, Lazar M. S, and Zhou S. Z. Spatial patterns underlying population differences in the background EEG. *Brain Topography* **2**(4), 275–284 (1990).

[3] Wang Y, Berg P, and Scherg M. Common spatial subspace decomposition applied to analysis of brain responses under multiple task conditions: a simulation study. *Clinical Neurophysiology* **110**(4), 604–614 (1999).

[4] Lemm S, Blankertz B, Curio G, and Müller K.-R. Spatio-spectral filters for robust classification of single trial EEG. *IEEE Transactions on Biomedical Engineering* **52**(9), 993 – 1002 (2004).

[5] Dornhege G, Blankertz B, Krauledat M, Losch F, Curio G, and Müller K.-R. Optimizing spatio-temporal filters for improving brain-computer interfacing. In: Weiss Y, Schölkopf B, and Platt J (Eds.), *Advances in Neural Information Processing Systems 18.* MIT Press, Cambridge, MA (2006).

[6] Bishop C. M. *Neural Networks for Pattern Recognition.* Oxford University Press, (1995).