Clustered Common Spatial Patterns

A. Llera, V. Gómez, H. J. Kappen
Donders Institute for Brain Cognition and Behaviour, R.U. Nijmegen, The Netherlands
Correspondence: A.Llera, Postbus 9010, 6500 GL Nijmegen, The Netherlands; E-mail: a.llera@donders.ru.nl

Abstract. We propose to cluster class-wise covariance matrices in order to identify different groups of covariances contributing to the same condition. Each cluster represents a different brain pattern associated with one class. Further, we present Clustered Common Spatial Patterns, a new algorithm that applies this technique prior to CSP. We show that CCSP can outperform CSP in a binary imagery movement task. Although in this work we consider only the case of CSP, this clustering technique could also be used to improve other feature extraction methods.

Keywords: Brain Computer Interface, Feature Extraction, Spatial Filters, Common Spatial Patterns, Clustering.

1. Introduction

The traditional training of CSP filters [Ramoser et al., 1998] uses the class spatial covariance matrices, $\tilde{C}_1$ and $\tilde{C}_2$, to construct m spatial filters (SF). We will denote this operation as $csp(\tilde{C}_1, \tilde{C}_2, m)$. In situations where more than one covariance structure contributes to a class, such as in presence of spatial shifts of the informative channels, $\tilde{C}_1$ or $\tilde{C}_2$ might over-represent the covariance structures from which more samples (trials) were observed. This can affect the generalization performance of the SF if covariances associated with infrequent trials become typical in the testing phase. Here we describe a new methodology to address this problem, Clustered Common Spatial Patterns (CCSP), and show that this technique has the potential to outperform standard CSP.

2. Clustered Common Spatial Patterns (CCSP)

CCSP performs per-class clustering of covariance matrices and combines the learned clusterings to learn SF. We illustrate the algorithm using K-means clustering [Bishop, 2007] and propose a simple way of combining the cluster centroids to construct the SF.

Consider a set of trials $\{X_{j,i} \in M_{n\times t} : j \in \{1,2\}, i \in \{1,\dots,s_j\}\}$, where j indexes the class label, s_j is the amount of trials of class j, n the number of electrodes and t the number of time samples per trial. Define two sets of $n \times n$ spatial covariance matrices $Z_j = \{C_{j,i} : i = \{1,\dots,s_j\}\}$, where $C_{j,i}$ is the covariance matrix of $X_{j,i}$. Given $(K_1, K_2) \in \mathbb{N}^2$, CCSP applies K_j-means clustering to Z_j resulting on K_j cluster centroids $\{C_j^1,\dots,C_j^{K_j}\}$. CCSP performs CSP by replacing the per-class covariance matrix by the per class mean cluster centroids. In other words, CCSP is equivalent to $csp(\frac{1}{K_1}\sum_{k=1}^{K_1} C_1^k, \frac{1}{K_2}\sum_{k=1}^{K_2} C_2^k, m)$.

3. Results

We use EEG data consisting of 70 train and 70 test trials from 8 subjects performing imagery movement collected using a 64 electrodes Biosemi system. The data were downsampled at 250 Hz, linearly detrended and bandpass-filtered in the frequency band 8-30 Hz. An automatic variance based routine was applied to remove noisy trials and channels from the train set.

In Fig. 1 we illustrate the idea of clustering the covariances. The training set covariance matrices from class 2 of one subject are projected onto their first two PCA components [Bishop, 2007] as small squares.

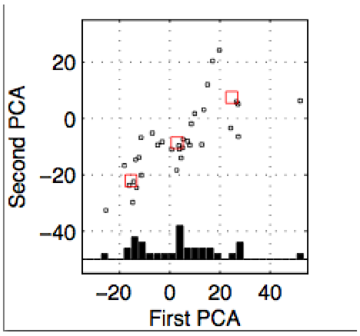

Figure1.Training set covariance matrices of class 2 projected onto their first two PCA dimensions (small squares) for one subject.The big squares represent 3 cluster centroids learned from the train covariances. The x-axis shows a histogram of the data projected onto the first PCA dimension.

Using $K_2 = 3$ we identify a clustering structure and we represent the learned cluster centroids as big squares. On the x-axis we present a histogram of the first PCA dimension of the plotted data. Note that the top cluster represents trials with a higher variance wrt to both dimensions which most probably can be identified as outliers. The lower clusters represent two groups of covariances with different number of elements that can be associated to two types of covariances corresponding to class 2.

Next we compare CCSP with CSP. For both methods we used the log-variance of the data projected onto 6 filters as features for classification. As classifier we used a SVM [Bishop, 2007]. Table 1 shows the classification results for CSP and CCSP with $m = 6$ and $K_1 = K_2 = 4$.

Subject	*1*	*2*	*3*	*4*	*5*	*6*	*7*	*8*
CSP	97.1	94.2	78.5	71.4	64.2	60	52.8	50
CCSP	100	92.8	90	77.1	54.2	67.1	77.1	58.5

Table1.Columns indicate subject number. Rows show the percentage of correctly classified test trials for CSP and CCSP with $K_1 = K_2 = 4$ respectively.

We can see that for this choice of parameters CCSP improves wrt CSP for 6 out of 8 subjects. In some cases (subjects 3 and 7) the increase in performance is notable. On the other hand, for subjects 2 and 5, CCSP performance decreases wrt CSP. We can conclude that clustering the train covariances can help us to learn better filters, and as a result CCSP can provide an efficient improvement wrt CSP.

4. Discussion

Clustering covariance matrices is not an easy task due to the high dimensionality of the space. To alleviate this problem, in this work we used the projection of the vectorized upper triangular parts of the covariance matrices onto their two first PCA dimensions as input to the clustering algorithm. Further, to avoid local minima, each clustering solution was chosen as the most likely out of 20 solutions obtained with different initializations.

There are several lines of ongoing research. First, we are studying how the quality of the clustering affects the learned filters (choice of (K_1, K_2), local minima, dimensionality reduction previous to clustering). Further, alternative ways of learning the filters after clustering are being investigated. For instance, one could learn one CSP filter for each cluster centroid and select the ones maximizing the variance between classes. Alternatively, one could learn filters using only the most dense clusters. Preliminary results suggest that these choices could improve the presented CCSP.

References

Bishop, C. M. Pattern Recognition and Machine Learning, Springer, 2007.

Ramoser H., Muller-Gerking J. and Pfurtscheller G. Optimal spatial filtering of single trial EEG during imagined hand movement. *IEEE Transactions on Rehabilitation and Engineering*, 8, 441—446, 1998.