ESANN 2012 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Bruges (Belgium), 25-27 April 2012, i6doc.com publ., ISBN 978-2-87419-049-0. Available from http://www.i6doc.com/en/livre/?GCOI=28001100967420.

Automatic selection of the number of spatial filters for motor-imagery BCI

Y. Yang[1,3], S. Chevallier[1,3], J. Wiart[2,3] and I. Bloch[1,3] *

1- Télécom ParisTech/CNRS LTCI - Signal and Image Processing Department
46 rue Barrault, Paris 75013 - France

2- Orange Labs R&D - SAFE/Wave Group
40 rue du Général Leclerc, Issy les Moulineaux 92794 - France

3- Whist Lab - Institut Télécom
46 rue Barrault, Paris 75013 - France

Abstract. Common spatial pattern (CSP) is widely used for constructing spatial filters to extract features for motor-imagery-based BCI. One main parameter in CSP-based classification is the number of spatial filters used. An automatic method relying on *Rayleigh quotient* is presented to estimate its optimal value for each subject. Based on an existing dataset, we validate the contribution of the proposed method through a study of the effect of this parameter on the classification performance. The evaluation on testing data shows that the estimated subject-specific optimal values yield better performances than the recommended value in the literature.

1 Introduction

Neuro-electrophysiologic studies reveal that both real movement and motor imagery of a specific body part induce an electro-encephalogram (EEG) rhythmic attenuation termed event-related desynchronization (ERD) in the μ (8-13Hz) and β (13-30Hz) bands over corresponding functional regions in the sensorimotor cortex [1]. Thus, the essential task of a motor-imagery based brain-computer interface (BCI) is to distinguish different spatial localizations of ERD for predicting different motor intentions. The common spatial pattern (CSP) algorithm is very effective in constructing optimal spatial filters that extract discriminative activity (i.e. ERD) and reduce feature dimensions in motor-imagery BCI [2]. This algorithm was firstly proposed for a binary discrimination and then extended to multi-class problems through various approaches (for details, see [3]).

One main parameter of CSP-based classification is the number of paired spatial filters, which determines the features used in classification and therefore affects the classification result. Most researchers choose the value of this parameter just based on their experience and often use a constant value for all subjects, which ignores the potential individual differences. Although it was mentioned in [4] that this parameter can be alternatively determined via cross validation, this work neither provided any detail nor experimental validation. Moreover, using exhaustive searching strategy to find the optimal value of this

*This work was partially supported by grants from China Scholarship Council and Orange Labs.

ESANN 2012 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Bruges (Belgium), 25-27 April 2012, i6doc.com publ., ISBN 978-2-87419-049-0. Available from http://www.i6doc.com/en/livre/?GCOI=28001100967420.

parameter in the whole range of values increases the computational time, particularly when the dimension of data is very large. Thus, the method proposed in this paper includes two steps: 1) a criterion based on *Rayleigh quotient* is applied for pre-selecting the range of this parameter, 2) an algorithm based on cross validation is then employed for more precise estimation of the optimal value of this parameter in the pre-selected range. Based on an existing dataset, we validate the importance of the estimation through studying the effect of this parameter on the classification, and then verify the effectiveness of the proposed method by comparing the classification results using the estimated optimal values with those obtained using the recommended fixed value in existing work in both binary-class and multi-class problems.

2 Methods

2.1 Pre-selection of paired spatial filters

CSP is a data-driven approach to construct spatial filters, $W = [w_1, ..., w_N]$, which decomposes the N-channel EEG $X = [x_1, x_2, ..., x_N]^T$ into N uncorrelated filtered signals $Z = [z_1, z_2, ..., z_N]^T$ through the transformation $z_j = w_j^T X$, $(j = 1, 2, ..., N)$ where w_j is a generalized eigenvector that satisfies:

$$w_j^T C^L w_j = \lambda_j^L, \quad w_j^T C^R w_j = \lambda_j^R, \quad \lambda_j^L + \lambda_j^R = 1 \tag{1}$$

where C^L, $C^R \in \mathbb{R}^{N\times N}$ are the estimated covariance matrices of two classes (i.e. 'L' and 'R') of N-channel EEG signals, respectively. As $\lambda_j^L + \lambda_j^R = 1$, w_j tends to yield a large variance of signal for one class and a small variance of signal for the other class. These contrary effects of w_j on two classes contribute to the discrimination. Usually, w_i and w_{N-i+1} according to i-th largest λ^L and λ^R, respectively, are used together as the i-th paired filters in CSP-based classification [4].

The discriminative activity S_d and common activity S_c between two classes are defined as $S_d = C^L - C^R$, $S_c = C^L + C^R$, respectively. Thus, the ratio between discriminative activity and common activity projected on the w_j spatial filter is the (*Rayleigh quotient*) $R(w_j)$ [4] and is obtained by:

$$R(w_j) = w_j^T S_d w_j / w_j^T S_c w_j = \left|\lambda_j^L - \lambda_j^R\right| / (\lambda_j^L + \lambda_j^R) = \left|2\lambda_j^L - 1\right| \tag{2}$$

For the i-th paired filters, $FD(i) = R(w_i) + R(w_{N-i})$ reflects their effectiveness in extracting the discriminative components from the original signal [5]. Usually the first m pairs of spatial filters according to the m largest $FD(i)$ are used. Too small or too large values of m will lead to poor classification performances (see Sect. 3.2), so that the optimal value of m should be estimated for each subject. A too small $FD(i)$ (typically $FD(i) < 0.1$) indicates that the i-th paired filters have a very weak ability of extracting discriminative components, and cannot improve classification results (see Sect. 3.2). As all paired filters are sorted in descending order of $FD(i)$, the $FD(i)$ values are used as a pre-selection criterion to shrink the range for seeking the optimal m value.

ESANN 2012 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Bruges (Belgium), 25-27 April 2012, i6doc.com publ., ISBN 978-2-87419-049-0. Available from http://www.i6doc.com/en/livre/?GCOI=28001100967420.

2.2 Refined estimation of the optimal number of paired filters

The optimality criterion for selecting paired spatial filters must satisfy two properties: (1) the number of paired spatial filters must be minimal, (2) it must yield the classification result that is equal or comparable to the best one, i.e. such that there is no statistical difference between them or their difference is less than a tolerance δ ($\delta = 0.015$ in this paper). Here, the classification performances are evaluated via the *kappa coefficient*: $\kappa = (P_o - P_e)/(1 - P_e)$, where P_o is the observed agreement between classifier and dataset labels, and P_e is the chance level for agreement (i.e. $P_e = 0.5$ for binary-class problems, $P_e = 0.25$ for four-class problems). Thus, a larger κ value indicates a good classification result [6].

Assuming the number of paired spatial filters with $FD(i) \geq 0.1$ is M, the optimal number of paired spatial filters is evaluated by checking each possible m value ($m \leq M$) to see whether its corresponding κ value is significantly larger than others obtained for smaller values of m. The paired difference test (paired t-test) is employed for the significance analysis [7]. In this case, if several m values yield equal or comparable classification results, the smallest one will be chosen as the most optimal. The algorithm of this procedure is described below.

Algorithm A: Selection of the optimal number of paired spatial filters

Let M denote the number of paired spatial filters with $FD(i) \geq 0.1$ and let $m \leq M$; $\kappa(m)$ is a set of κ for a given m evaluated with a 100 repetitions of 10-fold cross-validation ($\kappa(m) \in \mathbb{R}^{1\times 100}$), $\bar{\kappa}(m)$ is the mean value (over the 100 components), $t(a, b)$ represents the p-values of paired t-test between vectors a and b

$m_i \leftarrow 1$; $m_j \leftarrow 2$
while $m_j \leq M$ **do**
  if $\bar{\kappa}(m_j) > \bar{\kappa}(m_i) + \delta$ **and** $t(\kappa(m_i), \kappa(m_j)) < 0.05$ **then**
    $m_i \leftarrow m_j$
  endif
  $m_j \leftarrow m_j + 1$
endwhile
$m_{opt} \leftarrow m_i$
return the optimal parameter, m_{opt}

The optimal parameter m_{opt} is estimated offline from the training data for each subject, and then applied to the testing data or on-line applications for the same subject. This strategy can be extended to multi-class problems (see Sect. 3.4).

3 Experimental validation

3.1 Data description

The data used in this work are from BCI competition IV dataset IIa [8], which contains one training session and one testing session of 22-channel EEG data from 9 subjects who performed four classes cue-driven motor imagery (left hand,

ESANN 2012 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Bruges (Belgium), 25-27 April 2012, i6doc.com publ., ISBN 978-2-87419-049-0. Available from http://www.i6doc.com/en/livre/?GCOI=28001100967420.

right hand, both feet and tongue). Details about this dataset can be found in the associated technical document[1]. In this paper, we first use the data of left and right hands to investigate the effect of m on classification in Sect. 3.2 and then test the proposed automatic estimation strategy on binary-class (left vs. right hands) in Sect. 3.3 and multi-class data (the full dataset) in Sect. 3.4.

3.2 Sensitivity analysis of the number of paired spatial filters

A broad frequency band of 8-30Hz (μ and β bands) and the segment of 0.5-2.5s of EEG data after the cue onset were used in this study for calculating the transformation matrix W in CSP, and $FD(i)$ value for each paired spatial filters, and for training the classifier [2]. The *Fisher's linear discriminant analysis* (LDA), which is classically used with CSP, was employed here for the classification [4]. The effect of the number of spatial filters was studied on the training data using 100 repetitions of 10-fold cross-validations. The classification performances were measured by κ value. Algorithms of CSP, classifier training and evaluation (including calculating κ value) are performed with the BioSig toolbox[2].

The effects of the parameter m on the classification results and $FD(i)$ value of each paired spatial filters for all subjects are shown in Figure 1. It can be observed that (1) the performance of CSP-based classification is not proportional to m but has significant variations depending on m for all subjects; (2) the sensitivities of classification results to m are different between subjects: some (i.e. subjects 8, 9) are relatively low but most are relatively high; (3) adding the paired spatial filters with $FD(i) < 0.1$ does not improve the classification results: e.g. for subject 1, $M = 6$ and the κ value decreases if $m > 6$ is used; for subject 8, $M = 7$ and the κ value remains stable when $m > 7$. Those results prove that (1) it is critical to choose a right m value for each individual in CSP-based classification; (2) it is reasonable to estimate the optimal m in the range of $[1, M]$, where M is the number of the paired spatial filters with $FD(i) \geq 0.1$.

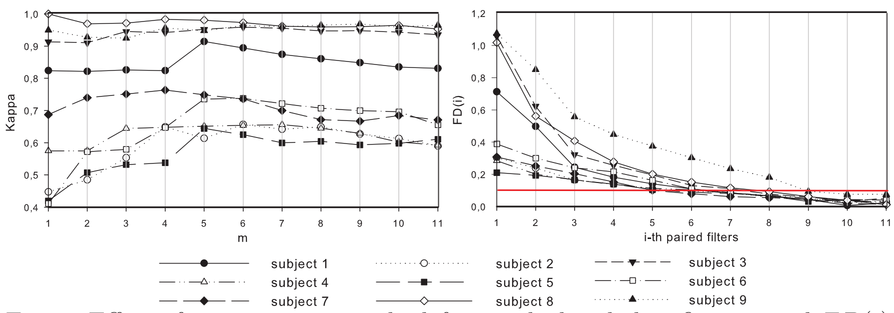

Fig. 1: Effect of parameter m on the left vs right hand classification and $FD(i)$ of each paired spatial filters for all subjects in the BCI competition IV dataset IIa. The horizontal line on the right plot indicates $FD = 0.1$.

[1] `http://bbci.de/competition/iv/desc_2a.pdf`
[2] http://biosig.sourceforge.net/

ESANN 2012 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Bruges (Belgium), 25-27 April 2012, i6doc.com publ., ISBN 978-2-87419-049-0. Available from http://www.i6doc.com/en/livre/?GCOI=28001100967420.

3.3 Comparison for binary-class discrimination

Table 1 lists the estimated m_{opt} values learned from the 100 repetitions of 10-fold cross-validations in the training data of two classes (i.e. left hand, right hand) and provides a comparison of the evaluation results on the independent testing data using m_{opt} and the classical value ($m = 3$) recommended in [2, 4]. The m_{opt} value varies for different subjects and the classification performances are better than those with the recommended value. For subjects 8 and 9, whose sensitivities to m are relatively low, one pair of filters can already yield fine performance, while others may need more pairs of filters.

	Subjects									Mean
	1	2	3	4	5	6	7	8	9	
m_{opt}	5	4	3	3	5	5	4	1	1	
κ (m_{opt})	0.75	0.22	0.96	0.40	0.11	0.35	0.70	0.94	0.86	0.59
κ ($m = 3$)	0.67	0.13	0.96	0.40	0.09	0.25	0.69	0.93	0.82	0.55

Table 1: Estimated m_{opt} values were obtained by 100 repetitions 10-fold cross-validations on the training data. The computational time was always less than 32s on a 2.66GHz PC with Matlab (2010Ra). The evaluation results were computed on the independent testing data using m_{opt} and the recommended value ($m = 3$).

3.4 Extension to multi-class problem

One Versus the Rest (OVR) CSP is a multi-class CSP approach that computes W for each class against all others and then projects the EEG signals on all the $2m \times P$ chosen spatial filters (P is the number of classes, here $P = 4$) to extract the features, and then performs a multi-class LDA classification [3]. Based on the pre-selection procedure in Sect. 2.1, each W generates a M value, thus $P \times M$ values are obtained. The largest M value (M_{max}) is chosen as the upper limit of possible m_{opt}. Then m_{opt} is estimated based on the classification results in the range of [1, M_{max}] using **Algorithm A** and then applied to the independent testing data. The comparison of results obtained with m_{opt} and with fixed recommended m is shown in Table 2 for the four-class problem of BCI competition IV dataset IIa. Using m_{opt} leads to better performances than using the fixed recommended m. As we used the broad frequency band (8-30Hz) of EEG signal in this work, it is difficult to make a comparison with the 1^{st} placed winner in BCI competition IV who extracted features from multiple narrow bands and reported the results based on searching the largest $Kappa$ over the entire time range of *the testing data* using a 2-s sliding window [9]. However, it makes more sense to compare with the 2^{nd} placed winner[3] who used the same frequency band, in order to validate the interest of using subject-specific m_{opt} with OVR CSP. The comparison showed that m_{opt} with OVR approach in CSP-classification needs less classifiers (only one multi-class LDA) and generates better mean performance.

[3] http://www.bbci.de/competition/iv/results/index.html

ESANN 2012 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Bruges (Belgium), 25-27 April 2012, i6doc.com publ., ISBN 978-2-87419-049-0. Available from http://www.i6doc.com/en/livre/?GCOI=28001100967420.

	Subjects									Mean
	1	2	3	4	5	6	7	8	9	
m_{opt}	4	3	1	2	3	3	2	4	2	
κ (m_{opt})	0.72	0.30	0.74	0.48	0.20	0.25	0.75	0.71	0.68	0.53
κ ($m = 3$)	0.69	0.30	0.71	0.47	0.20	0.25	0.74	0.71	0.50	0.51
κ (2^{nd})	0.69	0.34	0.71	0.44	0.16	0.21	0.66	0.73	0.69	0.52

Table 2: Estimated m_{opt} (obtained in less than 90s) and independent evaluation in a four-class problem using m_{opt} and fixed m, and comparison with the 2^{nd} placed winner in BCI competition IV who also used 8-30Hz data but fixed m ($m = 4$) and pair-wise approach with three LDA and one Bayesian classifiers.

4 Conclusion

The number of spatial filters used in feature extraction affects the classification results. An automatic strategy based on *Rayleigh quotient* and cross validation is proposed to estimate the subject-specific optimal m value. Experimental results show that the estimated optimal m values vary for different subjects and often yield better results than those obtained with the fixed recommended value for both binary-class and multi-class problems. The proposed strategy can be applied on the training data to estimate the optimal value of m for each subject and then use it for the long term on-line classification of the given classes for the same subject to achieve the best results.

References

[1] G. Pfurtscheller and F.H. Lopes da Silva. Event-related EEG/MEG synchronization and desynchronization: basic principles. *Clinical Neurophysiology*, 110(11):1842–1857, 1999.

[2] 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.

[3] G. Dornhege, B. Blankertz, G. Curio, and K.R. Müller. Increase information transfer rates in BCI by CSP extension to multi-class. *Advances in Neural Information Processing Systems*, 16:733–740, 2004.

[4] B. Blankertz, R. Tomioka, S. Lemm, M. Kawanabe, and K.R. Müller. Optimizing spatial filters for robust EEG single-trial analysis. *IEEE Signal Processing Magazine*, 25(1):41–56, 2008.

[5] J. Lv and M. Liu. Common spatial pattern and particle swarm optimization for channel selection in BCI. In *IEEE 3rd International Conference on Innovative Computing Information and Control. ICICIC'08.*, pages 457–457, 2008.

[6] G. Dornhege. *Toward brain-computer interfacing.* The MIT Press, 2007.

[7] R.F. Woolson, W.R. Clarke, and W.R. Clarke. *Statistical methods for the analysis of biomedical data.* Wiley New York, 1987.

[8] C. Brunner, M. Naeem, R. Leeb, B. Graimann, and G. Pfurtscheller. Spatial filtering and selection of optimized components in four class motor imagery EEG data using independent components analysis. *Pattern Recognition Letters*, 28(8):957–964, 2007.

[9] Z.Y. Chin, K.K. Ang, C. Wang, C. Guan, and H. Zhang. Multi-class filter bank common spatial pattern for four-class motor imagery BCI. In *Annual International Conference of the IEEE on Engineering in Medicine and Biology Society 2009 (EMBC 2009)*, pages 571–574, 2009.