Cogn Neurodyn (2010) 4:355–358
DOI 10.1007/s11571-010-9125-x

RESEARCH ARTICLE

Generalized optimal spatial filtering using a kernel approach with application to EEG classification

Qibin Zhao · Tomasz M. Rutkowski · Liqing Zhang · Andrzej Cichocki

Received: 28 February 2010 / Revised: 19 July 2010 / Accepted: 21 July 2010 / Published online: 3 August 2010

Abstract Common spatial patterns (CSP) has been widely used for finding the linear spatial filters which are able to extract the discriminative brain activities between two different mental tasks. However, the CSP is difficult to capture the nonlinearly clustered structure from the non-stationary EEG signals. To relax the presumption of strictly linear patterns in the CSP, in this paper, a generalized CSP (GCSP) based on generalized singular value decomposition (GSVD) and kernel method is proposed. Our method is able to find the nonlinear spatial filters which are formulated in the feature space defined by a nonlinear mapping through kernel functions. Furthermore, in order to overcome the overfitting problem, the regularized GCSP is developed by adding the regularized parameters. The experimental results demonstrate that our method is an effective nonlinear spatial filtering method.

Keywords EEG · BCI · Kernel method · CSP

Introduction

Recently, Brain Computer Interface (BCI), which aims to establish a direct communicate pathway between brain and computer, has been widely researched over the world [1]. EEG-based BCIs which rely on the motor imagery of users are of particular interest to the BCI community. Because of many facts such as low topographical resolution and high noise level, it has been a big challenge to extract effective features from EEG signals and perform classification. A lot of methods have been proposed for feature extraction and classification of EEG signals as well as parameters selection [2]. To this end, CSP method [3, 4, 5] has been proven to be very powerful in determining the spatial filters which can extract discriminative brain rhythms. However, it suffers from the limitation of assuming an absolutely linear relation between the source signals of the brain and the recorded EEG signals.

In order to make CSP applicable to nonlinearly structured data, kernel-based methods have been applied. The main idea of kernel-based methods is to map the input data to a feature space by a nonlinear mapping where inner products in the feature space can be computed by a kernel function without knowing the nonlinear mapping explicitly [6, 7, 8, 9, 10]. In [11], the nonlinear CSP was proposed, which is restricted in the equal number of trials from each class. A more flexible formulation of the kernel CSP was given in [12]. However, the complexity of the algorithm and the selection of kernel functions became new problems. More recently, a hybrid linear and kernel CSP [13] is proposed to use hybrid kernel functions which are mostly linear but can also account for small degrees of nonlinearity.

The conventional CSP algorithm can be solved by two step PCA procedure. And the most existing kernel CSP methods are based on kernel PCA [7]. A common limitation of these methods is that the class-covariance in the feature space must be nonsingular, which restricts it application to large EEG training samples. To reduce the time of calibration procedure for BCI application, the efficient training algorithm based on smaller number of

Q. Zhao (✉) · T. M. Rutkowski · A. Cichocki
Laboratory for Advanced Brain Signal Processing, Brain Science Institute, RIKEN, Saitama, Japan
e-mail: qbzhao@brain.riken.jp

L. Zhang
MOE-Microsoft Laboratory for Intelligent Computing and Intelligent Systems, Department of Computer Science and Engineering, Shanghai Jiao Tong University, Shanghai, China

EEG trials is necessary. Therefore, we propose the nonlinear CSP based on kernel methods by using the generalized singular value decomposition.

The remainder of the paper are organized as follows. In Sect. 2, we briefly introduce the classical CSP algorithm and represent it in another formula which can be solved by GSVD, and then the kernel-based GCSP algorithm is described in Sect. 3. In Sect. 4, experimental results are illustrated and discussed. Finally, the conclusions are provided in Sect. 5.

Linear CSP

Generally, a $M \times T$ matrix $\mathbf{X}_j$ represents the j-th trial of EEG signals, where M is the channels number and T is the samples per channel. The class-specific spatial covariance matrix can be obtained by

$$\mathbf{R}_c = \frac{1}{n_c} \sum_{\mathbf{X}_j \in \text{class c}} \frac{\mathbf{X}_j \mathbf{X}_j^T}{\text{trace}(\mathbf{X}_j \mathbf{X}_j^T)}, \quad c = 1, 2. \tag{1}$$

Here, n_c is the number of trials recorded under the c-th mental task. A normalization is applied to each $\mathbf{X}_j$ to make the sum of the energies of all signal channels is equal to 1. For two-class case, the total spatial covariance matrix can be formulated as

$$\mathbf{R}_t = \mathbf{R}_1 + \mathbf{R}_2. \tag{2}$$

The objective of CSP is to find the optimal spatial filter $\mathbf{w}$ which makes the average energy of one class is maximized while the other class is minimized (and vise versa). These filters are often expressed with the Rayleigh Quotient

$$\mathbf{w} = \arg\max_{\mathbf{w}} \frac{\mathbf{w}^T \mathbf{R}_c \mathbf{w}}{\mathbf{w}^T \mathbf{R}_t \mathbf{w}}. \tag{3}$$

$\mathbf{R}_c$ and $\mathbf{R}_t$ in (3) can be further formulated as

$$\mathbf{R}_c = \mathbf{H}_c \mathbf{H}_c^T \quad \text{and} \quad \mathbf{R}_t = \mathbf{H}_t \mathbf{H}_t^T, \tag{4}$$

with

$$\mathbf{H}_c = \left[\frac{\mathbf{X}_1}{\sqrt{n_c \times \text{trace}(\mathbf{X}_1 \mathbf{X}_1^T)}}, \ldots, \frac{\mathbf{X}_j}{\sqrt{n_c \times \text{trace}(\mathbf{X}_j \mathbf{X}_j^T)}} \right]_{\substack{\mathbf{X}_j \in \text{class c}, \\ 1 \le j \le n_c}} \tag{5}$$

and

$$\mathbf{H}_t = [\mathbf{H}_1, \quad \mathbf{H}_2], \tag{6}$$

where $\mathbf{H}_c \in \mathbf{R}^{M \times} p_c$, $p_c = T \times n_c$, p_c is the total number of sample points for all EEG trials belong to the c-class.

Kernel-based GCSP

In this section, we present a nonlinear extension of CSP based on kernel functions and the GSVD. By using the kernel method, we can work on the feature space through kernel functions, as long as the problem formulation depends only on the inner products between data points. Let's suppose that the input space is mapped into a Hilbert space through a nonlinear mapping function as:

$$\phi : \mathbf{R}^N \mapsto \mathcal{F}\, \mathbf{x} \mapsto \phi(\mathbf{x}). \tag{7}$$

In the nonlinear mapping space, the inner product is denoted by a kernel function,

$$< \phi(\mathbf{a}), \phi(\mathbf{b}) > \; = k(\mathbf{a}, \mathbf{b}). \tag{8}$$

We denote the nonlinear mapping of a matrix $\mathbf{X}$ which represents a single trial of EEG signals as

$$\mathbf{X} \mapsto \phi(\mathbf{X}), \quad i.e., \quad [\mathbf{x}_1 \ldots \mathbf{x}_n] \mapsto [\phi(\mathbf{x}_1) \ldots \phi(\mathbf{x}_n)]. \tag{9}$$

To apply the kernel method on CSP in the feature space instead of the original input space, the c-class covariance matrix in (1) can be expressed as

$$\mathcal{R}_c = \frac{1}{n_c} \sum_{\mathbf{X}_j \in \text{class c}} \frac{\phi(\mathbf{X}_j)\phi(\mathbf{X}_j)^T}{\text{trace}(\phi(\mathbf{X}_j)\phi(\mathbf{X}_j)^T)}, \quad c = 1, 2. \tag{10}$$

The normalized part of $\mathbf{X}_j$ in the feature space is denoted by the kernel function

$$h_{\mathbf{X}_j} = \text{trace}(\phi(\mathbf{X}_j)\phi(\mathbf{X}_j)^T) = \sum_{\mathbf{x}_k \in \mathbf{X}_j} k(\mathbf{x}_k, \mathbf{x}_k). \tag{11}$$

Similar with (4, 5, 6), we also have

$$\mathcal{R}_c = \mathcal{H}_c \mathcal{H}_c^T \quad \text{and} \quad \mathcal{R}_t = \mathcal{H}_t \mathcal{H}_t^T, \tag{12}$$

with

$$\mathcal{H}_c = \left[\frac{\phi(\mathbf{X}_1)}{\sqrt{n_c \sum_{\mathbf{x}_k \in \mathbf{X}_1} k(\mathbf{x}_k, \mathbf{x}_k)}}, \ldots, \frac{\phi(\mathbf{X}_j)}{\sqrt{n_c \sum_{\mathbf{x}_k \in \mathbf{X}_j} k(\mathbf{x}_k, \mathbf{x}_k)}} \right]_{\substack{\mathbf{X}_j \in classc, \\ 1 \le j \le n_c}}, \tag{13}$$

$$\mathcal{H}_t = [\mathcal{H}_1, \quad \mathcal{H}_2], \tag{14}$$

where $\mathcal{H}_c \in \mathcal{R}^{M \times p_c}, p_c = T \times n_c$.

Then the CSP in $\mathcal{F}$ space can also be expressed as Rayleigh Quotient

$$\varphi = \arg\max_{\varphi} \frac{\varphi^T \mathcal{R}_c \varphi}{\varphi^T \mathcal{R}_{\sqcup} \varphi}, \tag{15}$$

which can be solved by generalized eigenvalue problem

$$\mathcal{R}_c \varphi = \lambda \mathcal{R}_t \varphi. \tag{16}$$

Hence we can restrict the solution space for (16) to span $\{\phi(\mathbf{X})\}$. Let φ be represented as a linear combination of $\phi(\mathbf{x}_i)$

$$\varphi = \sum_{i=1}^{P} \alpha_i \phi(\mathbf{x}_i) = \mathbf{Y}\alpha, \tag{17}$$

where $\mathbf{Y} = [\phi(\mathbf{x}_1), ...\phi(\mathbf{x}_P)]$, $\alpha = [\alpha_1, ..., \alpha_P]^T$, and P is the total number of sample points for all EEG trials. According to (12) and (17), the equation (16) can be rewritten with left multiplied by $\mathbf{Y}^T$ as follows

$$(\mathbf{Y}^T\mathcal{H}_c)(\mathbf{Y}^T\mathcal{H}_c)^T\boldsymbol{\alpha} = \lambda(\mathbf{Y}^T\mathcal{H}_t)(\mathbf{Y}^T\mathcal{H}_t)^T\alpha. \tag{18}$$

Now α is the new eigenvector in the $\mathcal{F}$ space. The matrix $\mathcal{H}_c^T\mathbf{Y}$ can be computed and denoted by

$$\mathcal{K}_c^T = \mathcal{H}_c^T\mathbf{Y} = \begin{bmatrix} \frac{k(\mathbf{x}_1,\mathbf{x}_1)}{\sqrt{n_c h_{\mathbf{x}_1}}} & \cdots & \frac{k(\mathbf{x}_1,\mathbf{x}_P)}{\sqrt{n_c h_{\mathbf{x}_1}}} \\ \vdots & \vdots & \vdots \\ \frac{k(\mathbf{x}_{p_c},\mathbf{x}_1)}{\sqrt{n_c h_{\mathbf{x}_{p_c}}}} & & \frac{k(\mathbf{x}_{p_c},\mathbf{x}_P)}{\sqrt{n_c h_{\mathbf{x}_{p_c}}}} \end{bmatrix}, \tag{19}$$

where $h_{\mathbf{x}j} = h_{\mathbf{X}j}$ and $\mathbf{x}_j \in \mathbf{X}_j$. We define $N = n_1 + n_2$ is the total number of EEG trials for two-class case and $P = p_1 + p_2$ is the sample points for the both classes data set. Hence,

$$\mathcal{K}_t^T = \mathcal{H}_t^T\mathbf{Y} = \begin{bmatrix} \frac{k(\mathbf{x}_1,\mathbf{x}_1)}{\sqrt{n_c h_{\mathbf{x}_1}}} & \cdots & \frac{k(\mathbf{x}_1,\mathbf{x}_P)}{\sqrt{n_c h_{\mathbf{x}_1}}} \\ \vdots & \vdots & \vdots \\ \frac{k(\mathbf{x}_P,\mathbf{x}_1)}{\sqrt{n_c h_{\mathbf{x}_P}}} & & \frac{k(\mathbf{x}_P,\mathbf{x}_P)}{\sqrt{n_c h_{\mathbf{x}_P}}} \end{bmatrix}. \tag{20}$$

Then (18) is equivalent to

$$\mathcal{K}_c\mathcal{K}_c^T\alpha = \lambda\mathcal{K}_t\mathcal{K}_t^T\alpha. \tag{21}$$

In order to relieve the nonsingular restriction for $\mathcal{K}_c\mathcal{K}_c^T$ and $\mathcal{K}_t\mathcal{K}_t^T$, we apply the GSVD to the pair $(\mathcal{K}_c^T, \mathcal{K}_t^T)$. Then

$$\mathbf{U}^T\mathcal{K}_c^T\mathbf{X} = [\Gamma_c \quad 0] \quad \text{and} \quad \mathbf{V}^T\mathcal{K}_t^T\mathbf{X} = [\Gamma_t \quad 0], \tag{22}$$

where $\mathbf{U}$ and $\mathbf{V}$ are orthogonal and $\mathbf{X}$ is nonsingular, $\Gamma_c^T\Gamma_c + \Gamma_t^T\Gamma_t = \mathbf{I}$ and $\Gamma_c^T\Gamma_c, \Gamma_t^T\Gamma_t$ are diagonal matrices with nonincreasing and nondecreasing diagonal components respectively. Then the simultaneous diagonalizations of $\mathcal{K}_c\mathcal{K}_c^T$ and $\mathcal{K}_t\mathcal{K}_t^T$ can be obtained as

$$\mathbf{X}^T\mathcal{K}_c\mathcal{K}_c^T\mathbf{X} = \begin{bmatrix} \Gamma_c^T\Gamma_c & 0 \\ 0 & 0 \end{bmatrix} \quad \text{and} \quad \mathbf{X}^T\mathcal{K}_t\mathcal{K}_t^T\mathbf{X} = \begin{bmatrix} \Gamma_t^T\Gamma_t & 0 \\ 0 & 0 \end{bmatrix}. \tag{23}$$

The regularized GCSP criterion can be introduced through adding a regularization parameter η to (14)

$$\mathcal{H}_t = [\eta\mathcal{H}_1, \quad (1-\eta)\mathcal{H}_2], \quad 0 \le \eta \le 1. \tag{24}$$

Large values of η puts emphasis on the first class and vice versa. The appropriate value for η can be obtained by cross-validation.

The columns of $\mathbf{X}$ in (22) solves (21). Let α be the matrix obtained by the first r (default is 2) columns of $\mathbf{X}$. For each class of c, we can obtain the corresponding α. In the end, we combine the α for both two classes as the final spatial filters. Hence, for any new input EEG trial $\mathbf{Z}$, the nonlinear spatial filtering by j-th filters is given by

$$\begin{aligned} \varphi_j^T\phi(\mathbf{Z}) &= \boldsymbol{\alpha}_j^T\mathbf{Y}^T\phi(\mathbf{Z}) \\ &= \left[\sum_{i=1}^{P}\alpha_{ji}k(\mathbf{x}_i,\mathbf{z}_1),\ldots,\sum_{i=1}^{P}\alpha_{ji}k(\mathbf{x}_i,\mathbf{z}_T)\right]. \end{aligned} \tag{25}$$

Therefore, the energy of filtered EEG trial in (25) can be used to create the feature vectors as the input of classifier.

Experimental results

We conduct classification experiments on real EEG signals to discriminate between imagination of left hand movements (first class) and right hand movements (second class). Data are taken from two subjects S1 and S2. At second 2 of each trial a symbol indicating one specific mental task is displayed. At second 5, the screen is blank to relax the subject till the start of next trial. EEG signals are recorded from 6 electrodes with the sample rate of 256 Hz. The preprocess is performed to bandpass filter the EEG signals between 8 and 30 Hz. Each trial was split into non-overlapping time-segments of 1.5 s length prior to calculation of the spatial filters.

The Fig. 1 shows the data distribution of source signals filtered by the GCSP spatial filters. It is obvious that one class has the maximal variances while the others have the minimal variance in each direction. Therefore, the energies along the corresponding axes are enhanced by the kernel feature extractor. In this study, the Polynomial kernel function of $k(x, y) = (x, y)^d$, $d = 2$ and Gaussian kernel

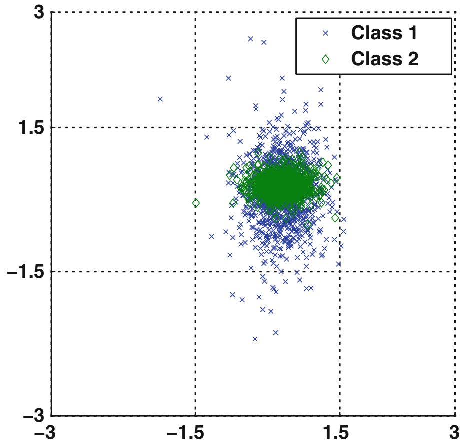

Fig. 1 The feature distribution of source signals filtered by nonlinear spatial filters obtained from GCSP algorithm

Table 1 Classification accuracies (%) of linear CSP and nolinear GCSP algorithm by cross-validation

Subjects	CSP	GCSP	
		Polynomial	Gaussian
S1 ($n = 20$)	57.2	60.6 ($\eta = 0.65$)	63.6 ($\eta = 0.65$)
S1 ($n = 40$)	71.0	75.6 ($\eta = 0.65$)	77.6 ($\eta = 0.65$)
S2 ($n = 20$)	65.7	66.2 ($\eta = 0.70$)	67.3 ($\eta = 0.70$)
S2 ($n = 40$)	83.9	83.3 ($\eta = 0.65$)	84.6 ($\eta = 0.65$)

The parameters n denotes the training number of EEG trials, and η denotes the regularization parameter

function of $k(x, y) = exp(-\|x - y\|^2/2\sigma^2)$ are applied in GCSP and linear support vector machine (SVM) classifiers are used to assess the classification performance. The optimal value for the parameter σ is determined through 5-fold cross-validation on training data. Table 1 gives the classification results of GCSP and linear CSP methods with small number of EEG training trials. In the case of subject S2, the linear CSP and GCSP resulted in roughly similar performance, while the results of subject S1 obtained by GCSP are generally better than the linear CSP. Experimental results demonstrate the superiority of nonlinear feature extractor empirically.

However, the disadvantages of the GCSP lie in the time and memory complexity of the algorithm. Consider the GSVD for the kernel matrix $\mathcal{K}_c, \mathcal{K}_t$ in Eq. (21). The kernel matrices have a relatively large dimension in typical EEG classification problems. To put it into perspective, assume 20 EEG trials for each class is given, each one with a length of 1000 samples, then $\mathcal{K}_c, \mathcal{K}_t$ will have about $20{,}000^2$ and $40{,}000^2$ elements respectively. GSVD of such two matrices imposes a high computational burden.

Conclusion

In this paper the kernel CSP approach based on GSVD is described as a nonlinear spatial filtering method. For real BCI applications, one tends to use as few training trials as necessary. The optimal kernel feature extractor proposed in this paper meets this need fairly well. One advantage of GCSP is that it can be applied regardless of singularity of the spatial covariance matrices both in the original space and in the feature space. In the future, the design of kernel functions specialized to each certain subject and new applications besides EEG signal classification would be further investigated.

Acknowledgments The work was supported in part by the Science and Technology Commission of Shanghai Municipality (Grant No. 08511501701) and the National Natural Science Foundation of China (Grant No. 60775007, 90920014).

References

1. Wolpaw JR, Birbaumer N et al (2002) Brain-computer interfaces for communication and control. Clin Neurophysiol 113:767–791
2. Long J, Li Y, Yu Z (2010) A semi-supervised support vector machine approach for parameter setting in motor imagery-based brain computer interfaces. Cogn Neurodynamics 1–10
3. Ramoser H, Muller-Gerking J, Pfurtscheller G (2000) Optimal spatial filtering of single trial EEG during imagined hand movement. IEEE Trans Rehabil Eng 8(4):441–446
4. Müller-Gerking J, Pfurtscheller G, Flyvbjerg H (1999) Designing optimal spatial filters for single-trial eeg classification in a movement tast. Clin Neurophysiol 110:787–798
5. Huang G, Liu G, Meng J, Zhang D, Zhu X (2010) Model based generalization analysis of common spatial pattern in brain computer interfaces. Cogn Neurodynamics. doi:10.1007/s11571-010-9117-x
6. Muller K, Mika S, Ratsch G, Tsuda K, Scholkopf B (2001) An introduction to kernel-based learning algorithms. IEEE Trans Neural Netw 12(2):181–201
7. Schölkopf B, Smola A, Müller K (1998) Nonlinear component analysis as a kernel eigenvalue problem. Neural Comput 10(5): 1299–1319
8. Baudat G, Anouar F (2000) Generalized discriminant analysis using a kernel approach. Neural Comput 12(10):2385–2404
9. Yang J, Frangi A, Yang J, Zhang D, Jin Z (2005) KPCA plus LDA: a complete kernel Fisher discriminant framework for feature extraction and recognition. IEEE Trans Pattern Anal Mach Intell (2005) 230–244
10. Mika S, Ratsch G, Weston J, Scholkopf B, Muller K (1999) Fisher discriminant analysis with kernels. Neural Netw Signal Process IX:41–48
11. Sun S, Zhang C (2006) An optimal kernel feature extractor and its application to EEG signal classification. Neurocomputing 69(13–15):1743–1748
12. Zhang J, Tang J, Yao L (2007) Optimizing spatial filters with kernel methods for BCI applications. In: Society of Photo-Optical Instrumentation Engineers (SPIE) conference series, vol 6790, 138
13. Nasihatkon B, Boostani R, Jahromi M (2009) An efficient hybrid linear and kernel CSP approach for EEG feature extraction. Neurocomputing