EUSIPCO 2013 1569746753

SPATIALLY SPARSE SPATIO-SPECTRAL FILTERS FOR FEATURE EXTRACTION IN BMI APPLICATIONS

Ibrahim Onaran *N. Firat Ince*

Department of Biomedical Engineering, University of Houston, TX 77204 USA

ABSTRACT

The common spatio-spectral pattern method is an extension of the traditional common spatial pattern technique that combines spectral filtering with the original spatial filtering. All recording channels are combined when extracting the variance as input features for a brain machine interface. This results in overfitting and robustness problems of the constructed system in presence of high number of channels. Here, we construct spatially sparse common spatio-spectral pattern method in which only a subset of all available channels is linearly combined when extracting the features. We utilized a modified version of the recently introduced recursive weight elimination technique to select a subset of electrodes for spatio-spectral projections. We evaluate the performance of the proposed method to distinguish between the movements of the first three fingers of the hand using electrocorticogram signals of the brain computer interfaces competition 2005. We observed that spatially sparse spatio-spectral filter outperforms both original common spatial pattern and non-sparse spatio-spectral filter and results in improved generalization in classification.

***Index Terms*—** Common spatial filters, common spatio spectral filters, brain computer interfaces, sparse projections.

1. INTRODUCTION

The functions of human hand such as grasping, lateral hip, pinch, etc. has a vital role in every aspect of the activities of the daily living. Due to interrupted neural pathways or amputation of upper limb, several people lose their hand function and have limitations in the activity of daily living. The brain controlled prosthetic hand, a neuroprosthetics, may bring many opportunities to the life of such subjects and can help them to regain their hand function.

Recent advances in electrode design and recording technology make it possible to record neural activity from many channels which can be used to control such hand prosthetics. With the increase in the number of channels, researchers can record signals from larger area of the brain or from smaller area with very high dense channel placement. This increases the computational demand of the neural decoding algorithms. To decrease the computational complexity of the processing step and increase the signal to noise ratio of the neural data, common spatial pattern (CSP) is widely used as a feature extraction technique in brain machine interface (BMI) applications. The CSP algorithm is also simple to implement for both invasive and noninvasive multichannel neural data [1, 2]. The common spatio-spectral pattern (CSSP) method of [3] is an extension of the traditional CSP that combines spectral filtering with the original spatial filtering. Although both CSP and CSSP methods are successfully used in various brain computer interface (BCI) applications they generally overfit the data when the number of training trials is limited [4, 5]. The sparseness of the spatial filter might have an important role to increase the robustness and generalization capacity of the decoder of the BMI system. Therefore, in the past years there is a growing interest on the construction of sparse spatial projections, so called sparse CSP methods, which use a limited number of channels for feature extraction [6–8]. It has been shown that such methods are superior to their non-sparse counterparts in terms of generalization capability [6–9].

In this paper, we construct spatially sparse spatio-spectral filters and study the performance of them in a BMI application. In particular, we utilized recently introduced recursive weight elimination (RWE) method [9] to select a subset of electrodes to generate spatio-spectral projections on multichannel invasive neural data. In order to investigate its generalization accuracy, we used this method as a feature extraction engine for the classification of electrocorticogram (ECoG) related to the movements of three different fingers. Such a decoder is expected to drive a robotic hand with three fingers. We show that spatial sparsification of the spatio-spectral filters increases the classification accuracy dramatically. In the next section we first refer to the details of the traditional CSP method and then show the relation between the spatio-spectral filtering with the original CSP formulation. Next, we describe the RWE method which is used to construct spatially sparse spatio-spectral projections. Finally we provide experimental results and our conclusion.

This research was supported in part by The National Science Foundation, award CBET-1067488

2. MATERIALS AND METHODS

2.1. Traditional CSP Method

In the CSP framework, the spatial filters are a weighted linear combinations of recording channels, which are tuned to produce spatial projections maximizing the variance of one class and minimizing the other. The spatial projection is computed using

$$X_{CSP} = W^T X_i$$

where the columns of W are the vectors representing each spatial projection and X_i is the multichannel ECoG data of the i^{th} trial.

The CSP algorithm creates variance imbalance to maximize the variance ratio of two different classes. The variance ratio is called Rayleigh quotient (RQ) [1] and can be written in the form of

$$\text{RQ}(w) = \frac{w^T A w}{w^T B w}$$

where A and B are the covariance matrices of two classes and w is the spatial filter that we would like to find.

Maximizing RQ can be expressed as an optimization problem as in the following equations.

$$\begin{aligned} \underset{w}{\text{maximize}} \quad & w^T A w \\ \text{subject to} \quad & w^T B w = 1. \end{aligned} \tag{1}$$

To solve the optimization problem expressed in Equation 1, we use Lagrange multipliers method [10] to obtain the equivalent problem in the form of $Aw = \lambda Bw$ which is generalized eigenvalue decomposition (GED) of the covariance matrices A and B. The solutions (w) to this problem are the joint eigenvectors of A and B and λ is the associated eigenvalue for the corresponding joint eigenvector.

2.2. Common Spatio-Spectral Filters (CSSP)

The common spatio-spectral pattern (CSSP) method of [3] is an extension of the traditional CSP that combines spectral filtering with the original spatial filtering. The multichannel data for the i^{th} trial is denoted as $X_i \in R^{C \times N}$ where C is the number of channels and N is the number of samples. First by considering only one coordinate in time, a delayed trial is obtained $X_{i,m}$, $m \in [0, \ldots, M]$, where m denotes the delay index and M is the amount of maximum delay. Then the traditional CSP algorithm is employed on the enlarged space that is obtained by concatenating the original channels and the channels which have delay indexes from 1 to M. Therefore, in this enlarged space the size of the covariance matrices are $C(M+1)$, hence, the length of the spatio-spectral filters. The projection is obtained from a particular delayed data as follows

$$Y_i = \sum_{m=0}^{M} W_m^T X_{i,m}$$

Here, W_m denotes the collection of spatial filters that is applied to signal which is delayed with the amount of m. It should be noted that the filter weights in W_m are obtained from the GED solution of the enlarged covariance matrices. Since each time sample of the spatially filtered signal Y_i is expressed as the linear combination of the original signal and its delayed versions, the whole process described so far can be interpreted as an application of the CSP algorithm to the finite impulse response (FIR) filtered multi-channel neural data, hence, it is named as common spatio-spectral pattern (CSSP). The advantage of this approach is that it extracts the spectral (frequency) information from the data that is specific to the subject and the channel. This advantage on the other hand comes with an additional (delay) parameter, M, to be tuned during the training phase. Moreover, the enlarged space, which results in a much larger covariance matrix of size $C(M+1)$ while solving the GED, causes overfitting and reduces the generalization capability of the classifier. Therefore, in practice a single delay is used which corresponds to a dimensionality of $2C$ [3]. As expected, this short filter also yields poor spectral resolution.

These drawbacks led us to find a spatially sparse solution to increase the robustness and generalization capability of the CSSP method with higher order spectral filters. In more detail, only a small number of channels with several temporal delays will be used to extract features. However, finding such a subset is infeasible and computationally complex. Consequently, a fast technique is needed to eliminate unnecessary channels. In the past few years, several greedy methods based on ℓ_0 norm such as backward elimination (BE), forward selection (FS) were proposed to find sparse spatial filters [6]. However, they have high computational complexity when the number of channels is high. Therefore, we used a modified version of recursive weight elimination method as described in [9]. It has been shown that the RWE has dramatically lower computational complexity and provides similar classification accuracy compared to other ℓ_0 norm based greedy sparse methods [9].

2.3. The Recursive Weight Elimination

The recursive weight elimination (RWE) approach is inspired from the work of Guyon et al. [11] which employs recursive feature elimination in an SVM framework. They assume that the coefficients of the weight vector are related to their contribution to the maximum margin of the SVM. Therefore they eliminate features iteratively corresponding to the minimum of the weight vector. Along the same line as [11], the RWE method assumes that the coefficients of the spatio-spectral filters are related to their contribution to the RQ. In this scheme, the RWE algorithm starts with a full size covariance matrices of the traditional CSP method. In the very first step, RWE solves general CSP problem and finds the weight vector w. The absolute value of the coefficients in w is sorted and the channel associated with the smallest value of the weight vector is eliminated. In the next step, the GED solution is obtained from the remaining channel set. This elimination procedure is iterated until the method reaches the desired cardinality (number of channels used in final projections). It is worth noting that using sparseness in CSSP framework was attempted in [12]. We have to emphasize that in [12] sparseness is sought in the length of the spectral FIR filter which was allowed to be arbitrarily long. On the contrary, we are seeking sparseness in the spatial domain.

In our exploration the goal is to analyze the effect of the sparseness in CSSP as a function of the number of channels. For this particular purpose, we modify the RWE search method such that a channel and its delayed versions are eliminated at each step. This is accomplished as follows. After solving the GED on the enlarged covariance matrices, the absolute value of the weight vector is computed. Then, for each channel the value corresponding to the maximum over its delayed weight indexes is stored. In the following step, as in the original RWE, the stored maximum values are sorted in descending order and the channel associated with minimum value is eliminated along with its delays. The procedure is iterated until desired cardinality over channels is reached. We also investigated the effect of the number of the maximum amount of delays (M) of the spectral filters in classification. In our experiments M ranged from 0 (regular sparse CSP method) to 6.

2.4. ECoG Dataset

We applied the sparse CSSP method on multiclass ECoG of BCI competitions IV. The ECoG data was recorded from three subjects during finger flexions and extensions [13] with a sampling rate of 1 kHz. The electrode grid was placed on the surface of the brain. Each electrode array contained $48(8 \times 6)$ or $64(8 \times 8)$ platinum electrodes. The finger index to be moved was shown with a cue on a computer monitor. The subjects moved one of their five fingers three to five times during the cue period. The ECoG data of each subject was sub-band filtered in the frequency range of $(40 - 200$ Hz). We used 1 s data following the movement onset in the analysis. The dataset contains around 146 trials for each subject. In this paper, we applied sparse CSSP filter to discriminate between the movements of three fingers only as they can achieve almost all the functionality that five fingers realize, and therefore widely used in robotic hand design [14] due to light weight requirements.

The signal was transformed into four virtual channels by taking first and last two eigenvectors of the GED solution. After computing the outputs of these four spatio-spectral filters, we calculated the energy of the output signal and converted it to log scale and used them as input features to lib-SVM classifier with an RBF kernel [15]. Since we are tackling a multiclass problem, we used the pairwise discrimination strategy of [2] for the three-class finger movement data. In other words, we constructed spatially sparse spatio-spectral filters tuned to contrast pairs of finger movements such as 1 vs. 2; 1 vs. 3; 2 vs. 3.

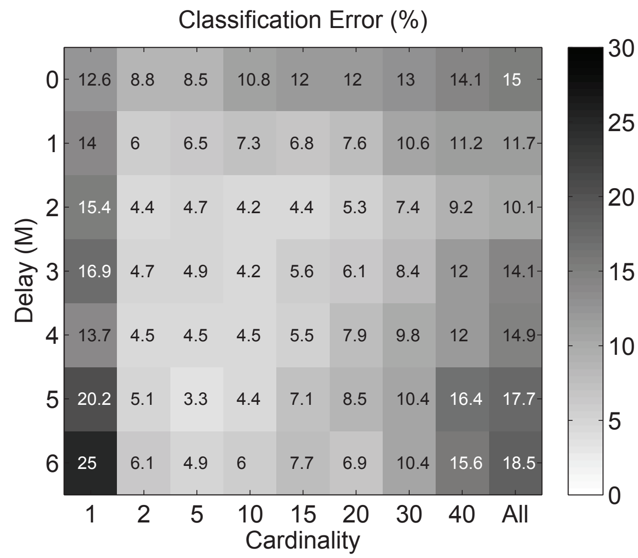

Fig. 1. The average classification errors over three subjects for different cardinalities and delays. The last column reflects the errors for full CSSP method. Note that the top right entry is the classification error of the traditional CSP method.

We studied the classification accuracy as a function of cardinality over channels and the amount of delay of the spectral filter. On the training data with the purpose of finding optimum sparseness level for the classification, we computed several sparse solutions, with decreasing cardinality. The filters were extracted with cardinalities 40, 30, 20, 15, 10, 5, 2, 1. We also varied the filter delay (M) from 0 to 6 by keeping all temporally delayed channels. Consequently, for zero delay the covariance matrix size was 64 by 64 whereas for the filter with 6 delays, the covariance matrix size had a dimensionality of 448 by 448 (64 channels x 7 taps = 448). A five fold cross validation was applied to the dataset to study the generalization accuracy. The classification accuracies are averaged over subjects.

3. RESULTS

In Fig. 1 the classification errors for different cardinality and filter lengths are given. In Fig. 1, the last error points correspond to the errors that are obtained by the original CSSP method working on all channels. The first row reflects the results that are obtained only from the spatial filters. Consequently, the error rate on the last column of the first row corresponds to the traditional CSP method. We observed that the minimum classification error (3.3%) was obtained with a cardinality of 5 and delay of 5. The standard CSP method

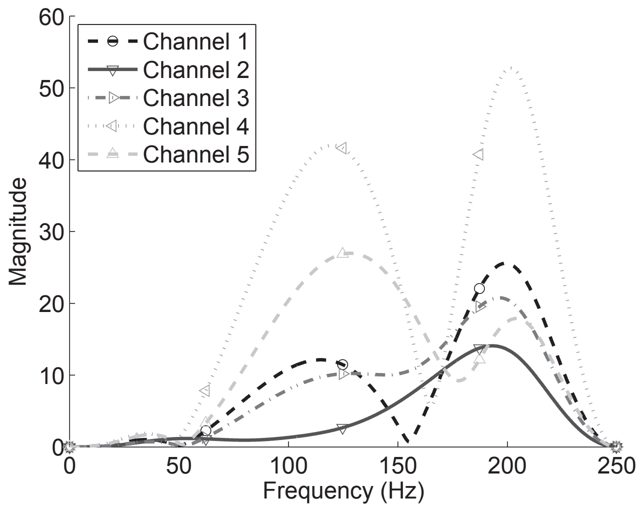

Fig. 2. 6 taps (M=5) sparse CSSP filter for cardinality 5. The effect of 40-200 Hz band filtering also included.

resulted in 15% error while distinguishing between the movements of the 3 fingers. For the CSSP method the error rate decreases with delays 1 and 2 and starts increase with delay 3. Using sparseness in the spatial domain improved the CSSP classification performance considerably, especially when the spatial filters are highly to moderately sparse. The best performances are obtained by spatial filters that are considerably sparse. Nevertheless, the spatially sparse CSSP method provided lower classification error rates for all delays. The standard CSP and CSSP methods had a minimum error rate of 15% and 10.1% (with a delay 2) respectively.

The frequency response of a filter obtained by spatially sparse CSSP method at cardinality 5 and amount of delay 5 is shown in Fig. 2. Each line represents the frequency response of a different channel. The effect of 40-200 Hz band filtering also included in the plot. The spectral filters that are formed by sparse CSSP are simply band-pass filters. The filters suppress the signal around 50 Hz and 160 Hz and tend to select higher bands in 40-200 Hz range. The suppression and amplification of the CSSP method for the indicated frequency bands may have an important role to improve the classification accuracies compared to traditional CSP method. We note that the CSSP method spectrally filters the data for each channel and subject independently, so it creates subject and channel specific FIR filters. Consequently, spatially sparsifying the CSSP solution also solves the overfitting problem.

4. CONCLUSION

The CSP and CSSP methods suffer from overfitting in the presence of high density recordings with small amount of training data. Since each delay expands the dimensionality of the covariance matrices used in GED, the CSSP method tends to overfit the training data more than the CSP method. Here, we constructed spatially sparse CSSP method in which only a subset of all available channels is linearly combined when extracting the features. For this particular purpose, we utilized a modified version of the recently introduced recursive weight elimination (RWE) technique to select a subset of spatio-spectral projections. We evaluate the performance of the proposed method to distinguish the movement of the first three fingers of the hand using electrocorticogram (ECoG) signals of the BCI competition 2005. We observe that spatially sparse spatio-spectral filters are superior to both original CSP and CSSP.

5. REFERENCES

[1] 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*, vol. 25, no. 1, pp. 41–56, 2008.

[2] Nuri Fırat İnce, Rahul Gupta, Sami Arica, Ahmed H. Tewfik, James Ashe, and Giuseppe Pellizzer, "High Accuracy Decoding of Movement Target Direction in Non-Human Primates Based on Common Spatial Patterns of Local Field Potentials," *PLoS ONE*, vol. 5, no. 12, pp. e14384, December 2010.

[3] S. Lemm, B. Blankertz, G. Curio, and K.-R. Müller, "Spatio-Spectral Filters for Improving the Classification of Single Trial EEG," *IEEE Transactions on Biomedical Engineering*, vol. 52, no. 9, pp. 1541–1548, September 2005.

[4] J. Farquhar, N. J. Hill, T. N. Lal, and B. Schlkopf, "Regularised CSP for Sensor Selection in BCI," in *Proceedings of the 3rd International Brain-Computer Interface Workshop and Training Course*, 2006.

[5] B. Reuderink and M. Poel, "Robustness of the Common Spatial Patterns Algorithm in the BCI-Pipeline," Tech. Rep. TR-CTI, Univ. of Twente, Enschede, July 2008.

[6] Fikri Göksu, Nuri Fırat İnce, and A. H. Tewfik, "Sparse Common Spatial Patterns in Brain Computer Interface Applications," in *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, May 2011, pp. 533–536.

[7] İbrahim Onaran, Nuri Fırat İnce, and A. Enis Çetin, "Sparse Spatial Filter via a Novel Objective Function Minimization with Smooth ℓ_1 Regularization," *Biomedical Signal Processing and Control*, vol. 8, no. 3, pp. 282–288, 2013.

[8] M. Arvaneh, Cuntai Guan, Kai Keng Ang, and Chai Quek, "Optimizing the Channel Selection and Clas-

sification Accuracy in EEG-Based BCI," *IEEE Transactions on Biomedical Engineering*, vol. 58, no. 6, pp. 1865–1873, June 2011.

[9] Fikri Göksu, Fırat İnce, and İbrahim Onaran, "Sparse Common Spatial Patterns with Recursive Weight Elimination," in *Conference Record of the Forty Fifth Asilomar Conference on Signals, Systems and Computers (ASILOMAR)*, November 2011, pp. 117–121.

[10] M.D. Weir, J.R. Hass, and F.R. Giordano, *Thomas' Calculus*, Thomas Series. Pearson Addison Wesley, 11^{th} edition, 2005.

[11] Isabelle Guyon, Jason Weston, Stephen Barnhill, and Vladimir Vapnik, "Gene Selection for Cancer Classification Using Support Vector Machines," *Machine Learning*, vol. 46, no. 1-3, pp. 389–422, 2002.

[12] G. Dornhege, B. Blankertz, M. Krauledat, F. Losch, G. Curio, and K.-R. Müller, "Combined Optimization of Spatial and Temporal Filters for Improving Brain-Computer Interfacing," *IEEE Transactions on Biomedical Engineering*, vol. 53, no. 11, pp. 2274–2281, November 2006.

[13] Kai J. Miller and G. Schalk, "Prediction of Finger Flexion 4th Brain-Computer Interface Data Competition," 2008, `http://www.bbci.de/competition/iv/desc_4.pdf`.

[14] Sung-yoon Jung, Sung kyun Kang, Myoung-Jun Lee, and Inhyuk Moon, "Design of Robotic Hand with Tendon-Driven Three Fingers," in *International Conference on Control, Automation and Systems, 2007. ICCAS '07.*, 2007, pp. 83–86.

[15] Chih-Chung Chang and Chih-Jen Lin, "A Library for Support Vector Machines," 2001, `http://www.csie.ntu.edu.tw/~cjlin/libsvm`.