

2014 AASRI Conference on Circuit and Signal Processing (CSP 2014)

Optimal Channel Selection for Robust EEG Single-trial Analysis

Kusuma Mohanchandra^{a,*}, Snehanshu Saha^b^aComputer Science & Engineering and MIRC, DSCE, Bangalore - 560078, India^bComputer Science & Engineering and CBIMMC, PESIT South, Bangalore - 560100, India* Tel.: +0-9886520551, kusuma-cs@dayanandasagar.edu

Abstract

EEG is an extensively used powerful tool for brain computer interface due to its good temporal resolution and ease of use. The signals captured by multichannel EEG recordings contribute to huge data and often lead to the high computational burden on the computer. An optimal number of electrodes that capture brain signals relevant to the purpose can be used, excluding the redundant and non-contributing electrodes. In this study, we propose an optimization technique on common spatial pattern for channel selection. The implementation of optimization is done as a sequential quadratic programming problem of fast convergence. Extensive experimentation is done to show that the proposed method induces large variance between two tasks of brain action related to sub vocalized speech.

© 2014 The Authors. Published by Elsevier B. V. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/3.0/>).

Peer-review under responsibility of Scientific Committee of American Applied Science Research Institute

Keywords: BC I; EEG; CSP; Characteristic value; Characteristic vector; Diagonalization;

1. Introduction

Brain-Computer Interface (BCI) is a field of dynamic scientific research which purports to convey human intentions to the external world at once from their thoughts, thus enhancing cognitive capabilities. The tremendous potential of BCI has a wide range of applications across various fields, both medical and non-medical. The BCI is used by population (Birbaumer et al., 1999) deprived of expressing through speech. It has proved useful for helping paralyzed or 'Locked-In' patients to develop ways of communication with the external world.

Since electroencephalography (EEG) is non-invasive, portable and has good temporal resolution (Wester, 2006), it is the most preferred technology for BCI. EEG is a recording of brain's electrical activity along the scalp. These voltages are the result of electrochemical signaling between neurons in the brain. The measured voltages are in microvolts and are typically recorded at multiple sites simultaneously. The EEG is recorded using an electrode cap, which has 32, 64, 128 or 256 electrodes and normally lasts 20–30 minutes. EEG signals from multiple electrodes have to be considered to achieve good performance. The massive multidimensional time series data requires large computational power for processing the information. So the need for dimensionality reduction of the data is crucial in reducing the storage and computational complexity. Relevant brain signals are obtained from an optimal number of electrodes placed on the scalps. Leaving out the redundant and noncontributing electrodes reduces the number of electrodes considerably. But, due to the volume conduction of the brain signals through the skull, scalp and other strata of the brain, the spatial resolution of EEG signals is very poor. A spatial filter such as a common spatial pattern (CSP) can be used to focus on this issue. An important attribute of the spatial filter is to minimize the number of channels on the scalp (Bashashati et al., 2007) and at the same time preserve every information necessary for classification.

In this study, selection of EEG channels that get both the spatial and temporal forms of EEG during sub-vocalization of two different words is computed using CSP. Neuroscience research has shown that mental rehearsal activates the frontal cortex as well as Broca's and Wernicke's (Blank et al., 2002) areas. Using the CSP approach, we obtain the optimal choice of the EEG channels and features that are suitable for discriminating the two cognitive (mental rehearsal of words) tasks. The classification of the two sub vocalized words are done initially using 64 electrodes and later for the optimal number of electrodes chosen. In (Blankertz et al., 2008) the EEG channel selection method is formulated as an optimization problem of CSP. Let $E_n \in R^{ch \times t}$ denote the matrix from preprocessed EEG signals of a single trial. The ch and t represent the number of channels and number of samples per channel respectively.

$$E_n \in R^{ch \times t} \xrightarrow{\text{transforms to}} X_n \in R^d \quad (1)$$

A set of feature vectors can be expressed as a matrix,

$$X \in R^{d \times n} \quad (2)$$

The CSP algorithm projects E_n to a spatial filter S such that,

$$S = W^T E, \text{ where } W \in R^{d \times ch} \text{ and } S \in R^{d \times t} \quad (3)$$

The CSP projection matrix W expressed in quadratic form is a candidate for quadratic optimization. The selection of appropriate weights using the characteristic values and characteristic vectors can be articulated as shown below

$$\max W^T \Sigma W, \text{ such that } W^T (\Sigma_1 + \Sigma_2) W = I \quad (4)$$

where I is an Identity matrix, Σ_1 and Σ_2 are the normalized covariance matrices obtained from equation (5). T denotes the transpose operator and $\text{trace}(A)$ computes the sum of the diagonal elements of the matrix A . The averaged normalized covariance matrix for each class are computed by averaging over all trials of each class separately.

$$\Sigma_1 = \underset{E_n \in C_1}{Exp} \left(\frac{E_n E_n^T}{\text{trace}(E_n E_n^T)} \right) \quad \text{and} \quad \Sigma_2 = \underset{E_n \in C_2}{Exp} \left(\frac{E_n E_n^T}{\text{trace}(E_n E_n^T)} \right) \quad (5)$$

The composite covariance matrix is the summation of the two resultant covariance matrices. This leads to

$$\Sigma_c = \Sigma_1 + \Sigma_2 = UDU^T \quad (6)$$

U is a matrix of orthonormal characteristic vectors, such that $UU^T = I$, and D is the diagonal matrix of characteristic values. The whitening transformation matrix is defined as

$$P = D^{-1/2} U^T \quad (7)$$

which transforms the average covariance matrix for fast convergence. The whitening transformation is applied to the original covariance matrices of the two classes as shown in equation (8). By simultaneous diagonalization, $\hat{\Sigma}_1$ and $\hat{\Sigma}_2$ achieve common characteristic vectors P and the sum of the corresponding characteristic values $\hat{\Sigma}_1$ and $\hat{\Sigma}_2$ will sum to the identity matrix.

$$\hat{\Sigma}_1 = P \Sigma_1 P^T \quad \text{and} \quad \hat{\Sigma}_2 = P \Sigma_2 P^T \quad \text{such that} \quad \hat{\Sigma}_1 + \hat{\Sigma}_2 = I \quad (8)$$

Any orthonormal matrix V obey the following,

$$V^T (\hat{\Sigma}_1 + \hat{\Sigma}_2) V = I \quad (9)$$

$$\text{as long as } \hat{\Sigma}_1 + \hat{\Sigma}_2 = I. \quad \text{Finally, } \hat{\Sigma}_1 = V \Lambda_1 V^T, \quad \hat{\Sigma}_2 = V \Lambda_2 V^T \quad (10)$$

The verification is shown later in our work. V is a set of characteristic vectors and Λ_1 and Λ_2 are diagonal matrices of characteristic values ordered in descending order. Lastly, the projection matrix W is computed as shown in equation (11), where the rows of the projection matrix W are the spatial filters and the columns of W^{-1} are the CSP.

$$W = P^T V \quad (11)$$

The original EEG signal transformed into uncorrelated, linearly independent components with the projection matrix is shown in equation (3).

2. Related work

Various relevant channel selection methods proposed in the literature are reviewed. In (Lei et al., 2009), the CSP filter bank is constructed through the division of recording channels and common spatial pattern ensemble (CSPE) classifier is applied to improve classifier performance. A filter approach using mutual information (MI) maximization for EEG channel selection has been proposed in (Lan et al., 2005), to rank the EEG channels so that the MI between the selected sensors and class labels is maximized. In (Ramoser et al., 2000), the EEG trials are weighted for each subject, where very low weights indicate those trials as artifacts or

outliers. In order to achieve accurate classification, it is necessary to find the frequency distribution of the brain signals associated with BCI tasks, so in (Higashi et al., 2013) a discriminating filter bank CSP (DFBCSP) with finite impulse response filters and the associated spatial weights are used to extract discriminative features. A sparse common spatial pattern (SCSP) algorithm is provided (Arvaneh et al., 2011) for EEG channel selection. The SCSP algorithm is formulated as an optimization problem that induces sparsity in the common spatial filters to select the least number of channels within a constraint of classification accuracy.

Despite various studies, finding the optimal number and positions of EEG electrodes in a BCI application is a rigid problem. Each individual's brain activity is different for the same task done, so huge data are needed to classify the task done. With proper training given to the subjects during data acquisition, more similar signal pattern can be produced, thereby reducing the number of channels. In this paper, we focus on selecting the optimal number of electrodes for processing the EEG signals recorded during subvocalization of two words. CSP is used for channel selection and Support Vector Machine for classification of the task performed.

3. Our Work

3.1. CSP algorithm as a generalized characteristic value problem

The common spatial pattern is used as spatial filters in BCI. CSP filters are good at detecting the range of sensorimotor rhythms and discerning the mental states. The spatial filters are designed to maximize the variance of signals of one class and at the same time reduce the variance of signals of the other class. The main idea is to use a linear transform to project the multichannel EEG data into low dimensional spatial subspace. Each row of the projection matrix consists of weights for the channels. CSP is based on simultaneous diagonalization of the covariance matrices of two classes. The diagonalization process leads to $\hat{\Sigma}_1 + \hat{\Sigma}_2 = I$. The whitening transformation matrix shown in equation (7) is used to equalize the variance in characteristic spaces. The process is shown beneath

$$P = D^{-1/2} U^T \Rightarrow D^{-1/2} P = U^T \Rightarrow P^T (D^{-1/2})^T = U \Rightarrow P^T D^{-1/2} = U \quad (12)$$

$$\begin{aligned} \text{Therefore, } \hat{\Sigma}_1 &= P \Sigma_1 P^T = D^{-1/2} U^T \Sigma_1 U D^{-1/2} \text{ and } \hat{\Sigma}_2 = P \Sigma_2 P^T = D^{-1/2} U^T \Sigma_2 U D^{-1/2} \\ \hat{\Sigma}_1 + \hat{\Sigma}_2 &= D^{-1/2} U^T \Sigma_1 U D^{-1/2} + D^{-1/2} U^T \Sigma_2 U D^{-1/2} = D^{-1/2} (U^T (\Sigma_1 + \Sigma_2) U) D^{-1/2} \\ &= D^{-1/2} (U^T D U) D^{-1/2} = D^{-1/2} \underbrace{U^T D^{1/2} D^{1/2} U}_{*} D^{-1/2} \end{aligned}$$

* $U^T D^{1/2}$ are interchangeable, since $D^{1/2}$ is diagonal, U is orthonormal and hence they commute. Since, U is orthonormal $U^T U = I$.

$$= \underbrace{D^{-1/2} D^{1/2}}_I \underbrace{U^T U}_I \underbrace{D^{1/2} D^{-1/2}}_I \quad \text{Therefore, } \hat{\Sigma}_1 + \hat{\Sigma}_2 = I \quad (13)$$

The resultant feature vectors of S are optimized for discriminating two classes of input signals.

3.2. Formulation of the optimization problem

CSP algorithm is formulated as an optimization problem using sequential quadratic programming. The local minimum is derived from the theory of the Hessian matrix, which is formed from the covariance matrix. The regularization term is added to the optimization problem to induce sparsity in the CSP. An $l_1 / (l_1 + lp)$ norm is used to induce sparsity in the optimization problem.

$$\min_{w_i} (1 - \alpha) \left\{ \sum_{i=m+1}^{2m} w_i \Sigma_1 w_i^T + \sum_{i=1}^m w_i \Sigma_2 w_i^T \right\} + \underbrace{\alpha \sum_{i=1}^{2m} \frac{\|w_i\|_1}{\|w_i\|_1 + \|w_i\|_p}}_{\text{Regularization term}}$$

Subject to : $w_i(\Sigma_1 + \Sigma_2) w_i^T = 1$, $i = \{1, 2, \dots, 2m\}$ and $p = 0.5$

$$w_i(\Sigma_1 + \Sigma_2) w_j^T = 0, \quad i, j = \{1, 2, \dots, 2m\} \quad i \neq j, \quad \text{and} \quad 0 \leq \alpha \leq 1 \quad (14)$$

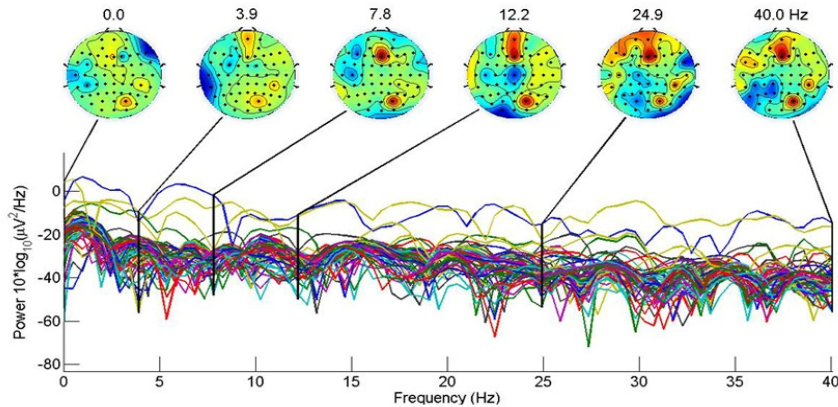
$$\text{The } l_p \text{ norm of } x, 0 < k \text{ is defined as } l_k = \|x\|_k = \left(\sum_{i=1}^n |x_i|^k \right)^{1/k} \quad (15)$$

Σ_1 and Σ_2 are the covariance matrices of class1 and class2. The weights w_i are computed using the constraints given in equation (14), such that $w_i \in \mathbb{R}^{1 \times n}$, $i = \{1, 2, \dots, 2m\}$, w_i represents the first and last m rows of CSP projection matrix, denoting the m largest characteristic values of Λ_1 and m largest characteristic values of Λ_2 . Optimal α is found from the local/global minima of the convex optimization problem.

4. Experiments

4.1. Data Acquisition

In this study the EEG data of three normal subjects during sub vocalization of two words “Food” (class1) and “Water” (class2) is used. The data were recorded using 64 channels placed in accordance to international 10-20 system. During the recording session, the subjects were instructed with visual cues to repeat one of the two words in their mind without moving their jaw or any speech articulators. The words /food/ and /water/ were specifically chosen due to their dissimilar muscle activations during real speech production. The EEG data on each subject consisted of 50 single trial for each task. The “20 single trial” data are used for training the classifier and the remaining “30 single trial” is used for testing the data.



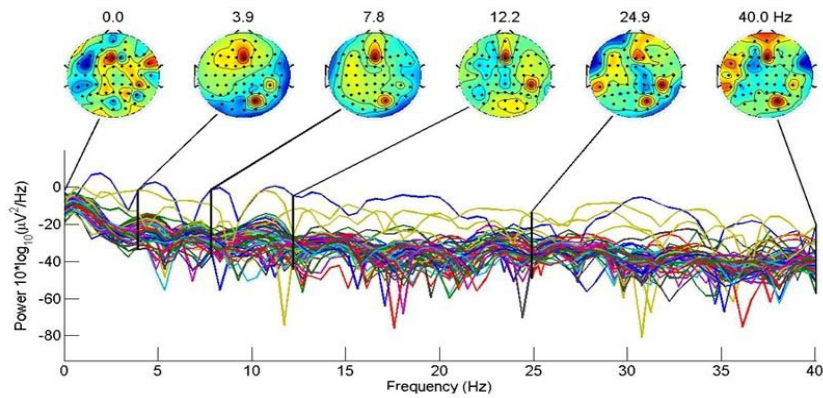


Fig. 1: The frequency components of average over 25 epochs/task for subject1 (a) task /food/, (b) task /water/.

4.2. Data Processing

Each trial of EEG data extracted from 64 channels is 1.5 seconds long, sampled at 250 Hz. The signals are band pass filtered between 0.5Hz and 40 Hz using an equiripple band pass filter. The equiripple band pass filter outperforms other filter in minimizing the maximum ripple in the passband and stopband. The preprocessed EEG signals from the training set are used for optimal channel selection. The optimal channels are computed from the equation (14). Using $m = \{1, 2, 3\}$ the optimal channels are selected from the first and last sets of sparse spatial filters. Consequently, the classification accuracy was evaluated using the test data on the SVM classifier. The SVM classifier is trained using the feature vectors of the training data set and then evaluated using the test data set.

5. Results

To visualize the event related pattern in time series, the EEG signal averaged over all trials of individual task across all subjects is used. The speech related potentials are found prominently in the frontal and anterior frontal cortex. The activity on the electrodes FP1, FPz, FP2, AF3 and AF4 of the international 10-20 electrode system are found to have more activity during the speech vocalization. It is observed that the alpha (8-15Hz) and beta (16-31Hz) frequency components contribute maximum to the activity. The scalp maps in figure (2) show the spatial patterns, before and after applying the spatial filters. The spatial filters are designed to maximize the variance between the two classes. The CSP maximizes the variance of signals belonging to class1, while minimizing the variance of the signal of class2. The figure 2(b) shows a strong discrimination in the pattern between the two classes of tasks.

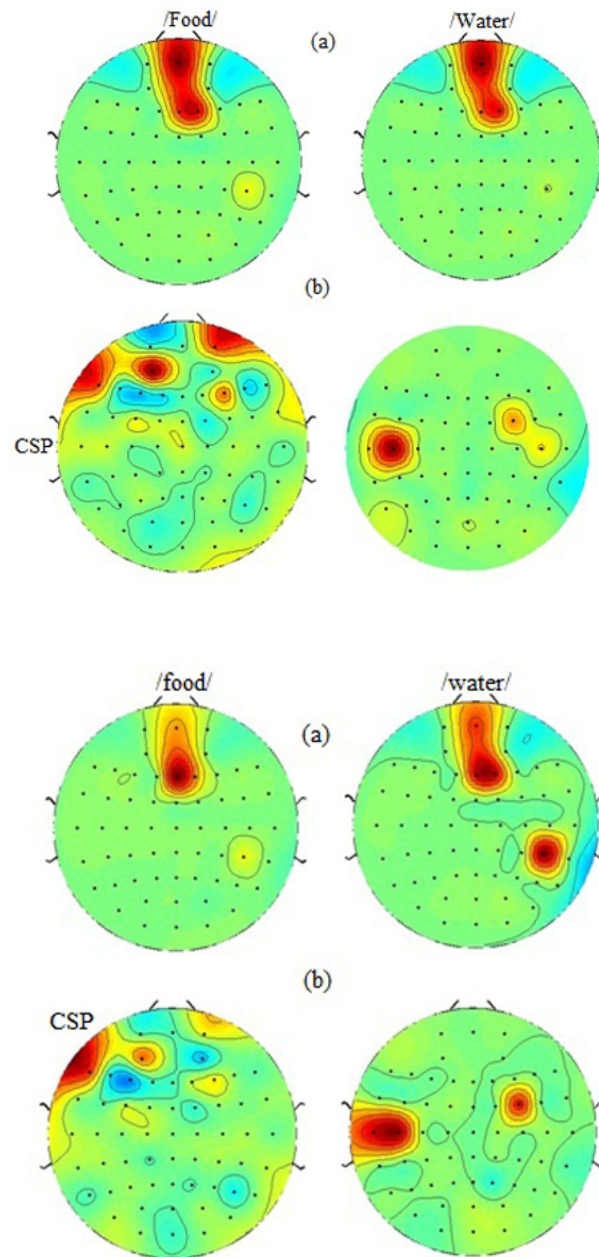


Fig. 2: The scalp maps showing the spatial patterns of the two classes (a) before (b) after applying the spatial filters.

6. Conclusion

The biggest challenges of EEG are its non-stationarity, low signal to noise ratio and a wide variety of noise sources. The high dimension of multichannel EEG demands large computational power and robust algorithms to give accurate results. In this work, we look into a fast converging channel reduction method based on common spatial patterns. The channel reduction is achieved as an optimization problem using sequential quadratic programming. A visualization of the spatial filters obtained from the proposed algorithm shows a large variance between the two classes. The EEG channels with large variances are considered and channels with low variances are discarded and an optimum number of channels are used for classification of the two tasks.

References

- [1] Birbaumer, N., Ghanayim, N., Hinterberger, T., Iversen, I., Kotchoubey, B., Kübler, A., ... & Flor, H. (1999). A spelling device for the paralysed. *Nature*, 398(6725), 297-298.
- [2] Wester, M. Unspoken speech - speech recognition based on electroencephalography. Master's thesis, Institut für Theoretische Informatik Universität Karlsruhe (TH), Karlsruhe, Germany, (2006).
- [3] Bashashati, A., Fatourehchi, M., Ward, R. K., & Birch, G. E. (2007). A survey of signal processing algorithms in brain-computer interfaces based on electrical brain signals. *Journal of Neural engineering*, 4(2), R32.
- [4] Blank, S. C., Scott, S. K., Murphy, K., Warburton, E., & Wise, R. J. (2002). Speech production: Wernicke, Broca and beyond. *Brain*, 125(8), 1829-1838.
- [5] Blankertz, B., Tomioka, R., Lemm, S., Kawanabe, M., & Muller, K. R. (2008). Optimizing spatial filters for robust EEG single-trial analysis. *Signal Processing Magazine, IEEE*, 25(1), 41-56.
- [6] Lei, X., Yang, P., Xu, P., Liu, T. J., & Yao, D. Z. (2009). Common spatial pattern ensemble classifier and its application in brain-computer interface. *J Electronic Science and Technology of China*, 7(1), 17-21.
- [7] Lan, T., Erdogmus, D., Adami, A., Pavel, M., & Mathan, S. (2006, January). Salient EEG channel selection in brain computer interfaces by mutual information maximization. In *Engineering in Medicine and Biology Society, 2005. IEEE-EMBS 2005. 27th Annual International Conference of the* (pp. 7064-7067). IEEE.
- [8] Ramoser, Herbert, Johannes Muller-Gerking, and Gert Pfurtscheller. "Optimal spatial filtering of single trial EEG during imagined hand movement." *Rehabilitation Engineering, IEEE Transactions on* 8.4 (2000): 441-446.
- [9] Higashi, Hiroshi, and Toshihisa Tanaka. "Simultaneous design of FIR filter banks and spatial patterns for EEG signal classification." *Biomedical Engineering, IEEE Transactions on* 60.4 (2013): 1100-1110.
- [10] Arvaneh, M., Guan, C., Ang, K. K., & Quek, C. (2011). Optimizing the channel selection and classification accuracy in EEG-based BCI. *Biomedical Engineering, IEEE Transactions on*, 58(6), 1865-1873.