A New BCI Classification Method based on EEG Sparse Representation

Younghak Shin, Seungchan Lee, Minkyu Ahn, Sung Chan Jun, Heung-No Lee*

Gwangju Institute of Science and Technology,
Gwangju, Republic of Korea

{shinyh,seungchan,frerap,scjun,heungno}@gist.ac.kr

Abstract

We propose a new sparse representation based classification (SRC) method for the motor imagery based Brain Computer Interface (BCI) system. For the success of SRC method, construction of a good dictionary matrix is critical. We provide a method based on common spatial pattern (CSP). The performance of the proposed method has been evaluated over the motor imagery data sets collected from five subjects. We observe that the proposed method gives higher classification accuracy results than the linear discriminant analysis (LDA) method does, one of the well-known classification methods for BCI systems.

1 Introduction

BCI system provides a new communication and control channel between people who have severe motor disabilities and an external device without any muscle movements. Many researchers have developed BCI systems using EEG signals. EEG signals can be easily acquired from the scalp; but they are very noisy and show nonstationary characteristics. Thus, powerful preprocessing methods are used for EEG signals such as the principal component analysis (PCA), the independent component analysis (ICA) and the common spatial pattern (CSP). Widely used classification methods in the EEG based BCI field have been adopted from the pattern recognition community, including the linear discriminant analysis (LDA) and the support vector machine (SVM).

In this paper, we aim to develop a new sparse representation based classification (SRC) method for the motor imagery based BCI system. The central problem of the sparse representation theory aims to search for the most compact representation of a given signal in terms of linear combination of atoms in an overcomplete dictionary[1]. When compact description of seemingly complex signals becomes possible, it can be used for a number of applications including noise reduction, dimensionality reduction, and pattern recognition. The sparse representation is achieved by the ell-1 minimization which is a suboptimal but closely attains the sparsest solution within a polynomial time if enough number of observation samples have been obtained. This ell-1 minimization approach can be utilized as a sparse representation tool in this paper for the classification purpose. For the success of this SRC method, construction of a good dictionary matrix is critical. We provide a method based on the CSP.

2 Methods

2.1 Experiment

In this study, we use data sets of BCI Competition III (Data set IVa)[2] which were recorded from five subjects. Subjects have taken the same procedure of a BCI experiment in which there are three classes, left and right hand, right foot of motor imagery movements. However, only data corresponding to right hand 'R' and right foot 'F' were provided for analysis. The data recording was made using BrainAmp amplifiers and a 128 channel Ag/AgCl electrode cap from ECI. 118 EEG channels were measured at positions of the extended international 10/20-system. Signals were band-pass filtered between 0.05 and 200Hz and then digitized at 1000Hz. For off line analysis signals were downsampled to 100Hz.

2.2 Preprocessing

We take a data segmentation for following analysis. We use 1000∼2000ms of signal samples (100 samples) after the Cue has been presented. Next, to eliminate the noise that is not related with sensorimotor rhythms (SMRs), we use a band-pass filter with 8∼15Hz cut off frequency.

To reduce the dimension of feature vector and make distinguishable features, we use the CSP method. CSP is a powerful signal processing technique that has been successfully applied for EEG-based BCIs[3].

Let $\mathbf{X} \in \mathbb{R}^{C\times T}$ be a segment of EEG signals where C is the number of EEG channels. In this study, C is 118, and T is the number of sampled time points collected in all the trials. We use 100 samples (one second). We have two classes of EEG training trials $\mathbf{X}_R \in \mathbb{R}^{C\times T}$ and $\mathbf{X}_F \in \mathbb{R}^{C\times T}$ each corresponding to the Right hand and Foot movement. Using the CSP method, we can obtain the CSP filters $\mathbf{W} \in \mathbb{R}^{C\times C}$. We call each column vector $\mathbf{w}_i \in \mathbb{R}^{C}(i = 1, 2, ..., C)$ of $\mathbf{W}$ a spatial filter. Among them, we use nCSP filters from the front and another set from the back. Then, we can make this as the CSP filtering matrix $\overline{\mathbf{W}} \in \mathbb{R}^{C\times 2n}$, i.e., $\overline{\mathbf{W}} := [\mathbf{w}_1, ..., \mathbf{w}_n, \mathbf{w}_{C-n+1}, \mathbf{w}_C]$. Given the two classes of EEG training signals, we can define the CSP filtered signals, i.e.,

$$\overline{\mathbf{X}}_R \in \mathbb{R}^{2n\times T} := \overline{\mathbf{W}}^{\mathrm{T}}\mathbf{X}_R, \quad \overline{\mathbf{X}}_F \in \mathbb{R}^{2n\times T} := \overline{\mathbf{W}}^{\mathrm{T}}\mathbf{X}_F \tag{1}$$

Next, we compute FFT of each set of T = 100 samples. Because the sampling rate is 100 samples/sec, the frequency resolution is 1 Hz. We then compute the frequency power of the Mu band (8∼15Hz). Thus, there are $N_f = 8$ columns in each matrix. They are:

$$\overline{\mathbf{X}}_R(f) \in \mathbb{R}^{2n\times N_f}, \quad \overline{\mathbf{X}}_F(f) \in \mathbb{R}^{2n\times N_f} \tag{2}$$

2.3 Linear Sparse Representation Model

Let N_t be the number of total training signals for each class $i = R, F$. We define the dictionary matrix $\mathbf{A}_i = [\mathbf{a}_{i,1}, \mathbf{a}_{i,2}, ..., \mathbf{a}_{i,N_t}]$ for $i = R, F$ where each column vector $\mathbf{a} \in \mathbb{R}^{m\times 1}$ having dimension $m = 2n\times N_f$ is obtained by concatenating the 2n rows of $\overline{\mathbf{X}}_R(f)$ and taking the transpose. Let's call this vectorization. The same procedure is repeated for the right foot part, $\overline{\mathbf{X}}_F(f)$. By combining the two matrices, we form the complete dictionary, $\mathbf{A} := [\mathbf{A}_R; \mathbf{A}_F]$. Thus, the dimension of $\mathbf{A}$ is $m \times 2N_t$. We apply the same procedure done to obtain the columns of the dictionary to the test signal. That is, the test signal is transformed to a vector $\mathbf{b} \in \mathbb{R}^{m\times 1}$ through the CSP filtering, FFT, and vectorization. Thus, the dimesion of $\mathbf{b}$ is the same as the dimension of the columns of the dictionary $\mathbf{A}$. Then, this test signal $\mathbf{b}$ can be sparsely represented as a linear combination of some columns of $\mathbf{A}$:

$$\mathbf{b} = \sum_{i=R,F} x_{i,1}\mathbf{a}_{i,1} + x_{i,2}\mathbf{a}_{i,2} + \cdots + x_{i,n_t}\mathbf{a}_{i,N_t} \tag{3}$$

where $x_{i,j} \in \mathbb{R}, j = 1, 2, ..., N_t$ are scalar coefficients. Then, we can represent this as a matrix algebraic form:

$$\mathbf{b} = \mathbf{A}\mathbf{x} \tag{4}$$

where $\mathbf{x} = [x_{R,1}, x_{R,2}, ..., x_{R,N_t} x_{F,1}, x_{F,2}, ..., x_{F,N_t}]^T \in \mathbb{R}^{2\cdot N_t}$.

# of CSP filter	al	ay	aw	aa	av
2 × 1	96.64	89.64	80.35	85.71	60.71
2 × 2	96.42	90.71	86.78	85.35	66.42
2 × 3	97.85	93.57	86.07	87.14	68.57
2 × 4	97.85	95.71	90.00	92.50	67.50
2 × 5	98.92	95.71	92.14	92.14	71.07
2 × 6	97.50	95.71	89.64	90.71	77.50

Table 1: Classification accuracy of SRC with different number of CSP filters

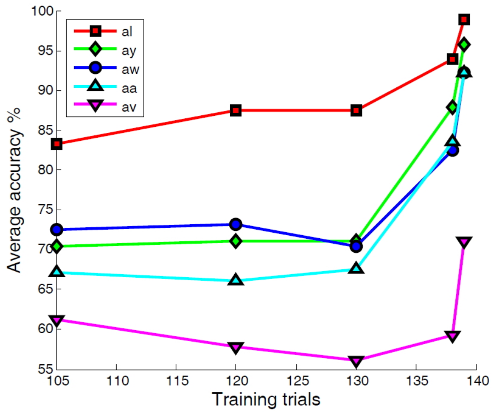

Figure 1: Average accuracy of SRC with different number of training signals

2.4 Sparse Representation by ell-1 Minimization

We have the number of total training signals $2N_t$ which is larger than the number of frequency power($m = 2n \times N_f$). The more trials, the larger the dictionary, and thus the better the sparse representation result will be. The column size is larger than the row size of the dictionary $\mathbf{A}$. Thus, the the linear equation(4) is under-determined($m < 2N_t$). Recent studies in the Compressed Sensing theory have shown that the ell-1 norm minimization, given below, can solve this under-determined system well in polynomial time [4]: $\min \|\mathbf{x}\|_1$ subject to $\mathbf{b} = \mathbf{A}\mathbf{x}$. In this paper, we use the basis pursuit method, one of the standard linear programming methods.

2.5 Sparse Representation based Classification

After solving the ell-1 minimization problem, the nonzero elements of $\mathbf{x}$ must be corresponding to the column of class i. Because the EEG signals are very noisy, the nonzero elements may appear in the indices corresponding to the column of another class. To make use of the sparse representation result, in a classification problem, we introduce the characteristic function δ[5]. For each class i, we define its characteristic function $\delta_i : \mathbb{R}^{2N_t} \to \mathbb{R}^{2N_t}$ which selects the coefficients associated with class i. For $\mathbf{x} \in \mathbb{R}^{2N_t}$, $\delta_i(\mathbf{x}) \in \mathbb{R}^{2N_t}$ is a new vector which is obtained by nulling all the elements of $\mathbf{x}$ that are associated with the other class. Then we can obtain the residuals $r_i(\mathbf{b}) := \|\mathbf{b} - \mathbf{A}\delta_i(\mathbf{x})\|_2$ for R and F. Then, the classification rule is given by: $\text{class}(\mathbf{b}) = \arg\min_i r_i(\mathbf{b})$. Thus, we determine the class i that has the minimum residuals.

3 Results and Discussion

We have analyzed five data sets, which have the same 140 trials for each class. We use the statistical k-fold cross-validation method to evaluate the average performance of the classifiers. Table 1 shows the classification accuracy of SRC with different number of CSP filters. Here, we

Classification Method	al	ay	aw	aa	av	Mean	Std.
LDA	92.50	92.50	91.07	91.07	63.92	86.21	12.48
SRC	98.92	95.71	92.14	92.14	71.07	90.00	10.95

$*$ P = 0.0428 < 0.05 (Using paired T-test)

Table 2: Classification accuracy of SRC and LDA

use 139 training signals for each class (leave-one-out). From these results, we choose the most important 10(2×5) CSP filters among 118.

Figure 1 shows the classification accuracy of SRC method when we use the 10 CSP filters. The SRC method shows good performance, especially when the number of training signals is large enough. We compare the classification accuracies of the LDA and the SRC method. Table 2 shows the comparison results when we use the leave-one-out method. In the LDA classification, we also use the CSP filtering, the FFT to extract features. For all subjects, we use the same 8∼15Hz frequency power, number of CSP filters (10) and time sample (1000∼2000ms). The results indicate that the accuracy of the SRC is better than that of the LDA method. It is statistically improved with the level of p which is less than 0.05, however, more datasets will be helpful for solid conclusion and it is currently under the investigation.

4 Conclusion

We have applied the idea of sparse representation as a new classification method to the motor imagery based BCI. The sparse representation method needs a well-designed dictionary matrix constructed from the training data. We use the CSP filtering and the FFT to produce the columns of the dictionary matrix. We have evaluated the proposed method over the BCI competition data sets. We observe that the performance of the proposed method is very satisfactory when the number of training signals is large enough. We have compared with the conventional approach such as LDA method, which is well known for the BCI system. For all subjects, our proposed method shows better classification accuracy than LDA method.

5 Acknowledgement

This work was supported by the National Research Foundation of Korea (NRF) grant funded by the Korea government (MEST) (Do-Yak Research Program, No. 2010-0017944) and NRF grant (No. 2010-0006135).

References

[1] K. Huang and S. Aviyente. Sparse representation for signal classification. *Neural Information Processing Systems*, 2006.

[2] Benjamin Blankertz. Berlin brain-computer interface. *http://www.bbci.de/*.

[3] Steven Lemm Motoaki Kawanabe Klaus-Robert Müller Benjamin Blankertz, Ryota Tomioka. Optimizing spatial filters for robust eeg single-trial analysis. *IEEE Signal Proc. Magazine*, 25(1):41–56, January 2008.

[4] J. Romberg E. Cande‘s and T. Tao. Stable signal recovery from incomplete and inaccurate measurements. *Comm. Pure and Applied Math.*, 59(8):1207–1223, 2006.

[5] Arvind Ganesh S. Shankar Sastry Yi Ma John Wright, Allen Y. Yang. Robust face recognition via sparse representation. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 31(2):210–227, February 2009.