Common Spatial Patterns in a few channel BCI interface

Pawel Górski

Faculty of Computer Science and Information Technology, West Pomeranian University of Technology, Szczecin, Poland

pagorski@wi.zut.edu.pl

Abstract: One source of EEG data quality deterioration is noise. The others are artifacts, such as the eye blinking, oculogyration, heart beat, or muscle activity. All these factors mentioned above contribute to the disappointing and poor quality of EEG signals. There are some solutions which allow increase of this signals quality. One of them is Common Spatial Patterns. Some scientific papers report that CSP can only be effectively used if there are many electrodes available. The aim of this paper is to use CSP method applied in the process of creating a brain computer interface in order to find out if there are any benefits of using this method in 3 channels BCI system.

Keywords: CSP, BCI, brain computer interface, classification

1. Introduction

A brain computer interface (BCI), is a communication system that enables users to send messages and commands to the external world without any movement. Instead, messages or commands are encoded in the brain activity and so that they can be transferred to the outside world and must be read directly from the brain by using a dedicated device, for example: Electroencephalograph.

In order to control a BCI system, different brain activity patterns are translated into commands after being identified by the system. In most existing BCI, this identification relies on a classification algorithm. The purpose of all BCI systems is to get as high accuracy as it is possible. To do this it requires special attention at all steps in the classification process, including the input data quality.

One source of data quality deterioration is noise. The others are artifacts, such as the eye blinking, oculogyration, heart beat, or muscle activity. Figure 1 shows one of the most popular artifact (artifacts are bolded), eye movement artifact (source: www.eegatlas-online.com).

When there are a lot of input channels, it is easy to detect and remove certain artifacts. But what, when only few channels of surface EEG signals are available for classification? There are some methods to mitigate the noise influence and artifacts. Some of them can be used to remove artifacts and noise, like Independent Component Analysis [1, 2, 3] or Common Spatial Patterns.

As was shown in many papers, if there are a lot of channels, Common Spatial Patterns be successfully used in BCI systems: for 55 [4], 56 [5], 118 [6, 7] up to 127 [8]. At the same time, in many papers it is emphasized that CSP is not suitable for direct usage in low channel interfaces [9]. The aim of this paper is to examine whether CSP transformation, per-
2. Common Spatial Patterns

The Common spatial patterns (CSP), proposed by Ramoser [5], is a method to use a linear transform to project the multi-channel EEG data into a low-dimensional spatial subspace with a projection matrix, of which each row consists of weights for channels. CSP maximizes the variance of band-pass filtered EEG signals from one class while minimizing the variance of EEG signals from the other class. It uses covariance to design Common Spatial Patterns based on the simultaneous diagonalization of the two covariance matrices [10]. Most of preprocessing methods requires only the training samples, but Common Spatial Patterns requires also the information to which condition the samples belong to. It is necessary to compute the linear transformation matrix [11].

Let $X \in \mathbb{R}^{N \times S}$ denotes a matrix that represents the EEG of a single-trial, where $N$ and $S$ denote the number of channels and number of measurement samples respectively [12]. The normalized spatial covariance of the EEG can be represented as:

$$C = \frac{EE'}{\text{trace}(EE')}$$

where:

' - denotes the transpose operator,
trace(x) gives the sum of diagonal elements of x.

The covariance matrices of each class, $C_1$ and $C_2$, are computed by averaging over multiple trials of EEG data. The composite spatial covariance can be factorized as:

$$C_1 = C_1^r + C_r = F_c\psi F_c^T$$

where:
- $F_c$ - is the matrix of eigenvectors
- $\psi$ - is the diagonal matrix of eigenvalues.

Note that throughout this section, the eigenvalues are assumed to be sorted in descending order.

The whitening transformation matrix:

$$P = \sqrt{\psi_c^{-1}} U_c$$

transforms the covariance matrices as:

$$C_1' = PC_1 P' \quad \text{and} \quad C_2' = PC_2 P'$$

Then $C_1$ and $C_2$ share common eigenvectors, the sum of corresponding eigenvalues for the two matrices are always one, such that:

$$C_1' = U\Lambda_1 U^T, C_2' = U\Lambda_2 U^T, \Lambda_1 + \Lambda_2 = I$$

where $I$ is the identity matrix. The eigenvectors with the smallest eigenvalues for $C_2$ and with the largest eigenvalues for $C_1$. The transformation of whitened EEG onto the eigenvectors corresponding to the largest eigenvalues in $\Lambda_1$ and $\Lambda_2$ is optimal for separating variance in two signal matrices.

With the projection matrix $W = (U P)'$, the decomposition (mapping) of a trial $E$ is given as

$$Z = WE$$

Where $Z$ can be seen as EEG raw signals including common and specific components of different tasks. The columns of $W^{-1}$ are the Common Spatial Patterns and can be seen as time-invariant EEG source distribution vectors [5].

The basic of CSP algorithm can deal with only two conditions [11].

### 3. Methods

#### 3.1. Data set

To explore advantages of Common Spatial Pattern, we conduct our experiments on data set which was submitted to the second BCI Competition by Department of Medical Informatics, Institute for Biomedical Engineering, Graz University of Technology [15]. EEG signal was recorded from a one female (25y). The task consist of performing motor imagery of the left or right hand in response to cue. Each of these cues provided information about the direction of the movements. The order of cues was random. For the one subject, 280 of 9 seconds of EEG signals was collected. The first 2s was quiet, at t=2s an acoustic stimulus was generated
and a cross + was displayed for 1s; then at \( t=3 \text{s} \), an arrow (left or right) was displayed as a cue. The EEG signals were measured over three bipolar EEG channels (C3, Cz and C4). The sample rate was 128Hz and the passband of the filter was from 0.5 to 30Hz. The data set was divided into two equal subsets - 140 trials each. The first one was intended for classifier training and the second intended for external classifier test. Since only data from the first subset was published with target values (1 - left hand, 2- right hand), only this subset could be used in the process of classifiers.

Figure 2. Position of EEG electrodes. C3,Cz and C4 are bolded.

3.2. Preprocessing

Data preprocessing describes any type of processing performed on raw data to prepare it for another processing procedure. Commonly used as a preliminary data mining practice, data preprocessing transforms the data into a format that will be more easily and effectively processed for the purpose of the user – for example, in a neural network. There are a number of different tools and methods used for preprocessing, including: sampling, which selects a representative subset from a large population of data; transformation, which manipulates raw data to produce a single input; denoising, which removes noise from data; normalization, which organizes data for more efficient access; and feature extraction, which pulls out specified data that is significant in some particular context.

In the preprocessing step two data sets were created. One, consist of original signal (three channels) and second one, that were created by transform original data set by Common Spatial Patterns algorithm described in previous section. Next each of above data sets were split into 7 subsets:

1. Three single channels sets (1,2 and 3)
2. Three double channels sets (1+2,1+3 and 2+3)
3. And one, complete three channels set.

3.3. Feature extraction

Next, for each data set we calculated features describing our signals. Single feature contains frequency band calculated from:
- 9 seconds of each trial in each signal,
- each of 12 frequency bands:
  a) alpha band (8-13Hz) and five sub-bands of alpha band (8-9Hz; 9-10Hz; 10-11Hz; 11-12Hz; 12-13Hz);
  b) beta band (13-30Hz) and also five sub-bands of beta band (13-17Hz; 17-20Hz; 20-23Hz; 23-26Hz; 26-30Hz).

In this way, the feature matrix for single channel was composed of 108, two channels of 216 and three of 324 features.

3.4. Feature selection

At was shown in [13, 14] often the genetic algorithm is used with success to reduce the number of input features. It has good results but the calculation time is too long. Because of this, in order to reduce the size of each matrix of features from 324 to 6, LASSO algorithm (Least Absolute Shrinkage and Selection Operator), proposed by Tibshirani [16], was applied. The LASSO optimization problem assumes linear dependency between input features and output values and is given as:

\[ L = \sum_i (y_i - \sum_p \beta_p \chi_{ip})^2 + \lambda \sum_p |\beta_p| \]  

where:
\( \chi_{ip} \) - \( p \)th predictor (feature) in the \( i \)th sample,
\( y_i \) - value of the response in this sample,
\( \beta \) - regression coefficient of the \( p \)th feature [17].

Since we wanted to analyze the classification results separately for each combination of signals comprises in each of the four sets, we applied LASSO 7 times for each set.

A linear SVM method was used in the classification process.

The classification threshold was set to 0.5 and hence, all classifier results greater than 0.5 were classified as class 2 (right hand) and results smaller or equal to 0.5 were classified as class 1 (left hand). The classifiers accuracy was tested with 10-fold cross-validation. The final accuracy measure of a given feature set was the mean value calculated on the basis of classification accuracy obtained for all validation sets.

4. Results

The main aim of the experiments described in the paper was to find out whether the preprocessing with CSP has an essential influence on the classification accuracy in 3 channels BCI system. Classification results for each trial is presented in tables 1 and 2. Classification accuracy is presented in 6 columns, first 5 of them are for each trial and the last one is the mean of them. First table shows the results for raw signal, the second one results for signal after CSP preprocessing.
Table 1. Classification accuracy on the raw signal

<table>
<thead>
<tr>
<th>Algorithm runs</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>mean</th>
</tr>
</thead>
<tbody>
<tr>
<td>Channel 1</td>
<td>0.72</td>
<td>0.74</td>
<td>0.74</td>
<td>0.74</td>
<td>0.74</td>
<td>0.73</td>
</tr>
<tr>
<td>Channel 2</td>
<td>0.66</td>
<td>0.69</td>
<td>0.66</td>
<td>0.66</td>
<td>0.66</td>
<td>0.67</td>
</tr>
<tr>
<td>Channel 3</td>
<td>0.74</td>
<td>0.74</td>
<td>0.75</td>
<td>0.74</td>
<td>0.74</td>
<td>0.74</td>
</tr>
<tr>
<td>Channels 1+2</td>
<td>0.75</td>
<td>0.74</td>
<td>0.74</td>
<td>0.72</td>
<td>0.76</td>
<td>0.74</td>
</tr>
<tr>
<td>Channels 2+3</td>
<td>0.88</td>
<td>0.88</td>
<td>0.86</td>
<td>0.84</td>
<td>0.89</td>
<td>0.87</td>
</tr>
<tr>
<td>Channels 1+3</td>
<td>0.89</td>
<td>0.91</td>
<td>0.90</td>
<td>0.91</td>
<td>0.91</td>
<td>0.91</td>
</tr>
<tr>
<td>Channels 1+2+3</td>
<td>0.90</td>
<td>0.91</td>
<td>0.90</td>
<td>0.90</td>
<td>0.90</td>
<td>0.90</td>
</tr>
</tbody>
</table>

As it can be seen the highest classification efficiency gained by a pair of channels 1 + 3. Both of single channels 1 and 2 (C3 and C4) achieve similar accuracy. Full set of channels, also has very good accuracy. Table number two, presents the results of the accuracy for the signal after CSP transformation.

Table 2. Classification accuracy after CSP preprocessing

<table>
<thead>
<tr>
<th>Algorithm runs</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>mean</th>
</tr>
</thead>
<tbody>
<tr>
<td>Channel 1</td>
<td>0.70</td>
<td>0.67</td>
<td>0.69</td>
<td>0.69</td>
<td>0.69</td>
<td>0.69</td>
</tr>
<tr>
<td>Channel 2</td>
<td>0.62</td>
<td>0.61</td>
<td>0.64</td>
<td>0.63</td>
<td>0.61</td>
<td>0.62</td>
</tr>
<tr>
<td>Channel 3</td>
<td>0.72</td>
<td>0.77</td>
<td>0.75</td>
<td>0.76</td>
<td>0.76</td>
<td>0.75</td>
</tr>
<tr>
<td>Channels 1+2</td>
<td>0.71</td>
<td>0.72</td>
<td>0.69</td>
<td>0.71</td>
<td>0.70</td>
<td>0.71</td>
</tr>
<tr>
<td>Channels 2+3</td>
<td>0.79</td>
<td>0.78</td>
<td>0.79</td>
<td>0.79</td>
<td>0.79</td>
<td>0.79</td>
</tr>
<tr>
<td>Channels 1+3</td>
<td>0.86</td>
<td>0.86</td>
<td>0.86</td>
<td>0.87</td>
<td>0.86</td>
<td>0.86</td>
</tr>
<tr>
<td>Channels 1+2+3</td>
<td>0.86</td>
<td>0.87</td>
<td>0.87</td>
<td>0.89</td>
<td>0.87</td>
<td>0.87</td>
</tr>
</tbody>
</table>

The third table presents the results in a more compact form. It sets, only the averaged results for both sets of inputs for each set of signals.

Table 3. Mean values of classification accuracy for raw signal and signal after CSP transformation for all subsets

<table>
<thead>
<tr>
<th>Channels</th>
<th>Original</th>
<th>CSP</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>mean</td>
<td>mean</td>
</tr>
<tr>
<td>Channel 1</td>
<td>0.73</td>
<td>0.69</td>
</tr>
<tr>
<td>Channel 2</td>
<td>0.67</td>
<td>0.62</td>
</tr>
<tr>
<td>Channel 3</td>
<td>0.74</td>
<td>0.75</td>
</tr>
<tr>
<td>Channel 1+2</td>
<td>0.74</td>
<td>0.70</td>
</tr>
<tr>
<td>Channel 2+3</td>
<td>0.87</td>
<td>0.79</td>
</tr>
<tr>
<td>Channel 1+3</td>
<td>0.91</td>
<td>0.87</td>
</tr>
<tr>
<td>Channel 1+2+3</td>
<td>0.90</td>
<td>0.87</td>
</tr>
</tbody>
</table>

As can be seen from table 3, for almost all cases, the classification accuracy is also higher, when the original signal was used. One exception is for channel number 3.

For both data sets, the classification accuracy obtained from the extracted features are worst for second channel (Cz).

Last figure, presents the accuracy on the chart. Only in one place the result for CSP is better than the original signal. It should also be noted that the fourth set of signals, achieved a similar classification accuracy, as the same signal C4 for raw signal and in the case of the CSP signal is even worse.
5. Conclusion

In this study we investigated the influence of Common Spatial Patterns transformation on the classification accuracy in three channel EEG. The results show that the mean of classification accuracy is better when original signal was used. This is proves the theory that Common Spatial Patterns should be used in many channels systems.

On the other hand, classification accuracy slightly increased for 3 channel after CSP preprocessing from 0.74 to 0.75%. Channel 3 was located beneath electrode position C4 situated over the motor cortex. So, the results may point to a promising area, but further studies and experiments are needed.

References

[8] Lei, X., Yang, P., Xu, P., Liu, T. J., Yao, D. Z.: Common Spatial Pattern Ensemble Classifier and
Common Spatial Patterns in a few channel BCI interface


