Common Spatial Pattern Filtering and Linear Discriminant Analysis for Classifying EEG signals

Mark B. Rosenberg

Abstract—**Common Spatial Pattern filtering was applied to training data from the third Brain-Computer Interface competition of the group Berlin Brain-Computer Interface. Testing data was classified with Linear Discriminant Analysis. The classification rate was 72%.**

Index Terms—**Common Spatial Pattern, Linear Discriminant Analysis, Brain-Computer Interface**

I. Introduction

The third Brain-Computer Interface competition [1] of the group Berlin Brain-Computer Interface obtained EEG signals of a subject performing motor imagery in two classes: left hand and right hand motor imagery. The goal of the competition participants was to perform supervised learning on a training set of EEG signals with known class labels in order to predict the class labels of the test set of EEG signals. There were 316 data points in the training set and 100 data points in the test set. Each data point was an EEG voltage signal represented by a 28 channel by 500 time step matrix. Fig. 1 shows one such data point.

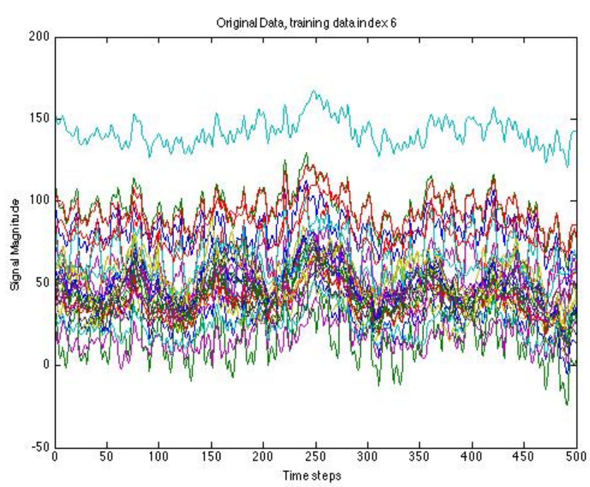

Figure 1

The 28 channels corresponded to the 28 locations on the subject's scalp where EEG signals were recorded. In order to classify the data, each data point should be represented as a point in feature space. In order to extract features from each data point, signal processing must first be performed on the data. The signal processing discussed in this paper will be Common Spatial Pattern (CSP) filtering, and the supervised learning classification method discussed will be Linear Discriminant Analysis (LDA).

Before applying signal-processing techniques to extract features from the data, pre-processing must first be performed in order to remove noise from the signal. Since the subject is awake and performing tasks, the spectrum of the brain wave signals is within the bands specified by alpha and beta brain waves, or from 7Hz to 30Hz. Furthermore there is power-line interference at odd multiples of 50Hz due to clipping. This is shown in Fig. 2 by the spectrum of training data point 6. The power line interference is 50Hz because the data was obtained in Europe.

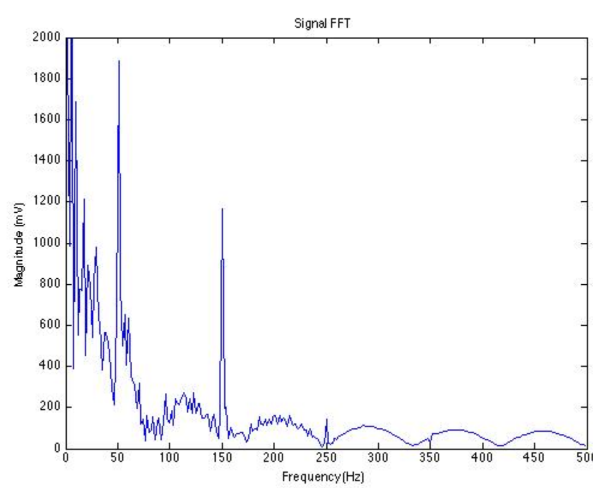

Figure 2

Since this noise occurs outside the frequency band containing the signal, notch filters won't be required to individually remove the power-line interference because a bandpass filter will be used on the frequency band where the signal is expected to be. The bandpass filter will also remove low frequency components, which effectively makes the signal zero-mean.

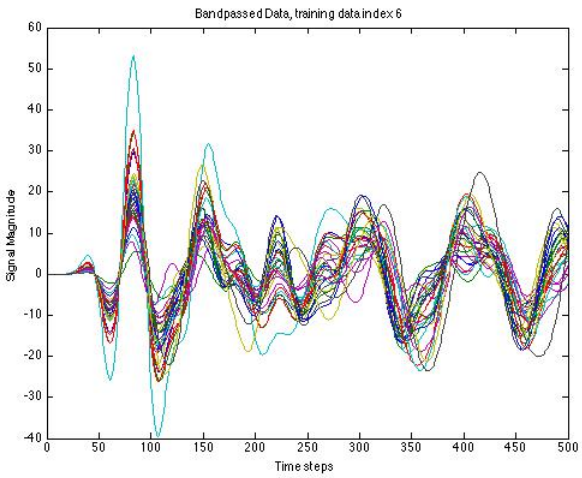

Figure 3

The design choices for the bandpass filter include filter type and filter parameters. An IIR Butterworth filter was chosen to reduce filter order and delay, and the passband was chosen to be from 7Hz to 30Hz because that's where the signal was expected to be. The effect of different passband boundaries on the classification rate were tested by iterating through two for loops, one for each end of the passband. The maximum classification rate over all CSP filter dimensions for 49 bandpass boundaries is shown in Table 1. The result of the passband filter from 7Hz to 30Hz on training data point 6 is shown in Fig. 3.

	30Hz	29Hz	28Hz	27Hz	26Hz	25Hz	24Hz	23Hz
7Hz	0.72	0.72	0.72	0.71	0.72	0.71	0.71	0.7
8Hz	0.76	0.75	0.75	0.75	0.74	0.75	0.74	0.74
9Hz	0.74	0.76	0.75	0.75	0.75	0.76	0.75	0.76
10Hz	0.73	0.74	0.74	0.75	0.75	0.74	0.76	0.76
11Hz	0.74	0.74	0.74	0.76	0.76	0.75	0.73	0.76
12Hz	0.75	0.75	0.74	0.74	0.74	0.73	0.71	0.72
13Hz	0.77	0.77	0.75	0.75	0.74	0.73	0.73	0.73
14Hz	0.75	0.75	0.76	0.75	0.75	0.72	0.74	0.71

Table 1: Maximum classification rate over all CSP filter dimensions for 49 bandpass boundaries.

II. Background

The reason it is desirable to have a high classification rate is so that the brain-computer interface can be used as part of a system where the user can provide input to a system simply by thinking, and can control that system through feedback. Such a system would have important applications for people with injuries to their nervous system that prevent them from controlling their bodies.

When the subject thinks about left or right hand motor imagery, a true signal in the brain is generated. The 28 locations on the scalp where EEG signals are recorded detect the result of that signal and noise. Some channels are in locations that have higher signal to noise ratios (SNRs) than other channels. The result of this configuration is that there is some correlation between the signals that were recorded in channel locations closer to each other, while other channel locations may have a very low SNR since the true brain signal is not generated near that particular area of the scalp. This motivates the design of the CSP filter, which is a square matrix that transforms the data from the 28-dimensional channel space to a 28-dimensional space that spans the same space but has non-uniform weights for each channel. Each column of the transformation matrix filters the data, so that each element of that column is a weight for that channel. The CSP filter is designed such that the weights in the first column maximize the variance of one training class and minimize the variance of the second training class, and the weights in the last column maximize the variance of the second training class and minimize the variance of the first training class.

III. Common Spatial Pattern Algorithm

First, the spatial covariance of each data point is computed and is normalized by its trace [2]. Then the average normalized covariance matrix is computed for each class from the normalized covariance matrices of the training data. The composite average normalized covariance matrix is obtained by summing the two average normalized covariance matrices for the two classes. By decomposing the composite covariance matrix into eigenvalues and eigenvectors, an operation multiplying the square root of the inverse of the diagonal eigenvalue matrix by the transpose of the eigenvector matrix results in a whitening transformation matrix. The whitening transformation reduces the significance of the covariance of a matrix. When it is applied to each of the two original covariance matrices by left multiplying by the whitening transformation matrix and right multiplying by its transpose, the resulting two matrices share the same eigenvectors, but with corresponding eigenvalues that sum to 1. Furthermore, when the eigenvectors of one of the transformed covariance matrices are ranked in descending order of corresponding eigenvalues, the corresponding eigenvalues from the transformed covariance matrix of the other class are ranked in ascending order. After the data is transformed by these eigenvectors, which comprise the CSP filter, the sum over time of the squared signal gives a measure of the power of that signal with respect to each of the 28 filter columns. Then since the resulting values are all positive, the log is taken, which has the effect of spreading out the data, as shown in Fig. 4.

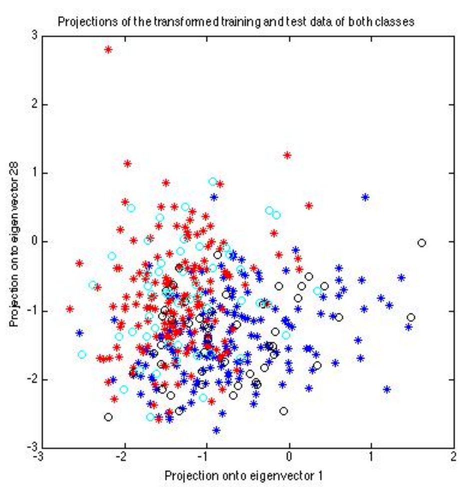

Figure 4: Training class 0 (blue asterisk), Training class 1 (red asterisk), Testing class 0 (black circle), Testing class 1 (cyan circle). Note that the class 0 data varies maximally along the first eigenvector and minimally along the last eigenvector, and vice versa.

Each transformed data point is a point in 28-space, where the variance of the training data for the first column in the CSP filter is maximized for the class whose top eigenvalue of the whitened transformed covariance matrix corresponded to eigenvector that comprised the first column of the CSP filter. The opposite is true for the variance of the second class with respect to that eigenvector: it is minimized. Similarly, for the last column of the CSP filter, the variances of the projections of the transformed data onto that vector have the opposite relationship. The result of this transformation is that choosing to only consider the transformation of the data onto the first few and last few CSP filter columns results in data that can be more easily classified by linear discriminant analysis. Since the number of first and last CSP filter columns chosen is a parameter we can choose, it is helpful to plot the classification rate for all choices from 1 to 14, as shown in Fig. 5.

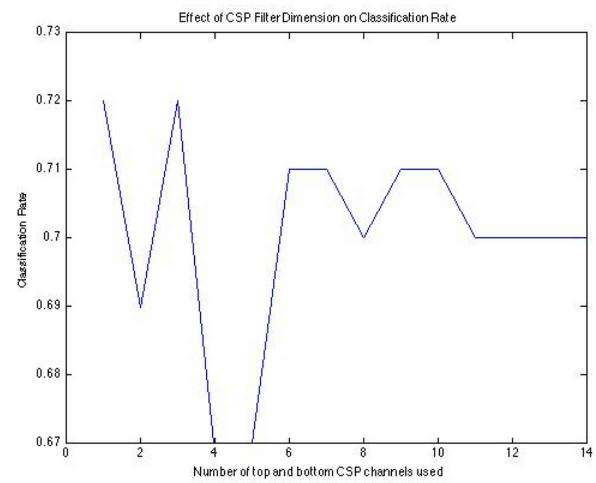

Figure 5

For the bandpass filter parameters that were tested, the highest classification rate typically occurred when only the first and last CSP filter columns were chosen. The plot of the first and last CSP filter columns indicates how much each channel is related to the differences between the two classes. If the magnitudes of the elements of these CSP filter columns were plotted as a colormap on the corresponding areas of the scalp, the result would indicate what areas of the brain are most related to the generation of each class signal. The plot of the first CSP filter column in Fig. 6 shows a high positive weight for channel 14, location C3, and the plot of the last CSP filter column in Fig. 7 shows a high positive weight for channel 18, location C4.

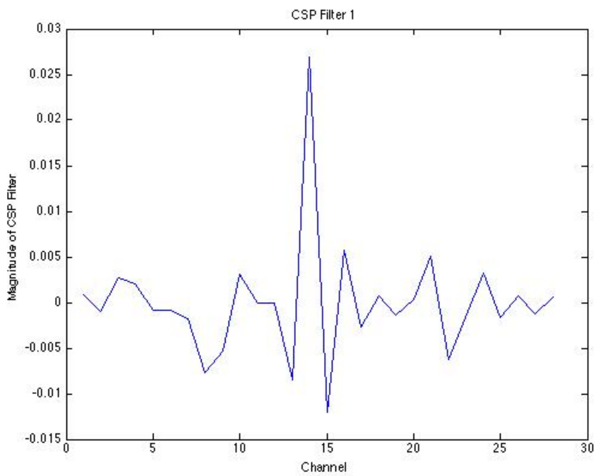

Figure 6

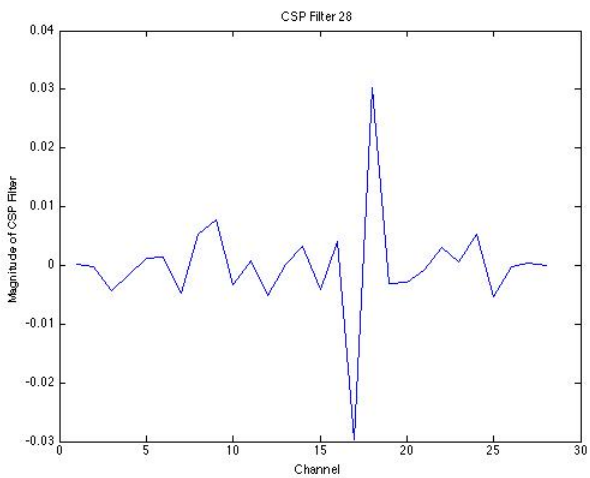

Figure 7

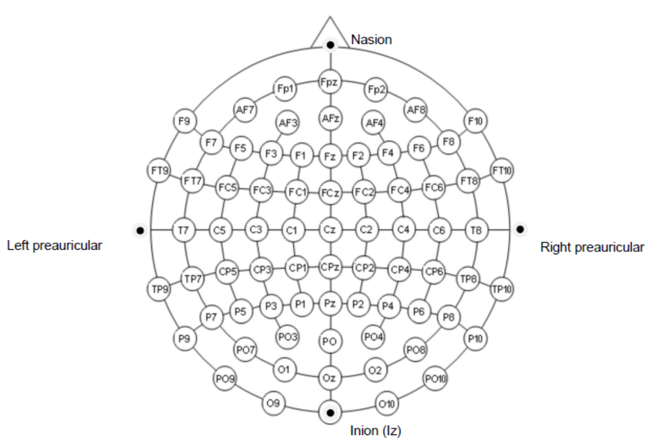

Figure 8: [3]

Fig. 8, which shows the channel locations on the scalp also shows that C3 and C4 are approximately the center of the left and right sides of the scalp, respectively, which is very significant validation of the CSP method since the two classes are left and right motor imagery, which we would expect to differ in terms of left-right symmetry of the brain signal.

IV. Linear Discriminant Analysis

After the CSP filter is determined by the training data, the test data is also bandpass filtered and then spatially filtered by the CSP filter. Linear discriminant analysis is used to determine the classification rate, which is the percentage of correct classifications. MATLAB's classify function was used, and code for Fisher's discriminant was also applied. For the Fisher method [4], the mean and covariance matrices of the transformed 28-dimensional training data are computed, and a multivariate Gaussian function is considered for each class. The test data, once transformed by the CSP procedure that relied on the training data, was evaluated for both Gaussian discriminant functions, and the data point was associated to the class with the greater valued discriminant function evaluated at that point. This method assumes the data has a multivariate Gaussian distribution, which is a reasonable assumption based on looking at plots of the projected transformed data on the first and last column of the CSP in Fig. 4.

V. Conclusions

The resulting predictions were compared with the testing labels by constructing a confusion matrix. A confusion matrix is a square matrix with side length equal to the number of classes. One axis is the predicted class, and the other axis is the actual class. The trace of this matrix is the number of data points correctly assigned, and when divided by the sum of the elements in the matrix results in the classification rate. For varying bandpass filter parameters, the maximum classification rate over all CSP filter dimensions ranged between 71% and 76% as shown in Table 1. For the 7Hz to 30Hz bandpass filter, the best classification rate occurred by choosing the first and last CSP filter columns, as shown in Fig. 5, and was 72%. The conclusion of these results is not only that this method performs better than random guessing, which would have a 50% classification rate on average, but coupled with the first and last CSP filter column weightings, it can be concluded that channel locations C3 and C4 are important in detecting the true brain hand motor signals with a high SNR.

REFERENCES

[1] BCI Competition III <http://www.bbci.de/competition/iii/> Accessed 11-22-2013

[2] Z. J. Koles, M. S. Lazar, S. Z. Zhou, “Spatial Patterns Underlying Population Differences in the Background EEG” in *Brain Topology*, Vol. 2, Num. 4, (1990).

[3] <http://www.gtec.at/var/plain_site/storage/images/media/images/support/mount_eeg_10_20_system/10664-3-eng-GB/mount_eeg_10_20_system.png> Accessed 11-22-2013

[4] R. A. Fisher, “The Use of Multiple Measurements in Taxonomic Problems" Annals Eugen. Vol. 7 (1936), pp. 179-188