Comparative Study of Robust Methods for Motor Imagery Classification based on CSP and LDA

Ana Julia Villar[1]

[1]Instituto Tecnológico de Buenos Aires, Computer Engineering Department, Buenos Aires, Argentina

Abstract— **Common spatial patterns analysis and linear discriminant analysis are popular algorithms for spatial filtering and classifying in motor imagery. These algorithms are very sensitive to noise and artifacts which affect the classification accuracy. To deal with this issue, it is proposed to replace the usual estimators of covariance and scale used in the algorithms for robust versions. The performance of the methods are evaluated and compared on EGG data from BCI competition data sets; results show that robust methods outperformed classical techniques for subjects with poor classification accuracy.**

Keywords— **BCI, motor imagery, CSP, LDA, robustness**

I. Introduction

Motor imagery (MI) is a neuronal activity that occurs when a subject voluntarily imagine making a movement, for example, moving the right hand. To imagine a movement produces a neuronal activity that is spatio-temporally similar to the activity generated during the real movement. The Brain-Computer Interface (BCI) protocol usually employed involves asking the patient to imagine various types of movements while the EEG signals are recorded. The characteristics used to quantify the EEG activity are extracted and subsequently classifiers can be applied to discriminate between two or more imagined movements, allowing each imagined activity to be assigned to a particular control signal.

Common spatial patterns analysis (CSP) [1] is a supervised spatial filtering method that is used to find a transformation that maximizes the separability between the EEG data of two conditions. Usually, only a few of most discriminative filters obtained are used for classification. Linear discriminant analysis (LDA) [2] is a classifier that provides acceptable accuracy without high computation requirements. It is usually applied to classify patterns into two classes, although it is possible to extend the method to multiples classes.

EEG signals are widely affected by a variety of large signal contaminations or artifacts, such as eyes movements and blinkings, heart and muscle activities, head and body movements, as well as external interferences due to power sources. This causes the EEG signals contain outliers, which are observations that deviate from the general pattern of the data [3]. The outliers can adversely affect the results obtained from the conventional estimation methods such as CSP and LDA. These algorithms are strongly influenced by outliers because they involve the use of sample covariance that is highly non-robust.

In this work, methods based on the regularized common spatial patterns analysis (RCSP) algorithm proposed by Yong et al. [4] are evaluated for three datasets taken from the BCI competitions III and IV using several robust covariance estimators: Minimum covariance determinant (MCD) [5], Stahel-Donoho estimator (DS) [6,7], MM-estimator [8]. The proposed estimators for the covariance matrix provide robust estimates with respect to individual samples, i.e. they ignore trial structure and downweights outlier samples rather than whole trials. Not only the estimation of covariance matrices, but also the sample variance estimates used in extracting the features from the projected EEG signals are also easily affected by even a single outlier [3]. In order to deal with this problem, the scale estimate is replaced with the median absolute deviation and mean absolute deviation estimates.

II. Material and methods

In BCI design the goal is to process the EEG signal to translate into the mental state of the user. The two main steps in the processing system are feature extraction and classification.

In motor imagery based BCI can be distinguished imaginary movements of the right hand from the left hand. To identify these two mental states from EEG signals, band power features are usually extracted in the μ (about 8−12 Hz) and β (about 16−24 Hz) frequency bands, for electrode localized over the motor cortex areas of the brain. Such features are then typically classified using LDA.

A. Classical and robust CSP

One way to extract features from multiple EEG channels is to use spatial filtering, as CSP, which is a technique to analyze multi-channel data based on recordings from two classes.

Let $\mathcal{X} = \{X_1, \cdots, X_M\}$ be a sample of M training EEG trials corresponding to two different mental states (classes 1 and 2), where $X_j \in \mathbb{R}^{N\times C}$ is the data matrix which corresponds to a trial j $(1 \le j \le M)$ of imaginary movement, with N the number of observations in each trial and C the number of channels. Let $\Sigma_i \in \mathbb{R}^{C\times C}$ be the spatial covariance matrix of the band-pass filtered signals in $\mathcal{X}$ from class i $(i = 1, 2)$.

CSP yields a data-driven supervised decomposition of the signal parameterized by a matrix $W \in \mathbb{R}^{m\times C}$ that projects a signal $X \in \mathbb{R}^{N\times C}$ in the original sensor space $X_{\text{CSP}} = XW^t$ where the m spatial filters (rows of W) $w_k \in \mathbb{R}^C$ $(1 \le k \le m)$ extremize the Rayleigh quotient $\text{J}(w) = w^t\Sigma_1 w / w^t\Sigma_2 w$.

For a given trial matrix X_j (centered and scaled) the normalized sample covariance matrix is obtained as $S_j = X_j{}^t X_j / \text{tr}(X_j{}^t X_j)$.

For each class i, an estimator of Σ_i is computed by averaging the covariances matrices of each trial as

$$\breve{\Sigma}_i = \frac{1}{|\mathcal{C}_i|}\sum_{j\in\mathcal{C}_i} S_j, \tag{1}$$

where $\mathcal{C}_i$ is the set of trials in $\mathcal{X}$ belonging to the class i and $|\mathcal{C}|$ denotes the cardinality of $\mathcal{C}$. Such computation of covariance matrices assumes that the signals have zero mean, which is true in practice for band-pass filtered signals.

The vector w in $\text{J}(w)$ can be achieved by solving the generalized eigenvalue problem

$$\Sigma_1 w = \lambda \Sigma_2 w. \tag{2}$$

Then the matrix W in X_{CSP} consists of the generalized eigenvectors w_k of Eq. 2 and $\lambda_{i_k} = w_k^t \Sigma_i w_k$ are the corresponding diagonal elements of Λ_i, while λ in Eq. 2 equals to $\lambda_{1_k}/\lambda_{2_k}$.

The sample covariance matrix is highly non-robust and has a breakdown point of 0, implying that is affected by even a single outlier. The simplest way to deal with this problem is to replace it with a robust estimate. In this study, the robust estimators used instead of S_j are MCD, DS-estimator and MM-estimator.

The MCD estimator is given by the subset of h $(> N/2)$ out of N data points with smallest covariance determinant. The location and scatter estimates are therefore the mean and a multiple of the covariance matrix computed on h such points. The DS-estimators of multivariate location and scatter are defined as a weighted mean and a weighted covariance matrix. The weights depend on a measure of outlyingness obtained by considering all univariate projections of the data. The idea of the multivariate MM-estimators is to estimate the scale by means of a very robust S-estimator, and then estimate the location and shape using a different loss function that yields better efficiency at the central model. The location and shape estimates inherit the breakdown point of the auxiliary scale.

The average estimates obtained in Eq. 1 can be substantially disturbed by outliers. It is possible to make them more robust by reducing the contributions of outlier covariances. An alternative iterative method is proposed by [9] using a weighted average where the weights controls the importance of each covariance depending on the distance from the center. Another weighted average, based on the number of outliers, can be obtained as follows. For a given X_j, let $\hat{\boldsymbol{\mu}}_j$ and $\hat{\Sigma}_j$ be robust estimates of location and scatter. Let $\{\boldsymbol{x}_1, \cdots, \boldsymbol{x}_N\}$ be the N rows of X_j, where $\boldsymbol{x}_i \in \mathbb{R}^C$ $(1 \le i \le N)$. A measure of outlyingness for a data point $\boldsymbol{x}_i$ is given by a robust version of the Mahalanobis distance $RD_i{}^2 = (\boldsymbol{x}_i - \hat{\boldsymbol{\mu}}_j)^t \hat{\Sigma}_j{}^{-1}(\boldsymbol{x}_i - \hat{\boldsymbol{\mu}}_j)$ with the usual cutoff value of $\chi^2_{C,0.975}$. For each row $\boldsymbol{x}_i$, RD_i is calculated to determine if $\boldsymbol{x}_i$ is an outlier; let n_j the number of rows of X_j which are not outliers. Thus, a robust estimator of Σ_i is given by $\tilde{\Sigma}_i = {}^1\!/_N \sum_{j\in\mathcal{C}_i} n_j \hat{\Sigma}_j$.

Then, replacing $\breve{\Sigma}_i$ by $\tilde{\Sigma}_i$ in Eq. 1, the proposed robust version of CSP will generate a weight matrix $\widetilde{W}$ as output.

B. Feature extraction

The classical measure for the selection of CSP filters is based on the eigenvalues in Eq. 2. The m filters $(m = 2p)$, corresponding to the p largest and the p lowest eigenvalues are used.

Once these filters are obtained, a CSP feature f_k for a signal X is defined as $f_k = \ln(w_k{}^t X^t X w_k) = \ln(\text{var}(w_k X))$, i.e., the m features used are simply the band power of the spatially filtered signals.

Each eigenvalue is the relative variance of the signal filtered with the corresponding spatial filter. This measure is not robust to outliers because it is based on simply pooling the sample covariance matrices in each class. A simple way to fix this issue proposed by Blankertz et al. [10]; they calculate the variance of the filtered signal within each trial and then calculate the corresponding ratio of medians: $\text{score}(w) = \frac{med_1}{med_1 + med_2}$, with $med_i = \text{median}(w^t X^t X w)$.

On the other hand, a robust scale estimate $\hat{s}$ of a sample $\mathcal{Y} = \{y_1, \cdots, y_N\}$ can be obtained using the median of the absolute deviations (MAD) of the sample from their median, given by $\hat{s} = \frac{1}{\Phi^{-1}(3/4)}\text{median}(|y_i - \text{median}(y_i)|)$.

Another scale estimator $\tilde{s}$ is the mean absolute deviation ($\overline{MAD}$), given by $\tilde{s} = \frac{\sqrt{\pi}}{N\sqrt{2}}\sum_{i=1}^N |y_i - \text{mean}(y_i)|$.

Thus, using any of the proposed variance estimators, for a signal X, the vector $F = [\, f_1 \cdots f_m]$ of features used in classification is found.

C. Classification

Linear discriminant analysis is typically carried out using Fisher's method. LDA assumes that the two classes are linearly separable. According to this assumption, it defines a linear discrimination function which represents a hyperplane in the feature space in order to distinguish the classes. The class to which the feature vector belongs will depend on the side of the plane where the vector is found.

For each X_j in the sample $\mathcal{X}$, the features vector F_j is obtained. Let Φ_i the set of these vectors such signals are in C_i, and let $\breve{\mu}_i$ and $\breve{\Sigma}_i$ be the sample mean and covariance obtained with the data in Φ_i. For a new signal $X \notin \mathcal{X}$, let π_i be the probability that the signal X to classify belongs to class i and F the features vector for X. Let $\breve{\Sigma}$ be the within groups covariance matrix given by the pooled version of the different scatter matrices $\breve{\Sigma} = \sum_{i=1}^{2} \pi_i \breve{\Sigma}_i$.

X is assigned to class l for which $D_l(F) = \min_{i=1,2} D_i(F)$, where $D_i^2(F) = (F - \breve{\mu}_i)^t \breve{\Sigma}^{-1} (F - \breve{\mu}_i) - 2 \ln \pi_i$.

D. Numerical experiments

a) Data sets

In order to compare the performance of the robust algorithms proposed with the classical CPS and LDA, three data sets of the BCI competitions, which contain motor imagery EEG signals, were used. The first two datasets were collected in a multiclass setting, with the subjects performing more than two different MI tasks; for them, the algorithms were evaluated on two-class problems by selecting only signals of left and right hand MI trials.

Datasets are: **IV-IIa** [11]: recorded using 22 electrodes from 9 subjects who performed left-hand, right-hand, foot and tongue MI. The training and testing sets containing 72 trials for each class; **III-IIIa** [12]: recorded using 60 electrodes from 3 subjects who performed left-hand, right-hand, foot and tongue MI. The training and a testing sets contain 45 trials per class for subject 1, and 30 trials per class for subjects 2 and 3; **III-IVa** [13]: recorded using 118 electrodes from 5 subjects who performed right hand and foot MI. A training set and a testing set were available for each subject, with different sizes (280 trials were available for each subject, among which 168, 224, 84, 56 and 28 composed the training set for subject A1, A2, A3, A4 and A5 respectively, the remaining trials composing the test set).

b) Data processing

For all data sets, EEG signals were band-pass filtered in 8-30 Hz, using a 5th order Butterworth filter. For each trial, the features are extracted from the time segment located from 0.5s to 2.5s after the cue instructing the subject to perform MI. The Matlab code of EEG signal processing/classification by Lotte, available at [14], was adapted to process the signals. Data were spatially filtered using the different variants of CSP presented in Section II.A. The robust statistical toolbox FSDA [15] was used to obtain the robust estimates of the covariance matrices by the functions "mcd" for MCD, "SDest" for DS-estimator and "MMmultcore" for MM-estimators. The filters were selected according to the p pairs of extreme eigenvalues and the p pairs of extreme scores. After spatially filtering the data, the log-variance was calculated in each trial using the different estimators of scale proposed in Section II.B, and LDA was trained on those features. The subjects' performance was calculated by filtering and classifying the testing data with the filters and LDA obtained in the manner mentioned before.

III. Results

Table 1 shows the best results of classifying the signals from the test sets using each of the techniques presented in Section II. The notation used is the following: CSP* for CSP, MCD* for MCD, DS* for Stahel-Donoho (using the weight functions: MDC (DSM*), Tukey's biweight (DST*) and Huber's (DSH*)) and MM* for MM-estimator; *1 for the scale estimator using MAD and *2 for $\overline{MAD}$; and *s when the m features are selected using scores instead of the extreme eigenvalues. Although in the BCI literature the usual number of features considered is between 2 and 6, in this work the number of pairs p (*pf*) was varied between 1 and 11 for the first dataset and between 1 and 30 for the last two ones; only the lowest p is reported in the case of ties for the best accuracies from the same method.

Results show that robust methods outperformed classical CSP for subjects with poor performances. In these cases, for most of them, the use of robust covariance matrices in CPS together the mean absolute deviation for scale and the features selected through the extreme eigenvalues produce the best classification accuracy. For the other subjects, this procedure provides an accuracy close to that of conventional method. This suggests that robust techniques are an alternative to improve the classification accuracy of CPS with LDA. However, determine the optimal method and the number of features suitable for classification is highly dependent on the individual.

IV. Conclusions

This study presents a comparison between the classical MI classification based on CSP and LDA and robust proposals. The methods are evaluated on MI data from 17 subjects. For all of them, better or similar performance in classification with classical CSP and RCSP (see [16]) are obtained. Results shows that for most subjects the best classifications are achieved using robust CSP.

The method with the best performance depends on the subject as it is seen from Table 1. Future work could be de-

Table 1 Classification accuracies obtained for each subject for the different methods. For each subject, the best result is displayed in bold characters

	BCI IV dataset IIa																	
	1		2		3		4		5		6		7		8		9	
	pf	%	pf	%	pf	%	pf	%	pf	%	pf	%	pf	%	pf	%	pf	%
CSP1	4	90.28	1	54.17	3	96.53	1	74.31	2	64.58	3	71.53	2	**84.03**	6	97.22	2	**93.75**
CSP1s	4	**92.36**	5	54.86	8	94.44	4	72.22	5	56.25	5	67.36	7	81.94	8	97.22	6	**93.75**
CSP2	3	88.19	1	54.17	5	96.53	1	75.00	2	65.97	3	69.44	5	81.94	5	96.53	1	92.36
CSP2s	4	90.28	2	57.64	3	95.14	4	76.39	4	56.94	5	68.75	3	76.39	9	96.53	11	93.06
MCD1	2	83.33	2	60.42	5	96.53	10	68.06	3	66.67	6	70.83	2	72.22	2	98.61	3	92.36
MCD1s	2	83.33	3	60.42	4	96.53	5	68.75	11	65.97	5	70.14	2	74.31	3	98.61	3	92.36
MCD2	3	88.19	3	**66.67**	7	**97.92**	1	76.39	6	67.36	6	67.36	2	79.17	8	97.92	3	93.06
MCD2s	3	88.19	3	**66.67**	4	**97.92**	2	74.31	3	67.36	6	68.06	2	79.17	7	98.61	2	93.06
DSM1	3	88.89	4	54.17	4	95.83	3	69.44	5	59.72	6	71.53	1	73.61	8	**99.31**	3	92.36
DSM1s	4	85.42	3	56.25	7	95.83	7	74.31	3	63.19	9	67.36	6	73.61	6	98.61	2	90.97
DSM2	3	89.58	2	55.56	5	96.53	10	70.83	3	61.81	4	**77.78**	2	78.47	8	98.61	3	93.06
DSM2s	5	88.89	7	58.33	3	96.53	6	70.14	3	61.11	8	73.61	4	73.61	8	98.61	2	92.36
DST1	1	88.19	3	57.64	6	96.53	8	72.22	2	64.58	9	70.83	3	75.69	6	**99.31**	3	91.67
DST1s	4	85.42	3	56.25	3	94.44	7	74.31	3	63.19	9	67.36	6	70.83	7	98.61	2	93.06
DST2	3	90.28	4	57.64	5	**97.92**	2	72.92	2	62.50	2	75.00	2	79.17	11	98.61	2	92.36
DST2s	5	88.89	7	58.33	3	96.53	6	70.14	7	61.81	8	73.61	4	76.39	8	**99.31**	4	**93.75**
DSH1	1	84.72	1	56.94	4	95.83	6	73.61	2	**73.61**	1	70.83	1	77.78	1	97.92	1	93.06
DSH1s	2	85.42	1	59.72	4	95.14	6	73.61	3	65.28	3	70.83	1	75.69	1	97.92	1	90.97
DSH2	3	88.19	1	56.25	3	95.83	6	75.00	2	65.28	1	75.69	11	75.69	5	97.22	1	92.36
DSH2s	2	90.28	1	63.89	4	96.53	6	76.39	1	63.19	2	69.44	4	75.69	6	97.22	1	92.36
MM1	2	85.42	3	56.25	5	95.83	1	72.92	2	67.36	2	66.67	1	75.69	5	98.61	2	91.67
MM1s	2	86.81	2	59.03	4	95.83	1	72.92	3	59.72	3	70.83	1	75.69	5	98.61	1	92.36
MM2	3	90.28	1	59.03	3	96.53	11	**77.78**	2	70.14	3	73.61	2	75.00	7	97.92	3	93.06
MM2s	2	91.67	2	57.64	4	97.22	11	**77.78**	10	62.50	4	62.50	3	77.08	6	**99.31**	1	93.06

	BCI III dataset IIIa						BCI III dataset IVa									
	1		2		3		1		2		3		4		5	
	pf	%	pf	%	pf	%	pf	%	pf	%	pf	%	pf	%	pf	%
CSP1	1	97.78	7	68.33	1	98.33	2	71.43	10	**100**	6	61.73	9	78.13	10	67.46
CSP1s	4	95.56	8	63.33	6	91.67	26	74.11	24	98.21	14	57.65	4	59.82	5	58.73
CSP2	2	97.78	6	73.33	2	**100**	2	74.11	10	**100**	5	57.14	3	78.13	10	76.19
CSP2s	5	95.56	16	61.67	7	91.67	12	74.11	8	98.21	6	55.61	6	61.61	5	58.73
MCD1	2	98.89	1	66.67	2	98.33	2	70.54	10	**100**	7	60.71	4	82.59	3	87.30
MCD1s	3	97.78	17	68.33	4	93.33	16	71.43	12	**100**	4	64.29	11	77.23	5	73.41
MCD2	2	98.89	14	73.33	2	**100**	2	74.11	8	**100**	8	64.29	6	80.80	7	82.94
MCD2s	3	98.89	16	68.33	4	91.67	7	67.86	22	**100**	4	64.29	10	78.13	5	74.21
DSM1	2	98.89	7	63.33	2	98.33	1	68.75	16	**100**	1	**73.98**	1	86.61	2	87.30
DSM1s	4	94.44	10	70.00	16	90.00	1	78.57	7	98.21	5	56.12	17	66.52	24	61.90
DSM2	3	**100**	10	66.67	2	98.33	1	69.64	27	**100**	3	71.94	3	85.71	2	82.54
DSM2s	7	94.44	7	66.67	11	86.67	1	**80.36**	16	**100**	7	57.14	20	71.88	9	56.35
DST1	2	98.89	1	61.67	2	98.33	1	73.21	8	**100**	2	69.90	1	**87.95**	1	**88.89**
DST1s	6	94.44	7	58.33	7	93.33	1	76.79	2	**100**	18	59.18	1	53.57	10	71.83
DST2	2	98.89	20	71.67	2	98.33	1	75.00	3	98.21	3	**73.98**	2	86.61	3	84.92
DST2s	4	93.33	8	61.67	10	88.33	1	77.68	6	98.21	17	58.67	8	54.91	11	67.46
DSH1	2	98.89	10	68.33	2	98.33	1	76.79	25	**100**	4	70.92	1	86.16	2	85.71
DSH1s	4	97.78	11	71.67	5	93.33	29	67.86	17	96.43	15	58.16	3	80.36	3	53.57
DSH2	1	97.78	13	**75.00**	2	98.33	1	72.32	26	**100**	2	66.33	1	80.36	2	82.94
DSH2s	3	96.67	12	73.33	9	93.33	2	67.86	6	98.21	12	61.22	3	81.70	21	57.54
MM1	2	96.67	17	68.33	2	**100**	9	74.11	1	96.43	6	71.43	2	86.61	2	79.37
MM1s	2	96.67	16	71.67	3	90.00	11	65.18	2	94.64	6	69.90	13	70.54	11	63.49
MM2	3	**100**	12	70.00	2	98.33	7	75.00	24	98.21	4	70.41	2	85.27	2	76.59
MM2s	2	96.67	13	73.33	3	95.00	7	67.86	4	96.43	17	67.86	12	73.66	15	62.70

veloping techniques to determine from the training data which is the most convenient method among the proposed in this paper for classifying MI signals.

Conflict of Interest

The author declare that she has no conflict of interest.

References

1. Hoffman, A., Farkaš, I. (2013) Using common spatial patterns for EEG feature selection. Technical report TR-2013-040. Comenius University in Bratislava.
2. Fisher, R. A. (1936) The use of Multiple Measurements in Taxonomic Problems. Annals of Eugenics **7** (2): 179–188.
3. Maronna, R.A., Martin D. et al (2006), Robust Statistics, Theory and Methods, Wiley, New York.
4. Yong, X., Ward, R. K., et al (2008) Robust Common Spatial Patterns for EEG Signal Preprocessing. 30th Annual International Conference of the IEEE Engineering in Medicine and Biology Society: 2087 - 2090
5. Rousseeuw, P. J. (1984) Least medians of squares regression. Journal of American Statistical Association, vol. 79: 851–857
6. Stahel, W.A. (1981) Breakdown of covariance estimators. Research Report 31, Fachgruppe für Statistik, E.T.H. Zürich, Switzerland.
7. Donoho, D.L. (1982) Breakdown Properties of Multivariate Location Estimators, Ph.D. dissertation, Harvard University
8. Lopuhaä, H. P. (1992) Highly efficient estimators of multivariate location with high breakdown point. The annals of Statistics 1992 Vol 20 No. 1: 398-413.
9. Kawanabe, M., Vidaurre, C. (2009) Improving BCI performance by modified common spatial patterns with robustly averaged covariance matrices. World Congress on Medical Physics and Biomedical Engineering, Munich, Germany Vol. 25/9 IFMBE Proceedings: 279-282.
10. Blankertz, B., Tomioka, R. et al (2008) Optimizing Spatial Filters for Robust EEG Single-Trial Analysis. IEEE signal processing magazine, vol. XX: 41-56.
11. Naeem, M., Brunner, C. et al (2006) Separability of four-class motor imagery data using independent components analysis, Journal of Neural Engineering, Vol. 3, no. 3: 208-216.
12. Schlögl, A., Lee, F. et al (2005) Characterization of four-class motor imagery EEG data for the BCI-competition 2005. Journal of Neural Engineering, Vol. 2, no. 4: 14-22.
13. Blankertz, B., Muller, K. et al (2006) The BCI competition III: Validating alternative approaches to actual BCI problems. IEEE Trans. Neur. Syst. Rehab, Vol. 14: 153-159.
14. RCSP toolbox at sites.google.com/site/fabienlotte/code-and-softwares
15. FSDA at ec.europa.eu/jrc/en/scientific-tool/fsda-matlab-code/downloads
16. Lotte, F. Guan, C. T. (2011) Regularizing Common Spatial Patterns to Improve BCI Designs: Unified Theory and New Algorithms. IEEE Transactions on Biomedical Engineering, vol. 58, no. 2: 355-362