ESANN 2016 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Bruges (Belgium), 27-29 April 2016, i6doc.com publ., ISBN 978-287587027-8.
Available from http://www.i6doc.com/en/.

Unsupervised Cross-Subject BCI Learning and Classification using Riemannian Geometry

Samaneh Nasiri Ghosheh Bolagh[1], Mohammad Bagher Shamsollahi[1]
Christian Jutten[2], Marco Congedo[2]

1- Biomedical Signal and Image Processing Lab (BiSIPL)
Department of Electrical Engineering, Sharif University of Technology, Tehran, Iran

2- Grenoble Image Parole Signal Automatique (GIPSA) Laboratory
CNRS, University of Grenoble-Alpes, Grenoble Institute of Technology, Grenoble, France

Abstract. The inter-subject variability poses a challenge in cross-subject Brain-Computer Interface learning and classification. As a matter of fact, in cross-subject learning not all available subjects may improve the performance on a test subject. In order to address this problem we propose a subject selection algorithm and we investigate the use of this algorithm in the Riemannian geometry classification framework. We demonstrate that this new approach can significantly improve cross-subject learning without the need of any labeled data from test subjects.

1 Introduction

Predicting the stimulus from concurrent brain neuroimaging data is an approach that can be used to understand underlying mental processes [1]. It is well known that brain electroencephalographic (EEG) signals are subject-specific, thus brain decoding models are traditionally designed individually: training and test data belongs to the same subject [1]. However, collecting sufficient labeled data from new subjects for learning a classifier is an expensive and time consuming procedure, jeopardizing the participants' willingness to use the system [2]. Given the abundance of labeled data from other subjects, it is tempting to use them for training a classifier. Nevertheless, not all such training subjects may improve the performance on a given test subject because of inherent inter-subject variability [2]. To address this problem, we propose an algorithm, named Rank Of Subjects (ROS), which ranks training subjects and selects a relevant subset. The proposed method has the same structure as the subject selection algorithm proposed in [1]. The relevant subjects are selected through the maximization of the accuracy that is obtained by applying the classifier trained on data from all subjects to the unlabeled data of the test subject.

In [3] a multi-class BCI classification framework based on Riemannian Geometry has been proposed. The idea is to use appropriate forms of data covariance matrices as features and classifying using the simple concepts of distance and center of mass of these covariance matrices on a Riemannian manifold [4]. Thanks to the use of an appropriate metric, this approach has displayed excellent cross-subject generalization capabilities, besides robustness to EEG artifacts, outliers and mislabeling [4].

In this article, we address the problem of cross-subject training without assuming knowledge of labels from the test subjects. The proposed method is a

ESANN 2016 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Bruges (Belgium), 27-29 April 2016, i6doc.com publ., ISBN 978-287587027-8. Available from http://www.i6doc.com/en/.

combination of previously proposed subject selection algorithms and some fundamental concepts of Riemannian Geometry. The performance of our algorithm is assessed on the open "DecMeg2014" Kaggle dataset (www.kaggle.com).

2 Rank Of Subjects

Due to inter-subject variability, the distribution of data in the training set may be more or less different from the test subject. A cross-subject classifier may be trained using only those training subjects that are similar to the test subject, according to some criterion [1]. We propose to rank the training subjects and select a subset of them according to Algorithm 1. In this algorithm, first we construct virtual labels for the test subject by training a SVM (support-vector machine) classifier using labeled data from all training subjects. Then, the most relevant training subjects are selected by maximizing the accuracy obtained on the virtual labels. Accuracy is returned by the function $J(X_k)$ when the trained SVM classifier on X_k is applied on test data. Finally, among the available $\mathbf{N}_s$ training subjects, we select the subset of size $\mathbf{N}_R$ featuring highest accuracy.

Algorithm 1 Rank Of Subjects (ROS)

Input: $\mathbf{X}_t^u$ Unlabeled Signal from Target Subject
Input: $\left\{\mathbf{X_s^1}\right\}_{s=1}^{N_s}$ Labeled Signal from $\mathbf{N}_s$ Source Subjects
Output: $\left\{\mathbf{X_s^1}\right\}_{s=1}^{N_R}$ Relevant Subjects

Train a model by concatenated labeled data from all source subjects
Generate virtual label $\mathbf{V}_u$ by applying trained model on target data
Initialize $\mathbf{accuracy}_0 = 0$, $K = 1$, $\mathbf{X}_0 = \{\}$,
while $\mathbf{n} < \mathbf{N_s}$ **do**
  Step 1 (Inclusion)
  $[\mathbf{x}^+, \mathbf{accuracy_{max}}] = \underset{\mathbf{x}\in[\{\mathbf{X_s^1}\}_{s=1}^{N_s} - \mathbf{X_k}]}{\arg\max} J(\mathbf{X_k} + \mathbf{x})$
  $\mathbf{X_{k+1}} = \mathbf{X_k} + \mathbf{x}^+$, $k = k + 1$
  $accuarcy_k = accuracy_{max}$
  Step 2 (Exclusion)
  if $k > 2$ **then**
    $[\mathbf{x}^-, \mathbf{accuracy_{max}}] = \underset{\mathbf{X_k}}{\arg\max}\, J(\mathbf{X_k} - \mathbf{x})$
    if $accuracy_{max} > accuarcy_{k-1}$ **then**
      $\mathbf{X_{k-1}} = \mathbf{X_k} - \mathbf{x}^-$, $k = k - 1$
      $accuarcy_{k-1} = accuracy_{max}$
      go to step 2
    else
      go to step 1
    end if
  end if
end while
$Best\_N = \underset{\mathbf{s}\in[\mathbf{1}:\mathbf{N_s}]}{\arg\max}\, accuracy_k$
$ROS = \mathbf{X}_{Best\_N}$
return the first $\mathbf{N}_R$ subjects

ESANN 2016 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Bruges (Belgium), 27-29 April 2016, i6doc.com publ., ISBN 978-287587027-8. Available from http://www.i6doc.com/en/.

3 Spatial Filtering

In order to enhance the signal-to-noise ratio (SNR) of EEG evoked potentials, we use a standard spatial filtering approach. Let $\mathbf{X}_i \in \Re^{C\times N}$ denote a trial indexed by i, with C the number of channels, N the number of time samples and y_i the class label of the trial. For each class, a set of N_{filter} spatial filters are built. The spatial filters are an adaptation of the well-known common spatial pattern to evoked-potential data [6], here implemented following the winner of the "DecMeg2014" competition [8]. The average trial of the class k is denoted by $\mathbf{P}^{(k)}$ [8]:

$$\mathbf{P}^{(k)} = \frac{1}{|\mathcal{I}^{(k)}|} \sum_{i\in\mathcal{I}^{(k)}} \mathbf{X}_i, \tag{1}$$

where $\mathcal{I}^{(k)} = \{i \mid y_i = k\}$ is the set of indices of the trials belonging to class k. The spatial filter $\mathbf{w} \in \Re^{C\times 1}$ for each class k solves the following optimization problem:

$$\mathbf{w}^* = \arg\max_{\mathbf{w}} \frac{\mathbf{w}^T\mathbf{P}^{(k)}{\mathbf{P}^{(k)}}^T\mathbf{w}}{\mathbf{w}^T\mathbf{X}\mathbf{X}^T\mathbf{w}}, \tag{2}$$

where $\mathbf{X}$ is the matrix holding the continuous EEG recording (here we consider the concatenation of all trials from all classes). This equation is a generalized Rayleigh quotient, the solutions to which can be found as the eigenvector matrix of matrix $(\mathbf{X}\mathbf{X}^T)^{-1}(\mathbf{P}^{(k)}\mathbf{P}^{(k)T})$. By construction, the C resulting eigenvectors are ranked by SNR. For each class, only the first N_{filter} vectors are selected. Therefore, for each class k, the spatial filters is $\mathbf{W}^{(k)} \in \Re^{C\times N_{filter}}$. Spatial filtering operation is simply done by linear projection of the trial by the matrix $\mathbf{W} = [\mathbf{W}^{(0)}, \mathbf{W}^{(1)}, \ldots, \mathbf{W}^{(K)}] \in \Re^{C\times(K*N_{filter})}$, which is the aggregation of the K spatial filters for each class in a single matrix, such as

$$\mathbf{Z}_i = \mathbf{W}^T\mathbf{X}_i. \tag{3}$$

The usual covariance matrices do not hold the temporal structure of ERP trials. Therefore, in order to keep all spatial and temporal information a special estimation of the covariance matrix is used [4]. We build a new trial $\tilde{\mathbf{Z}}_i \in \Re^{(2*K*N_{filter})\times T}$ by concatenation of the spatially filtered response obtained by averaging several single trial responses of one class $\mathbf{P}^{(k)}$ and the spatially filtered trial $\mathbf{Z}_i$ [8]:

$$\tilde{\mathbf{Z}}_i = \left[\ {\mathbf{W}^{(0)}}^T\mathbf{P}^{(0)}, {\mathbf{W}^{(1)}}^T\mathbf{P}^{(1)}, \ldots, {\mathbf{W}^{(K)}}^T\mathbf{P}^{(K)}, \mathbf{Z}_i\ \right]^T. \tag{4}$$

These "super" trials are used to build the feature covariance matrices. The feature covariance matrices are obtained simply by using a Sample Covariance Matrix (SCM) estimator [7], such as

$$\mathbf{\Sigma}_i = \frac{1}{N}\tilde{\mathbf{Z}}_i\tilde{\mathbf{Z}}_i^T. \tag{5}$$

The covariance matrices in (5) belong to the space of Symmetric Positive Definite (SPD) matrices. The space of SPD matrices forms a Riemannian manifold of

ESANN 2016 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Bruges (Belgium), 27-29 April 2016, i6doc.com publ., ISBN 978-287587027-8. Available from http://www.i6doc.com/en/.

non-positive curvature [7]. Therefore, we can use tools issued from differential geometry on Riemannian manifolds to manipulate them.

Riemannian Distance: For any two covariance matrix $\mathbf{\Sigma}_1$ and $\mathbf{\Sigma}_2$, the Riemannian distance according to the Riemannian metric is given by [9]

$$\delta_R(\mathbf{\Sigma}_1, \mathbf{\Sigma}_2) = \|\log\left(\mathbf{\Sigma}_1^{-1/2}\mathbf{\Sigma}_2\mathbf{\Sigma}_1^{-1/2}\right)\|_F = \left[\sum_{c=1}^{C}\log^2\lambda_c\right]^{1/2} \tag{6}$$

where $\lambda_c, c = 1 \dots C$ are the real eigenvalues of $\mathbf{\Sigma}_1^{-1/2}\mathbf{\Sigma}_2\mathbf{\Sigma}_1^{-1/2}$ and C the number of channels. This distance is *Affine-invariant* [3], meaning that is *invariant* with respect to similar and congruent transformation, but also to inversion.

Riemannian Mean: The Riemannian geometric mean of I covariance matrices (denoted by $\mathfrak{G}(.)$), also called Fréchet or Karcher mean, is the point on the manifold minimizing the dispersion [3], i.e.

$$\mathfrak{G}\left(\mathbf{\Sigma}_1, \dots, \mathbf{\Sigma}_I\right) = \underset{\mathbf{\Sigma}}{\operatorname{argmin}} \sum_{i=1}^{I} \delta_R^2\left(\mathbf{\Sigma}, \mathbf{\Sigma}_i\right). \tag{7}$$

There is no closed form expression for this mean for $I > 2$, however a gradient descent in the manifold can be used in order to find the solution [5].

Algorithm 2 Rank Of Subjects and Minimum Distance to Riemannian Mean (ROS+MDRM)

Input: $\mathbf{X}_t^u$ Unlabeled Signal from Target Subject
Input: $\left\{\mathbf{X_s^1}\right\}_{s=1}^{N_s}$ Labeled Signal from $\mathbf{N}_s$ Source Subjects
Input: Source Subjects who are more relevant to Target Subject as selected by the ROS algorithm
Output: The label of target subject
Train a classifier on labeled data from relevant source subjects
Initialize target labels by applying the designed classifier to Target Subject
while $y_{i,(n+1)} = y_{i,n}$ **do**
    Train spatial filters given the labels $y_{i,n}$ of the trial i at the iteration n.
    Apply the spatial filters on the trials
    Estimate special form covariance matrices $\mathbf{\Sigma}_i$
    Obtain K mean covariance matrix for each class: $\mathbf{\Sigma}_{\mathfrak{G}}^{(0)}, \dots, \mathbf{\Sigma}_{\mathfrak{G}}^{(K)}$, Eq.(7)
    Classify each trial according to the Riemannian distance
    $\hat{\mathbf{y}} = \underset{\mathbf{k}}{\operatorname{argmin}}\, \delta_{\mathbf{R}}(\mathbf{\Sigma}_{\mathfrak{G}}^{(\mathbf{k})}, \mathbf{\Sigma_i})$, Eq.(6)
end while
return $\hat{\mathbf{y}}$

4 Results

We evaluate the proposed method on the magnetoencephalography (MEG) data set used in the "DecMeg2014" Kaggle competition. This dataset is comprised of 16 subjects. During the experiment participants observed two visual stimuli: **face** and **scrambled face**. Approximately 580-590 trials have been recorded for each subject. The duration of each trial was 1.5 seconds, with the stimulus presented at time 0.5 second. The class label was either Face (class 1) or Scramble Face (class 0).

As pre-processing, the original MEG recording (306 channels) has been downsampled at 250Hz and high-pass filtered at 1Hz. We use the SVM^{light} toolbox

ESANN 2016 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Bruges (Belgium), 27-29 April 2016, i6doc.com publ., ISBN 978-287587027-8. Available from http://www.i6doc.com/en/.

for training SVM classifier and constructing virtual test labels. The standard SVM uses the default parameter tuning. In the ROS algorithm, we select the eight most relevant subjects (half of available training subjects). We compare result of standard SVM classifier with ROS and without ROS. The performance of the proposed method is evaluated by means of percent classification accuracy on the test data using a leave-one-out procedure (LOO).

The first column of Table 1 reports the accuracy of a standard SVM classifier. In this method we pool all training data from all training subjects and train a SVM classifier on such labeled data and then apply on test data. This is a generic classifier and it achieves an accuracy of 69.85 % . For the second column (SVM+ROS), we train SVM classifier on training data from a selected subset of training subjects as per the ROS algorithm; this generic classifier reaches 71.47% accuracy. In the third column (MDRM), for each class of test data, a set of four spatial filters are built as per Eq.(1), where in the equation $P^{(k)}$ is constructed based on virtual labels. We estimate covariance matrices by Eq.(5). Each trial is classified according to the Riemannian distance from two mean covariance matrices for each class $\mathbf{\Sigma}_{\mathfrak{G}}^{(0)}$ and $\mathbf{\Sigma}_{\mathfrak{G}}^{(1)}$. This procedure is iterated until convergence. Results of the combination ROS+MDRM are provided in the last column. The combination ROS+MDRM achieves a much higher performance (**77.94%**) as compared to the other methods. The performance improvement is **2.14%** compared to MDRM [8]. The results show **10 %** improvement compared to [10] which used another cross-subject transfer learning approach to classify this same dataset.

Subject	**SVM**	**SVM+ROS**	**MDRM** [8]	**ROS+MDRM**
1	77.27	79.28	64.3	**84.67**
2	68.94	70.94	72.7	**74.74**
3	62.46	62.93	**63.3**	60.72
4	79.63	85.11	86.5	**91.24**
5	67.06	70.11	74.2	**76.11**
6	66.50	66.81	**73.8**	69.38
7	71.77	78.89	79.9	**88.60**
8	67.74	69.24	76.6	**78.20**
9	76.09	74.73	79.6	**86.19**
10	68.81	70.32	**78.9**	75.25
11	72.47	73.98	59.9	**72.62**
12	74.23	73.68	**83.7**	79.18
13	68.62	73.28	73.4	**81.95**
14	69.59	66.81	**87.0**	76.68
15	68.62	72.20	**87.0**	81.72
16	58.14	61.17	**72.5**	69.83
Mean ±(std)	69.85 ± (5.5)	71.47 ± (6.32)	75.8 ± (8.3)	**77.94** ± (7.84)

Table 1: Classification accuracy in LOO Cross-Validation

ESANN 2016 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Bruges (Belgium), 27-29 April 2016, i6doc.com publ., ISBN 978-287587027-8. Available from http://www.i6doc.com/en/.

5 Conclusion

In this paper we have proposed a method for brain decoding with cross-subject learning. It is well-known that EEG and MEG signals are very specific to each subject. As a result, establishing a generic model featuring high classification performance is a very difficult task due to large inter-variability between subjects. We proposed a ROS algorithm for ranking automatically the relevant training subjects and selecting the best training subset. By combination of this method with Riemannian Geometry we have simply classified each trial based on distance of each trial from the two class-related mean covariance matrices. By evaluating the proposed algorithm on the MEG dataset of the "DecMeg2014" competition we have shown that this approach can outperform other methods. In future works the methods will be tested on other datasets and we will study how the mismatching between training and test subjects can be reduced.

Acknowledgment

This research has been partially supported by the European project ERC-2012-AdG-320684-CHESS.

References

[1] Lotte, F., Guan, C. "Regularizing Common Spatial Patterns to Improve BCI Designs: Unified Theory and New Algorithms,"Biomedical Engineering, IEEE Transactions on , vol.58, no.2, pp.355,362, Feb. 2011

[2] Lotte, F., Guan, C. "Learning from other subjects helps reducing Brain-Computer Interface calibration time,"Acoustics Speech and Signal Processing (ICASSP), 2010 IEEE International Conference on , vol., no., pp.614,617, 14-19 March 2010

[3] Barachant, A., Bonnet, S., Congedo, M., Jutten, C., "Multiclass Brain-Computer Interface Classification by Riemannian Geometry,"Biomedical Engineering, IEEE Transactions on , vol.59, no.4, pp.920,928, April 2012

[4] Congedo, M. "EEG Source Analysis ", HDR presented at Doctoral School EDISCE, University of Grenoble, 2013.

[5] Congedo, M., Afsari, B., Barachant, A., Moakher, M. "Approximate Joint Diagonalization and Geometric Mean of Symmetric Positive Definite Matrices, ", PLoS ONE 10(4): e0121423 2015

[6] Rivet, B., Souloumiac, A., Attina, V., Gibert, G., "xDAWN Algorithm to Enhance Evoked Potentials: Application to Brain Computer Interface,"Biomedical Engineering, IEEE Transactions on , vol.56, no.8, pp.2035,2043, Aug. 2009

[7] A. Barachant, S. Bonnet, M. Congedo and C. Jutten, "Classification of covariance matrices using a Riemannian-based kernel for BCI applications,"in NeuroComputing, vol. 112, p. 172-178, 2013

[8] Barachant, A. "MEG decoding using Riemannian Geometry and Unsupervised classification," 2014

[9] Moakher, M. "A differential geometric approach to the geometric mean of symmetric Positive-Definite matrices,"SIAM J. Matrix Anal. Appl., vol. 26, no. 3, pp. 735-747, 2005

[10] Olivetti, E. Kia, S.M. Avesani, P., "MEG decoding across subjects,"Pattern Recognition in Neuroimaging, 2014 International Workshop on , vol., no., pp.1,4, 4–6, June 2014