Learning from other Subjects Helps Reducing Brain-Computer Interface Calibration Time

Fabien Lotte, Cuntai Guan

HAL Id: inria-00441670
https://hal.inria.fr/inria-00441670

Submitted on 17 Dec 2009

LEARNING FROM OTHER SUBJECTS HELPS REDUCING BRAIN-COMPUTER INTERFACE CALIBRATION TIME

Fabien LOTTE, Cuntai GUAN

Institute for Infocomm Research (I²R), Singapore

ABSTRACT

A major limitation of Brain-Computer Interfaces (BCI) is their long calibration time, as much data from the user must be collected in order to tune the BCI for this target user. In this paper, we propose a new method to reduce this calibration time by using data from other subjects. More precisely, we propose an algorithm to regularize the Common Spatial Patterns (CSP) and Linear Discriminant Analysis (LDA) algorithms based on the data from a subset of automatically selected subjects. An evaluation of our approach showed that our method significantly outperformed the standard BCI design especially when the amount of data from the target user is small. Thus, our approach helps in reducing the amount of data needed to achieve a given performance level.

Index Terms— Brain-Computer Interfaces (BCI), subject-to-subject transfer, regularization

1. INTRODUCTION

Brain-Computer Interfaces (BCI) are communication systems that enable subjects to send commands to computers by using only their brain activity [1]. Most existing BCI are based on ElectroEncephaloGraphy (EEG) as the measure of brain activity [1]. A major limitation of BCI is that many examples of the subject's EEG signals must be recorded in order to calibrate the BCI, which is inconvenient and time consuming. Indeed, due to large inter-subject variabilities, each BCI must be tuned specifically for the corresponding subject.

To our best knowledge, few studies have addressed the problem of calibration time reduction for BCI. Exceptions are the works in [2, 3]. In [2], the calibration time is reduced by using data recorded from the same user, during previous BCI sessions. As such, this method cannot be used for a new BCI user, with no previous data available. In [3], a BCI is first built using very few signals from the target user. Then the BCI is adapted online using unsupervised learning. The limitation is that this BCI will have initially limited performance, becoming good only after a significant adaptation time.

In this paper, we propose to use EEG data from other subjects in the calibration process of the BCI. This aims at building an initially good BCI by using only a small amount of data from the new target user, hence reducing the calibration time. This paper is organized as follows: Section 2 describes the motivation and the principle of our approach. Then, Section 3 describes the evaluation of the proposed method and the obtained results. Finally, Section 4 concludes the paper.

2. METHOD

Our hypothesis is that despite large inter-subject variabilities, it is still possible to find common information in the EEG signals of some subjects. Thus, such information from additional subjects could be used to design a robust BCI even when little data from the target subject is available.

In this paper, we propose to use data from other subjects in the learning process of the Common Spatial Patterns (CSP) [4] and Linear Discriminant Analysis (LDA) algorithms [5]. These are indeed among the most popular and efficient algorithms used for BCI design [5, 6, 4]. However, it should be mentioned that our approach can be easily applied to other similar machine learning techniques.

2.1. The CSP and LDA algorithms

The CSP algorithm aims at learning spatial filters which can maximize the variance of the EEG signals from one class, while minimizing the variance of the EEG signals from the other class, hence achieving optimal discrimination [4]. More formally, when training the CSP, the obtained spatial filters w are the filters which maximize the following function:

$$\frac{wC_1w^T}{w(C_1 + C_2)w^T} \qquad (1)$$

where C_i is the spatial covariance matrix of the EEG signals from class i. This optimization problem is solved by jointly diagonalizing the two matrices C_1 and $(C_1 + C_2)$ [4]. Once the filters w have been obtained, CSP feature extraction consists in spatially filtering the EEG signals using the w and then computing the variance of the resulting signals.

The LDA classifier uses a linear hyperplane to separate feature vectors from two classes [5]. The intercept b and normal vector a of this hyperplane are computed as follow:

$$b = -\frac{1}{2}(\mu_1 + \mu_2)C^{-1}(\mu_1 - \mu_2)^T \; and \; a = C^{-1}(\mu_1 - \mu_2)^T \qquad (2)$$

with μ_1 and μ_2 being the mean feature vector for each class and C the covariance matrix of both classes. With LDA, for an input feature vector x, the classification output is $a^T x + b$.

2.2. Learning from other subjects

Both the CSP and LDA approaches are based on the estimation of the class covariance matrices. Much data is required to obtain proper estimates of these matrices, explaining why much training data is required to design a robust BCI. In order to train the CSP and LDA with less training data, hence reducing the BCI calibration time, we propose to incorporate information from other subjects as a regularization term in the estimation of the covariances matrices:

$$\tilde{C}_t = (1 - \lambda)C_t + \lambda \frac{1}{|S_t(\Omega)|} \sum_{i \in S_t(\Omega)} C_i \qquad (3)$$

where C_t is a covariance matrix estimated from the data of the target subject, Ω is a set of subjects whose data has been recorded previously, $|A|$ is the number of elements in set A, λ is the regularization parameter ($\lambda \in [0, 1]$), $S_t(\Omega)$ is a subset of subjects from Ω, and C_i is a covariance matrix estimated from the data of subject i. The aim of this regularization scheme is to obtain a better and more robust estimate of the covariance matrix by using covariance matrices from other subjects for which much data is available. This should lead to better classification performances, especially when little data is available for the target subject.

With our approach, incorporating information from other subjects into the training of CSP or LDA simply consists in replacing the covariance matrices from Eq. 1 and 2 by their regularized estimates from Eq. 3. To train the LDA classifier, the mean feature vectors μ are similarly regularized:

$$\tilde{\mu}_t = (1 - \lambda)\mu_t + \lambda \frac{1}{|S_t(\Omega)|} \sum_{i \in S_t(\Omega)} \mu_i \qquad (4)$$

This approach raises two questions: 1) How to select a good subset of additional subjects $S_t(\Omega)$ and 2) how to efficiently select the value of the regularization parameter λ? These two questions are addressed in the following sections.

2.3. Selecting a relevant subset of additional subjects

Even if data from many subjects is available, all these subjects may not be relevant to use due to the huge inter-subject variability. Rather, we should select a subset of subjects whose data can be used to classify the target subject's data. To do so, we propose the subject selection algorithm described in Algorithm 1. In this algorithm, the function $accuracy = trainThenTest(trainingSet, testingSet)$ returns the accuracy obtained when training the CSP and LDA on the data set *trainingSet* and testing them on data set *testingSet*. The function $(best_i, max_{f(i)}) = max_i \ f(i)$

returns $best_i$, the value of i for which $f(i)$ is maximum (this maximum value is denoted $max_{f(i)}$).

Input: D_t: training EEG data from the target subject.
Input: $\Omega = \{D_s\}, s \in [0, N_s]$: set of EEG data from the N_s other subjects available ($D_t \ni \Omega$).
Output: $S_t(\Omega)$: a subset of relevant subjects whose data can be used to classify the data D_t of the target subject.
Output: $selectedAcc$: the accuracy obtained when classifying D_t using a BCI trained on the subset of selected additional subjects $S_t(\Omega)$.
$selected_0 = \{\}$;
$remaining_0 = \Omega$;
$accuracy_0 = 0; n = 1$;
while $n < N_s$ **do**
    Step 1: $(bestSubject, bestAccuracy) = max_{s \in remaining_{n-1}} \quad trainThenTest(selected_{n-1} + \{D_s\}, D_t)$;
    $selected_n = selected_{n-1} + \{D_{bestSubject}\}$;
    $remaining_n = remaining_{n-1} - \{D_{bestSubject}\}$;
    $accuracy_n = bestAccuracy$;
    $n = n + 1$;
    Step 2: if $n > 2$ **then**
        $(bestSubject, bestAccuracy) = max_{s \in selected_n} \quad trainThenTest(selected_n - \{D_s\}, D_t)$;
        if $bestAccuracy > accuracy_{n-1}$ **then**
            $selected_{n-1} = selected_n - \{D_{bestSubject}\}$;
            $remaining_{n-1} = remaining_n + \{D_{bestSubject}\}$;
            $accuracy_{n-1} = bestAccuracy$;
            $n = n - 1$;
            go to Step 2;
        else
            go to Step 1;
        end
    end
end
$(bestN, selectedAcc) = max_{n \in [1, N_s]} \quad accuracy_n$;
$S_t(\Omega) = selected_{bestN}$;
    Algorithm 1: Subject selection algorithm

In short, this algorithm sequentially selects the subject to add or to remove from the current subset of subjects, in order to maximize the accuracy obtained when training the BCI on the data from this subset of subjects and testing it on the training data of the target subject. This algorithm has the same structure as the Sequential Forward Floating Search algorithm [7], used to select a relevant subset of features. This ensures the convergence of our algorithm as well as the selection of a good subset of additional subjects.

2.4. Selecting an appropriate regularization parameter

Another problem to solve is the selection of the regularization parameter λ in Eq. 3 and 4. This parameter specifies the proportion of data from the additional subjects that will be

used in the learning process. Here, selecting λ using cross-validation (CV) is not recommended because 1) CV is a very time consuming approach, and 2) our method is aimed at being used with a small training set for the target subject. Therefore, it would not be reasonable to waste some precious training data when dividing the data set using CV. Rather, we propose the following computationally efficient heuristic to select an appropriate value for λ:

1. Compute the classification accuracy $targetAcc$ obtained when using only the EEG signals from the target subject to train the BCI, by using Leave-One-Out Validation (LOOV) on the target subject training set. Indeed, LOOV removes only a single training example for evaluation.

2. Compute the classification accuracy $selectedAcc$ obtained when classifying the target subject training set with a BCI trained using only the EEG signals from the selected subjects in $S_t(\Omega)$ (see Algorithm 1).

3. Compute the accuracy $RandAcc$ obtained by a classifier performing at the chance level (e.g., 50% for 2 classes).

4. Determine the value of λ as follows:

$$\lambda = \begin{cases} 1 & targetAcc \leq RandAcc \\ 0 & targetAcc \geq selectedAcc \\ \frac{selectedAcc - targetAcc}{100 - RandAcc} & \text{otherwise} \end{cases}$$
(5)

This heuristic considers that if the data from the target subject is not appropriate to train a satisfactory BCI ($targetAcc \leq RandAcc$), then we should rely only on the data from the other subjects ($\lambda = 1.0$). If the classification performance expected when using only the data from the target user is better than that expected when using only the other subjects ($targetAcc \geq selectedAcc$), then we should rely only on the data from the target subject ($\lambda = 0.0$). The remaining case corresponds to the accuracy expected when using only the other subjects being higher than that expected when using only the target subject ($targetAcc < selectedAcc$). In this case, the higher the difference between these two expected accuracies, the higher lambda should be as we should give more confidence to the EEG signals from the other subjects.

3. EVALUATION

3.1. EEG Data Set

Our approach was evaluated on data set 2a from BCI competition IV [8], provided by the Graz group [9]. This set comprises EEG signals from 9 subjects who performed left hand, right hand, foot and tongue Motor Imagery (MI). The EEG signals were recorded using 22 EEG channels. For the purpose of this study, only EEG signals corresponding to left and right hand MI were used. EEG signals were band-pass filtered in the 8-30 Hz frequency band using a 5th order Butterworth filter. Indeed, this frequency band contains the main frequencies involved in MI [6].

A training and a testing set were available for each subject. Both sets contained 72 trials for each class, the duration of a trial being 7 sec. Subjects performed MI within the time interval of t=3 to 7 sec of each trial. In this work, we only considered the discrete classification of the trials, i.e., we assigned a class to each trial and not to each trial sample. In order to do so, we extracted the CSP features from the time segment t=3.5 to 5.5 sec of each trial, as this was the segment used by the BCI competition winner [8]. Concerning the CSP, we used the 3 most discriminative pairs of filters (corresponding to the highest eigenvalues) for feature extraction.

3.2. Results

We compared the performances obtained by our approach, which includes information from other subjects in the CSP and LDA training, with the standard BCI design, which trains the CSP and LDA using only the EEG data from the target subject. We also compared our approach with a regularized variant of CSP and LDA, based on diagonal loading. More precisely, we used regularized covariance matrices $\tilde{C}_i = C_i + \kappa I$ in Eq. 1 and 2, with I being the identity matrix and κ the regularization parameter, here automatically selected using Ledoit and Wolf's method [10]. To our best knowledge, this BCI design, based on diagonal loading in CSP and LDA, has not been studied in the literature.

To assess whether our approach could reduce the number of training examples required from the target subject, we performed these comparisons with different sizes for the training set of the target subject. When evaluating the three approaches on a given subject, this subject was used as the target subject and the 8 other subjects were used as the set of additional subjects Ω. Evaluations were performed while using only the N first trials from each class from the target subject's training set. The data used for the additional subjects in Ω was their full training set. Once the BCI trained, it was evaluated on the test set of the target user, for which all data were used, i.e, 72 trials per class. The average classification accuracies obtained (over all subjects) are displayed on Figure 1.

On the data set used, our results suggested that incorporating information from other subjects in the training of CSP and LDA can improve the classification accuracy of the BCI, especially when the amount of training data for the target subject is small. When a sufficient amount of data was available for training, e.g., for $N = 72$, our BCI design and the standard BCI design reached similar performances. However, when the amount of available data was very limited, our method outperformed the standard design. In the most striking case, for $N = 10$, our method reached an average accuracy nearly 10 % higher than that obtained with the standard design. Interestingly enough, a paired t-test revealed that the average classification accuracy (over all subjects and all N values) obtained by our BCI design was statistically higher than that obtained by both the standard BCI design

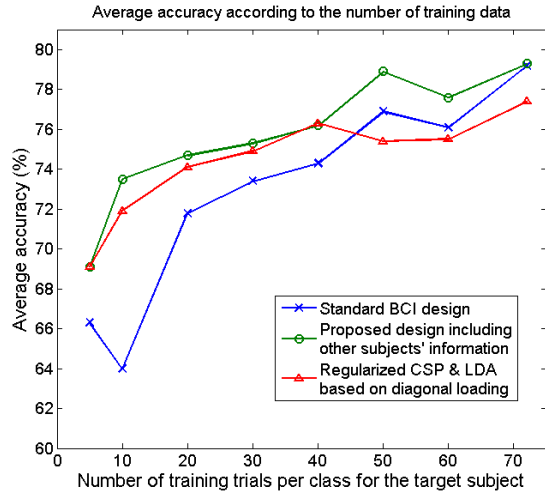Average accuracy according to the number of training data

Fig. 1. Average classification accuracies according to the number of training data available for the target subject.

($p < 0.001$) and the regularized BCI design based on diagonal loading ($p < 0.05$).

Results also suggested that this approach could be used to reduce the BCI calibration time for a new user, by reducing the amount of data required from this user. For example, on the data set used, in order to obtain the same classification accuracy as the one obtained by the standard design with 30 training trials per class, our approach needs only 10 trials per class (see Figure 1). Moreover, Blankertz et al suggested that approximately 40 trials per class were necessary to obtain reasonable BCI performances [11]. In this study, our approach obtained the same performance as the standard BCI design with 40 trials per class while using only 20 trials per class.

Finally, it should be noted that our approach is also computationally efficient. Indeed, on this data set, the average runtime was 92.9 s for N=72 data per class for the target subject, 55.3 s for N=40 and 36.4 s for N=10. The machine used was an Intel Core 2 Quad CPU @ 2.83 GHz with 4 Go RAM, Windows Vista and MATLAB 7.4.

4. CONCLUSION

In this paper, we proposed a method to reduce the BCI calibration time for a new user thanks to the incorporation of information from other subjects into the learning process of CSP and LDA. More precisely, we proposed to regularize the mean vectors and covariance matrices used in the CSP and LDA training with the mean vectors and covariance matrices estimated from a subset of relevant subjects. We proposed an algorithm to select automatically this relevant subset of subjects for a given target user, as well as a method to select automatically the extent of the regularization. The whole approach is easy to implement and computationally efficient. An evaluation of this approach on data set 2a from BCI competition IV showed that it could outperform both the standard BCI design and a regularized BCI based on diagonal loading, both using only the data from the target user. This improvement was generally higher when the amount of data from the user was small. This suggested that our approach can be used to train a BCI for a new user with less data from this user than the standard approach, hence reducing the calibration time.

Future work will be dedicated to study this method with a larger pool of subjects, including disabled subjects (e.g., stroke patients), in order to assess whether patients may also benefit from this approach. Indeed, it may be more difficult to find common information between stroke patients than between healthy users.

5. REFERENCES

[1] J.R. Wolpaw, N. Birbaumer, D.J. McFarland, G. Pfurtscheller, and T.M. Vaughan, "Brain-computer interfaces for communication and control," *Clinical Neurophysiology*, vol. 113, no. 6, pp. 767–791, 2002.

[2] M. Krauledat, M. Tangermann, B. Blankertz, and K.-R. Müller, "Towards zero training for brain-computer interfacing," *PLoS ONE*, vol. 3, no. 8, pp. e2967, 2008.

[3] Y. Li, C. Guan, H. Li, and Z. Chin, "A self-training semi-supervised svm algorithm and its application in an eeg-based brain computer interface speller system," *Pattern Recognition Letters*, vol. 29, no. 9, pp. 1285–1294, 2008.

[4] B. Blankertz, R. Tomioka, S. Lemm, M. Kawanabe, and K.-R. Müller, "Optimizing spatial filters for robust eeg single-trial analysis," *IEEE Signal Proc Magazine*, vol. 25, no. 1, pp. 41–56, 2008.

[5] F. Lotte, M. Congedo, A. Lécuyer, F. Lamarche, and B. Arnaldi, "A review of classification algorithms for EEG-based brain-computer interfaces," *Journal of Neural Engineering*, vol. 4, pp. R1–R13, 2007.

[6] G. Pfurtscheller and C. Neuper, "Motor imagery and direct brain-computer communication," *proceedings of the IEEE*, vol. 89, no. 7, pp. 1123–1134, 2001.

[7] P. Pudil, F. J. Ferri, and J. Kittler, "Floating search methods for feature selection with nonmonotonic criterion functions," *Pattern Recognition*, vol. 2, pp. 279–283, 1994.

[8] B. Blankertz, "BCI competition IV website," .

[9] M. Naeem, C. Brunner, R. Leeb, B. Graimann, and G. Pfurtscheller, "Seperability of four-class motor imagery data using independent components analysis," *Journal of Neural Engineering*, vol. 3, pp. 208–216, 2006.

[10] O. Ledoit and M. Wolf, "A well-conditioned estimator for large-dimensional covariance matrices," *Journal of Multivariate Analysis*, vol. 88, no. 2, pp. 365–411, 2004.

[11] B. Blankertz, F. Losch, M. Krauledat, G. Dornhege, G. Curio, and K.-R. Müller, "The berlin brain-computer interface: Accurate performance from first-session in bci-naive subjects," *IEEE Trans Biomed Eng*, vol. 55, no. 10, pp. 2452–2462, 2008.