Minimizing Subject-dependent Calibration for BCI with Riemannian Transfer Learning

Salim Khazem, Sylvain Chevallier, Quentin Barthélemy, Karim Haroun, Camille Noûs

HAL Id: hal-03202360
https://hal.uvsq.fr/hal-03202360

Submitted on 19 Apr 2021

Minimizing Subject-dependent Calibration for BCI with Riemannian Transfer Learning

Salim Khazem[1], Sylvain Chevallier[1], Quentin Barthélemy[2], Karim Haroun[1] and Camille Noûs[3]

Abstract— Calibration is still an important issue for user experience in Brain-Computer Interfaces (BCI). Common experimental designs often involve a lengthy training period that raises the cognitive fatigue, before even starting to use the BCI. Reducing or suppressing this subject-dependent calibration is possible by relying on advanced machine learning techniques, such as transfer learning. Building on Riemannian BCI, we present a simple and effective scheme to train a classifier on data recorded from different subjects, to reduce the calibration while preserving good performances. The main novelty of this paper is to propose a unique approach that could be applied on very different paradigms. To demonstrate the robustness of this approach, we conducted a meta-analysis on multiple datasets for three BCI paradigms: motor imagery, event-related potentials (P300) and SSVEP. Relying on the MOABB open source framework to ensure the reproducibility of the experiments and the statistical analysis, the results clearly show that the proposed approach could be applied on any kind of BCI paradigm and in most of the cases to significantly improve the classifier reliability. We point out some key features to further improve transfer learning methods.

I. INTRODUCTION

Brain-computer interfaces allow to interact with a computer or a robotic system without relying on muscular capacity. As EEG measurements could be made with relatively inexpensive or easily available devices when compared to other kind brain imaging techniques, it is a common choice to conduct studies on many subjects. While the BCI research is very active and diverse, the EEG-based BCI still suffers from important limitations [1]. Thus, they are mainly used for lab experiment and practical applications are scarce. Indeed, it is possible to enhance the preprocessing or the classification algorithms to better handle the individual variability [2]. Advanced signal processing and machine learning techniques are available but could not reduce the EEG-based BCI limitations completely.

An important leverage for improving the reliability of BCI is to improve the human-computer interfaces aspects [3]. The existing protocols are suboptimal and could be enhanced [4]. To address these issues, it is possible to build better model of the user and to better characterize what the user is learning [3]. One of the main sources of EEG variability is the user, due to his/her fatigue, mood change or attention variations.

The BCI protocols generate an important cognitive fatigue, especially if the user is maintaining a high attentional focalization during the interactions. To enforce the BCI reliability, it is common to include a calibration task for training preprocessing and machine learning algorithms. This calibration could be long; it is a source of fatigue and distraction, even before the starting to use the BCI. While setting up a subject independent calibration or a calibration-free system [5] is possible, the performances are not as good as subject-dependent calibration systems.

To reduce – and ideally to remove – calibration it is possible to rely on transfer learning. This is an important topic in machine learning and many approaches have been adapted and investigated in BCI [2]. As for classification algorithms, Riemannian approaches that rely on covariance matrices, have achieved robust and efficient discrimination [6], [7], [8]. This geometrical approach of the EEG feature representation allows to define metric spaces with chosen invariances. One main interest of the so-called Riemannian BCI is that the same classification algorithms could be applied to virtually any type of BCI. Indeed, transfer learning using covariance matrices has been investigated and achieves high performance [9], but nonetheless requires calibration data.

The proposed algorithm named MDWM, grounded in the manifold of covariance matrices, extends the works presented in [10] on SSVEP. Our contributions are twofold: MDWM works with small number of training data and it could be applied for different BCI paradigms. In this paper, we generalize the MDWM transfer learning approach to different paradigms and we conduct a meta-analysis on P300 datasets, Motor Imagery datasets and SSVEP datasets. After a short presentation of the method in Section II, this approach is evaluated on multiple datasets in Section III. Section IV concludes this report.

II. METHODS

A. Riemannian BCI

Common preprocessing and classification algorithms for EEG-based BCI rely on estimation of covariance matrices. These matrices are symmetric positive-definite and are thus restricted on a specific region of the space of real matrices. Furthermore, these matrices have a characteristic geometry and, endowed with a proper metric, it is possible to benefit from invariance properties to make robust classification [7].

The most simple yet effective Riemannian classifier for BCI is the minimum distance to mean (MDM). This algorithm brings a significant improvement when compared

[1]Salim Khazem, Sylvain Chevallier and Karim Haroun are with LISV, Université de Versailles Saint-Quentin, University Paris-Saclay, Saclay, France sylvain.chevallier@uvsq.fr

[2]Quentin Barthélemy is with Foxstream, Vaulx-en-Velin, France q.barthelemy@foxstream.fr

[3]Camille Noûs is with Cogitamus, CNRS, France camille.nous@cogitamus.fr

to Euclidean metrics [11] but does not achieve the highest classification performance amongst the Riemannian approaches [12]. Most notably, MDM is robust to noise and works well with online implementation [13]. This is thus a very good choice to serve as a basis for a transfer learning approach that should be used with minimal number of calibration samples.

From a set $\{X_1, \ldots, X_m\}$ of EEG epochs with c channels extracted from m trials, we can estimate the associated covariance matrices $\{\Sigma_1, \ldots, \Sigma_m\}$, using $\Sigma_i = 1/c\ X_i^{\mathsf{T}} X_i$. It is advised to use more advanced covariance estimators, such as described in [11]. In the space of covariance matrices, the shortest curve between two matrices Σ_i and Σ_j is known as the geodesic:

$$\Sigma_i \#_\lambda \Sigma_j = \Sigma_i^{\frac{1}{2}} \left(\Sigma_i^{-\frac{1}{2}} \Sigma_j \Sigma_i^{-\frac{1}{2}} \right)^\lambda \Sigma_i^{\frac{1}{2}}\ ,\ \text{with } \lambda \in [0, 1]. \quad (1)$$

The length of this geodesic is the Riemannian distance.

The first step of the MDM is to compute the average covariance matrix $\bar{\Sigma}^k$ for each of the $k = 1, \ldots, K$ classes [10]. Then, it is possible to assign a newly seen covariance matrix Σ to the class with the minimum distance to mean, in the sense of the Riemannian distance.

B. Transfer Learning

The proposed transfer learning approach consists in using as little as possible calibration data from a user, relying on augmented models that make use of information available from other users. As the proposed approach combine difference source EEG to estimate any target EEG signal, this approach is called Minimum Distance to Weighted Mean (MDWM) [10].

This is accomplished by combining the estimated centers of class, using the data available from the dataset of source subjects $\{\bar{S}^k\}$ averaged in $\bar{D}^k$, and the data available from the target subject $\bar{T}_k$ as: $\bar{A}^k = \bar{T}^k \#_\lambda \bar{D}^k$, using Eq. (1). The parameter λ controls the trade-off between target and source data. For $\lambda = 0$, only the few data $\bar{T}^k$ from the target subject are used (ie, no transfer scheme). For $\lambda = 1$, the calibration of the target subject is completely discarded and we rely entirely on $\bar{D}^k$ data extracted from the source subjects (ie, calibration free scheme).

The proposed algorithm MDWM, illustrated in Fig. 1, uses only a few calibration data by combining them with existing data from the source subjects. A weight can be added for each source subjects, defined for example as the similarity with the target subject [10]. In this study, uniform weights are chosen throughout the experiments.

C. MOABB

The choice of MOABB framework [12] is grounded in the need for reproducibility: this open-source project allows to download many EEG datasets and to easily evaluate any algorithm on those datasets [14]. The objective of MOABB is to yield reliable machine learning framework and sound statistical analysis to compare classification algorithms. Working on publicly available datasets, it is thus possible to reproduce results.
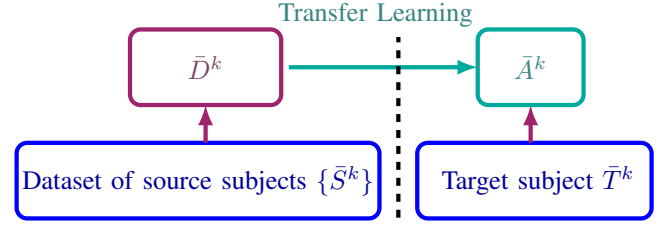

Fig. 1. Illustration of the cross-subject transfer learning: combining small calibration from the target subject with data extracted from source subjects.

Dataset name	Paradigm	#Classes	#Channels	#Subj.
SSVEP Exo [13]	SSVEP	4	8	12
Nakanishi 2015 [15]	SSVEP	12	8	10
Zhou 2016 [16]	MI	3	14	4
BNCI-004-2014 [17]	MI	2	22	9
BNCI-001-2014 [18]	MI	2	22	9
BNCI-008 2014 [19]	P300	2	8	8
BNCI-009 2014 [20]	P300	2	16	10
BNCI-003 2015 [21]	P300	2	8	10

TABLE I
LIST OF DATASETS: SSVEP, MOTOR IMAGERY AND P300 SPELLER.

While MOABB is a powerful tool, it still lacks the support to do cross-subject transfer learning. It is a more global issue as no existing framework allows to fairly evaluate transfer learning algorithms. This works is a first attempt to improve MOABB with transfer learning support and we will contribute to this open source project. Our analysis focuses three kinds of paradigms, SSVEP, Motor Imagery (MI) and P300 Speller. For the MI, the choice was based on two typical paradigms signal, left-hand vs right-hand imagery. The number of channels and subjects for each paradigm is described in Table I.

We have benchmarked MDWM on seven different datasets (2 for SSVEP, 2 for MI and 3 for P300) and shows that it performs significantly better than reference machine learning pipelines (CCA, CSP+LDA and xDAWN+LDA).

D. Analysis on transfer learning

Our objective being to minimize the calibration step of the target subject by using data from source subjects. To evaluate the influence of the number of samples used from the target subject, we use different numbers of training samples from the target subject $n \in \{5, 30, 55\}$. To evaluate the influence of the trade-off between source and target subjects, we report MDWM performances for different values of $\lambda \in \{0, 0.1, 0.3, 0.7\}$.

For each dataset, the evaluation is conducted as follows: the target subject data is separated from the other source subjects in a given dataset (following a leave-one-subject-out scheme), the trials from the target subject are separated into n trials for training (considered as calibration data) and the rest of the trials are used for testing. The model is trained on all the data available from the source subjects and the n trials from the target subject. The model is evaluated on the rest of the trials from the target subject. This process is

Fig. 3. F1-score depending on the number of samples available for P300 datasets.

CSP+LDA for MI, xDAWN+LDA for P300, and CCA for SSVEP) across datasets and paradigms, we use a meta-analysis. It formulates a null hypothesis H_0, which we are trying to reject, that there is no difference between the two methods. Then, we want to answer the following question: under H_0, what is the probability to obtain the difference in the means that we observed in the experimental data?

To obtain results with comparable metrics across paradigms, all evaluations are performed with the same classification metric: the balanced accuracy, more appropriate when classes are imbalanced like in P300. Meta-analysis gives two values: a standardized mean difference (SMD) and a p-value p. SMD is a normalized measure to assess the difference between the means of the classification metrics obtained by the two methods. It allows to understand the classification delta between both methods. The p-value is the probability of observing such difference considering the null hypothesis H_0 as true. It allows to reject H_0 when it is very low. Statistical significance is $\star$ for $p < 0.05$, $\star\star$ for $p < 0.01$ and $\star\star\star$ for $p < 0.001$, using Stouffer's method to combine p-values obtained with Wilcoxon signed-rank test on each separate dataset.

III. RESULTS AND DISCUSSIONS

In these experiments, we have used MDWM with $\lambda = 0.7$ and n that is two times the number of classes. Results of meta-analysis are shown in Fig. 2. The top plot shows that MDWM outperforms CSP+LDA for MI in all three datasets, with a sizable SMD and a good statistical significance. On middle plot, for the case of P300 speller paradigm, we can notice that MDWM outperforms xDAWN+LDA on two datasets over 3, with a noticeable SMD and a high statistical significance. On the bottom plot, MDWM outperforms CCA for both SSVEP datasets, with a high SMD and a high statistical significance. These results are a clear validation of the proposed transfer learning method, on 8 datasets containing a total of 72 subjects.

The Fig. 3 shows that by increasing the number of samples, the accuracy increases. We notice that by having a λ equal to 0.7, which correspond to rely mostly on data from source subjects, we obtain fairly good results. It is thus possible to reduce the calibration burden without losing to much precision.
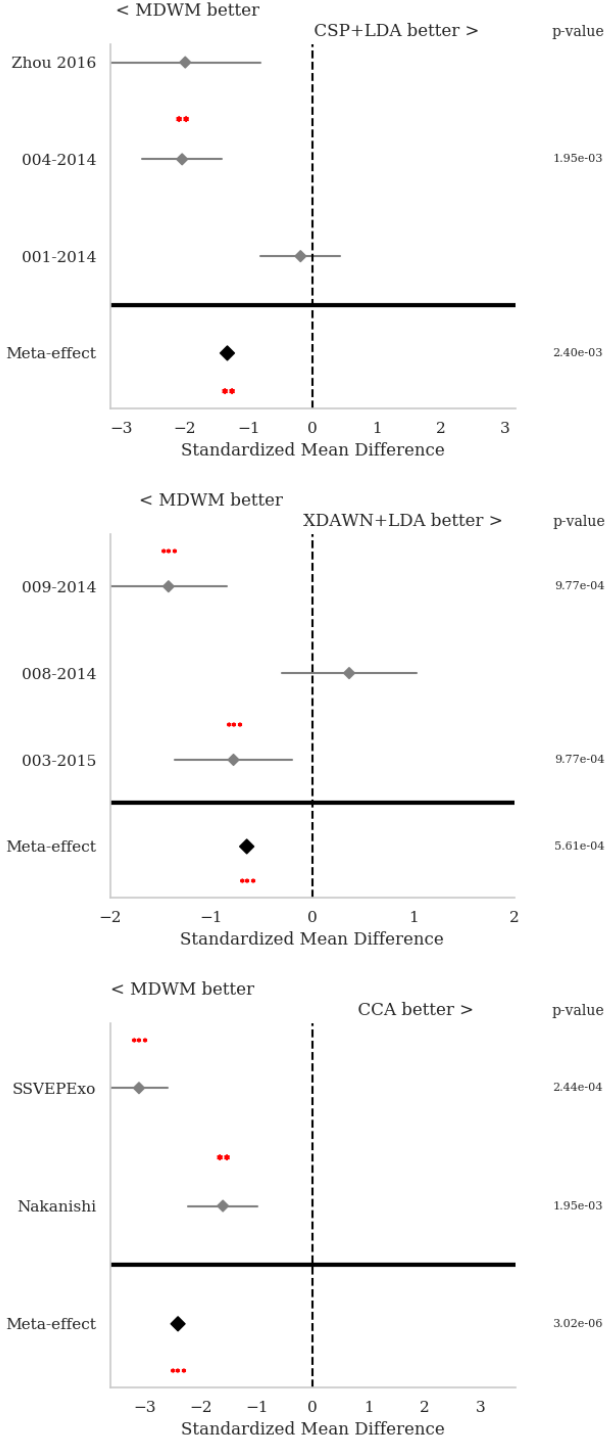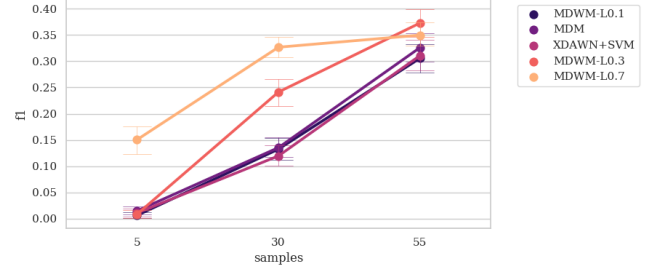

Fig. 2. Meta-analysis on several datasets, for different BCI paradigms comparing MDWM against CSP+LDA on MI (top), with xDAWN+LDA on P300 (middle), and with CCA on SSVEP (bottom).

repeated 10 times, with different trials kept as the n training samples

E. Meta-analysis on classification accuracy

In order to understand if the proposed method (MDWM) is consistently better than the state-of-the-art method (*i.e.*,

As a discussion from these transfer learning results, we could report that not all subjects have an equivalent contribution for reducing the calibration. The "transfer-ability" of a subject is an open question and there are interesting investigation on this matter in [9]. What is interesting here is that the MDWM is sufficiently robust to aggregate results from all subjects without distinction. Indeed, a proper selection of source subjects to train the MDWM could really improve the performances.

We also investigate if the calibration could be completely bypassed, by make use of the resting state EEG. It is common to start the EEG recording by asking the subject to close his/her eyes or to relax a little bit with eyes closed or opened. This resting EEG could be used to characterize the similarity between two subjects and thus to select best matching subject to train MDWM. We have investigated if resting EEG convey as much information as task related data for selecting similar subjects and it is the case. This subject will be the focus of another study.

The performances of MDWM could be improved by carefully selecting the λ value. A possible choice is to set λ as a function of the n, the number of calibration trial. Thus, the more data is available from the subject, the lower could be λ. We plan to investigate different function to set up automatically λ.

As we have pointed out, the reproducibility of results is a crucial issue for signal processing and machine learning approaches in BCI. MOABB, or other open source frameworks, are important to ensure that a fair evaluation of existing algorithms is available. While no framework proposes to benchmark transfer learning approaches yet, we will contribute improving MOABB on this point.

IV. Conclusion

In this contribution, we have described a transfer learning approach called MDWM and we have demonstrated that it could be applied on three different BCI paradigms, namely motor imagery, P300 and SSVEP. The MDWM uses a combination of a few calibration data from a target user and data available from other subjects. We have benchmarked MDWM on 7 different datasets with different number of electrodes and positions, and we show that it performs significantly better than reference machine learning pipelines (CSP, xDAWN and CCA). This is, to the authors' knowledge, the first attempt to systematically evaluate one classification pipeline on three different BCI paradigms. These results are promising and are a first step towards a reliable transfer learning for reducing the calibration phase.

Environmental impact

The approach taken in this submission tried to reduce the number of experiments made on servers, in order to reduce its environmental impact. The authors communicated through Slack, git and overleaf. Based on Shift Project and ADEME reports, we estimate that the impact of publication is roughly equivalent to 73 gCO_2.

References

[1] M. C. Thompson. Critiquing the Concept of BCI Illiteracy. *Science and Engineering Ethic*, 2018.

[2] F Lotte, L Bougrain, A Cichocki, M Clerc, M Congedo, A Rako-tomamonjy, and F Yger. A Review of Classification Algorithms for EEG-based Brain-Computer Interfaces: A 10-year Update. *J Neural Eng*, 2018.

[3] F Lotte, C Jeunet, J Mladenović, B N'Kaoua, and L Pillette. *A BCI challenge for the signal processing community: considering the user in the loop*, pages 1–33. IET, 2018. Chapter 8.

[4] C Jeunet, E Jahanpour, and F Lotte. Why standard brain-computer interface (BCI) training protocols should be changed: An experimental study. *J Neural Eng*, 13, 2016.

[5] S Fazli, C Grozea, M Danóczy, B Blankertz, F Popescu, and K-R Müller. Subject independent EEG-based BCI decoding. In *NIPS*, pages 513–521, 2009.

[6] F. Yger, M. Berar, and F. Lotte. Riemannian approaches in brain-computer interfaces: a review. *IEEE Trans Neural Syst Rehabil Eng*, 25(10):1753–1762, 2017.

[7] M. Congedo, A. Barachant, and R. Bhatia. Riemannian geometry for EEG-based brain-computer interfaces; a primer and a review. *Brain-Computer Interfaces*, 4:1–20, 2017.

[8] J. Xu, V. Jayaram, B. Schölkopf, and M. Grosse-Wentrup. Feature extraction from the Hermitian manifold for brain-computer interfaces. In *NER*, pages 965–968, 2019.

[9] P. L. C. Rodrigues, C. Jutten, and M. Congedo. Riemannian procrustes analysis: Transfer learning for brain–computer interfaces. *IEEE Trans Biomed Eng*, 66(8):2390–2401, 2019.

[10] E K Kalunga, S Chevallier, and Q Barthélemy. Transfer learning for SSVEP-based BCI using Riemannian similarities between users. In *EUSIPCO*, pages 1685–1689, 2018.

[11] S. Chevallier, E. Kalunga, Q. Barthélemy, and F Yger. *Brain Computer Interfaces Handbook: Technological and Theoretical Advances*, chapter 19 - Riemannian classification for SSVEP based BCI : offline versus online implementations. CRC Press, 2018.

[12] V Jayaram and A Barachant. MOABB: trustworthy algorithm benchmarking for BCIs. *J Neural Eng*, 15(6):066011, 2018.

[13] E. K. Kalunga, S. Chevallier, Q. Barthélemy, K. Djouani, E. Mona-celli, and Y. Hamam. Online SSVEP-based BCI using Riemannian geometry. *Neurocomputing*, 191:55–68, 2016.

[14] MOABB, https://github.com/neurotechx/moabb.

[15] M Nakanishi, Y Wang, Y-T Wang, and T-P Jung. A comparison study of canonical correlation analysis based methods for detecting steady-state visual evoked potentials. *PloS One*, 10(10):e0140703, 2015.

[16] B Zhou, X Wu, Z Lv, L Zhang, and X Guo. A Fully Automated Trial Selection Method for Optimization of Motor Imagery Based Brain-Computer Interface. *PloS One*, 11(9):1–20, 2016.

[17] R Leeb, Felix L, C Keinrath, R Scherer, H Bischof, and G Pfurtscheller. Brain–computer communication: motivation, aim, and impact of exploring a virtual apartment. *IEEE Trans Neural Syst Rehabil Eng*, 15:473–482, 2007.

[18] M Tangermann, K-R Müller, A Aertsen, N Birbaumer, C Braun, C Brunner, R Leeb, C Mehring, K Miller, G Mueller-Putz, G Nolte, G Pfurtscheller, H Preissl, G Schalk, A Schlögl, C Vidaurre, S Waldert, and B Blankertz. Review of the BCI Competition IV. *Front Neurosci*, 6:55, 2012.

[19] A Riccio, L Simione, F Schettini, A Pizzimenti, M Inghilleri, M Olivetti Belardinelli, D Mattia, and F Cincotti. Attention and P300-based BCI performance in people with amyotrophic lateral sclerosis. *Front Hum Neurosci*, 7:732, 2013.

[20] P. Aricò, F. Aloise, F. Schettini, S. Salinari, D. Mattia, and F. Cincotti. Influence of P300 latency jitter on event related potential-based brain–computer interface performance. *J Neural Eng*, 11(3), 2014.

[21] C Guger, S Daban, E Sellers, C Holzner, G Krausz, R Carabalona, F Gramatica, and G Edlinger. How many people are able to control a P300-based brain–computer interface? *Neurosci Lett*, 462(1):94 – 98, 2009.