Channel and Trials Selection for Reducing Covariate Shift in EEG-based Brain-Computer Interfaces

He He and Dongrui Wu

Abstract—*Objective*: This paper aims at reducing the calibration effort of EEG-based brain-computer interfaces (BCIs). More specifically, in the context of cross-subject classification, we correct covariate shift of EEG data from different subjects, so that a classifier trained on auxiliary subjects can also be applied to a new subject, without any labeled trials from the new subject. *Methods*: We propose two approaches to enhance the performance of a state-of-the-art Riemannian space transfer learning (TL) algorithm: 1) trials selection, which resamples trials from the auxiliary subjects so that they become more consistent with those of the new subject; and, 2) channel selection, which reduces the number of channels and hence makes the Riemannian space computations more accurate and efficient. *Results*: We tested the proposed approaches on two motor imagery datasets. The results verified that they can enhance the performance of the state-of-the-art TL algorithm. *Conclusion and significance:* Our proposed approaches make the state-of-the-art TL algorithm more effective and efficient.

I. INTRODUCTION

Brain computer interfaces (BCIs) [1], [2] enable a user to interact with his/her surroundings by using brain signals, including electroencephalogram (EEG), magnetoencephalogram, electrocorticography, and so on. Early BCIs were developed only for the disabled, allowing them to communicate with the environment without the involvement of muscles [3]. For example, a severely paralyzed patient can control a powered exoskeleton or wheelchair by imagining the movement of his/her body with the help of a BCI system. Recently, there has been a growing interest of BCI research on able-bodied users, such as playing video games and controlling unmanned aerial vehicles [4], [5]. EEG is the most popular form of BCI input as it is easy and safe to acquire, and offers high temporal resolution.

However, EEG signals are very weak and can be easily contaminated by various artifacts and noise [6]. Therefore, sophisticated signal processing and machine learning algorithms are needed to clean and decode EEG signals. One of the most popular approaches for enhancing the signal-to-noise ratio of EEG is common spatial pattern (CSP) filtering [7]–[11]. Then, discriminant features are extracted and fed into a classifier. Riemannian space classifiers [12]–[15], which integrate feature extraction and classification, have becoming popular in the last decade. They take the symmetric positive definite (SPD) spatial covariance matrices

H. He and D. Wu are with the Key Laboratory for Image Processing and Intelligent Control, Ministry of Education, and the School of Artificial Intelligence and Automation, Huazhong University of Science and Technology, Wuhan, China. `hehe91@hust.edu.cn`, `drwu@hust.edu.cn`.

of the EEG trials as features and directly classify them by measuring the distances on the Riemannian manifold.

Although sophisticated signal processing and classification/regression algorithms have been proposed, there are still many challenges in real-world applications of EEG-based BCIs [2], [16]. Because of individual differences, algorithms trained on auxiliary subjects may not be directly applied to a new subject, because people show different neural responses even in the same task. Therefore, BCI systems usually need to be calibrated before each use, which is time-consuming and inconvenient.

Transfer learning (TL) [17], which leverages information from related domains to improve the learning performance in a target domain, is a promising approach for shortening or even eliminating the calibration process. TL has been successfully used in EEG-based BCIs [18]–[22]. It can compensate data shifts among different subjects, including covariate shift, concept shift, and prior shift.

Covariate shift means the distribution of the input changes across datasets, which is the most commonly occurred and studied problem in TL. Recently, Zanini et al. [23] proposed a Riemannian alignment (RA) approach to cope with covariate shift in BCIs. It first computes a reference matrix for each subject, then centers the trials of each subject with respect to his/her reference matrix, so that the trials from different subjects are better aligned. The key step is to compute the reference matrix for each subject. [23] used "resting trials", i.e., epochs between two successive imagery tasks, to compute the reference matrix. Our recent work [24] discussed different choices of the reference matrix, and demonstrated that the Riemannian mean of the imagery trials outperformed that of the resting trials. To our knowledge, RA using mean covariance matrix of the imagery trials is the state-of-the-art in Riemannian-based TL for EEG-based BCIs.

This paper proposes two approaches to further enhance this state-of-the-art RA:

1) Trials Selection (TS), which reduces the covariate shift between the target subject and the auxiliary subjects by selecting the most similar trials from the auxiliary subjects.
2) Channel Selection (CS), which selects a subset of channels such that the dimensionality of the spatial covariance matrices decreases. This may be beneficial since the computations in the Riemannian space become more accurate and efficient.

The remainder of this paper is organized as follows: Section II introduces some essential background knowledge.

Section III presents our enhancements to RA. Section IV compares the performance of our proposed approaches with RA on two MI datasets. Finally, Section V draws conclusion.

II. BACKGROUND

This section first introduces CSP filtering, based on which channel can be performed. Then, the Riemannian geometry and a state-of-the-art TL approach are described.

A. Common Spatial Pattern (CSP) Filtering

Let $X_i \in \mathbb{R}^{C \times T}$ be an EEG trial, where C is the number of channels and T the number of time samples. Consider a binary classification problem. We first compute the mean covariance matrix of the trials in Class p by:

$$\bar{\Sigma}_p = \frac{1}{n_p} \sum_{i=1}^{n_p} X_{p,i} X_{p,i}^T, \quad p = 0, 1 \tag{1}$$

where $X_{p,i}$ is the ith trial in Class p, and n_p is its number of trials.

CSP projects an EEG trial into a few subcomponents which have the maximum difference between the variances of the two classes, i.e.,

$$W_0 = \arg \max_W \frac{\mathrm{tr}(W^T \bar{\Sigma}_0 W)}{\mathrm{tr}(W^T \bar{\Sigma}_1 W)}, \tag{2}$$

where $W_0 \in \mathbb{R}^{C \times M}$ is a filter matrix whose columns are the individual filters, M is the number of filters, and $\mathrm{tr}(\cdot)$ is the trace of a matrix.

W_0 maximizes the variance for Class 0 while minimizing it for Class 1. In practice, we often construct a CSP filter matrix $W_* = [W_0, W_1] \in \mathbb{R}^{C \times 2M}$, where

$$W_1 = \arg \max_W \frac{\mathrm{tr}(W^T \bar{\Sigma}_1 W)}{\mathrm{tr}(W^T \bar{\Sigma}_0 W)}, \tag{3}$$

i.e., W_1 maximizes the variance for Class 1 while minimizing it for Class 0.

W_* is the concatenation of the *2M* eigenvectors associated with the *M* largest and *M* smallest eigenvalues of the matrix $\bar{\Sigma}_1^{-1} \bar{\Sigma}_0$ (or $\bar{\Sigma}_0^{-1} \bar{\Sigma}_1$).

B. Riemannian Geometry

Since this paper investigates TL approaches in the Riemannian space, we next introduce some basic Riemannian space concepts.

1) Riemannian Distance: The Riemannian distance between two SPD matrices P_1 and P_2 is the minimum length of a curve connecting them on the Riemannian manifold:

$$\delta_R(P_1, P_2) = \| \log(P_1^{-1} P_2) \|_F = \left[\sum_{r=1}^{R} \log^2 \lambda_r \right]^{\frac{1}{2}}, \tag{4}$$

where the subscript F denotes the Frobenius norm, and λ_r ($r = 1, 2, \cdots, R$) are the real eigenvalues of $P_1^{-1} P_2$.

2) Riemannian Mean: The Riemannian mean of a set of SPD matrices is their geometric mean in the Riemannian space, instead of the arithmetic mean (Euclidean mean). Specifically, it is defined as the matrix minimizing the sum of the squared Riemannian distances:

$$\varrho(P_1, \cdots, P_N) = \arg \min_P \sum_{n=1}^{N} \delta_R^2(P, P_n), \tag{5}$$

There is no closed-form solution to (5), and it is usually computed by an iterative gradient descent algorithm [25].

3) Minimum Distance to Riemannian Mean (MDRM): The MDRM classifier [12]–[14] first computes the covariance matrix of each EEG trial, and then the Riemannian mean of each class using the labeled training trials. Next, it assigns each test trial to the class whose Riemannian mean is the closest to its covariance matrix, i.e.,

$$g(\Sigma) = \arg \min_{c=1,2,\cdots,C} \delta_R(\Sigma, \bar{\Sigma}^c), \tag{6}$$

where $\bar{\Sigma}^c$ is the Riemannian mean of Class c, Σ is the covariance matrix of the test trial, and $g(\Sigma)$ is the prediction of its class label.

C. Riemannian Space Alignment (RA)

Zanini et al. [23] first proposed RA to align covariance matrices of trials from different sessions/subjects in the Riemannian space. The basic idea of RA is to *"center the covariance matrices of every session/subject with respect to a reference covariance matrix so that what we observe is only the displacement with respect to the reference state due to the task."* Therefore, the reference covariance matrix plays a significant role in RA, and [23] used the resting trials to compute the reference matrix for each session/subject, under the assumption that *"different source configurations and electrode positions induce shifts of covariance matrices with respect to a reference (resting) state, but that when the brain is engaged in a specific task, covariance matrices move over the SPD manifold in the same direction."*

More specifically, RA first computes the covariance matrices of the resting trials, $\{R_i\}_{i=1}^n$, and then their Riemannian mean:

$$\tilde{R} = \arg \min_R \sum_{i=1}^{n} \delta_R^2(R, R_i), \tag{7}$$

where $\tilde{R}$ is the reference matrix.

Then, $\tilde{R}$ is used to transform the data of the corresponding subject, making data from different subjects more consistent:

$$\tilde{\Sigma}_i = \tilde{R}^{-1/2} \Sigma_i \tilde{R}^{-1/2}, \tag{8}$$

where Σ_i and $\tilde{\Sigma}_i$ are the covariance matrix before and after alignment.

We [24] discussed different choices of the reference matrix in RA, and demonstrated that using the Riemannian mean of the imagery trials achieved better performance than using the Riemannian mean of the resting trials. Therefore, we used the former approach to compute the reference matrix in RA in this paper.

III. PROPOSED ENHANCEMENTS TO THE RA

This section proposes two enhancements to the RA. The first selects a subset of labeled trials from the auxiliary subjects based on the Riemannian distance. The second selects a subset of channels to make the computations in the Riemannian space more accurate and efficient.

A. Trials Selection (TS)

Covariate shift, or sample selection bias, refers to the input data distribution discrepancy between training and test data. We propose a TS approach, in Algorithm 1, to resample the training data. In order to reduce the input data distribution discrepancies between the target subject and the auxiliary subjects, we select the auxiliary trials located closest to the target trials in the Riemannian space.

Algorithm 1: The TS algorithm.

Input: N^t unlabeled trials from the target subject, $\{\mathbf{x}_i^t\}_{i=1}^{N^t}$;
N^s labeled trials from the auxiliary subjects, $\{\mathbf{x}_j^s\}_{j=1}^{N^s}$;
k, the number of selected neighbours for each target trial;

Output: The selected auxiliary trials.

for $i = 1, ..., N^t$ **do**
    for $j = 1, ..., N^s$ **do**
        Compute the Riemannian distance between the covariance matrices of $\mathbf{x}_i^t$ and $\mathbf{x}_j^s$, by (4);
    end
    Select k nearest neighbours from the N^s auxiliary trials for $\mathbf{x}_i^t$;
end
Combine all selected auxiliary trials;
Remove duplicate trials.

B. Channel Selection (CS)

Both RA and TS operate on the covariance matrices in the Riemannian space, whose computational cost increases with the number of channels. Two CS approaches are used to alleviate it.

1) Manual Selection: Riemannian approaches decode EEG signal by computing covariance matrices, which mainly reflect the spatial distribution of neural activities. For example, imagining the movements of a body part (hand, foot, tongue, etc.) would cause modulations of brain rhythms in the involved cortical areas. So, the contributions of different EEG channels vary across locations and tasks.

The first CS approach (CS1) manually selects the channels located in the areas most related to the corresponding MI tasks, which usually contain more information and less noise.

2) CSP-based Selection: Since CSP filters carry the channel weighting information, we also used a CSP-based CS approach (CS2) [26].

Let $W_* \in \mathbb{R}^{C \times 2M}$ be the CSP filtering matrix, and w_i be its ith row. Then, the score for Channel c is

$$SC(c) = \frac{\| w_c \|_1}{\| W \|_1}, \quad c = 1, ..., C \quad (9)$$

The channels with the maximum scores are then selected.

IV. EXPERIMENTS

This section describes the performance of our proposed approaches on two MI datasets in offline unsupervised classification. For each dataset, we picked one subject as the target subject (test set) each time, and combined the remaining subjects as auxiliary subjects (training set). The target subject did not use any labeled training data from himself/herself.

A. Datasets

Two public MI datasets from BCI Competition IV[1] were used. Both were recorded with a cue-based BCI paradigms: Each subject sat in front of a computer, wearing a BCI headset and being prepared for the visual cues that would appear on the computer screen. An arrow pointing to a certain direction was presented, and the subject was asked to perform the corresponding MI task in this period. When the visual cue disappeared, the subjects relaxed and waited for the next trial.

The first dataset[2] (Dataset 1 [27]) was recorded from 59 EEG channels at 100 Hz. It includes seven healthy subjects. Each subject did the experiments for three phases: calibration, evaluation, and special feature. Here we only used the calibration phase data, which provided complete label information. In the calibration phase, each subject was instructed to perform two of the three MI tasks (left hand, right hand, and foot), with 100 trials for each task.

The second MI dataset (Dataset 2a[3]) was recorded using 22 EEG channels and three EOG channels at 250Hz. It includes nine subjects. Each subject performed a training phase and an evaluation phase. In both phases the subjects were instructed to perform four different MI tasks, namely the imagination of the movement of the left hand, right hand, both feet, and tongue. Here we only used the EEG channels and selected two classes (left hand and right hand) from the calibration data.

The EEG signals from both datasets were preprocessed using the Matlab EEGLAB toolbox [28], following the guideline in [6]. First, a band-pass filter (linear phase Hamming window FIR filter, with 6dB cut-off frequencies at [8, 30] Hz), whose order was defined as the filter length minus one and made mandatorily even, was applied to remove muscle artifacts, line-noise contamination and DC drift. Next, we extracted EEG signals between $[0.5, 3.5]$ seconds after the cue appearance as our trials for Dataset 1 and EEG signals between $[0.5, 2.5]$ seconds for Dataset 2a.

---

[1]http://www.bbci.de/competition/iv/.
[2]http://www.bbci.de/competition/iv/desc_1.html.
[3]http://www.bbci.de/competition/iv/desc_2a.pdf.

B. Algorithms

Each algorithm had three stages:

1) *Preprocessing stage*: It first temporally filters the EEG data, then divides the continuous signals into epochs, as described in the previous subsection.
2) *Middle stage*: It performs CS as described in Section III-B, RA in Section II-C, or TS in Section III-A, according to the specific algorithm.
3) *Classification stage*: It classifies the spatially filtered data by MDRM.

All algorithms had the same preprocessing and classification stages. They are distinguished only by the middle stage.

Because Dataset 1 has more channels than Dataset 2a, we applied both CS and TS to Dataset 1, but only TS to Dataset 2a. Two CS approaches were used for Dataset 1:

1) CS1: We manually selected channels {'Fz', 'FC3', 'FC1', 'FCz', 'FC2', 'FC4', 'C5', 'C3', 'C1', 'Cz', 'C2', 'C4', 'C6', 'CP3', 'CP1', 'CPz', 'CP2', 'CP4', 'P1', 'Pz', 'P2', 'PO1' }, located in the areas most related to the activations of left hand and right hand movements.
2) CS2: We selected the channels based on the CSP filters, as introduced in Section III-B.2.

C. Evaluation of TS

We first tested TS on Dataset 2a, and compared its performance with algorithms using all trials from the auxiliary subjects. More specifically, four approaches were compared:

1) Raw, which directly feeds the filtered data into the MDRM classifier, i.e., it does nothing in the middle stage.
2) TS, which performs only TS in the middle stage.
3) RA, which performs only RA in the middle stage.
4) RA-TS, which performs both RA and TS in the middle stage.

To study how the performance of TS changes with k, we started with $k = 20$, trained TS and RA-TS models, and evaluated their performances on the test set. Then, we increased k by 10 in each iteration, and updated the models and results, until $k = 100$. Raw and RA always used all labeled trials of auxiliary subjects to train their models. The results are shown in Fig. 1, where the first nine subfigures show the performances on the individual subjects and the last shows the average performance across all subjects. In each subfigure, the horizontal axis denotes k, and the vertical axis denotes the test accuracy. Observe that:

1) TS outperformed Raw on 8 out of the 9 subjects, and the two had comparable performances on the remaining subject, suggesting that TS generally reduced covariate shift between the target and auxiliary subjects.
2) On average RA-TS outperformed RA, suggesting that TS can also enhance the performance of RA.
3) The performance improvement of RA-TS over RA was smaller than that of RA over Raw. This may be because
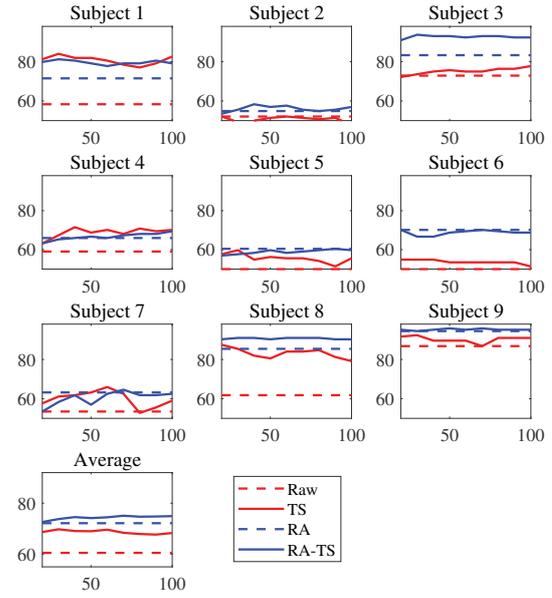
Fig. 1. Classification accuracies (%) on Dataset 2a. Horizontal axis: k; vertical axis: test accuracy.

RA had reduced some discrepancies between the target and auxiliary subjects.

D. Evaluation of CS

In this section we first compared the performances of algorithms using CS1 with the one using all channels. More specifically, the following four approaches were compared:

1) Raw, which is the same as before.
2) CS1, which performs only CS1 in the middle stage.
3) RA, which performs only RA in the middle stage.
4) CS1-RA, which performs both CS1 and RA in the middle stage.

Their classification accuracies are shown in Table I. CS1 performed slightly worse than Raw, but CS1-RA outperformed RA, suggesting that CS1 can enhance the performance of RA.

TABLE I
CLASSIFICATION ACCURACIES (%) OF CS1 ON DATASET 1.

Subject	Raw	CS1	RA	CS1-RA
1	51.0000	50.0000	**68.5000**	68.0000
2	50.0000	50.0000	**57.5000**	57.0000
3	50.0000	50.0000	52.0000	**62.5000**
4	51.0000	50.0000	**61.0000**	**61.0000**
5	50.0000	50.0000	74.0000	**76.0000**
6	60.5000	51.0000	63.5000	**69.5000**
7	50.0000	50.0000	**69.5000**	**69.5000**
avg	51.7857	50.1429	63.7143	**66.2143**

CS could not only improve the classification accuracy, but also reduce the computational cost. We compared the computational cost of RA and CS1-RA. The platform was a

ThinkPad laptop with Intel Core i5-6200U CPU@2.30GHz, 4GB memory, and 190 GB SSD, running 64-bit Windows 10 and Matlab 2017a. The results are shown in Table II. `CS1-RA` was 4.14-5.04 times faster than `RA`, and also it had smaller standard deviation.

TABLE II
THE COMPUTATION TIME (SECONDS) OF RA AND CS1-RA.

	RA	CS1-RA
Mean	12.4743	2.5861
std	0.7859	0.3800

Next, we tested the performance of CS when different numbers of channels were selected. We used CS2 to select channels and increased the number of selected channels by 5 in each iteration, from 15 to 35. As the main purpose was to verify that CS2 can enhance the performance of RA, we only compared `RA` with `CS2-RA`. The results are shown in Table III. `CS2-RA` always outperformed `RA` on average, and it also slightly outperformed `CS1-RA`. However, there was not a clear and consistent pattern between the classification performance and the number of selected channels.

TABLE III
CLASSIFICATION ACCURACIES (%) OF CS2 ON DATASET 1.

	RA	CS2-RA				
Subject	59 (All)	15	20	25	30	35
1	68.50	69.50	**71.00**	70.50	68.50	67.50
2	57.50	**63.50**	58.00	57.50	56.00	56.50
3	52.00	65.50	64.00	67.00	**69.50**	68.50
4	61.00	57.00	59.00	60.50	63.50	**65.50**
5	74.00	77.50	78.50	74.00	**82.00**	81.50
6	63.50	69.50	**72.50**	67.50	68.50	69.00
7	69.50	70.00	73.00	76.50	75.00	**77.50**
avg	63.71	67.50	68.00	67.64	69.00	**69.43**

E. Evaluation of CS and TS

We also studied how CS, RA, and TS affected each other. First, we used CS1 to select channels and compared the following approaches:

1) `Raw`, which is the same as before.
2) `TS`, which performs only TS in the middle stage.
3) `CS1`, which performs only CS1 in the middle stage.
4) `CS1-TS`, which performs both CS1 and TS in the middle stage.
5) `RA`, which performs only RA in the middle stage.
6) `RA-TS`, which performs both RA and TS in the middle stage.
7) `CS1-RA`, which performs both CS1 and RA in the middle stage.
8) `CS1-RA-TS`, which performs all three operations in the middle stage.

The results are shown in Fig. 2. The last subfigure shows the average performances across all subjects. Observe that:

1) Both `TS` and `CS1` performed slightly worse than `Raw`, suggesting that TS and CS did not work when they were independently used on datasets with a relatively large number of channels.
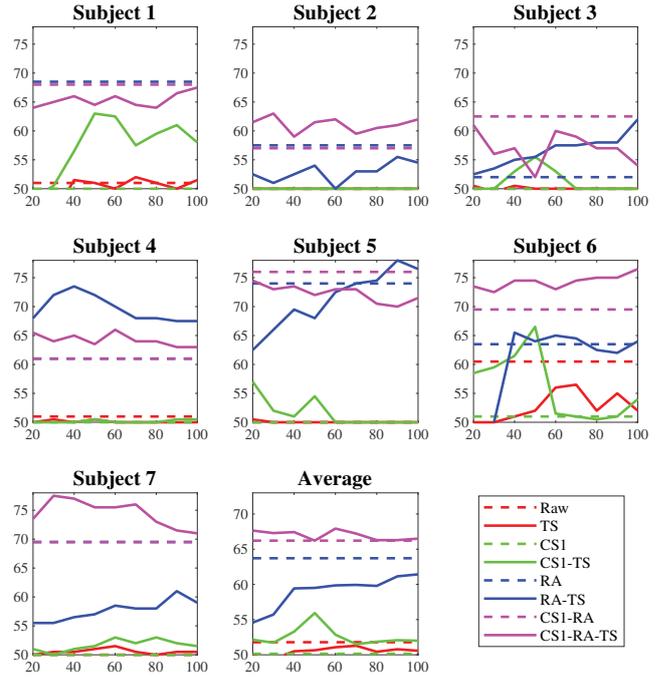
Fig. 2. Classification accuracies (%) on Dataset 1. The horizontal axis: k; vertical axis: test accuracy.

2) `CS1-TS` outperformed both `CS1` and `TS`, suggesting that more useful trials were selected when applying CS before TS. This might be because CS improved the computational accuracy in the Riemannian space.
3) `CS1-RA` outperformed both `CS1` and `RA`, suggesting again that CS improved the computational accuracy in the Riemannian space.
4) `CS1-RA-TS` performed the best among all algorithms, suggesting that it is beneficial to integrate TS, CS and RA.

Next, we compared the eight algorithms when CS2 was used to select channels. Due to the page limit, we only present their average performances across all seven subjects for each number of selected channels. The results are shown in Fig. 3. Both `CS2-RA` and `CS2-RA-TS` outperformed `RA`. However, `CS2-RA-TS` may not outperform `CS2-RA` when the number of selected channels was large.

V. CONCLUSION

The transition of BCIs from laboratories to the real world is hindered significantly by individual differences, which require a time-consuming calibration process to collect subject-specific labeled data. Transfer learning, which leverages labeled data from auxiliary subjects to learn a model for a new subject, is a promising solution to this problem. However, data discrepancies among different subjects may result in negative transfers. So, it is important to reduce the data shift between the new subject and the auxiliary subjects.

This paper investigated approaches for handling covariate shift, one of the main causes of data shift. More specifically, we used two approaches to enhance the performance of
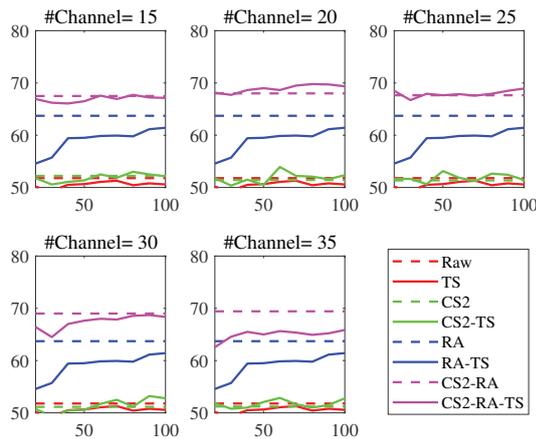
Fig. 3. Average classification accuracies (%) across all subjects on Dataset 1. Horizontal axis: k; vertical axis: test accuracy.

a state-of-the-art RA approach: 1) TS, which resamples trials from the auxiliary subjects so that they become more consistent with those of the new subject; and, 2) CS, which reduces the number of channels and hence makes the Riemannian space computations more accurate and efficient. We first validated the effectiveness of TS and CS separately. Particularly, two CS approaches (manually selection based on the channel locations, and automatic selection using CSP) were compared. Experiments showed that both enhanced the performance of RA on average, and the automatic selection approach was more stable. We also showed that integrating TS and CS can sometimes result in additional performance improvement.

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] B. J. Lance, S. E. Kerick, A. J. Ries, K. S. Oie, and K. McDowell, "Brain-computer interface technologies in the coming decades," *Proc. of the IEEE*, vol. 100, no. 3, pp. 1585–1599, 2012.

[3] G. Pfurtscheller, G. R. Müller-Putz, R. Scherer, and C. Neuper, "Rehabilitation with brain-computer interface systems," *Computer*, vol. 41, no. 10, pp. 58–65, 2008.

[4] L. F. Nicolas-Alonso and J. Gomez-Gil, "Brain computer interfaces, a review," *Sensors*, vol. 12, no. 2, pp. 1211–1279, 2012.

[5] J. van Erp, F. Lotte, and M. Tangermann, "Brain-computer interfaces: Beyond medical applications," *Computer*, vol. 45, no. 4, pp. 26–34, 2012.

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

[7] Z. J. Koles, M. S. Lazar, and S. Z. Zhou, "Spatial patterns underlying population differences in the background EEG," *Brain Topography*, vol. 2, no. 4, pp. 275–284, 1990.

[8] J. Müller-Gerking, G. Pfurtscheller, and H. Flyvbjerg, "Designing optimal spatial filters for single-trial EEG classification in a movement task," *Clinical Neurophysiology*, vol. 110, no. 5, pp. 787–798, 1999.

[9] H. Ramoser, J. Muller-Gerking, and G. Pfurtscheller, "Optimal spatial filtering of single trial EEG during imagined hand movement," *IEEE Trans. on Rehabilitation Engineering*, vol. 8, no. 4, pp. 441–446, 2000.

[10] D. Wu, J.-T. King, C.-H. Chuang, C.-T. Lin, and T.-P. Jung, "Spatial filtering for EEG-based regression problems in brain-computer interface (BCI)," *IEEE Trans. on Fuzzy Systems*, vol. 26, no. 2, pp. 771–781, 2018.

[11] H. He and D. Wu, "Spatial filtering for brain computer interfaces: A comparison between the common spatial pattern and its variant," in *Proc. 8th IEEE Int'l. Conf. on Signal Processing, Communications and Computing*, Qingdao, China, September 2018, pp. 1–6.

[12] A. Barachant, S. Bonnet, M. Congedo, and C. Jutten, "Multiclass brain-computer interface classification by Riemannian geometry," *IEEE Trans. on Biomedical Engineering*, vol. 59, no. 4, pp. 920–928, 2012.

[13] M. Congedo, A. Barachant, and A. Andreev, "A new generation of brain-computer interface based on Riemannian geometry," *arXiv: 1310.8115*, 2013.

[14] F. Yger, M. Berar, and F. Lotte, "Riemannian approaches in brain-computer interfaces: a review," *IEEE Trans. on Neural Systems and Rehabilitation Engineering*, vol. 25, no. 10, pp. 1753–1762, 2017.

[15] D. Wu, V. J. Lawhern, B. J. Lance, S. Gordon, T.-P. Jung, and C.-T. Lin, "EEG-based user reaction time estimation using Riemannian geometry features," *IEEE Trans. on Neural Systems and Rehabilitation Engineering*, vol. 25, no. 11, pp. 2157–2168, 2017.

[16] S. Makeig, C. Kothe, T. Mullen, N. Bigdely-Shamlo, Z. Zhang, and K. Kreutz-Delgado, "Evolving signal processing for brain-computer interfaces," *Proc. of the IEEE*, vol. 100, no. Special Centennial Issue, pp. 1567–1584, 2012.

[17] S. J. Pan and Q. Yang, "A survey on transfer learning," *IEEE Trans. on Knowledge and Data Engineering*, vol. 22, no. 10, pp. 1345–1359, 2010.

[18] V. Jayaram, M. Alamgir, Y. Altun, B. Scholkopf, and M. Grosse-Wentrup, "Transfer learning in brain-computer interfaces," *IEEE Computational Intelligence Magazine*, vol. 11, no. 1, pp. 20–31, 2016.

[19] D. Wu, V. J. Lawhern, W. D. Hairston, and B. J. Lance, "Switching EEG headsets made easy: Reducing offline calibration effort using active weighted adaptation regularization," *IEEE Trans. on Neural Systems and Rehabilitation Engineering*, vol. 24, no. 11, pp. 1125–1137, 2016.

[20] D. Wu, V. J. Lawhern, S. Gordon, B. J. Lance, and C.-T. Lin, "Driver drowsiness estimation from EEG signals using online weighted adaptation regularization for regression (OwARR)," *IEEE Trans. on Fuzzy Systems*, vol. 25, no. 6, pp. 1522–1535, 2017.

[21] D. Wu, "Online and offline domain adaptation for reducing BCI calibration effort," *IEEE Trans. on Human-Machine Systems*, vol. 47, no. 4, pp. 550–563, 2017.

[22] H. He and D. Wu, "Transfer learning enhanced common spatial pattern filtering for brain computer interfaces (BCIs): Overview and a new approach," in *Proc. 24th Int'l. Conf. on Neural Information Processing*, Guangzhou, China, November 2017.

[23] P. Zanini, M. Congedo, C. Jutten, S. Said, and Y. Berthoumieu, "Transfer learning: a Riemannian geometry framework with applications to brain-computer interfaces," *IEEE Trans. on Biomedical Engineering*, vol. 65, no. 5, pp. 1107–1116, 2018.

[24] H. He and D. Wu, "Transfer learning for brain-computer interfaces: A Euclidean space data alignment approach," *IEEE Trans. on Biomedical Engineering*, 2019, in press.

[25] P. T. Fletcher and S. Joshi, "Principal geodesic analysis on symmetric spaces: Statistics of diffusion tensors," *Lecture Notes in Computer Science*, vol. 3117, pp. 87–98, 2004.

[26] J. Meng, G. Liu, G. Huang, and X. Zhu, "Automated selecting subset of channels based on CSP in motor imagery brain-computer interface system," in *Proc. IEEE Int'l Conf. on Robotics and Biomimetics*, Guilin, China, Dec. 2009, pp. 2290–2294.

[27] B. Blankertz, G. Dornhege, M. Krauledat, K. R. Muller, and G. Curio, "The non-invasive Berlin brain-computer interface: Fast acquisition of effective performance in untrained subjects," *NeuroImage*, vol. 37, no. 2, pp. 539–550, 2007.

[28] A. Delorme and S. Makeig, "EEGLAB: an open source toolbox for analysis of single-trial EEG dynamics including independent component analysis," *Journal of Neuroscience Methods*, vol. 134, pp. 9–21, 2004.