Transfer learning in BCI-EEG

Nadia ALOUI, Stephane BONNET

University Grenoble Alpes, CEA, LETI, Minatec, F-38000 Grenoble, France.
nadia.aloui@cea.fr, stephane.bonnet@cea.fr

Résumé – L'apprentissage par transfert traite du problème de variabilité inter-individus. En interface cerveau-machine, cela désigne l'ensemble de techniques qui vise à exploiter la connaissance acquise sur d'autres sujets ayant réalisé la même tâche mentale que celle en cours d'analyse chez un nouveau sujet. La méthode TCA (analyse en composantes de transfert) recherche ainsi un espace de dimension réduite où les caractéristiques des deux domaines (source, cible) sont proches au sens d'un certain critère. Dans ce papier, nous proposons une nouvelle méthode pour appliquer la TCA en lien avec la géométrie Riemannienne dans le cas de système BCI-EEG. L'approche proposée permet de travailler nativement avec des caractéristiques de type matrice de covariance. Les résultats sont démontrés sur un jeu de compétition BCI et sont supérieurs à ceux obtenus lorsqu'on applique TCA sur des caractéristiques vectorielles de type (log)variance.

Abstract – Transfer learning deals with the problem of variability between subjects. In brain-computer interface (BCI), it aims at using the knowledge of previous subjects having performed the same kind of experiments. Transfer Component Analysis (TCA) is a transfer learning method that searches for a low-dimensional feature space which reduces the difference between domain distributions. In this paper, we propose a novel manner to apply the TCA method in the Riemannian geometry. First, the method, originally proposed to be applied in the Euclidean space, is tested to align the variance of the spatially filtered EEG epochs of different subjects. Then, the method is adapted to align covariance matrix based feature. Our approach is applied on a publicly available MI BCI-database. Results indicate that our method gives better results than the traditional TCA that is based on feature vectors.

1 Introduction

Brain-computer interface (BCI) transfer learning is currently being actively investigated to reduce calibration times in EEG experiments [1]. Indeed EEG is known to be highly non stationnary either in a cross-session or in a cross-subject manner. In the cross-session case, both fluctuating mental states and electrode contact impedance cause the EEG distribution to vary between sessions. In the cross-subject case, variability can be explained by head morphology and proficiency to perform a task. Transfer learning allows us to use the knowledge of previous subjects having performed the same kind of experiments. Zanini et al. [1] have proposed a Riemannian alignement by recentering the covariance matrices of each subject/session with respect to a reference matrix. The same idea was originally proposed in [2] for classification purpose. The authors in [3] have proposed a Riemannian version of Procrustes analysis (RPA). The method, after recentering the covariance matrices of both source/target domains, further matches their dispersion and applies a rotation transformation. TCA has rarely been used for BCI application. In [4], TCA was applied to EEG dataset for emotion recognition. In [5], the authors proposed to reuse the TCA projection matrix computed on a subset of data to transform new data.

Tang et al [6] have recently extended the TCA method to matrix form in order to make the method usable in the Riemannian framework. But, the authors did not go for a full Riemannian geometry treatment of the information as they use the Euclidean distance in the kernel function.

In this paper, we propose to use the TCA method in the Riemannian framework but in its traditional vector form. To do so elegantly, we make change in the kernel function used in TCA by including the Riemannian distance.

In this study, we are investigating transfer learning on a motor imagery (MI)-BCI competition dataset [8] to demontrate the benefits of the proposed approach.

2 Methods

2.1 Maximum Mean Discrepancy

The Maximum Mean Discrepancy (MMD) **[9]** between the N_1 **source observations** (domain 1) and N_2 **target observations** (domain 2) is computed as the distance between their means

$$\mathrm{MMD} = \|\mathbf{m}_1 - \mathbf{m}_2\|^2_{\mathcal{H}} \tag{1}$$

where $\mathbf{m}_c = \sum_n y_{nc}\psi(\mathbf{u}_n)$ and $y_{nc} = 1/N_c$ if the n-th observation belongs to the c-th domain and 0 otherwise. Here $\psi(\mathbf{u})$ denotes some transformation applied to the feature vector $\mathbf{u}$. We note $N = N_1 + N_2$ the total number of observations.

Rearranging terms, one can write the MMD distance as

$$\mathrm{MMD} = \mathrm{tr}(\mathbf{KL}) \tag{2}$$

The symmetric kernel matrix $\mathbf{K}$ has for element $K_{ij} = \langle\psi(\mathbf{u}_i), \psi(\mathbf{u}_j)\rangle_{\mathcal{H}}$, the inner product in $\mathcal{H}$ between the

i-th and j-th observations.
The kernel matrix $\mathbf{K}$ can be written as

$$\begin{aligned} K &= [\mathbf{K}_{S,S}, \mathbf{K}_{S,T}; \mathbf{K}_{T,S}, \mathbf{K}_{T,T}] && (3) \\ &= [\mathbf{K}_S; \mathbf{K}_T] && (4) \end{aligned}$$

where $\mathbf{K}_{S,S}$,$\mathbf{K}_{T,T}$ and $\mathbf{K}_{S,T}$ are respectively the kernel matrices defined on the source domain, target domain and cross domains.
As an example of kernel functions, the linear kernel is simply

$$K_{ij}^L = \mathbf{u}_i^T \mathbf{u}_j \quad (5)$$

Another popular kernel is the Gaussian kernel with elements

$$K_{ij}^G = \exp\left[-\frac{d_E^2(\mathbf{u}_i, \mathbf{u}_j)}{2\sigma^2}\right] \quad (6)$$

where $d_E(\mathbf{u}_i, \mathbf{u}_j) = \|\mathbf{u}_i - \mathbf{u}_j\|_2$ the ℓ_2 norm between feature vectors. The matrix $\mathbf{L}$ contains the domain information with $L_{ij} = \sum_c y_{ic} y_{jc}$ if the i-th and j-th observations belong to the same domain and $L_{ij} = -1/(N_1\, N_2)$ otherwise.

2.2 Transfer Component Analysis

Transfer Component Analysis (TCA) [10] aims at finding a projection/mapping matrix $\mathbf{W}$ ($N \times d$) such that the linear MMD of the projected feature set $\mathbf{Z} = \mathbf{KW}$ is minimized. Using Eq. (2), the linear MMD of the projected feature set is written as

$$\mathrm{MMD} = \mathrm{tr}((\mathbf{ZZ^T})\mathbf{L}) = \mathrm{tr}(\mathbf{W^T KLKW}) \quad (7)$$

TCA also imposes that the matrix $\mathbf{W}$ should be found such that the covariance matrix of the projected features, $\mathbf{Z}^T\mathbf{HZ}$, is equal to identity matrix. $\mathbf{H}$ denotes the $N \times N$ centering matrix. The final minimization problem for TCA is then set up as

$$\begin{aligned} \phi(\mathbf{W}) &= \mathrm{tr}(\mathbf{ZZ}^T\mathbf{L}) + \lambda\|\mathbf{W}\|_F^2 \\ &= \mathrm{tr}\left[\mathbf{W}^T\left(\mathbf{KLK} + \lambda \mathbf{I}_N\right)\mathbf{W}\right] && (8) \end{aligned}$$

s.t. $\mathbf{Z}^T\mathbf{HZ} = \mathbf{W}^T\mathbf{KHKW} = \mathbf{I}_d$.
where $\lambda\|\mathbf{W}\|_F^2$ is a regularization term used to ensure numerical stability.
TCA matrix solution $\mathbf{W}^*$ contains the N-dimensional eigenvectors corresponding to the d lower eigenvalues of the generalized eigenvalue problem of $(\mathbf{KLK} + \lambda\mathbf{I}_n, \mathbf{KHK})$.

2.3 Classification

A last ingredient consists in exploiting the labels of the source examples. Let denote the $N_1 \times d$ (source) feature matrix $\mathbf{Z}_s = \mathbf{K}_S\mathbf{W}$ obtained after TCA and denote the associated label $N_1 \times 1$ vector $\mathbf{y}_s$.
A supervised classification rule is built on the source training set to predict source label $\hat{\mathbf{y}} = f_\theta(\mathbf{z})$. Because source and target distributions are close (in the sense of MMD), the decision rule can be applied safely to the target set. In this paper, we use a simple linear discriminant analysis (LDA) classifier but other classifiers like logistic regression or SVM can be used instead.

2.4 Focus on the feature

In BCI, each bandpass filtered EEG epoch is summarized by its (spatial) covariance matrix

$$\mathbf{C} = \frac{1}{N_t - 1}\mathbf{XX}^T \quad (9)$$

where $\mathbf{X}$ denotes the $N_c \times N_t$ matrix of the epoched EEG time series, N_c and N_t denote the number of electrodes and the number of samples respectively. Note that other covariance estimate like the one from Ledoit-Wolf can be used here.
A popular feature to resume the EEG epoch is to consider the variance of the spatially filtered EEG epoch, $v = \mathbf{a}^T\mathbf{Ca}$ where $\mathbf{a}$ is the $N_c \times 1$ spatial filter. If several filters are used, one can build the feature vector $\mathbf{u}$ that will represent the log variance of the different filtered signals. For motor-imagery, CSP filters are often used [7] and we use the Gaussian kernel $\mathbf{K}^G$ (6) in the TCA framework.The TCA features would then be $\mathbf{Z} = \mathbf{K^G W}$. Using the kernel trick, we observe that it is actually not necessary to know the nonlinear mapping ψ since only the dot product is needed to build the kernel matrix $\mathbf{K}$. This gives us the opportunity to use Riemannian geometry concepts and directly work on covariance matrices instead of the feature vectors.
In this study, we will use the kernel

$$K_{ij}^R = \exp\left[-\frac{d_R^2(\mathbf{C}_i, \mathbf{C}_j)}{2\sigma^2}\right] \quad (10)$$

to generate TCA features as $\mathbf{Z} = \mathbf{K^R W}$. Here $d_R^2(\mathbf{C}_i, \mathbf{C}_j)$ computes the distance between two symmetric positive definite (SPD) matrices, aka the spatial covariance matrices in a given frequency band. In this work, we consider the usual affine invariant Riemannian metric [2] but other metrics can be chosen. The σ parameter will be selected as the mean (Euclidean/Riemanian) distance among the source and target features [11].
A slight modification of this method is to first whiten the covariance matrices for each domain. This can be achieved by computing the Fréchet mean of all the covariance matrices of one domain (without any label consideration) [2].
The kernel is modified as

$$K_{ij}^{wR} = \exp\left[-\frac{d_R^2\left(\widetilde{\mathbf{C}}_i, \widetilde{\mathbf{C}}_j\right)}{2\sigma^2}\right] \quad (11)$$

where $\widetilde{\mathbf{C}}_k = \mathbf{G}_k^{-1/2}\mathbf{C}_k\mathbf{G}_k^{-1/2}$ and $\mathbf{G}_k$ is the domain barycenter the k-th observation belongs to. This new kernel is identical to the one in (10) except when observations come from different domains.

3 Results

3.1 Dataset

We used the dataset IIa from BCI competition IV [8] consisting of the EEG data of 9 subjects performing 4 types of motor imagery tasks namely the movement imagination of the left

hand, the right hand, feet and tongue.
The experimental paradigm is defined as follows : A trial begins with a fixation cross and an acoustic warning tone. Then, a visual cue associated with an imagery motor task is shown on the computer screen. The subject, seated in front of the screen, performs the demanded task before a short break.
EEG signals were recorded at a sampling frequency of 250 Hz using 22 electrodes and 3 EOG electrodes. Measurements were performed in two sessions of two different days. Each session consists of 72 trials per task (class).
In this study, we focus on left and right hand movement imagery tasks. As in [12], session 1 is used as training session and session 2 as a test session. EEG signals are band-pass filtered in the frequency band of [8-30] Hz, then, epoched to [0.5-2.5] second interval relative to the cue time apparition. Variance and covariance features are extracted on the obtained two-second epochs before applying the TCA method.

3.2 Subject selection

Cross-subject TL can be influenced by both variability between subjects and bad accuracy of the source subjects. Thus, we focus only on 'good subjects' [1], those with good accuracies and good generalization between sessions.
To identify such subjects, we simply learned a LDA classifier on session 1 and applied it on the second session (without any transfer). Features are log-variance after CSP filtering. The mean classification accuracy between sessions given in Table 1 shows that the best results are obtained with subjects S1, S3, S7, S8 and S9. The same subjects were obtained in [1] when covariance matrices were used as features for the classification of the four imagery motor tasks.
In the following, results will be given on this subset of subjects. We note that some subjects are 'illiterate' in the motor imagery paradigm.

3.3 Selection of source domain

For the assessment of our approach, we need to build pairs of source and target data. For each subject (seen as a target), the source is selected as the subject that has the most similar feature distribution to it, in terms of MMD criterion. Fig. 1 shows the MMD between all possible source-target pair when using the recentered covariance matrices as features. The red squares correspond to the closest source in terms of MMD. As expected, the features distribution computed in session 1 is closer to the one computed in session 2 for the same subject than for any other subjects. This shows that the MMD criterion is relevant for the selection of source domain.
The sources selected for the classification based on the recentered covariance feature are represented in Fig. 1 with green squares. For instance the target Subject 1/session 2 will be paired with the source Subject 7/session 1. Note that the MMD criterion depends on the relevance of the features. For the log variance feature vectors, the target Subject 7/session 2 is closer to Subject 3/session 1 than itself in session 1.

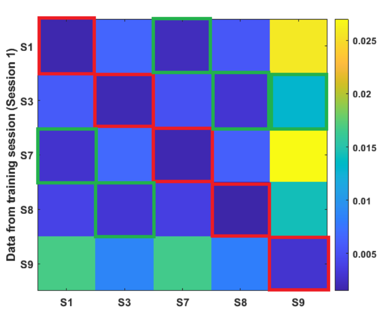

FIGURE 1 – Selection of pairs of source and target when using the recentered covariance matrices as features.

TABLE 1 – Mean classification accuracy using session 1 as training set and session 2 as test set, and vice-versa. Here no TCA is used to transfer between sessions.

Subject	**8**	**3**	**9**	**1**	**7**	4	6	5	2
S1S2	93.05	97.22	92.36	90.97	81.94	70.13	66.66	54.86	54.16
S2S1	98.61	91.66	93.75	81.25	75.69	64.58	67.36	58.33	50
Mean accuracy	95.83	94.44	93.05	86.11	78.81	67.35	67.01	56.59	52.08

3.4 Hyperparameter selection

For the hyperparameter selection $\{\lambda, d\}$, we adopted an empirical approach. We vary the parameters values and we assess for each value the classification accuracy. This can be seen as a wrapper method. Best hyperparameter values are selected as those giving the best median accuracy among all subjects.
The parameters λ was varied in $\{0.005, 0.01, 0.05, 0.1\}$ and the reduced dimension size d was varied in $\{2, 3, 4, 5, 6\}$
Fig. 2 illustrates the described approach when Riemannian kernel $\mathbf{K}^R$ is used in TCA. Table 2 gives the selected values for the different feature types where $log(var)$ refers to the case where log-variance of the CSP filtered signals are used as features ; cov and $wcov$ refer to the covariance features and recentered covariance features respectively.

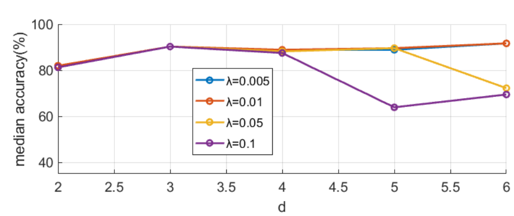

FIGURE 2 – Median accuracy after TCA when using the covariance matrices of EEG epochs as features

TABLE 2 – Selected values of hyperparameters

Features	log(var)	cov	wcov
d	3	3	3
λ	0.1	0.01	0.01

3.5 Performance accuracy

A box plot for the variance based features before and after applying TCA method is given in Fig. 3. We observe that TCA has dramatically improved the results : the median accuracy value has for instance increased from 50% to almost 80%. In

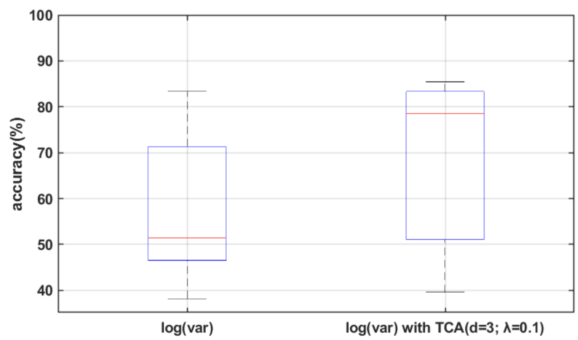

FIGURE 3 – Transfer learning with TCA (right) and without (left). TCA is applied on log variance feature vectors (log(var)).

Fig. 4, we illustrate the LDA classification accuracies obtained with different types of features when applying the TCA method. We observe that covariance-based features (recentered or not) result in better accuracy that the variance-based features. Recentering the covariance matrices has improved the classification performance : from a minimum performance value of 45%, we get a minimum value of 70% thanks to the Riemannian recentering operation.

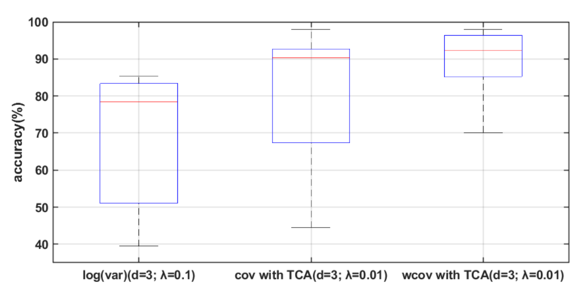

FIGURE 4 – TCA classification accuracy for different input feature type.

4 Conclusion

The paper successfully applies TCA transfer learning concepts in the context of brain-computer interface. Covariance matrices are easily handled using the kernel trick.However, it should be noted that the TCA method requires the selection of some hyperparameters. The approach, used in this paper, uses all of the test data which may lead to some bias in the results. In the near future, an approach for hyperparameter selection that uses a subset of test samples will be investigated. Moreover, the TCA method will be tested on other datasets. Cross-session variability will also be investigated.

Références

[1] P. Zanini et al., *Transfer learning : a Riemannian geometry framework with applications to Brain-Computer Interfaces*, IEEE Trans Biomed Eng, vol. 65, pp. 1107-1116, 2018.

[2] A. Barachant et al., *Classification of covariance matrices using a Riemannian-based kernel for BCI applications*, Neurocomputing, Vol. 112, pp. 172-178, 2013.

[3] P. Rodrigues et al., *Riemannian Procrustes Analysis : Transfer Learning for Brain-Computer Interfaces*, IEEE Trans Biomed Eng, vol. 66, pp. 2390-2401, 2018.

[4] W.-L. Zheng et al., *Transfer Components between Subjects for EEG-based Emotion Recognition*, 2015.

[5] O. Demsy et al., *Inter-Subject Transfer Learning Using Euclidean Alignment and Transfer Component Analysis for Motor Imagery-Based BCI*, IEEE International Conference on Systems, Man, and Cybernetics (SMC),17-20 October, 2021. Melbourne, Australia.

[6] X. Tang et al., *Transfer Learning : Rotation Alignment with Riemannian Mean for Brain-Computer Interfaces and Wheelchair Control*, IEEE Transactions on Cognitive and Developmental Systems, 20xx.

[7] B. Blankertz et al., *Optimizing Spatial filters for Robust EEG Single-Trial Analysis*, IEEE Signal Processing Magazine, vol. 25, no. 1, pp. 41-56, 2008.

[8] C. Brunner et al., *BCI Competition 2008, Graz data set A*, pp.1-6, 2008.

[9] K.M. Borgwardt et al., *Integrating structured biological data by kernel maximum mean discrepancy*, Proceedings of the 14th International Conference on Intelligent Systems for Molecular Biology, 49-57, 2006.

[10] S.J. Pan et al., *Domain Adaptation via Transfer Component Analysis*, IEEE Transactions on Neural Networks, vol. 22, no. 2, pp. 199-210, Feb. 2011.

[11] G. Matasci et al., *Semi supervised Transfer Component Analysis for Domain Adaptation in Remote Sensing Image Classification*, IEEE Transactions on Geoscience and Remote Sensing, vol. 53, no. 7, pp. 3550-3564, July 2015.

[12] A.Cruz et al. *Spatial filtering based on Riemannian distance to improve the generalization of ErrP classification*, Neurocomputing,Volume 470,pp 236-246,2022.