Inter-Subject Transfer Learning Using Euclidean Alignment and Transfer Component Analysis for Motor Imagery-Based BCI

Orvin Demsy[1], David Achanccaray[1] and Mitsuhiro Hayashibe[2]

***Abstract*— Brain-computer interface (BCI) requires calibration phase to learn user-specific decoder that translates brain signals into desired commands. Calibration phase can be lengthy and exhausting for motor imagery-based (MI) BCI, and can potentially reduce the effectiveness of BCI system. Transfer learning (TL) approaches have been implemented in the field of BCI to tackle this problem. Transfer learning based on domain adaptation reduces domain discrepancy between two domains. Transfer component analysis (TCA) is one domain adaptation technique that maps features of both domains into a new space while simultaneously reducing their domain discrepancy. Recently, Euclidean alignment (EA) is used as transfer learning technique in BCI preprocessing step to align electroencephalography (EEG) trials between subjects. This paper proposed a combination of both EA and TCA (EA-TCA) as a TL approach for inter-subject transfer learning. This paper also proposed TCA method that can reuse existing projection matrix to transform new target data (TCA-W). The efficacy of EA to this method is also observed (EA-TCA-W). Results indicate that EA-TCA and EA-TCA-W outperform classification accuracy of those without EA by 3.32% and 6.50%, respectively. This concludes that EA can improve performance of both conventional TCA and proposed TCA-W.**

I. Introduction

Brain-computer interface (BCI) allows direct communication from the brain signals to external devices bypassing peripheral nerves and muscles. Typical steps required for BCI system to work are signal acquisition, preprocessing, feature extraction, classification, and communication to devices. Electroencephalography (EEG) is the most popular way to non-invasively acquire brain signals due to its safety, practicality, and portability [1].

Brain signal patterns are modulated to operate BCI system. These modulation techniques are also called paradigms, some popular ones are steady-state visually evoked potential (SSVEP), P300, and motor imagery (MI) [1]. This study will focus on the MI paradigm. MI task could be imagination of left, right hand, or feet movement. MI-based BCI has experienced rapid growth over the past few years, due to its usefullness for stroke-rehabilitation [2].

However, MI-based BCI system requires long calibration process due to the inherent characteristic of brain signals that are non-linear and non-stationary. This means brain signals are highly varied not only between subjects, but even with the same subject on different sessions. Several transfer learning techniques, such as Riemannian alignment and weighted transfer learning, have been implemented to BCI systems to tackle this problem [3].

Transfer learning (TL) is a method in machine learning where knowledge/information of one domain (source) can be used to influence performance of other domain (target). TL may have different terminologies in different discipline of study such as knowledge reuse, pre-training, fine-tuning, and domain adaptation [4]. Domain adaptation refers to reducing difference between domains, thus making two data distributions more similar. TL has been successfully applied in other applications, such as image, text-related tasks, bioinformatics, and recommender systems [5].

TL in MI-based BCI can be applied to typical BCI steps shown in Fig. 1. Namely, preprocessing, feature extraction and classification step [6]. In the preprocessing step, Zanini et al. [7] proposed Riemannian alignment, which aims to center the covariance of resting state EEG signals of all domains to a reference covariance matrix. Another approach is called Euclidean alignment (EA) [8], which aims to align EEG trials across subjects, such that their data distributions are more similar. Unlike Riemannian alignment, EA has low computational cost and can be applied directly on EEG trials, such that subsequent signal processing can be applied directly on them. In the feature extraction step, most TL approaches focus on modifying common spatial patterns (CSP). Dai et al. [9] proposed transfer kernel CSP, which learns domain-invariant kernel and finds largest energy difference between two classes. Liyanage et al. [10] used domain adaptation technique named transfer component analysis (TCA), to transfer cross-session knowledge for MI-based BCI. In the classification step, Azab et al. [6] proposed weighted TL. This approach modifies regularization term in logistic regression classifier such that the cross-entropy loss between target and source data are minimized.

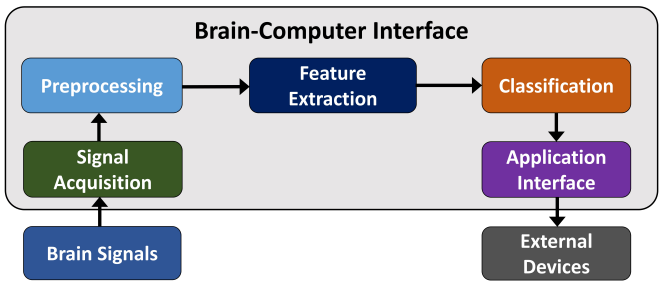


Fig. 1. Typical BCI steps, where transfer learning can be applied independently to preprocessing, feature extraction and classification step

[1]Orvin Demsy and David Anchanccaray are with the Department of Robotics, Tohoku University, Sendai, Japan, demsy.orvin.r5@dc.tohoku.ac.jp, david.ad@dc.tohoku.ac.jp

[2]Mitsuhiro Hayashibe is with Graduate School of Biomedical Engineering and Department of Robotics, Tohoku University, Sendai, Japan hayashibe@tohoku.ac.jp

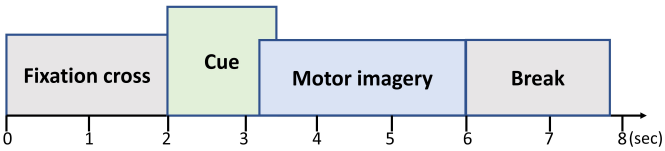


Fig. 2. MI tasks timing scheme

The study of using TCA techniques for BCI application was rarely done. To the best of our knowledge, there were only two studies that implements TCA to BCI data, which were to MI dataset and to affective BCI dataset. TCA successfully handled non-stationarity issue in EEG data by finding latent space between subjects [10], [11]. However, those works only used TCA. This means TL is being done only in feature extraction step, while applying TL on multiple BCI steps has proven successful in increasing performance [12]. This paper combines EA in preprocessing step and TCA in feature extraction step as a TL approach for MI-based BCI.

II. Methods

A. Dataset Description

The BCI Competition IV Dataset 2a was used to evaluate proposed methods [13]. The dataset contains EEG data of 9 subjects performing MI task. The experiment used 22 EEG and 3 EOG electrodes recorded at 250 Hz. Subjects were seated in front of a computer screen where different cues are shown. There are four kinds of cues, each associated with different tasks, namely, left hand, right hand, foot, or tongue. Two different sessions, training and evaluation were recorded on two different days, each session consists of 6 runs, and each run consists of 48 trials (12 trials for each task). In this paper only left and right hand data of training session are used, thus each subject will have total of 144 trials, with 72 trials of each class. Fig. 2 shows timing scheme of each trial.

A 50th-order FIR band-pass filter with cut-off frequencies at [8, 30] Hz is applied. EEG signals from 0.5 to 3.5 seconds after cue onset are extracted as signals that correspond to MI task.

B. Transfer Learning

Table I describes common notation used in transfer learning setting. Domain is denoted by $\mathbb{D} = \{\mathcal{X}, P(X)\}$ where $X \in \mathcal{X}$, and task is denoted by $\mathbb{T} = \{\mathcal{Y}, f(\cdot)\}$. There exists source domain $\mathbb{D}_s = \{\mathcal{X}_s, P(X_s)\}$ from which knowledge is extracted and transferred to target domain $\mathbb{D}_t = \{\mathcal{X}_t, P(X_t)\}$. Various condition of source and target domains lead to different cases of TL.

In BCI, there are four different TL cases, namely cross-subject, cross-session, cross-task, and cross-device. This paper focuses on cross-subject TL [3].

C. Common Spatial Pattern

Common spatial pattern is a commonly used feature extraction method for two classes MI-based BCI [2]. This method seeks to find spatial filters such that variance of one class is minimized while variance of other class is maximized [14].

TABLE I
Notations used in transfer learning

Notation	Description	Notation	Description
X	Input feature	$\mathbb{D}$	Domain
$\mathcal{X}$	Feature space	$\mathbb{T}$	Task
$P(X)$	Marginal probability	s	Source
$\mathcal{Y}$	Label space	t	Target
$f(\cdot)$	Predictive function		

EEG signals are projected as follows:

$$Z = W^T X \tag{1}$$

where, Z is the projected EEG signals, W is the spatial filter, and X is EEG signals, lastly, the features can be extracted by computing log-variance of Z as follows:

$$f_p = \log\left(\frac{\text{var}\,(Z_p)}{\sum_{i=1}^{2m}\text{var}\,(Z_i)}\right) \quad p = 1:2m \tag{2}$$

where $\text{var}\,(Z_i)$ is the variance of Z in row i, and $2m$ is the total number of filters in W.

D. Euclidean Alignment (EA)

Euclidean alignment (EA) is an alignment technique such that distribution between two subjects becomes more similar. Hence, classifier for target subject can be trained using any source subject data [8].

In order to perform alignment, EA needs reference matrix $\bar{R}$, which is defined as,

$$\bar{R} = \frac{1}{n}\sum_{i=1}^{n} X_i X_i^T, \tag{3}$$

where i is the i-th EEG trial, and n is the total trials. $\bar{R}$ computes the mean of all covariances from a subject, then alignment can be computed by,

$$\tilde{X}_i = \bar{R}^{-1/2} X_i, \tag{4}$$

After EA is applied to all subjects, mean covariance of all subjects will be equal to identity matrix. This implies a more uniform and similar distribution across all subjects, which will be beneficial for TL [8].

E. Transfer Component Analysis (TCA)

TCA is a domain adaptation technique that seeks to reduce dissimilarity between two domains by finding a transformation $\phi(\cdot)$ such that two domains are more similarly distributed in Reproducing Kernel Hilbert Space (RKHS). TCA also preserves important characteristic of both domains, in other words, $\phi(\cdot)$ maps both features into a new space that is invariant across two domains [10], [15].

Domain adaptation setting assumes that marginal probability of source and target data are different $P\,(X_s) \neq P\,(X_t)$ but their conditional probability are the same $P\,(Y_s \mid X_s) = P\,(Y_t \mid X_t)$. TCA seeks $\phi(\cdot)$ such that marginal probability

of features in projected space are similar $P(X'_s) \approx P(X'_t)$, this is achieved by minimizing maximum mean discrepancy (MMD) distance $MMD(X'_s, X'_t)$ [16]. By using kernel trick, MMD distance can be defined as $tr(KL)$, where K

$$K = \begin{bmatrix} K_{s,s} & K_{s,t} \\ K_{t,s} & K_{t,t} \end{bmatrix} \in \mathbb{R}^{(n_s+n_t)\times(n_s+n_t)} \tag{5}$$

is the kernel matrix, with $K_{s,s}$, $K_{t,t}$, $K_{t,s}$ are Gram matrices of source data, target data, and cross data in projected space, respectively, n_s and n_t are number of source and target data. L is multiplier to K where $L_{ij} = 1/n_s^2$ if $x_i, x_j \in X_s$, else $L_{ij} = 1/n_t^2$ if $x_i, x_j \in X_t$, and $L_{ij} = -\frac{1}{n_t n_s}$ otherwise.

By decomposing matrix K and defining resulting kernel, there exists matrix $W \in \mathbb{R}^{(n_s+n_t)\times m}$, that maps features to m-dimensional space. Then, the objective function of TCA to learn $\phi(\cdot)$ can be defined as:

$$\begin{aligned} \min_{W} \quad & \operatorname{tr}\left(W^T W\right) + \mu \operatorname{tr}\left(W^T KLKW\right) \\ \text{s.t.} \quad & W^T KHKW = I, \end{aligned} \tag{6}$$

where $I \in \mathbb{R}^{m\times m}$ is identity matrix, and $H = I_{n_s+n_t} - \frac{1}{n_s+n_t}\mathbf{1}\mathbf{1}^T$ is the centering matrix, where $\mathbf{1} \in \mathbb{R}^{n_s+n_t}$ is column vector of ones, and $I_{n_s+n_t} \in \mathbb{R}^{(n_s+n_t)\times(n_s+n_t)}$ is the identity matrix. Solutions of W are the m leading eigenvectors of $(I + \mu KLK)^{-1}KHK$ [15].

F. Evaluation Scheme

The dataset consists of nine subjects. In the evaluation, each subject will act a as target once. The remaining eight subjects will be source sj, where j = $(1, ..., m)$, where m is total source. Let X and y be raw EEG signals and its corresponding label, where $X \in \mathbb{R}^{C\times t}$, C is the number of channel and t is number of time samples. Data from a single source is denoted by $(X_{sj}, y_{sj})_{i=1}^{n_{sj}}$. Two sets of data from the current target $(X_{t1}, y_{t1})_{i=1}^{n_{t1}}$ and $(X_{t2}, y_{t2})_{i=1}^{n_{t2}}$ are also required for evaluation, where i correponds to the i-th trial, and n_{sj}, n_{t1}, n_{t2} corresponds to total trial of source data, target data 1 and target data 2, respectively. During evaluation n_{sj}, n_{t1}, n_{t2} are randomly picked, thus for each method the evaluation is repeated 20 times. Note that $(X_{t1}, y_{t1})_{i=1}^{n_{t1}} \neq (X_{t2}, y_{t2})_{i=1}^{n_{t2}}$.

The evaluation is divided into two phases, distance measuring phase, and classification phase. Fig. 3 shows flowchart of a distance measuring phase. MMD, a non-parametric distance measure between data distributions, is used to compute distance [16]. Let X'_{t1} and X'_{sj} be CSP features obtained from (X_{t1}, y_{t1}) and (X_{sj}, y_{sj}), respectively. In distance measuring phase, MMD computes distance between X'_{t1} to each X'_{sj}. Then, three most similar source subjects ssk, where k = (1, 2, 3), are chosen and passed on to classification phase. Raw EEG signals and correponding labels of k-th similar source subject are denoted by X_{ssk} and y_{ssk}, respectively.

Fig. 4 shows flowchart of the classification phase. In this phase X_{ssk}, X_{t1}, X_{t2} are processed, depending on each method's algorithm. CSP features are denoted by X'. CSP filters are computed using (X_{ssk}, y_{ssk}) and applied on X_{ssk}, X_{t1}, X_{t2}, resulting in X'_{ssk}, X'_{t1}, X'_{t2}, respectively. First and last three spatial filters of CSP are chosen. TCA features are denoted by X'', such that X''_{ssk}, X''_{t1}, and X''_{t2}, mean TCA features of X'_{ssk}, X'_{t1}, and X'_{t2}, respectively. Projection matrix after performing TCA is denoted by W. Support vector machine (SVM) is used as classifier in all methods. The highest accuracy between ssk is chosen as the result.

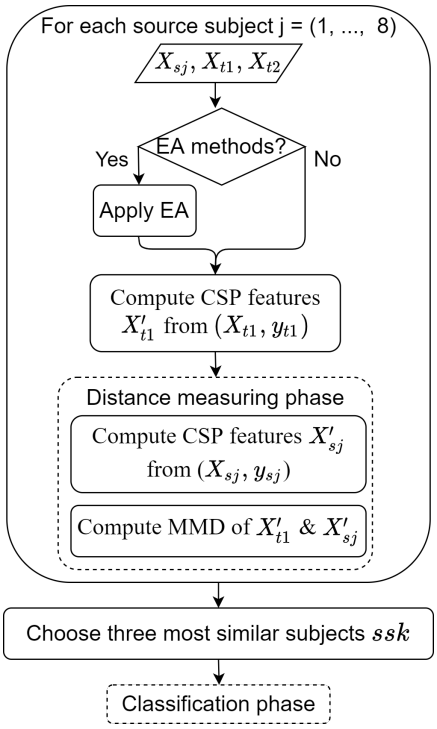


Fig. 3. Flowchart of distance measuring phase, where distance between target subject to each source subject is measured to obtain three most similar source subjects

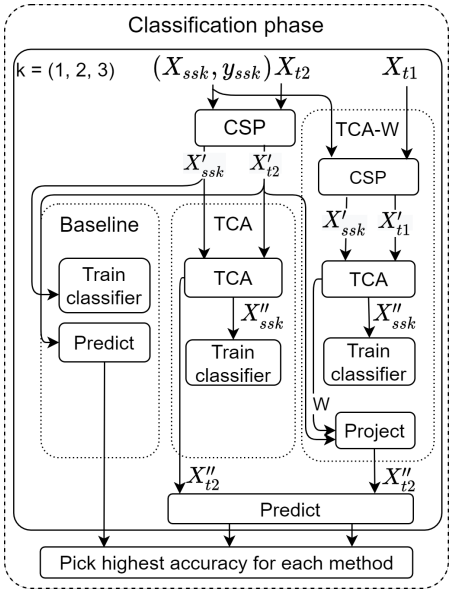


Fig. 4. Flowchart of classification phase, where baseline, TCA, and TCA-W methods process data of target subject and three most similar source subjects.

In the baseline method, classifier is trained only with X'_{ssk}, and used to predict X'_{t2}. TCA is the conventional TCA method. X''_{ssk} and X''_{t2} are obtained from computing TCA on X'_{ssk} and X'_{t2}. Classifier is trained on X''_{ssk} and used to predict X''_{t2}. TCA-W is the proposed TCA, where projection matrix stored when TCA features X''_{ssk}, X''_{t1} are computed. Hence, when there is a new set of features X'_{t2}, existing W can be used to project it into X''_{t2}. Classifier is trained only once using X''_{ssk}.

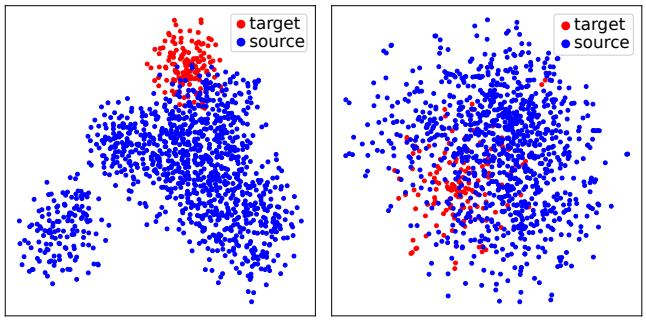
The effect of EA on each method is observed. Since EA only modifies the raw EEG signals, each algorithm can be implemented directly on them. Methods using EA are denoted by EA-baseline, EA-TCA, and EA-TCA-W, respectively.

III. Results and Discussions

All results shown in this section are evaluated following description in section II-F. The variable n_{sj}, n_{t1} and n_{t2} are set to 30, 30, 50, respectively. This means 30 source trials and 30 target trials are used to compute distance between target and source in distance measuring phase. Then 50 target trials are predicted using respective methods. First, the effect of EA on data is observed using t-Stochastic Neighbor Embedding (t-SNE) as a visualization technique [17]. In TCA, kernel type and dimensionality of latent space, are parameters which can be adjusted. Meanwhile, SVM can also use different type of kernels. Hence, different number of TCA dimensionality and best pair of TCA and SVM kernel are evaluated. Next, the effect of number source trials on accuracy is also investigated. Lastly, all methods are evaluated based on best parameters observed in each subsection.

A. EA Visualization on All Subjects

EA seeks to align data of different domains, such that their distributions are more similar. Fig. 5 shows effect of applying EA on the data. Fig. 5a and 5b shows data distribution before and after applying EA, respectively. Data of target subject are colored red, while rest of source subjects are colored blue.



(a) Before EA (b) After EA

Fig. 5. t-SNE visualization of data before EA (a) and after EA (b), when one subject acts as target, and remaining subjects as source

As can be seen, in Fig. 5a, data distribution of target (red) is more clustered to specific location, this implies marginal probability difference between target and source. While in Fig. 5b, after applying EA, the target data is more sparse and appears to be similarly distributed to source data (blue). This shows the effectiveness of EA to make two data distributions more similar [8].

TABLE II
Mean accuracy of four methods using TCA for different dimensionality (dim)

dim	TCA	EA-TCA	TCA-W	EA-TCA-W	Mean
	Accuracy				
1	55.39	63.53	64.64	69.46	63.26
2	69.14	74.62	68.27	74.30	71.58
3	70.14	73.72	68.13	74.98	71.74
4	**70.64**	**74.66**	**69.14**	74.60	72.26
5	70.22	73.87	69.10	74.91	72.03
6	70.33	73.93	69.11	**74.93**	72.08

B. TCA dimension

The dimensionality of latent space is a parameter in TCA that can be adjusted. In order to achieve similar dimensionality to CSP features, TCA dimension from 1 to 6 are observed for four methods that use TCA. Table II shows how different TCA dimension will affect accuracy. The boldfaced numbers indicate highest accuracy in each method. The results suggest that for all methods, low dimensions (1 or 2) are unlikely to be a good representation of latent space between subjects, while higher dimensions (greater than 3) show better results. In the subsequent evaluation, TCA dimension of 4 is chosen.

C. TCA and SVM Kernels

TCA kernel and SVM kernel are parameters that can be adjusted. Fig. 6 shows effect of combinations of TCA and SVM kernels on accuracy. This evaluation is done only on the four methods that use TCA. It can be seen that for TCA and EA-TCA, both combinations of linear and radial basis function (rbf) kernel for TCA and SVM do not significantly affect accuracies. On the other hand, for TCA-W and EA-TCA-W, using rbf as SVM kernel can drop accuracies to approximately 50%. This implies that for proposed methods, classifier with rbf kernel failed to find good feature representation in the mapped space [18].

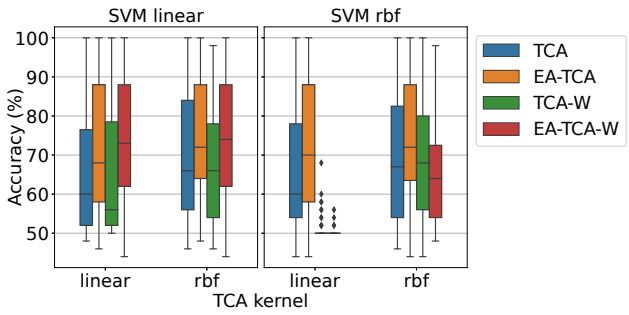


Fig. 6. Accuracy of four methods using TCA with respect to different SVM and TCA kernel combinations

TABLE III
MEAN ACCURACY (%) OF BASELINE, TCA, AND TCA-W WITH AND WITHOUT EA

Subject	Baseline	EA-Baseline	TCA	EA-TCA	TCA-W	EA-TCA-W
1	70.80	86.30	77.20	86.90	71.40	**88.40**
2	51.10	**56.70**	51.30	53.40	52.00	52.50
3	97.40	**98.30**	95.90	96.90	95.70	96.50
4	51.90	**72.30**	56.70	70.10	55.30	72.00
5	51.60	55.50	53.10	53.50	50.70	**58.00**
6	52.50	**71.20**	64.30	69.10	62.90	69.00
7	59.80	72.80	63.20	71.00	59.10	**75.10**
8	88.30	90.00	**96.10**	91.20	95.50	91.10
9	68.60	**76.90**	73.60	69.20	72.90	71.40
Mean	65.78 ±16.20	75.56 ±13.49	70.16 ±16.01	73.48 ±14.53	68.39 ±16.26	74.89 ±13.96

Overall, results suggest that pair of linear SVM kernel and rbf TCA kernel could maximize accuracies of these four methods.

D. Varying Number of Source Trials

Accuracy of each method by varying number of source trials is investigated. When acting as a source, each subject has maximum of 144 trials. Fig. 7 shows comparison of accuracy of each method with respect to different number of source trials, starting from 10 to 130, with interval of 30.

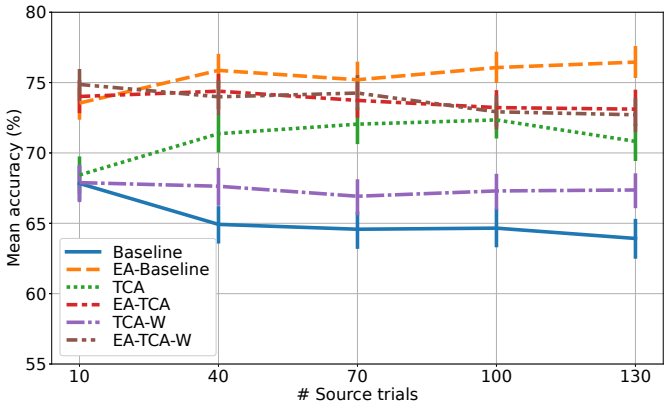


Fig. 7. Comparison of mean accuracy of each method by varying number of source trials

With only 10 source trials, TCA and TCA-W performs similarly to baseline method, this implies that extremely low source trials might prevent TCA from learning a good invariant space between subjects. Meanwhile, methods with EA show a significant improvement even only with 10 source trials. TCA and TCA-W begin to outperform baseline method as number of source trials are increased. It is also obvious that methods with EA perform significantly better than those without EA. Results also suggest that mean accuracy of each method is relatively stable after approximately greater than 30 source trials are used, which implies that even with few source trials, satisfactory classification accuracy can be achieved, especially for methods with EA.

E. Result Comparison of All Methods

Table III shows mean accuracy (%) of all methods on each subject using parameter values defined in previous subsections. Boldfaced numbers are highest accuracies obtained in each subject. Saphiro-Wilk test was performed to test normality of accuracies in each method. Since results suggest that none of them is normally distributed, Wilcoxon signed-rank test was performed to check the statistical significance between methods.

It can be seen that TCA outperformed the baseline by 4.38% ($p < 0.05$). The proposed TCA-W also slightly outperformed the baseline by 2.61% ($p < 0.05$). This shows that both conventional TCA and proposed TCA-W successfully reduced domain discrepancy between subjects.

Furthermore, applying EA to both TCA methods are proven to be successful. Compared to the conventional TCA and TCA-W, EA-TCA and EA-TCA-W improved the mean accuracy by 3.32% ($p < 0.05$) and 6.50% ($p < 0.05$). Compared to baseline method, EA-baseline also improved mean accuracy by 9.78% ($p < 0.05$).

The efficacy of proposed TCA-W is investigated further. With respect to TCA, TCA-W decreased the accuracy by 1.77% ($p < 0.05$), and with respect to EA-TCA, EA-TCA-W increased the accuracy by 1.4% ($p < 0.05$). Despite not showing a significant increase in accuracy, the computational cost for TCA-W was quite satisfactory. Preliminary experiment showed that for 20 repetitions of each method, the proposed TCA-W reduced the processing time by 0.42s compared to conventional TCA. The evaluation was done using a computer with Intel Core i7-8750H CPU @ 2.20Ghz.

Next, statistical significant values between EA-baseline, EA-TCA, and EA-TCA-W are compared. EA-baseline and EA-TCA were different ($p < 0.05$). EA-baseline and EA-TCA-W were not different ($p > 0.05$). EA-TCA and EA-TCA-W were different ($p < 0.05$). This results showed that among methods that uses EA, EA-baseline and EA-TCA-W shows no significant difference.

To the best of our knowledge, so far implementation of TCA has been using specific set of data from both domain, which means, each time a new set of data comes, new projection matrix and classifier are computed. The proposed TCA-W and EA-TCA-W can work with incremental target data such that projection matrix and classifier are computed once and reused for projection and prediction. This can potentially reduce processing time.

Note that, the CSP spatial filters are built using only 30

source trials. The filters are then applied to source data to train the classifier. Next, target data is also filtered and predicted using the trained classifier. This setting suggests the possibility of using only source data to build filters for target subject, thus completely avoiding calibration phase for target subject. In the section II-F, X_{t1} is required for each method only to choose most similar source to current target. In fact, when dealing with TL case with multiple sources, several studies pool together all sources data into one big source [8], [11]. Hence, distance measuring phase is not required since there is only one source.

IV. Conclusion

This paper evaluates the feasibility of applying transfer learning techniques in two BCI steps, namely, preprocessing and feature extraction step. Euclidean alignment (EA) is implemented in the preprocessing step, while transfer component analysis (TCA) is implemented in the feature extraction step. A TCA method which can reuse existing projection matrix and classifier (TCA-W) is also evaluated. All methods are evaluated using MI-based BCI dataset. Results indicate that both TCA and proposed TCA-W outperform accuracy of baseline method. Next, the effect of applying EA to each method is also evaluated. Results indicate that methods with EA outperform accuracies of those without EA. The proposed TCA-W also demonstrates faster processing time compared to conventional TCA. Few numbers of source trials are used to train classifier, without even needing the target data. This suggests TL methods evaluated in this paper could reduce the calibration time in MI-based BCI.

References

[1] C. S. Nam, A. Nijholt, and F. Lotte, *Brain-Computer Interfaces Handbook Technological and Theoretical Advances*. Boca Raton: CRC Press, LLC, 2018, vol. 53, no. 9.

[2] D. Achanccaray, S. Izumi, and M. Hayashibe, "Visual-electrotactile stimulation feedback to improve immersive brain-computer interface based on hand motor imagery," *Computational Intelligence and Neuroscience*, vol. 2021, Feb 2021.

[3] D. Wu, Y. Xu, and B.-L. Lu, "Transfer Learning for EEG-Based Brain-Computer Interfaces: A Review of Progress Made Since 2016," pp. 1–16, 2020. [Online]. Available: http://arxiv.org/abs/2004.06286

[4] Q. Yang, Y. Zhang, W. Dai, and S. J. Pan, *Transfer Learning*, 2020.

[5] F. Zhuang, Z. Qi, K. Duan, D. Xi, Y. Zhu, H. Zhu, H. Xiong, and Q. He, "A Comprehensive Survey on Transfer Learning," pp. 1–31, 2019. [Online]. Available: http://arxiv.org/abs/1911.02685

[6] A. M. Azab, L. Mihaylova, K. K. Ang, and M. Arvaneh, "Weighted Transfer Learning for Improving Motor Imagery-Based Brain-Computer Interface," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 27, no. 7, pp. 1352–1359, 2019.

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

[8] H. He and D. Wu, "Transfer Learning for Brain-Computer Interfaces: A Euclidean Space Data Alignment Approach," *IEEE Transactions on Biomedical Engineering*, vol. 67, no. 2, pp. 399–410, 2020.

[9] M. Dai, D. Zheng, S. Liu, and P. Zhang, "Transfer kernel common spatial patterns for motor imagery brain-computer interface classification," *Computational and Mathematical Methods in Medicine*, vol. 2018, pp. 1–9, 03 2018.

[10] S. R. Liyanage, J. S. Pan, H. Zhang, K. K. Ang, C. Guan, J. X. Xu, and T. H. Lee, "Stationary transfer component analysis for brain computer interfacing," *Proceedings of the IASTED International Conference on Engineering and Applied Science, EAS 2012*, no. Eas, pp. 335–340, 2012.

[11] Wei-Long Zheng, Y.-Q. Zhang, J.-Y. Zhu, and B.-L. Lu*, "Transfer Components between Subjects for EEG-based Emotion Recognition," 2015.

[12] D. Wu, R. Peng, J. Huang, and Z. Zeng, "Transfer Learning for Brain-Computer Interfaces: A Complete Pipeline," pp. 1–9, 2020. [Online]. Available: http://arxiv.org/abs/2007.03746

[13] C. Brunner and R. Leeb, "BCI Competition IV 2008 – Graz data set A," pp. 1–6, 2008.

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

[15] S. J. Pan, I. W. Tsang, J. T. Kwok, and Q. Yang, "Domain adaptation via transfer component analysis," *IEEE Transactions on Neural Networks*, vol. 22, no. 2, pp. 199–210, 2011.

[16] A. Gretton, K. Borgwardt, M. J. Rasch, B. Scholkopf, and A. J. Smola, "A kernel method for the two-sample problem," 2008.

[17] L. van der Maaten and G. Hinton, "Visualizing data using t-sne," *Journal of Machine Learning Research*, vol. 9, no. 86, pp. 2579–2605, 2008. [Online]. Available: http://jmlr.org/papers/v9/vandermaaten08a.html

[18] C.-J. L. C.-W. Hsu, C.-C. Chang, "A practical guide to support vector classification," https://www.csie.ntu.edu.tw/∼cjlin/, 2016, accessed: 2021–04-22.