ESANN 2024 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Bruges (Belgium) and online event, 9-11 October 2024, i6doc.com publ., ISBN 978-2-87587-090-2. Available from http://www.i6doc.com/en/.

Geometric Deep Learning to Enhance Imbalanced Domain Adaptation in EEG

Shanglin Li[1,2], Motoaki Kawanabe[1,3] and Reinmar Kobler[1,3]

1- Advanced Telecommunications Research Institute International, Kyoto, Japan
2- Nara Institute of Science and Technology, Nara, Japan
3- RIKEN Artificial Intelligence Project, Tokyo, Japan

Abstract. Electroencephalography (EEG) based brain-computer interfaces (BCIs) face great challenges in generalizing across different domains (i.e., sessions and subjects) without costly supervised calibration. To avoid supervised calibration, transfer learning, particularly unsupervised domain adaptation, has been a popular approach. In this work, we focus on a geometric deep learning framework previously proposed for EEG-based mental imagery BCIs. The framework aligns marginal feature distributions in latent space, assuming identical label distributions across domains. Here, we propose a novel approach integrating data augmentation and clustering techniques to align the latent distributions under label shifts.

1 Introduction

A brain-computer interface (BCI) enables direct communication between the brain and external devices, offering great potential for rehabilitation and communication [1]. Despite their capabilities, electroencephalography (EEG) based BCIs currently suffer from low signal-to-noise ratio, insufficient specificity, and domain shifts (e.g., changes in the data distribution).

Domain shifts have been traditionally mitigated by collecting labeled calibration data and training domain-specific models [1]. However, this approach is resource-intensive and time-consuming. As an alternative, unsupervised domain adaptation (UDA) learns a model from labeled source domains that performs effectively on different (but related) unlabeled target domains [1]. Within the BCI field, UDA primarily addresses inter-session and inter-subject transfer learning (TL) problems [2], aiming to achieve robust generalization across domains (i.e., sessions and subjects) without supervised calibration.

In our previous work, we developed a geometric deep learning framework, denoted TSMNet [3], to perform statistical alignment on the symmetric, positive definite (SPD) manifold. TSMNet jointly learns a convolutional feature extractor and tangent space mapping (TSM) on the SPD manifold equipped with the affine invariant Riemannian metric that is well-suited for EEG data due to its inherent invariance to linear mixing of latent sources [4]. Many UDA frameworks, including TSMNet, align the marginal feature distributions, implicitly assuming identical label distributions across domains. However, label shifts are frequently encountered in practice, and marginal feature alignment under label shifts can increase the generalization error [5]. Recent approaches frame this alignment problem as an imbalanced multi-source and multi-target UDA problem [6].

This paper introduces an extension to TSMNet, enhancing its capability to simultaneously address feature and label shifts. To maintain the TSMNet training scheme and online extendability, we limit our approach to a source-free unsupervised domain adaptation (SFUDA) problem, where the pre-trained source model is available instead of raw training data.

2 Preliminaries

Imbalanced multi-source multi-target UDA. Let x denote the input data, y the corresponding output labels, p_{s_i} the i-th source domain probability distributions, and q_{t_j} the j-th target domain probability distributions. We assume that all the $p_{s_i}(x)$ can be different from each other and different from $q_{t_j}(x)$. In the multi-source, multi-target unsupervised domain adaptation scenario, we assume that all the $p_{s_i}(y)$ are equal to each other but different from $q_{t_j}(y)$ (Figure 1a). We additionally assume that the distribution shifts in p_{s_i} and q_{t_j} are dominated by translations on the covariance matrices. Given N source domains $\{\mathcal{D}_{s_i}\}_{i=1}^{N}$ and M target domains $\{\mathcal{D}_{t_j}\}_{j=1}^{M}$, each source domain $\mathcal{D}_{s_i} = \{(x_{s_i}, y_{s_i})\}_{i=1}^{l_{s_i}}$ with l_{s_i} labeled examples, and each target domain $\mathcal{D}_{t_j} = \{(x_{t_j})\}_{j=1}^{l_{t_j}}$ with l_{t_j} unlabeled examples. The goal is to transfer the knowledge learned from $\mathcal{D}_s$ to $\mathcal{D}_t$ and learn a target prediction function $h_t : x_t \to y_t$ with only target data $\{(x_{t_j})\}_{j=1}^{l_{t_j}}$ and the source prediction function $h_s : x_s \to y_s$.

Riemannian geometry and TSMNet. The smooth manifold of real $D \times D$ SPD matrices $\mathcal{S}_D^+ = \{Z \in \mathbb{R}^{D\times D} : Z^T = Z, Z \succ 0\}$ together with an inner product on the tangent space $\mathcal{T}_Z\mathcal{S}_D^+$ at each point Z forms a Riemannian manifold. Here, we use the affine invariant Riemannian metric as the inner product. Tangent spaces have Euclidean structure with easy-to-compute distances, which locally approximate Riemannian distances on $\mathcal{S}_D^+$. For a set of SPD points $\mathcal{Z} = \{Z_j \in \mathcal{S}_D^+\}_{j\leq n}$, the Fréchet mean is defined as the SPD point that minimizes the average squared Riemannian distances.

Typical TSM models use a feature extractor f_θ to map preprocessed EEG epochs into points $Z \in \mathcal{S}_D^+$, then use a tangent space mapping function m_ϕ to project $\mathcal{Z}$ to the tangent space at the Fréchet mean G_Z and use parallel transport so that the Fréchet mean becomes the identity matrix. The mapping function m_ϕ yields output in the Euclidean space, so any standard classifier g_ψ can be used. TSMNet extends typical TSM models by learning the parameters $\Theta = \{\theta, \phi, \psi\}$ in an end-to-end fashion (Figure 1b). Additionally, m_ϕ serves as a domain-specific tangent space mapping function, which keeps multiple parallel domain-specific batch normalization layers to compute domain-specific Fréchet means and variances during training. These learnable statistics are substituted with target domain statistics during testing, thus aligning marginal feature distributions using domain-specific Fréchet means and variances, transforming domain-specific inputs into domain-invariant outputs.

TSMNet benefits from the advantage of latent representation space alignment where classes are more linearly separable. Nonetheless, while beneficial for accuracy, the increased linear separability of classes also renders the alignment more susceptible to the effects of label shifts [7].

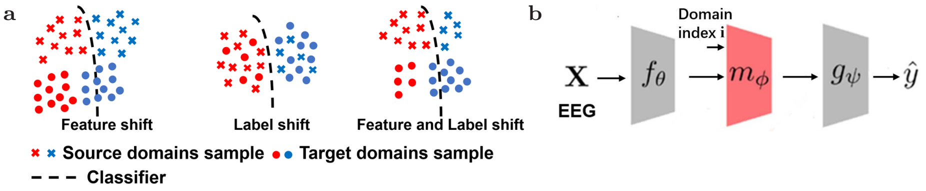

Figure 1: **a**, Imbalanced UDA. **b**, Overview of TSMNet model architecture.

3 Methods

Theoretical motivation. According to [5], denote $h \in \mathcal{H}$ as the hypothesis, the target error $\epsilon_t(h)$ cannot be effectively minimized by merely aligning marginal feature distributions and minimizing the source error $\epsilon_s(h)$. Denote d_{JS} as the Jensen-Shannon divergence between two distributions, [5] proposes:

$$\epsilon_s(h) + \epsilon_t(h) \geq \frac{1}{2}\left(d_{JS}(p(y), q(y)) - d_{JS}(p(x), q(x))\right)^2$$

Generally, if label shifts $d_{JS}(p(y), q(y))$ are significant, minimizing the divergence between the feature distributions $d_{JS}(p(x), q(x))$ and the source task error $\epsilon_s(h)$ will enlarge the target task error $\epsilon_t(h)$. Considering the SFUDA scenario which we only have access to pre-trained models, we address label shifts by dividing our methods into source domain training and target domain alignment.

Data augmentation for source domain training. A mini-batch balanced (MB) sampler is commonly used to compensate for label shifts in deep learning [6]. It over-samples minority classes and ensures that the training data are balanced within each mini-batch. Although over-sampling is effective, it can lead to over-fitting the minority classes. To mitigate this issue, we propose a data augmentation method called mix-up mini-batch balance (MUMB). MUMB reduces direct over-sampling by sampling linear mix-ups of domain-specific and label-specific samples to get balanced mini-batches. Additionally, we use the LDAM loss [8], which has been introduced to increase the margin of the minority classes as a label-dependent regularization technique.

Clustering for target domain alignment. The lack of label information in the target domain prevents the direct application of techniques used in the source domain. Since we assume that our distribution shifts are primarily driven by translations on $\mathcal{S}_D^+$, our trained classifier will likely be biased for the target domain after aligning the marginal feature distributions. Assuming that the target domain latent data are clustered according to the considered classes, some labels will likely be mapped to the correct side of the classifier. To exploit this, we propose to use the initial labels to estimate class-specific means and refine them with k-means clustering (align pseudo labels). The refined means are used to estimate a balanced Fréchet mean (Figure 2). This balanced Fréchet mean is then used inside m_ϕ to align the target domain distribution.

Clustering initial centroids. Initial centroids are crucial in clustering because they influence the quality of the resulting clusters. Following [9], we derived initial centroids from predicted labels (i.e., employ the original TSMNet to obtain predicted labels and calculate the class-specific Fréchet means to establish initial centroids). For the inter-session TL problem, we propose to use class-specific

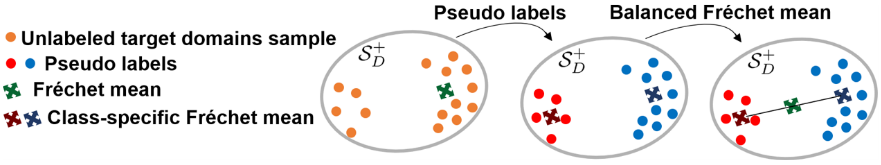

Figure 2: Clustering for target domain alignment.

source domain Fréchet means as the initial centroids because the expected distribution shifts across sessions are typically small. We note that this approach is not suitable for the inter-subject TL problem because differences in brain structures and variations in the performed task additionally drive the shifts.

4 Experiments

In this proof-of-concept experiment with simulations and mental imagery BCI data, we used balanced datasets and artificially introduced label shifts. Specifically, all source domains shared a label distribution inverse to all target domains. We quantified the imbalanced ratio as the ratio of the minority to majority class counts, with other classes slowly exponentially increasing from this base minority class ratio. To ensure a fair comparison across different ratios, we standardized the sample size and gradient steps within each dataset. For the training process, we evaluated the original TSMNet, a mini-batch balanced (MB) sampler with the LDAM loss (MB+LDAM), and our proposed method with the LDAM loss (MUMB+LDAM). For target alignment, we evaluated the original TSMNet marginal feature alignment (MFA) [3] and K-means with initial centroids based on predicted labels (Kmean-P) or source domain means (Kmean-S). We implemented a cross-validation scheme that either leaves one session (inter-session TL) or one subject (inter-subject TL) as a test set, used balanced accuracy as the performance metric, and maintained the original TSMNet hyperparameters and architecture. To reduce random effects, we averaged results over 20 repetitions.

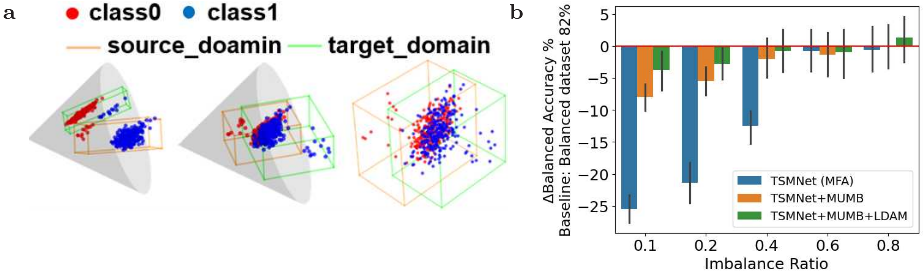

Figure 3: Simulations. **a**, generated SPD data (imbalance ratio = 0.1) (left), recentered around the identity matrix (mid), and tangent space mapping (right). These steps are performed by the original TSMNet with MFA. Results for TSMNet with MFA are colored in blue in b. **b**, simulation results. Barplots summarize the grand average balanced accuracy score relative to the baseline (TSMNet with MFA fitted to a balanced dataset, i.e., imbalance ratio = 1). Error bars indicate bootstrapped 95% confidence intervals (over groups).

Simulations. We simulated 2D binary classification problems across different imbalance ratios and created feature shifts between source and target domains by controlling the separation between class clusters. A total of 20 groups of source domains and their corresponding target domains were generated in the tangent space of the identity matrix using the scikit-learn `make_classification` function, with each domain consisting of 4000 normally distributed observations. We first used parallel transport and then used the pyRiemann `unupper` function to project the generated data from tangent space onto the SPD manifold so that each domain Fréchet mean becomes a distinct point on the SPD manifold. An example is visualized in Figure 3a, where one can see the negative impact of MFA in the presence of label shifts. Since the generated data (2×2 SPD matrices) reside on the SPD manifold, we skipped the feature extractor f_θ in TSMNet. The results are summarized in Figure 3b. TSMNet is robust to mild label shifts, maintaining a relatively stable model performance. However, we observed a significant performance decline as the imbalance ratio intensified, similar to the findings reported by [7]. This significant performance decline underscores the detrimental impact of label shifts on the efficacy of marginal feature alignment methods. The proposed MUMB method provided a remedy to the label shifts, and the LDAM loss further improved the performance to closely match those of a balanced dataset (baseline result).

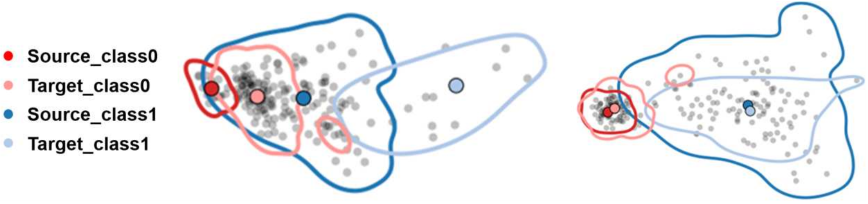

Figure 4: Subject1 from BNCI2015001 PCA visualization of the TSMNet with MFA (left) and TSMNet+MUMB+LDAM+Kmean-S after alignment (right).

Mental imagery. We evaluated four mental imagery datasets [10] (9 subjects/2 sessions/4 classes/22 channels), [11] (12/2-3/2/13), [12] (4/3/3/14) and [13] (9/5/2/3). We used MOABB [14] to preprocess the datasets. The preprocessing steps included resampling the EEG signals to 250 or 256 Hz, applying temporal filters to filter EEG within the 4 to 36 Hz frequency range, and extracting 3-second segments associated with specific class labels. We investigated the performance of each method under a certain imbalance ratio of 0.2. The results are summarized in Table 1, distinguishing between inter-session TL and inter-subject TL. PCA visualizations suggest that our proposed method approximately aligned class-conditional distributions (Figure 4). As expected, class-specific source domain Fréchet mean initialization (Kmean-S) is ineffective for inter-subject TL, yet it achieved the best performance in inter-session TL. In both scenarios, we observed a large variance in the accuracy among subjects and a significantly higher performance score alongside a systematic increase in both the t-value and score.

5 Conclusion and Discussion

We proposed a novel approach combining data augmentation and clustering techniques to align feature distributions under label shifts for TSMNet and other explicit marginal feature alignment methods on the SPD manifold. Our simu-

Training	**Alignment**	**Inter-session**			**Inter-subject**		
		Score ↑	**t**	**p**	**Score ↑**	**t**	**p**
Baseline	Kmean-P	1.1±11.7	2.5	0.016	-0.7±10.0	-0.7	0.475
Baseline	Kmean-S	2.3±12.5	4.7	<0.001	-1.7±11.8	-1.1	0.276
MB+LDAM	MFA	6.9±11.5	10.4	$\ll$0.001	7.6±12.3	6.1	$\ll$0.001
MB+LDAM	Kmean-P	8.0±13.0	10.5	$\ll$0.001	7.9±13.5	5.7	$\ll$0.001
MB+LDAM	Kmean-S	8.2±13.2	10.4	$\ll$0.001	2.0±14.8	0.9	0.342
MUMB+LDAM	MFA	8.9±11.6	13.1	$\ll$0.001	8.3±12.9	7.0	$\ll$0.001
MUMB+LDAM	Kmean-P	9.7±13.0	13.0	$\ll$0.001	**8.7±14.2**	6.2	$\ll$0.001
MUMB+LDAM	Kmean-S	**9.9±12.9**	13.2	$\ll$0.001	3.0±15.4	1.4	0.162

Table 1: Mental imagery results (4 datasets, artificially introduced imbalance ratio 0.2). The reported score summarizes balanced accuracy (mean±std, 20 repetitions, std over subjects) relative to baseline (i.e., the original TSMNet with MFA). Paired t-tests with degrees of freedom 33 were computed to identify significant differences between the baseline and other methods. P-values were corrected for multiple comparisons (8 tests, false discovery rate correction).

lation results underscore the importance of addressing label shifts in marginal feature alignment methods. We observed a systematic performance increase in the considered mental imagery EEG datasets. Altogether, the proposed methods extend the TSMNet framework to learning scenarios with label shifts, laying the groundwork for future applications with inherent label shifts (e.g., sleep staging).

References

[1] Lotte et al. A review of classification algorithms for eeg-based brain–computer interfaces: a 10 year update. *Journal of neural engineering*, 15(3):031005, 2018.

[2] Wu et al. Transfer learning for eeg-based brain–computer interfaces: A review of progress made since 2016. *IEEE Transactions on Cognitive and Developmental Systems*, 14(1):4–19, 2020.

[3] Kobler et al. Spd domain-specific batch normalization to crack interpretable unsupervised domain adaptation in eeg. *Advances in Neural Information Processing Systems*, 35:6219–6235, 2022.

[4] Congedo et al. Riemannian geometry for eeg-based brain-computer interfaces; a primer and a review. *Brain-Computer Interfaces*, 4(3):155–174, 2017.

[5] Zhao et al. On learning invariant representations for domain adaptation. In *International conference on machine learning*, pages 7523–7532. PMLR, 2019.

[6] Tan et al. Class-imbalanced domain adaptation: An empirical odyssey. In *Computer Vision–ECCV 2020 Workshops: Glasgow, UK, August 23–28, 2020, Proceedings, Part I 16*, pages 585–602. Springer, 2020.

[7] Bakas et al. Latent alignment with deep set eeg decoders. *arXiv preprint arXiv:2311.17968*, 2023.

[8] Cao et al. Learning imbalanced datasets with label-distribution-aware margin loss. *Advances in neural information processing systems*, 32, 2019.

[9] Wang et al. Unsupervised domain adaptation via structured prediction based selective pseudo-labeling. In *Proceedings of the AAAI conference on artificial intelligence*, volume 34, pages 6243–6250, 2020.

[10] Tangermann et al. Review of the bci competition iv. *Frontiers in neuroscience*, 6:55, 2012.

[11] Faller et al. Autocalibration and recurrent adaptation: Towards a plug and play online erd-bci. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 20(3):313–319, 2012.

[12] Zhou et al. A fully automated trial selection method for optimization of motor imagery based brain-computer interface. *PloS one*, 11(9):e0162657, 2016.

[13] Leeb et al. Brain–computer communication: motivation, aim, and impact of exploring a virtual apartment. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 15(4):473–482, 2007.

[14] Jayaram et al. Moabb: trustworthy algorithm benchmarking for bcis. *Journal of neural engineering*, 15(6):066011, 2018.