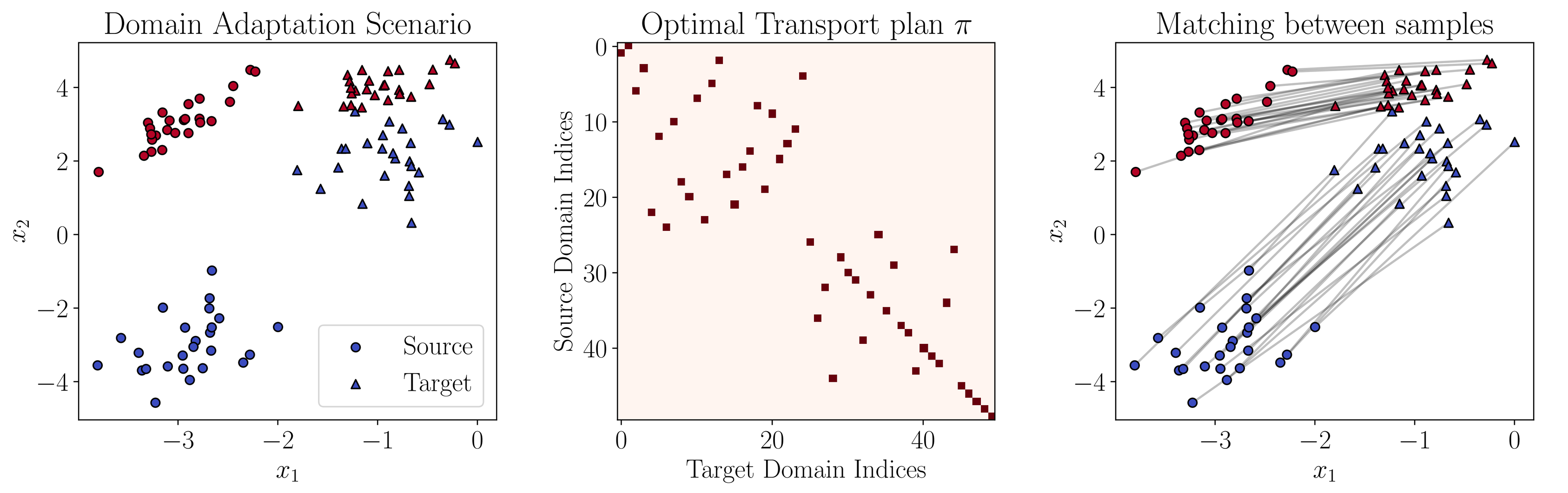
Multi-Source Domain Adaptation through Wasserstein Barycenters

Eduardo Montesuma [1] Fred Ngolè Mboula [1] Antoine Souloumiac [1]

[1]Université Paris-Saclay, CEA, List, F-91120 Palaiseau France

Optimal Transport

Discrete Optimal Transport deals with empirical distributions, $\hat{P}(\mathbf{x}) = \frac{1}{n}\sum_{i=1}^{n}\delta(\mathbf{x}-\mathbf{x}_i^{(P)})$, $\mathbf{x}_i^{(P)} \overset{i.i.d.}{\sim} P$. One searches for an OT plan, $\pi \in \Pi(P,Q) = \{\pi \in \mathbb{R}_+^{n\times m} : \sum_{i=1}^{n}\pi_{ij} = m^{-1} \text{ and } \sum_{j=1}^{m}\pi_{ij} = n^{-1}\}$. For instance,



Given P and Q, and for a ground-cost $C_{ij} = c(\mathbf{x}_i^{(P)}, \mathbf{x}_j^{(Q)})$, OT is a linear program on $\{\pi_{ij}\}$:

$$\pi^{\star} = \underset{\pi\in\Pi(P,Q)}{\text{argmin}} \sum_{i=1}^{n}\sum_{j=1}^{m}\pi_{ij}C_{ij}.$$

Based on the transport plan $\pi^{\star}$, [1] proposed to use the Barycentric mapping for domain adaptation:

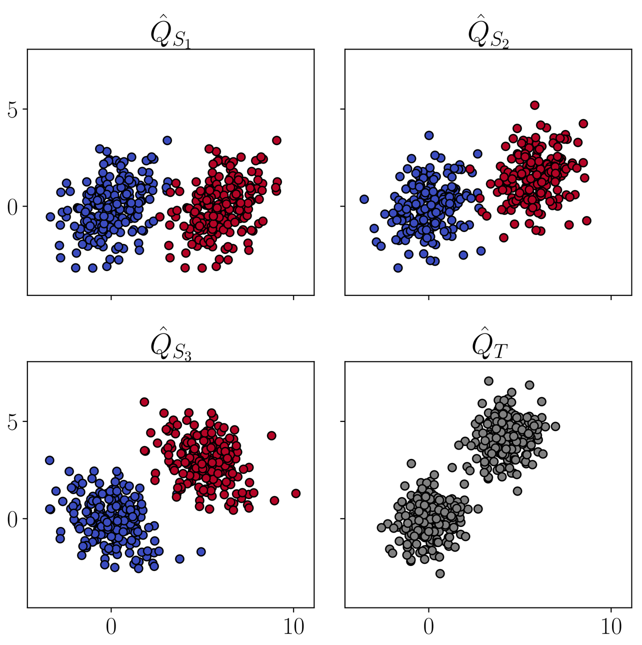
$$T_{\pi^{\star}}(\mathbf{x_i}^{(P)}) = m\sum_{j=1}^{m}\pi_{ij}\mathbf{x}_j^{(Q)}$$

Wasserstein Barycenters and Multi-Source Domain Adaptation

In multi-source domain adaptation, one has access to a set of source distributions $\mathcal{Q}_S = \{\hat{Q}_{S_\ell}\}_{\ell=1}^{N_S}$, and a target distribution $\hat{Q}_T$. **The goal** is to adapt knowledge from $\mathcal{Q}_S$ to $\hat{Q}_T$. **The challenge** comes from **distributional shift:** $\hat{Q}_{S_i} \neq Q_{S_j}$, $i \neq j$ and $Q_{S_i} \neq Q_T$, $\forall i$. Furthermore, **the target domain is unlabeled**. Henceforth we denote,

$$h_T = \underset{h\in\mathcal{H}}{\text{argmin}}\, \mathcal{R}_{Q_T}(h),$$

$$h_\alpha = \underset{h\in\mathcal{H}}{\text{argmin}} \sum_{\ell=1}^{N_S}\alpha_\ell\mathcal{R}_{Q_{S_\ell}}$$

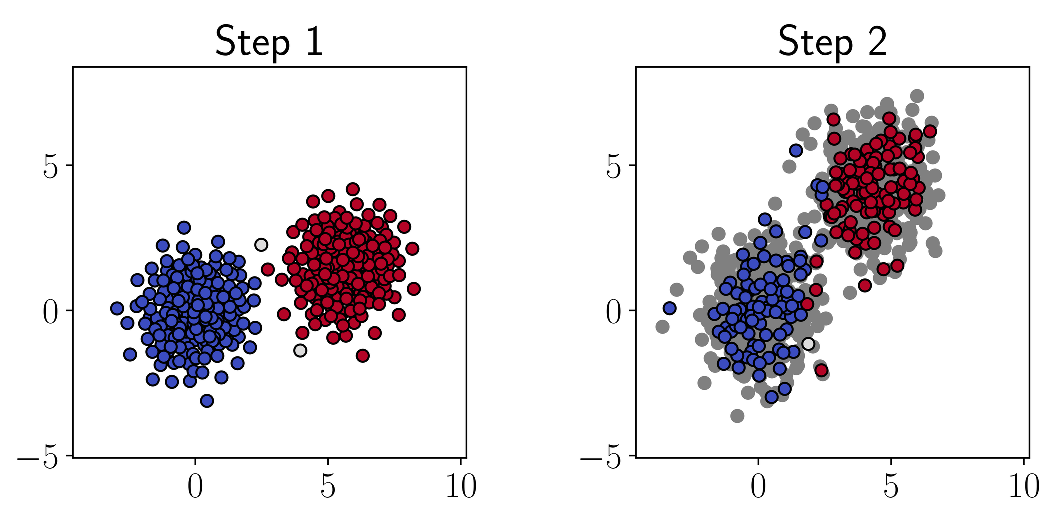


Wasserstein Barycenter Transport [3]. Based on domain adaptation theory [2], for classifiers h_T and h_α,

$$\Delta_T(h_T, h_\alpha) \leq 2\sum_{\ell=1}^{N_S}\alpha_\ell W_1(Q_{S_\ell}, B) + W_1(Q_T, B) + \sum_{\ell=1}^{N}\alpha_\ell\lambda_\ell,$$

In [3], we make this bound tight by minimizing the highlighted terms on the r.h.s.,

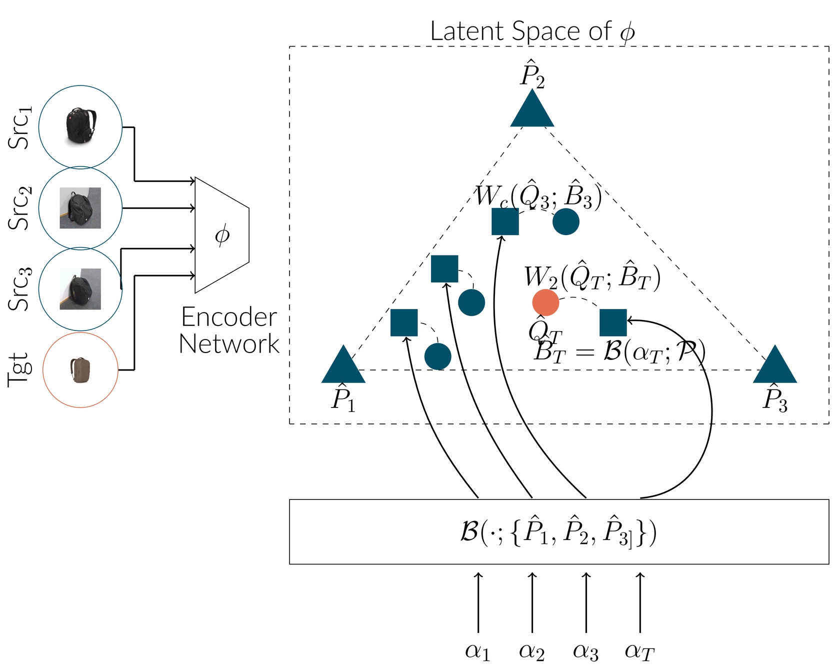
$$\hat{B} = \underset{B}{\text{argmin}}\, \underbrace{\alpha_1 W_1\Big(\quad, B\Big) + \alpha_2 W_1\Big(\quad, B\Big) + \alpha_3 W_1\Big(\quad, B\Big)}_{\text{Step 1: Barycenter Calculation}} + \underbrace{W_1\Big(\quad, B\Big)}_{\text{Step 2}}$$



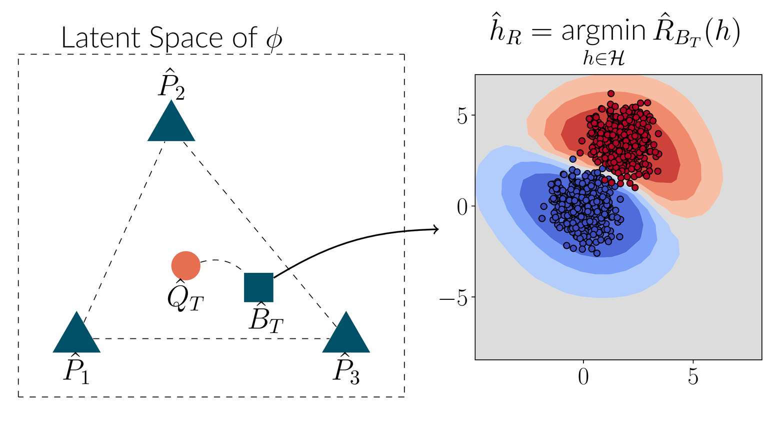
Dataset Dictionary Learning [4]. Let $\mathcal{Q} = \{\hat{Q}_{S_\ell}\}_{\ell=1}^{N_S} \cup \{\hat{Q}_T\}$ correspond to N_S labeled sources and an unlabeled target. Let $\mathcal{A} = [\alpha_1, \cdots, \alpha_{N_S}, \alpha_{N_S+1}]$, and $\mathcal{P} = \{\hat{P}_k\}_{k=1}^{K}$. The $\hat{P}_k$'s are an empirical approximation of the point clouds that interpolate distributional shift and $\alpha_T := \alpha_{N_S+1}$. For $N = N_S+1$, DaDiL minimizes ,

$$(\mathcal{P}^{\star}, \mathcal{A}^{\star}) = \underset{\mathcal{P},\mathcal{A}\in(\Delta_K)^N}{\text{argmin}} \frac{1}{N}\sum_{\ell=1}^{N}\mathcal{L}(\hat{Q}_\ell, \mathcal{B}(\alpha_\ell;\mathcal{P})),$$

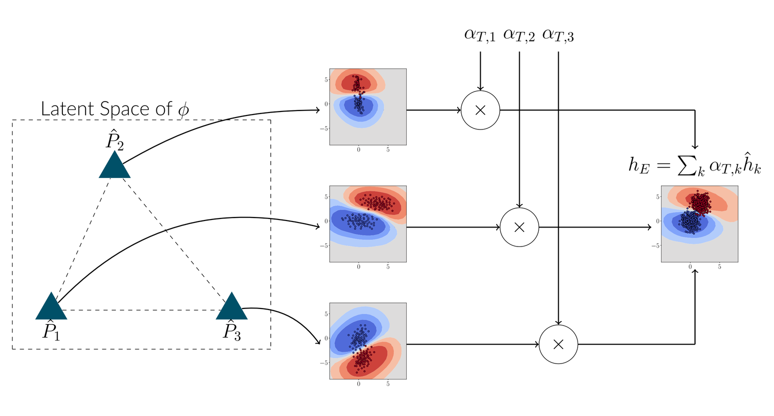
where, $\mathcal{L}(\hat{Q}_\ell, \hat{B}_\ell) = W_c(\hat{Q}_\ell, \hat{B}_\ell)$ for the sources, and $\mathcal{L}(\hat{Q}_T, \hat{B}_T) = W_2(\hat{Q}_T, \hat{B}_T)$, for the target. Based on DaDiL we have two strategies for MSDA,



- **DaDiL-Reconstruction.** Reconstruction of distributions through Wasserstein barycenters.
- **DaDiL-Ensembling:** Ensemble of classifiers fit on atom data.



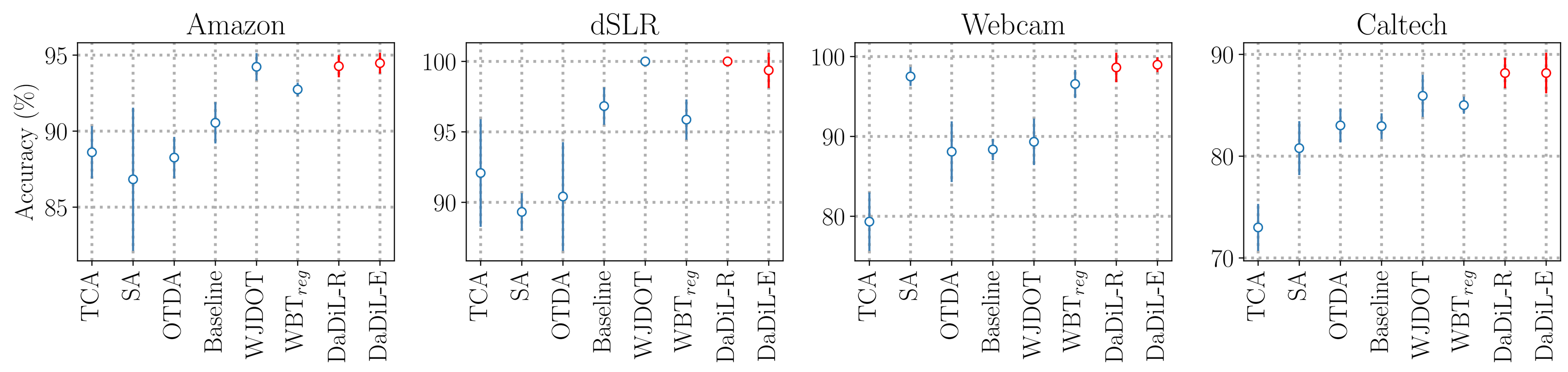
DaDiL-R



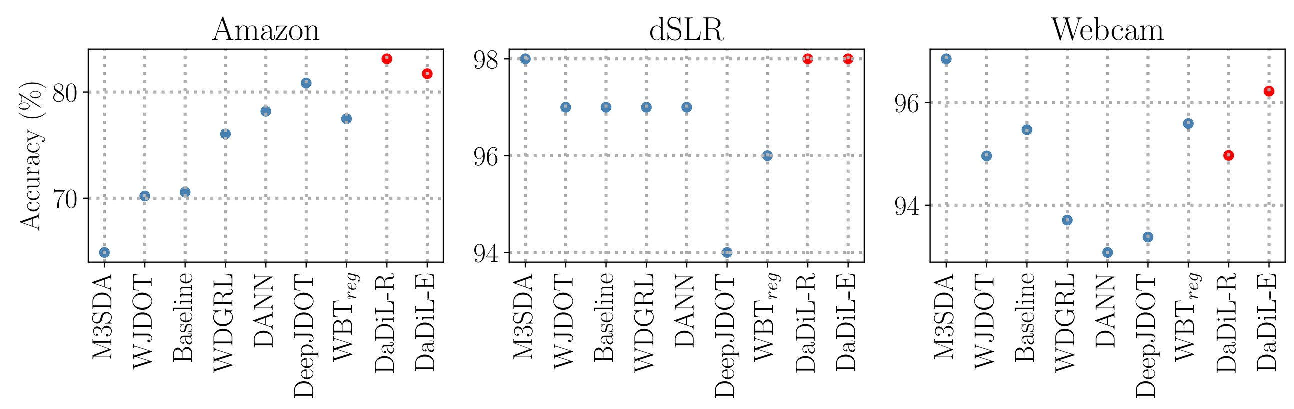
DaDiL-E

Empirical Results

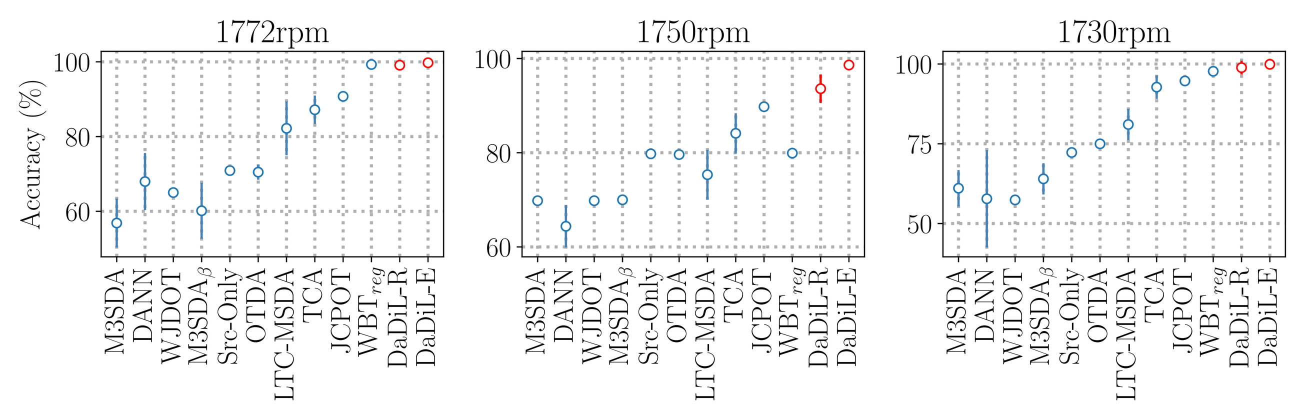
Caltech-Office 10



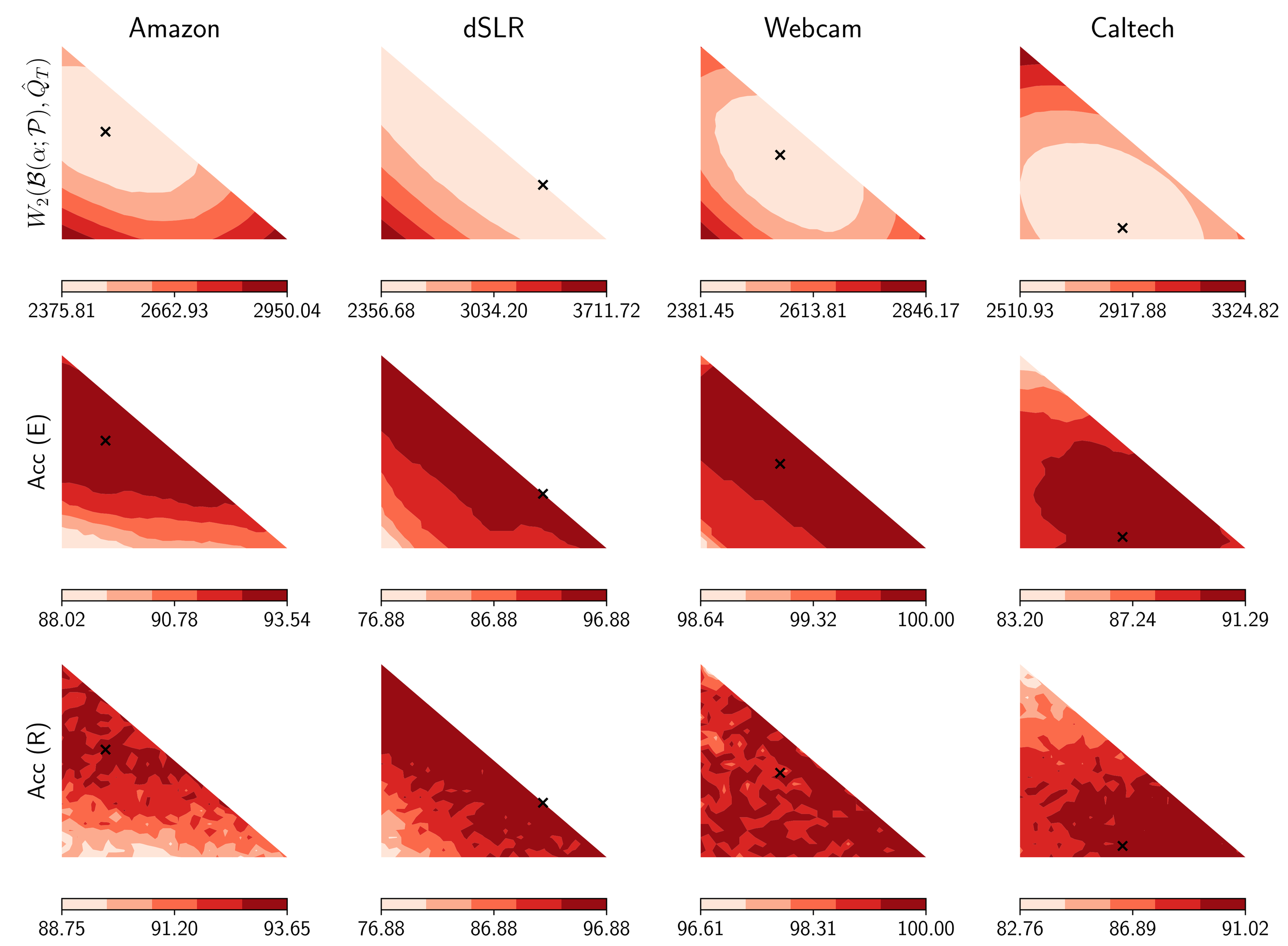
Refurbished Office 31



CWRU



Atom Interpolations



Perspectives

- **Novelty.** We introduce the usage of (free-support) Wasserstein barycenters in MSDA [3, 4].
- **Challenges.** Estimating OT in high-dimensional spaces.
- **Future Works.** Applying MSDA on fault diagnosis [5], federated dictionary learning [6], dataset distillation [7].

References

[1] Nicolas Courty, Rémi Flamary, Devis Tuia, and Alain Rakotomamonjy. Optimal transport for domain adaptation. IEEE transactions on pattern analysis and machine intelligence, 39(9):1853–1865, 2016.

[2] Ievgen Redko, Emilie Morvant, Amaury Habrard, Marc Sebban, and Younes Bennani. Advances in domain adaptation theory. Elsevier, 2019.

[3] Eduardo Fernandes Montesuma and Fred Ngolè Mboula. Wasserstein barycenter for multi-source domain adaptation. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) June 2021.

[4] Eduardo Fernandes Montesuma, Fred Ngolè Mboula and Antoine Souloumiac. Multi-Source Domain Adaptation through Dataset Dictionary Learning in Wasserstein Space. 26th European Conference on Artificial Intelligence. October 2023

[5] Eduardo Fernandes Montesuma, Michela Mulas, Fred Ngolè Mboula, Francesco Corona, and Antoine Souloumiac. Multi-Source Domain Adaptation for Cross-Domain Fault Diagnosis of Chemical Processes. arXiv:2308.11247. August 2023

[6] Fabiola Espinoza Castellon, Eduardo Fernandes Montesuma, Fred Ngolè Mboula, Aurélien Mayoue, Antoine Souloumiac, Cdric Gouy-Pailler. Federated Dataset Dictionary Learning for Multi-Source Domain Adaptation. arXiv:2309.07670. September 2023

[7] Eduardo Fernandes Montesuma, Fred Ngolè Mboula, Antoine Souloumiac. Multi-Source Domain Adaptation meets Dataset Distillation through Dataset Dictionary Learning. arXiv:2309.07666. September 2023

Learn more!

DaDiL Paper | WBT Paper | Portfolio

Fault Diagnosis | Federated DaDiL | Distillation