Multi-Source Domain Adaptation through Dataset Dictionary Learning in Wasserstein Space

Eduardo Fernandes Montesuma Fred Ngolè Mboula Antoine Souloumiac

Université Paris-Saclay, CEA, List, F-91120 Palaiseau, France

Project page: `https://eddardd.github.io/demo-dadil/`
Paper: `https://arxiv.org/pdf/2307.14953.pdf`

26th European Conference on Artificial Intelligence - October 2023

Summary

Introduction

Problem Setting

- Domain Adaptation (DA): Source distribution $\neq$ Target distribution
- We have access to labeled source domain samples, and unlabeled target domain samples
- Multi-Source DA: Training distribution is composed of multiple, heterogeneous distributions.

Challenge: how to exploit similarities between the sources for MSDA?

[Crammer, Kearns and Wortman, 2008]

Background

Optimal Transport

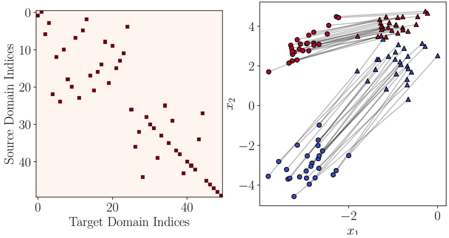
OT plans: $\Pi(P,Q) = \{\pi \in \mathbb{R}^{n\times m} : \sum_i \pi_{ij} = m^{-1}, \text{ and, } \sum_j \pi_{ij} = n^{-1}\}$

- Effort of transportation,

$$\mathcal{L}_{OT}(\pi) = \sum_{i,j} \pi_{ij} \underbrace{\|\mathbf{x}_i^{(P)} - \mathbf{x}_j^{(Q)}\|_2^2}_{\text{Ground-Cost } C_{ij}}$$

- OT Problem

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

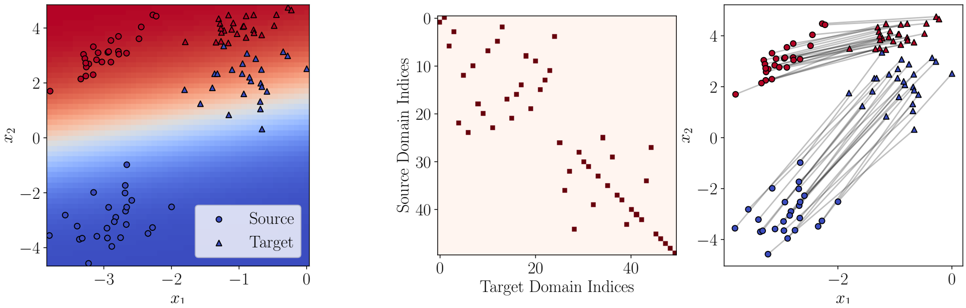


(L) OT plan (R) Matched points

Wasserstein distance: $W_c(\hat{P}, \hat{Q}) = \sum_{i,j} \pi^\star_{ij} C_{ij}$

[Montesuma, Mboula and Souloumiac, 2023]

Optimal Transport & Domain Adaptation



[Courty et al., 2017]

Dictionary Learning

From a set of vectors $\mathbf{X} \in \mathbb{R}^{N \times d}$, DiL learns a set of atoms $\mathcal{P} \in \mathbb{R}^{K \times d}$ and representations $\mathcal{A} \in \mathbb{R}^{N \times K}$ s.t.,

$$\underset{\mathcal{P},\mathcal{A}}{\operatorname{argmin}} \frac{1}{N} \sum_{i=1}^{N} \underbrace{\mathcal{L}(\mathbf{x}_\ell, \mathcal{P}^T \alpha_\ell)}_{\text{Reconstruction Loss}} + \underbrace{\lambda_A \Omega_A(\mathcal{A}) + \lambda_P \Omega_P(\mathcal{P})}_{\text{Regularization}},$$

Example: Principal Component Analysis,

 $= \underbrace{\alpha_{\ell 1}}_{\text{Representation}}$  $\underbrace{}_{\text{Atom}}$ $+\alpha_{\ell 2}$  $+ \alpha_{\ell 3}$ ,

$\mathcal{L}(\mathbf{x}_\ell, \mathcal{P}^T \alpha_\ell) = \Big\|$  $-$  $\Big\|_2^2$

Dataset Dictionary Learning

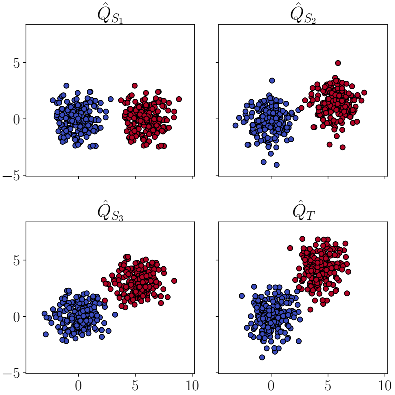
Dataset Dictionary Learning - Setting

- Set of datasets $\hat{Q}_\ell \in \mathcal{Q}$,

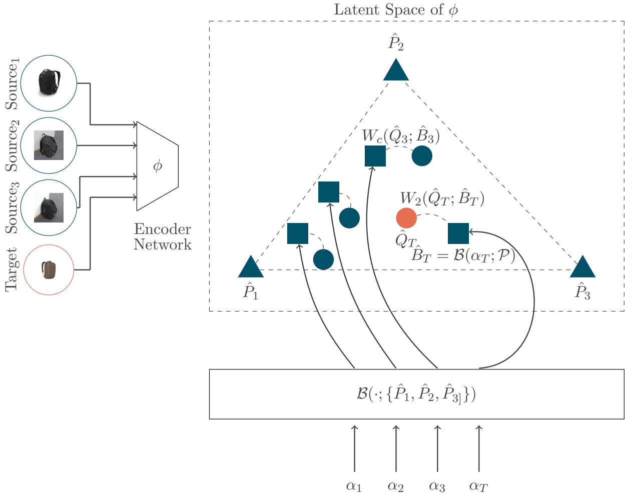
$$\hat{Q}_\ell(\mathbf{x},\mathbf{y}) = \frac{1}{n_\ell}\sum_{i=1}^{n_\ell}\delta\left((\mathbf{x},\mathbf{y}) - \underbrace{(\mathbf{x}_i^{(Q_\ell)},\mathbf{y}_i^{(Q_\ell)})}_{\text{Real Samples}}\right)$$

- Each atom is a virtual dataset

$$\hat{P}_k(\mathbf{x},\mathbf{y}) = \frac{1}{n}\sum_{j=1}^{n}\delta\left((\mathbf{x},\mathbf{y}) - \underbrace{(\mathbf{x}_j^{(P_k)},\mathbf{y}_j^{(P_k)})}_{\text{Synthetic Samples}}\right)$$



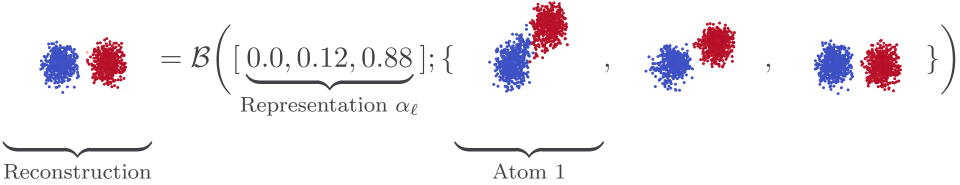
Dataset Dictionary Learning - Framework



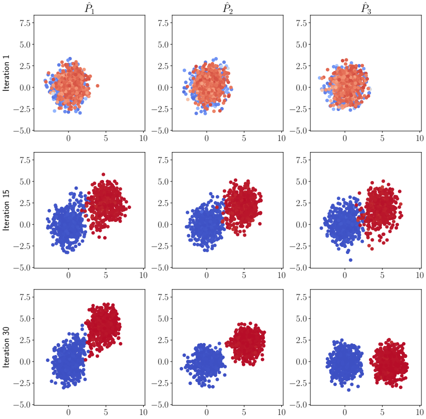
Dataset Dictionary Learning - Framework

For a set of datasets $\mathcal{Q} = \{\hat{Q}_\ell\}_{\ell=1}^N$, atoms $\mathcal{P} = \{\hat{P}_k\}$ and $\mathcal{A} = \{\alpha_\ell\}_{\ell=1}^N$, $\sum_k \alpha_{\ell,k} = 1$, we propose the Dataset Dictionary Learning (DaDiL) framework through,

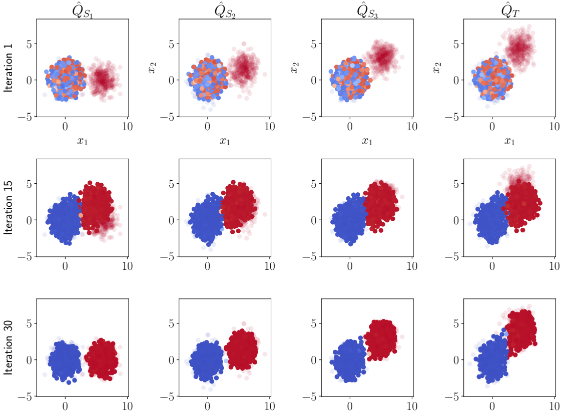
$$(\mathcal{P}^\star, \mathcal{A}^\star) = \underset{\mathcal{P},\mathcal{A}}{\text{argmin}} \frac{1}{N}\sum_{\ell=1}^{N} \mathcal{L}(\hat{Q}_\ell, \mathcal{B}(\alpha_\ell; \mathcal{P})),$$



Dataset Dictionary Learning

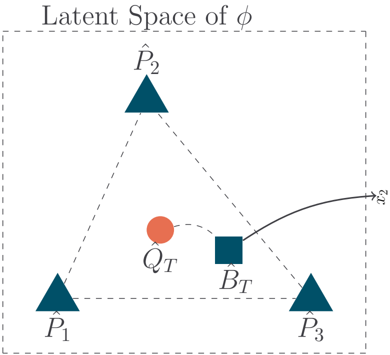


Evolution of Atoms

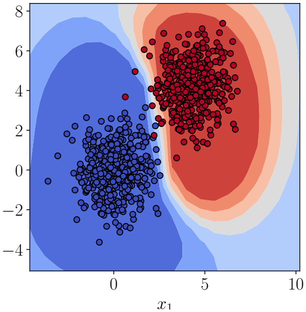


Evolution of Reconstructions

Dataset Dictionary Learning - DaDiL-R

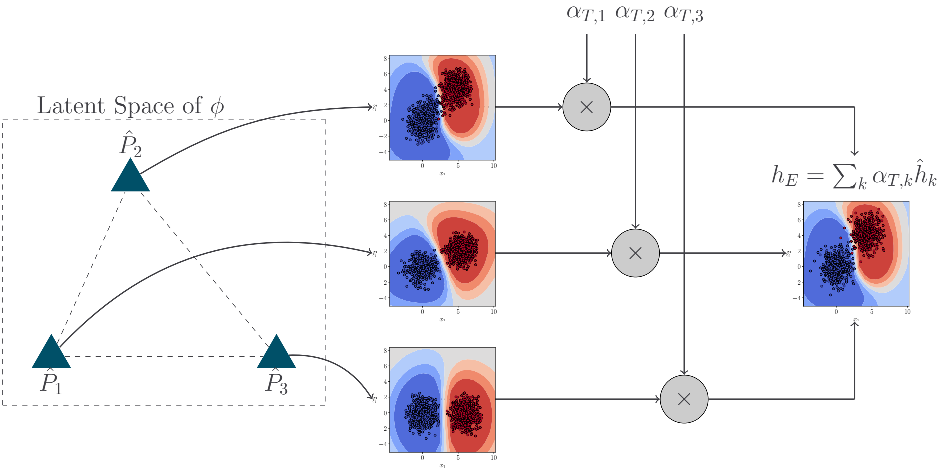


$$\hat{h}_R = \underset{h \in \mathcal{H}}{\text{argmin}}\ \hat{R}_{B_T}(h)$$



$$\mathcal{R}_{Q_T}(h) \leq \mathcal{R}_{B_T}(h) + \underbrace{W_2(\hat{Q}_T, \hat{B}_T)}_{\text{Reconstruction Error}} + \underbrace{\sqrt{2(\log 1/\delta)/\xi'}\left(\sqrt{1/n} + \sqrt{1/n_T}\right)}_{\text{Sample Complexity } \mathcal{O}(n^{-1/2})} + \underbrace{\min_{h \in \mathcal{H}} \mathcal{R}_{Q_T}(h) + \mathcal{R}_{B_T}(h)}_{\text{Adaptation Complexity}},$$

Dataset Dictionary Learning - DaDiL-E



Dataset Dictionary Learning - DaDiL-E

Theoretical bound,

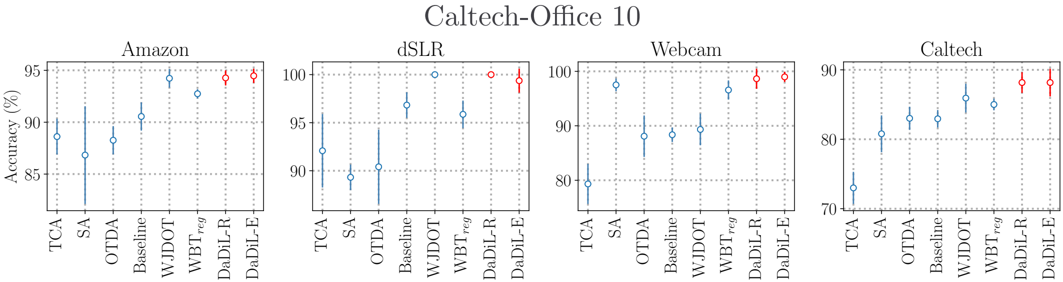
$$\mathcal{R}_{Q_T}(\hat{h}_\alpha) \leq \mathcal{R}_\alpha(\hat{h}_\alpha) + \underbrace{\mathrm{W}_2(\hat{Q}_T, \mathcal{B}(\alpha;\mathcal{P}))}_{\text{Reconstruction Error}} + \underbrace{\sum_{k=1}^{K}\alpha_k \mathrm{W}_2(\hat{P}_k, \mathcal{B}(\alpha;\mathcal{P}))}_{\text{Dictionary Geometry}}$$

$$+ \underbrace{\sum_{k=1}^{K}\alpha_k \sqrt{2\log {}^{1/\delta}/\xi'}\left(\sqrt{{}^{1}/n} + \sqrt{{}^{1}/n_T}\right)}_{\text{Sample Complexity}}$$

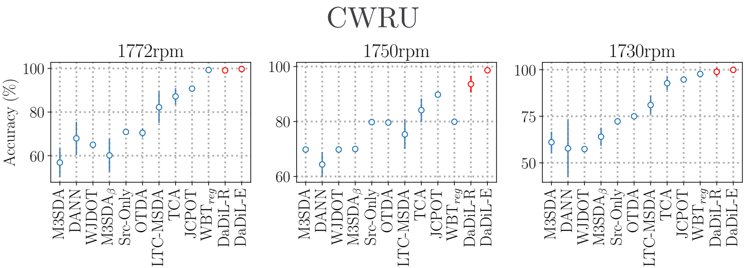
$$+ \underbrace{\sum_{k=1}^{K}\alpha_k \left(\min_{h\in\mathcal{H}}\mathcal{R}_{P_k}(h) + \mathcal{R}_{Q_T}(h)\right)}_{\text{Adaptation Complexity}},$$

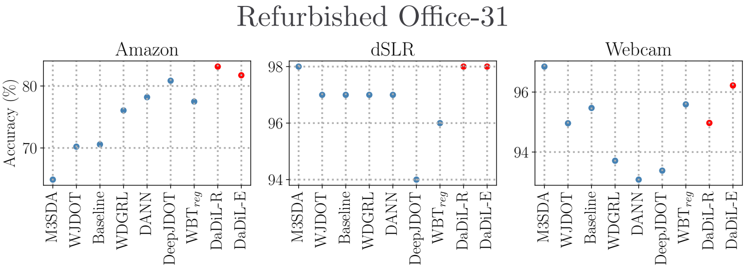
Experiments

Multi-Source Domain Adaptation

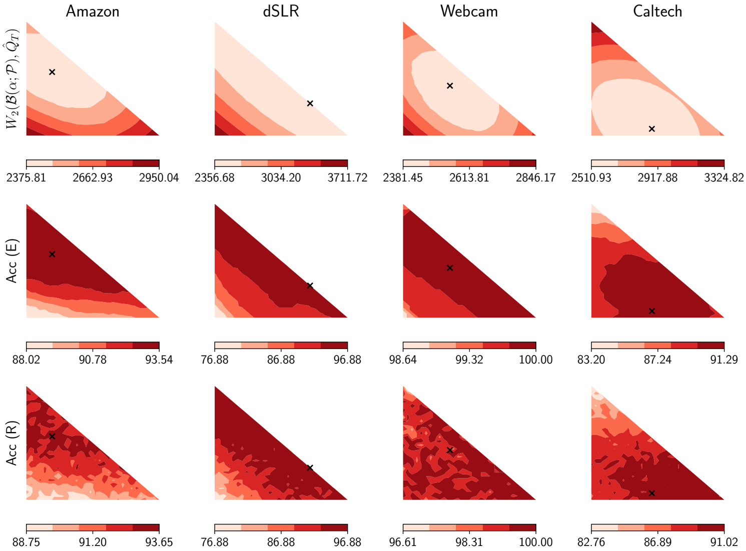


Multi-Source Domain Adaptation





Atom Interpolations



Conclusion

Final Remarks

- We propose a novel DiL framework over point clouds.
- Based on our dictionary we provide 2 theoretically driven MSDA methods.
- Our methods achieve SOTA, and atom interpolations generalize to the target.

Paper

Code Demo

Future Works

References

[Crammer, Kearns and Wortman, 2008] Learning from multiple sources. Journal of Machine Learning Research, 9(8)

[Courty et al., 2017] Optimal transport for domain adaptation. IEEE Trans. Pattern Anal. Mach. Intell, 1, 1-40.

[Montesuma, Mboula and Souloumiac, 2023] Recent advances in optimal transport for machine learning. arXiv preprint arXiv:2306.16156.