TBSI 清华–伯克利深圳学院
Tsinghua-Berkeley Shenzhen Institute

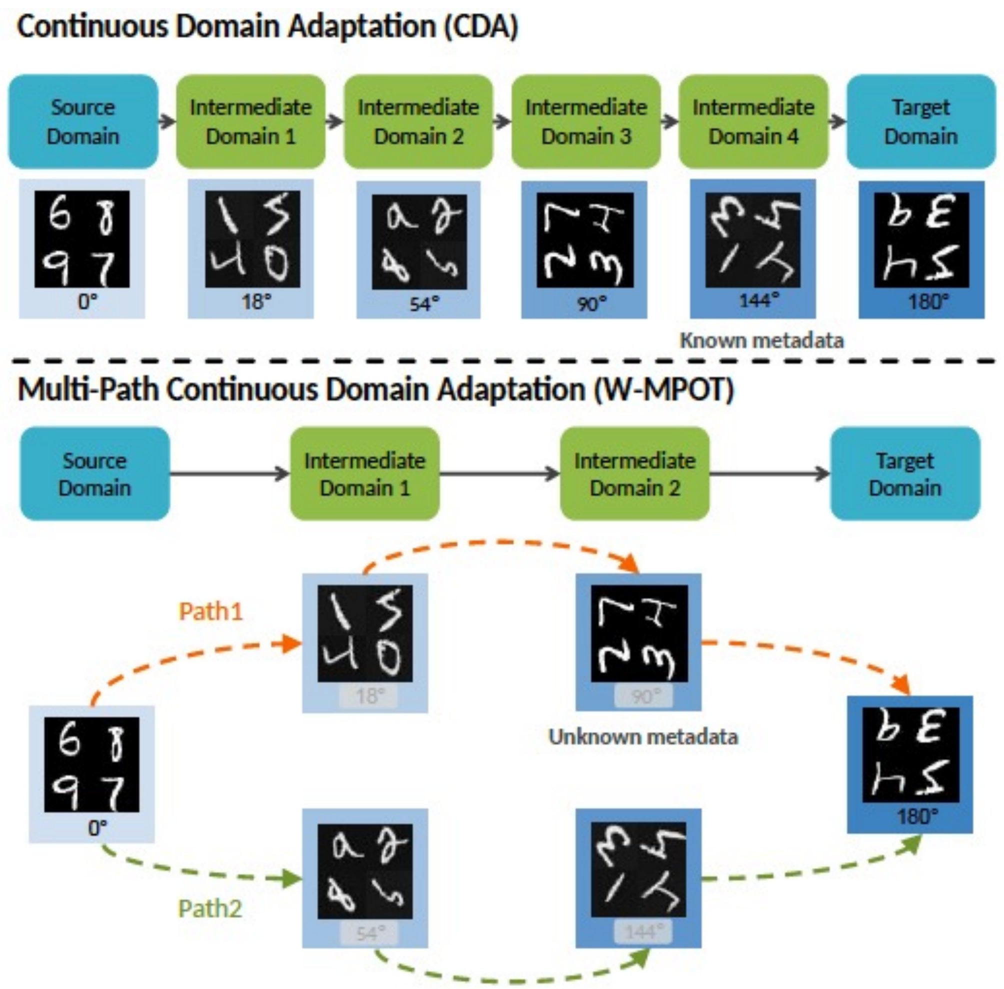
Multi-Path Continuous Domain Adaptation with Wasserstein-based Transfer Curriculum

Hanbing Liu
Advisor: Prof. Yang Li
2023/10/19

Contents

Introduction

- **Substantial domain shift**
  - Machine learning based models
  - Continuous domain adaptation
- **Transfer order and Cumulative errors**
  - Missing metadata
  - Progressively adaptation errors



Introduction

- **Related works**
  - Unsupervised Domain Adaptation: learning domain-invariant representations by aligning the source and target distributions
  - Continuous Domain Adaptation: self-training, adversarial algorithms, Optimal transport
  - Intermediate Domain Selection: domain discriminator

Contents

Problem definition

Known data

$\boldsymbol{\mathcal{X}} \subset \mathbb{R}^{\boldsymbol{d}}$: feature space, $\boldsymbol{\mathcal{Y}} \subset \mathbb{R}$: label space

- Source domain $D_S = \{(\boldsymbol{x_j}, y_j)\}_{j=1}^{N_S}$
- Target domain $D_T = \{\boldsymbol{x_j}\}_{j=1}^{N_T}$
- Intermediate Domain set $\mathcal{D}_I = \{D_{I_1}, D_{I_2}, D_{I_3}, \dots, D_{I_K}\}$ $k = 0, 1, \dots, K$; ordered sequence $\widehat{\mathcal{D}}_I$
- Intermediate domain $D_{I_k} = \{\boldsymbol{x_j}\}_{j=1}^{N_{I_k}}$
- $\mu_S, \mu_{I_k}, \mu_T \in \mathcal{P}(\mathcal{X})$: probability measures on $\mathbb{R}^d$

Objective

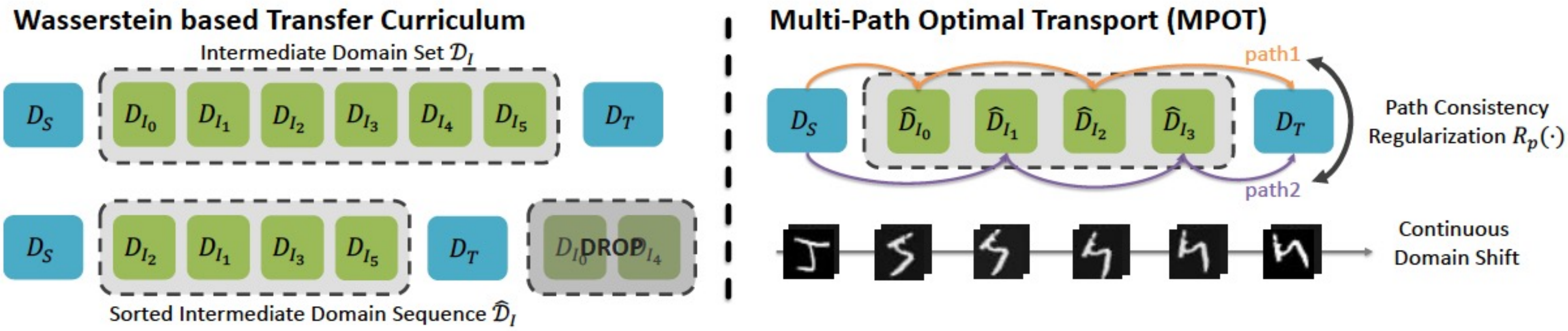
- $\{\hat{y}_j\}_{j=1}^{N_T}$ in target domain

Contents

- Introduction
- Problem definition
- Methodology
- Experiments
- Conclusion

Methodology

- **Wasserstein based transfer curriculum**
- **Multi-path optimal transport**



Methodology

➢ Wasserstein based transfer curriculum

- Wasserstein distance plays an important role in deriving the generalization bound in domain adaptation

- Two candidate intermediate domains D_{I_0} and D_{I_1}
- Assume that the optimal transfer order $D_S \to D_{I_0} \to D_{I_1} \to D_T$,
- Another possible transfer order $D_S \to D_{I_1} \to D_{I_0} \to D_T$

$$\epsilon_{\mu_T}(h,f) \leq \epsilon_{\mu_S}(h,f) + 2A \cdot \mathcal{W}_1(\mu_S, \mu_{I_0}) + 2A \cdot \mathcal{W}_1(\mu_{I_0}, \mu_{I_1}) + 2A \cdot \mathcal{W}_1(\mu_{I_1}, \mu_T) + \epsilon, \tag{2}$$

$$\epsilon_{\mu_T}(h,f) \leq \epsilon_{\mu_S}(h,f) + 2A \cdot \mathcal{W}_1(\mu_S, \mu_{I_1}) + 2A \cdot \mathcal{W}_1(\mu_{I_1}, \mu_{I_0}) + 2A \cdot \mathcal{W}_1(\mu_{I_0}, \mu_T) + \epsilon, \tag{3}$$

- **better domain transfer order** $D_S \to D_{I_0} \to D_{I_1} \to D_T$ **will lead to tighter generalization bound**

Methodology

➢ Wasserstein based transfer curriculum

- measure the closeness between each intermediate domain D_{I_k} and the source domain D_S

$$W_k = \min_{\gamma} \langle \gamma_k, \mathbf{M}_k \rangle_F + \lambda \cdot \Omega(\gamma_k)$$
$$\text{s.t. } \gamma_k \mathbf{1} = \mu_S \quad \gamma_k^T \mathbf{1} = \mu_{I_k}, k = 1, \ldots, K,$$

- intermediate domain that is further from the source domain than the target domain is discarded
- The remaining N domains in the intermediate domain set are then sorted in order of W_k
- Obtain a domain series $\widehat{\mathcal{D}}_I = \widehat{D}_{I_0} \rightarrow \widehat{D}_{I_1} \rightarrow \widehat{D}_{I_2} \rightarrow \ldots \rightarrow \widehat{D}_{I_N}, \quad W_0 \leq W_1 \leq W_2 \ldots \leq W_N.$
- By utilizing the w-distance to sort multiple intermediate domains, we eliminate the need for meta-information

Methodology

TBSI 清华-伯克利深圳学院 Tsinghua-Berkeley Shenzhen Institute

➢ Multi-Path Optimal Transport

- Given the sorted sequence of intermediate domain series $\widehat{\mathcal{D}}_I$, the source domain is initially mapped to the first intermediate domain $\widehat{D}_{I_0}$ using direct optimal transport
- For the following intermediate domains, the probabilistic coupling γ_n between the domain $\widehat{D}_{I_{n-1}}$ and the subsequent domain $\widehat{D}_{I_n}$ is calculated using continuous optimal transport

$$\gamma_n = \underset{\gamma \in \mathbb{R}^{N_S \times N_T}}{\operatorname{argmin}} \left\langle \gamma, \mathbf{M}^{[n-1,n]} \right\rangle + \lambda \Omega(\gamma) + \eta_t R_t(\gamma) \quad \text{s.t. } \gamma \mathbf{1} = \mu_{SI_{n-1}} \quad \gamma^T \mathbf{1} = \mu_{I_n} \quad \gamma \geq 0, \tag{5}$$

$$\begin{aligned} R_t(\gamma) &= \mathcal{B}_{I_n}(D_S^{I_{n-1}}) - \mathcal{B}_{I_{n-1}}(D_S^{I_{n-2}}) \\ &= \left\| N_S \cdot \gamma \cdot \widehat{D}_{I_n} - N_S \cdot \gamma_{n-1} \cdot \widehat{D}_{I_{n-1}} \right\|_F^2, \end{aligned} \tag{6}$$

清华-伯克利深圳学院 TBSI Tsinghua-Berkeley Shenzhen Institute

Multi-Path Optimal Transport

- To address the challenge of accumulated errors in long transfer sequences, we further introduce a path consistency regularizer $R_p(\cdot)$ by comparing it with another transfer path

$$R_p(\gamma, \gamma_{p_2}) = \left\| N_S \cdot \gamma \cdot \widehat{D}_{I_n} - N_S \cdot \gamma_{p_2} \cdot \widehat{D}_{I_n} \right\|_F^2, \quad (7)$$

- Where γ_{p_2} is the transport plan of the second possible path which is utilized to refine the γ of path 1
- From $\widehat{D}_{I_N}$ to the target domain D_T will be conducted using MPOT

$$\begin{aligned} \gamma_{N+1} = \underset{\gamma \in \mathbb{R}^{N_s \times N_t}}{\operatorname{argmin}} & \left\langle \gamma, \mathbf{M}^{[N,N+1]} \right\rangle + \lambda \Omega(\gamma) \\ & + \eta_t R_t(\gamma) + \eta_p R_p(\gamma, \gamma_{p_2}) \\ \text{s.t. } \gamma \mathbf{1} = & \mu_{S I_N} \quad \gamma^T \mathbf{1} = \mu_T \quad \gamma \geq 0, \end{aligned} \quad (8)$$

Methodology

➢ Multi-Path Optimal Transport

Algorithm 1: Bidirectional Optimization algorithm in MPOT

Input: Transport matrix of Path 1 γ_{p_1}, Transport matrix of Path 2 γ_{p_2}, step size α, cost matrix of Path 1 $\mathrm{M}^{[1]}$, cost matrix of Path 2 $\mathrm{M}^{[2]}$, weight of Path 1 λ_1 and Path 2 λ_2, iteration times c

Output: Refined transport matrix from N-th intermediate domain to target domain γ'_{N+1}

Initialize: $\gamma_0 \in (0, +\infty)^{N_S \times N_T}$

for $c \leftarrow 0, 1, \ldots$ **do**

- $\mathrm{M}_c^{[1]} = \alpha \mathrm{M}^{[N,N+1]} + \alpha \nabla \mathcal{J}(\gamma_c, \gamma_{p_2})$
- $\mathrm{M}_c^{[2]} = \alpha \mathrm{M}^{[N,N+1]} + \alpha \nabla \mathcal{J}(\gamma_c, \gamma_{p_1})$
- $\mathrm{M}_c = \lambda_1 \mathrm{M}_c^{[1]} + \lambda_2 \mathrm{M}_c^{[2]}$
- $\gamma_{c+1} = \mathrm{Sinkhorn}\left(\mathrm{M}_c, 1 + \alpha\lambda, \mu_{SI_N}, \mu_T\right)$

$\gamma'_{N+1} = \gamma_\infty$

Contents

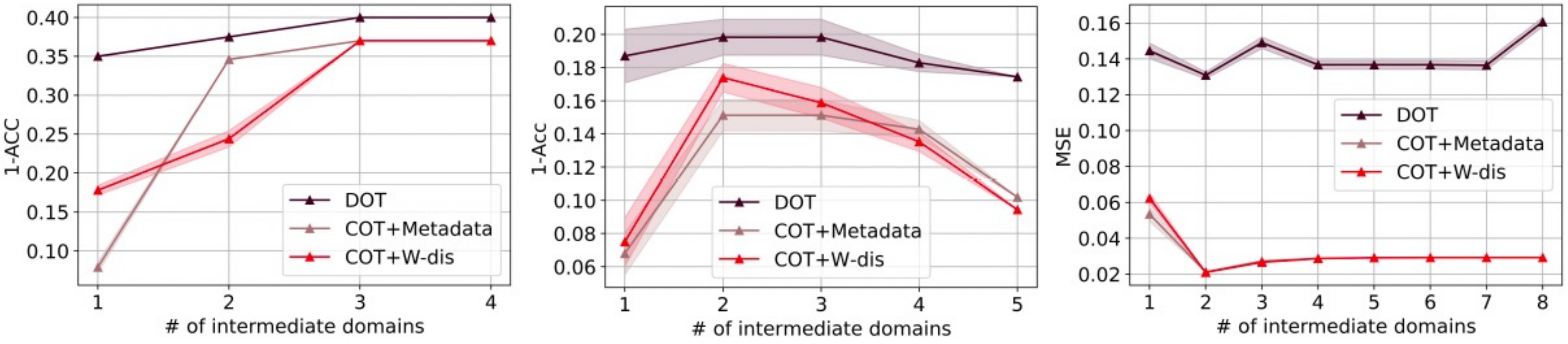
- Introduction
- Problem definition
- Methodology
- **Experiments**
- Conclusion

Experiments

➢ Dataset

- ADNI: The Alzheimer's Disease Neuroimaging Initiative, 2D MRI images, under different age
- Battery Charging-discharging Capacity: under different SoC
- Rotated MNIST: under different angle

Experiments



(a) Rotated MNIST (b) ADNI (c) Battery

TABLE I
MSE OR ACCURACY FOR THREE DATASETS OF DIFFERENT ALGORITHMS

Method	ADNI (↑)	Battery (↓)	ROT MNIST (↑)
AGST [43]	57.3	0.3534	76.2
Gradual ST [11]	64.5	0.1068	87.9
CDOT [15]	82.6	0.0209	75.6
W-MPOT(p2→p1)	<u>86.7</u>	0.0199	<u>88.3</u>
W-MPOT(p1→p2)	86.5	<u>0.0197</u>	87.2
W-MPOT(p1 + p2)	**88.3**	**0.0185**	**89.1**

Contents

Conclusion

- Summary
  - Present a comprehensive framework W-MPOT for Continuous Domain Adaptation (CDA), addressing the challenge of significant domain shift and missing metadata
  - The Wasserstein-based Transfer Curriculum efficiently determines the order of intermediate domains in CDA
  - By enforcing consistency along multiple adaptation paths, MPOT minimizes the impact of errors and enhances the overall robustness and stability of the adapted model
- Future work
  - Deriving the generalization bound of error for W-MPOT
  - Try other transfer algorithm like reinforcement learning

Thanks
Q&A

Hanbing Liu
Advisor: Prof. Yang Li
2023/10/19