On Scalable and Efficient Computation of Large Scale Optimal Transport

Yujia Xie, Minshuo Chen, Haoming Jiang,
Tuo Zhao, Hongyuan Zha

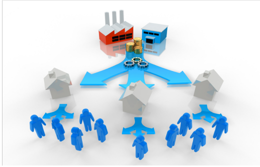
School of Computational Science and Engineering
H. Milton Stewart School of Industrial and Systems Engineering
Georgia Institute of Technology

Jun. 13, 2019

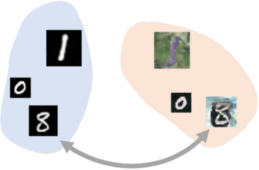
Optimal Tranport (OT)

The OT problem aims to **align** data from multiple sources.

Resource Allocation: We want to assign a set of assets to a set of receivers so that an optimal economic benefit is achieved.



Domain Adaptaion: We collect multiple datasets from different domains, and we need to learn a model from a source dataset, which can be further adapted to target datasets.

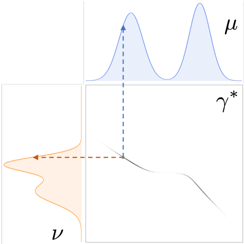


Both applications can be formulated as OT problems.

Optimal Tranport

Formulation OT aims to find an optimal joint distribution γ^* of μ and ν , which minimizes the expectation on some cost function c, i.e.,

$$\gamma^* = \arg\min_{\gamma} \mathbb{E}_{(X,Y)\sim\gamma}[c(X,Y)],$$
$$\text{subject to} \quad X \sim \mu, \quad Y \sim \nu.$$



γ^* is referred as the **optimal transport plan**, suggesting the way to transport between μ and ν with minimum cost.

Existing Methods Discretization + Linear Programming
The number of grids needs to **scale exponentially** w.r.t. dimension

SPOT

- OT: $\gamma^* = \arg\min_\gamma \mathbb{E}_{(X,Y)\sim\gamma}[c(X,Y)], \quad \text{s.t.} \quad \mathrm{X} \sim \mu, \mathrm{Y} \sim \nu.$
- Approximate γ^* by an implicit generative model $G(Z)$,

$$G(Z) = \left[\frac{G_X(Z)}{G_Y(Z)}\right] \approx \left[\frac{X}{Y}\right],$$

where $\quad Z \sim \rho, X \sim \mu, Y \sim \nu.$

- Substitute $G(Z)$ into OT problem, we can rewrite the problem as

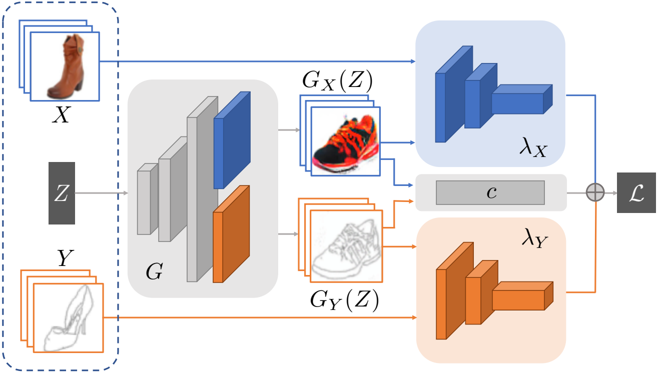
$$\underset{G}{\arg\min} \;\; \mathbb{E}_{Z\sim\rho}[c(G_X(Z), G_Y(Z))],$$

$$\text{subject to} \;\; \mathcal{W}_1(G_X(Z), \mu) = 0, \;\; \mathcal{W}_1(G_Y(Z), \nu) = 0.$$

where $\mathcal{W}_1(G_X(Z), \mu)$ denotes the standard Wasserstein metric between a random vector $G_X(Z)$ and a distribution μ. Here we use the fact that $\mathcal{W}_1(G_X(Z), \mu) = 0$ indicates $G_X(Z) \sim \mu$.

SPOT

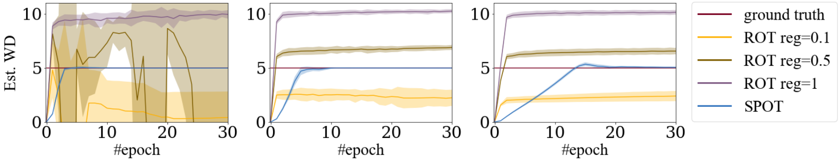
$$\min_{G \in \mathcal{G}} \max_{\lambda_X \in \mathcal{F}_X^1, \lambda_Y \in \mathcal{F}_Y^1} \mathbb{E}_{Z \sim \rho}[c(G_X(Z), G_Y(Z))] + \eta\big(\lambda_X(G_X(Z), X) + \lambda_Y(G_Y(Z), Y)\big),$$



Computing Wasserstein Distance (WD)

WD is the expected cost of optimal transport plan,

$$\mathcal{W} = \mathbb{E}_{(X,Y)\sim\gamma^*}[c(X,Y)].$$



LR $=10^{-3}$ LR $=10^{-4}$ LR $=10^{-5}$

Here, ROT is the state-of-the-art method (Seguy, 2018).

Generate Paired Samples

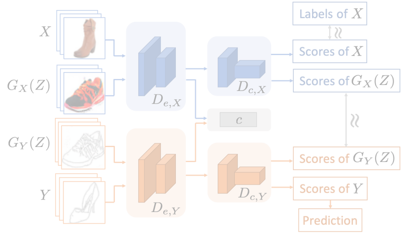
Photos-Monet

Domain Adaptation (DA)

Setting:

Goal: predict the labels of $\{y_j\}$.

New DA method – **DASPOT**



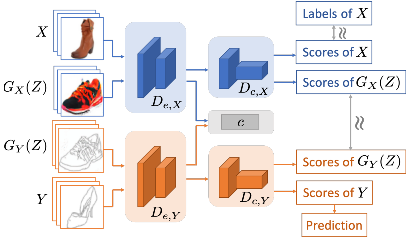
Source Target	MNIST USPS	USPS MNIST	SVHN MNIST	MNIST MNISTM
ROT (Seguy, 2018)	72.6%	60.5%	62.9%	–
StochJDOT (Damodaran, 2018)	93.6%	90.5%	67.6%	66.7%
DeepJDOT (Damodaran, 2018)	95.7%	96.4%	**96.7%**	92.4%
DASPOT	**97.5%**	**96.5%**	96.2%	**94.9%**

Domain Adaptation (DA)

Setting:

Goal: predict the labels of $\{y_j\}$.

New DA method – **DASPOT**

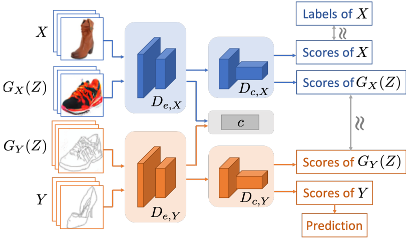


Source Target	MNIST USPS	USPS MNIST	SVHN MNIST	MNIST MNISTM
ROT (Seguy, 2018)	72.6%	60.5%	62.9%	–
StochJDOT (Damodaran, 2018)	93.6%	90.5%	67.6%	66.7%
DeepJDOT (Damodaran, 2018)	95.7%	96.4%	**96.7%**	92.4%
DASPOT	**97.5%**	**96.5%**	96.2%	**94.9%**

Domain Adaptation (DA)

Setting:

New DA method – **DASPOT**



Goal: predict the labels of $\{y_j\}$.

Source Target	MNIST USPS	USPS MNIST	SVHN MNIST	MNIST MNISTM
ROT (Seguy, 2018)	72.6%	60.5%	62.9%	–
StochJDOT (Damodaran, 2018)	93.6%	90.5%	67.6%	66.7%
DeepJDOT (Damodaran, 2018)	95.7%	96.4%	**96.7%**	92.4%
DASPOT	**97.5%**	**96.5%**	96.2%	**94.9%**

Thank you!