

# 最適輸送を用いた教師なしドメイン適応問題における 輸送先クラスラベル推定に基づく識別部分空間学習

北村 優輝士<sup>1,a)</sup> 福永 拓海<sup>2</sup> 笠井 裕之<sup>1,2</sup>

**概要:** ドメイン適応問題は、ソース領域のラベルを利用することでラベル情報が不足しているターゲット領域の分類器の予測精度を向上する問題に利用される。この問題に対して、従来より、最適輸送問題を用いて分類性能を向上させる手法が提案されてきたが、近年、クラスラベル情報を考慮した計量学習により、部分空間へデータを射影した上で輸送を行う手法も提案されている。しかしながら、ターゲットデータについてはクラスラベルが存在しないことから部分空間の計量学習を行えないという問題がある。そこで本稿では、ターゲットデータについても計量学習を行って最適な部分空間を求めることを目的として、ターゲットデータのラベル推定と、推定ラベルに基づく計量学習による最適輸送を用いたドメイン適応手法を提案する。具体的には、ターゲットデータの確率的クラスラベル推定と推定ラベルを用いた確率的計量学習による部分空間学習により最適輸送ドメイン適応手法を提案する。

## Discriminative subspace learning with target class label estimation in unsupervised domain adaptation with optimal transport

### 1. はじめに

ドメイン適応問題は、ソース領域のラベルを利用することで、ラベル情報が不足している関連ターゲット領域の分類器の予測精度を向上する問題として定義される。ドメイン適応問題はコンピュータビジョンや医療診断、リコメンドシステムなどの様々な分野で利用されている。ドメイン適応問題においては、ターゲットラベル情報の条件に基づき、ラベル情報の一部を利用できる半教師有り学習とラベル情報を全く使用できない教師無し学習の2種類の学習方法が検討されている。本稿では、教師なしドメイン適応を対象にするが、学習データとテストデータが一致しないドメインシフトを引き起こし、学習性能を低下するということが知られている。このような問題に対して、ドメインシフトに対応した教師無し学習のドメイン適応を扱うために

多くの手法が提案されている。基本的な手法は固有ベクトルを用いてターゲットとソース領域を表現する Subspace Alignment (SA)[1] や二次統計量である分散・共分散を用いた the Correlation Alignment (CORAL)[2] である。また、深層学習を用いたドメイン適応も関心を集めており、コンピュータビジョンでその成功を収めている [3]。

一方、これらの手法と異なったアプローチとして最適輸送問題を利用したドメイン適応も提案されている。最適輸送問題は始点から終点へのコストを最小化する問題として定められ、確率分布間の距離を表す Wasserstein 距離として知られている。最適輸送問題のドメイン適応問題への適用手法として、輸送元と輸送先の関係を表す輸送計画を用いて、ドメインシフトの影響を低減した最適輸送問題を用いた手法として Optimal Transport for Domain Adatation (OTDA) が提案されている [4], [5]。本手法により、異なったラベルを持つソースデータが同じターゲットデータに対して輸送される状況を防ぐことが可能となり、分類性能が向上している。また最近の研究では、SA と OTDA の考えを統合し、それぞれの領域の計量学習を導入した Metric Learning in Optimal Transport (MLOT) が提案されている [6]。この手法により、それぞれのドメインのラベルを考

<sup>1</sup> 早稲田大学 基幹理工学部 情報通信学科  
Department of Communications and Computer Engineering,  
School of Fundamental Science and Engineering, Waseda  
University

<sup>2</sup> 早稲田大学大学院 基幹理工学研究科 情報理工・通信専攻  
Department of Computer Science and Communications Engineering,  
Graduate School of Fundamental Science and Engineering,  
Waseda University

a) makikitamu@toki.waseda.jp

慮した距離と低次元表現可能な部分空間の学習を可能にしている。

しかしながら、MLOTではターゲットデータがクラスラベルを持たないため、ターゲットデータについては従来の計量学習を行えないという問題がある。これに対し、最適輸送によりソースとターゲット間で対応関係が輸送行列で表現されていることに着目し、この対応関係をクラスラベルにも適用可能することを考える。そして、ターゲットデータのクラスラベルを確率的に推定し、ターゲットデータについても計量学習を行なうことで、識別性能を向上する部分空間を求めることを考える。そこで本稿では、最適輸送によりターゲットクラスラベルを推定し、その推定ラベルに基づきターゲットデータ空間における計量学習を提案する。

本稿で用いる表記法を整理する。 $\mathbb{R}^n$  は  $n$  次元実数ベクトル全体の集合、 $\mathbb{R}^{m \times n}$  は  $m$  行  $n$  列の行列全体の集合を表す。ベクトルは  $\mathbf{a}, \mathbf{b}, \dots$  で表し、行列は  $\mathbf{A}, \mathbf{B}, \dots$  で表す。サイズが同じ行列  $\mathbf{A}, \mathbf{B}$  に対するフロベニウスノルムを  $\langle \mathbf{A}, \mathbf{B} \rangle = \text{tr}(\mathbf{A}^T \mathbf{B})$  で表す。同一空間内のベクトル  $\mathbf{a}, \mathbf{b} \in \mathbb{R}^n$  の内積を  $\mathbf{a}^T \mathbf{b}$  で表す。行列  $\mathbf{A} \in \mathbb{R}^{m \times n}$  が与えられたとき、その第  $i$  列ベクトルを  $\mathbf{a}_i$  で表す。

## 2. 関連研究

本節では、提案手法で用いる最適輸送問題、最適輸送を利用したドメイン適応、および計量学習について説明する。

### 2.1 最適輸送問題

最適輸送問題は、始点から終点への輸送コストを最小化する問題である。二つの経験的確率分布  $\mu = \sum_i \mathbf{a}_i \delta_{\mathbf{x}_i}$ ,  $\nu = \sum_i \mathbf{b}_i \delta_{\mathbf{y}_i}$  と輸送コスト  $C_{i,j}$  が与えられる時、最適輸送問題は次で定義される。

$$\min_{\mathbf{T} \in \mathcal{U}(\mu, \nu)} \langle \mathbf{T}, \mathbf{C} \rangle,$$

ここで制約条件  $\mathcal{U}(\mu, \nu)$  は

$$\mathcal{U}(\mu, \nu) = \{ \mathbf{T} \geq \mathbf{0} : \mathbf{T} \mathbf{1}_n = \mu, \mathbf{T}^T \mathbf{1}_m = \nu \}$$

で定められる。最適輸送問題は確率分布間の距離を表現可能な Wasserstein 距離 [7], [8] として知られており、その性質からグラフ最適輸送問題 [9], シーケンスマッチング [10], 敵対的学習 [11], [12] などの、機械学習における様々な分野で応用されている。

### 2.2 最適輸送を利用したドメイン適応 (OTDA)

OTDA は最適輸送問題を用いたドメイン適応を扱う手法である。従来最適輸送問題に二つの正則化項を加えた関数を最小化する問題であり、次に定義される。

$$\min_{\mathbf{T} \in \mathcal{U}(\mu, \nu)} \langle \mathbf{T}, \mathbf{C} \rangle - \lambda_e \Omega_e(\mathbf{T}) + \lambda_c \Omega_c(\mathbf{T}), \quad (1)$$

ここで  $\Omega_e(\mathbf{T})$  はシャノンエントロピー項、 $\Omega_c(\mathbf{T})$  はグループ正則化項を表す。最適輸送問題では、エントロピー正則化を施すことで Sinkhorn アルゴリズムを適用することが可能であり、従来の線形計画問題よりも高速に解くことが可能である [13]。グループ正則化項は二つの異なったラベルを持つソースデータが同一ターゲットデータに対して輸送される状況を防いでいる。問題 (1) は制約付き最適化問題であるため解くことが難しい。[14] ではグループ正則化項として、クラス  $c$  を有する行のみを含むような、輸送行列  $\mathbf{T}$  の第  $j$  列目のノルムを用いて、 $\Omega_c(\mathbf{T}) = \sum_j \sum_c \|\mathbf{T}(L_c, j)\|_1^{\frac{1}{2}}$  を採用し、特殊な制約付き非凸最適化問題を解いているが、非凸最適化問題であるため大域的最適解を求めるのは依然難しい。そこで一般化座標勾配法 [15] を利用することで、Sinkhorn アルゴリズムを利用して問題 (1) を解くことができる [5]。その際、グループ正則化項としてグループラッソ正則化やラプラス正則化が利用される。

### 2.3 計量学習

計量学習はデータのクラスラベルの情報を反映した距離を学習する問題である。ユークリッド空間における二点間の距離はその差分の  $l_2$  ノルムで表される。しかし、実世界のデータはラベル情報も含むため、ユークリッド距離が最適であるとは限らない。そこで、ラベル付きデータについて、同一クラスラベルデータ同士を密接させ、異なるクラスラベルデータ同士を乖離させる計量空間を求めることが重要である。そこで計量学習では、クラス識別を考慮したマハラノビス距離を表現する半正定値行列  $\mathbf{M}$  を求める問題として定式化される。しかしながら、一般的に半正定値行列  $\mathbf{M}$  の算出は難しいことが知られていることから、多くの計量学習の手法では、半正定値行列を  $\mathbf{M} = \mathbf{L}^T \mathbf{L}$  に分解し、 $\mathbf{L}$  を学習するアプローチを採用している。

$$\begin{aligned} d(\mathbf{x}_i, \mathbf{x}_j) &= (\mathbf{x}_i - \mathbf{x}_j)^T \mathbf{M} (\mathbf{x}_i - \mathbf{x}_j) \\ &= (\mathbf{x}_i - \mathbf{x}_j)^T \mathbf{L}^T \mathbf{L} (\mathbf{x}_i - \mathbf{x}_j) = \|\mathbf{L}(\mathbf{x}_i - \mathbf{x}_j)\|_2^2. \end{aligned}$$

射影行列  $\mathbf{L}$  を学習する計量学習の手法はいくつか知られている。クラス分布の形状についての仮定を持たずにノンパラメトリックな分類モデルを学習する Neighborhood Components Analysis (NCA)[16] や、フィッシャーの線形判別と Locality Preserving Projection (LPP)[17] を組み合わせることでマルチモーダルなデータについても学習を可能とした Local Fisher Discriminant Analysis (LFDA)[18], カーネル回帰を用いることでリープワンアウト回帰誤差を直接最小化する Metric Learning for Kernel Regression (MLKR)[19] などが挙げられる。

本稿の提案手法では、特に、計量学習としてラージマージン最近傍法に着目する [20]。ラージマージン最近傍法は、あるノードデータ  $\mathbf{x}_i$  とその近傍の同クラスデータ (ターゲットネイバー  $\mathbf{x}_j$ ) とのマハラノビス距離を小さくし、あ

るノードデータ  $\mathbf{x}_i$  と近傍の別クラスデータ (インポスター  $\mathbf{x}_l$ ) の距離についてはマハラノビス距離を大きくすることで、ラベル間の混在を減らすような計量を導出する。インポスター  $\mathbf{x}_l$  とは、ノード  $i$  との距離の二乗  $\|\mathbf{L}(\mathbf{x}_i - \mathbf{x}_l)\|_2^2$  が、ノード  $i$  とそのターゲットネイバー  $\mathbf{x}_j$  の距離とマージン  $\xi$  の和  $\|\mathbf{L}_s(\mathbf{x}_i - \mathbf{x}_j)\|_2^2 + \xi$  よりも小さい別クラスデータである。すなわち、インポスターは次の不等式を満たす別クラスデータである。

$$\|\mathbf{L}(\mathbf{x}_i - \mathbf{x}_l)\|_2^2 < \|\mathbf{L}(\mathbf{x}_i - \mathbf{x}_j)\|_2^2 + \xi.$$

学習の指標として、マージンの最大化とインポスターの個数削減を目的としている。ターゲットネイバーとのマハラノビス距離の減少はマージンの最大化に、インポスターとのマハラノビス距離の増大はインポスターの個数削減に寄与している。これらを考慮し、損失関数  $\Omega_l(\mathbf{L})$  は、次で定義される。

$$\Omega_l(\mathbf{L}) = (1 - \mu) \cdot \epsilon_{\text{pull}}(\mathbf{L}) + \mu \cdot \epsilon_{\text{push}}(\mathbf{L}).$$

ここで、 $\epsilon_{\text{pull}}(\mathbf{L})$  は以下に定義される。

$$\epsilon_{\text{pull}}(\mathbf{L}) = \sum_i \sum_{j \in \mathcal{N}_i} \|\mathbf{L}(\mathbf{x}_i - \mathbf{x}_j)\|_2^2. \quad (2)$$

また、 $\epsilon_{\text{push}}(\mathbf{L})$  は以下に定義される。

$$\epsilon_{\text{push}}(\mathbf{L}) = \sum_i \sum_{j \in \mathcal{N}_i} \sum_{l \in \mathcal{I}_i} (\|\mathbf{L}(\mathbf{x}_i - \mathbf{x}_j)\|_2^2 + \xi - \|\mathbf{L}(\mathbf{x}_i - \mathbf{x}_l)\|_2^2). \quad (3)$$

ここで、 $\mathcal{N}_i$  はノード  $i$  のターゲットネイバー  $\mathbf{x}_j$  のノードインデックス集合であり、 $\mathcal{I}_i$  はノード  $i$  のインポスター  $\mathbf{x}_l$  のノードインデックス集合である。 $\epsilon_{\text{pull}}(\mathbf{L})$  は射影空間におけるマージンの最大化を促し、 $\epsilon_{\text{push}}(\mathbf{L})$  は、射影空間におけるインポスターの個数削減に寄与する。

## 2.4 計量学習を導入した最適輸送におけるドメイン適応 (MLOT)

最適輸送を用いたドメイン適応を考える手法が提案されているが、最適輸送で用いられる輸送コスト  $\mathbf{C}$  はユークリッドノルムを用いて定義された距離であることが多い。そのため、ソース・ターゲットデータのクラスラベル情報を考慮していない。そこで、前述の計量学習であるラージマージン最近傍法を導入した手法が MLOT である。具体的には、ラベル  $\mathbf{y}^s$  を有する  $m$  個のソースデータ  $\mathbf{X}^s = [\mathbf{x}_1^s, \mathbf{x}_2^s, \dots, \mathbf{x}_m^s]$  から  $n$  個のターゲットデータ  $\mathbf{X}^t = [\mathbf{x}_1^t, \mathbf{x}_2^t, \dots, \mathbf{x}_n^t]$  のラベル  $\mathbf{y}^t$  を予測すると想定すると、それぞれの部分空間へ射影した際の距離を考えるため、射影行列  $\mathbf{L}_s, \mathbf{L}_t$  を用いて、輸送コストは  $\mathbf{C}(\mathbf{L}_s, \mathbf{L}_t)_{i,j} = \|\mathbf{L}_s \mathbf{x}_i^s - \mathbf{L}_t \mathbf{x}_j^t\|_2^2$  で定められる。ソースドメインに対する計量学習については、射影行列  $\mathbf{L}_s$  に対する正則化項  $\Omega_{l_s}(\mathbf{L}_s)$  を導入する。この問題の

定式化は次に定義される。

$$\min_{\substack{\mathbf{T} \in \mathcal{U}(\boldsymbol{\mu}, \boldsymbol{\nu}) \\ \mathbf{L}_s \in \mathbb{R}^{m \times m}}} \langle \mathbf{T}, \mathbf{C}(\mathbf{L}_s, \mathbf{L}_t) \rangle + \lambda_e \Omega_e(\mathbf{T}) + \lambda_c \Omega_c(\mathbf{T}) + \lambda_{l_s} \Omega_{l_s}(\mathbf{L}_s).$$

なお、ターゲットデータはクラスラベルを有さないため計量学習を適用せず、主成分分析を用いて  $\mathbf{L}_t$  を一度だけ算出する。

## 3. 提案手法

MLOT では、ターゲットデータ空間においてクラスラベルがないため計量学習を行えない。これに対し、最適輸送によりソースとターゲット間で対応関係が存在し、それは輸送行列で表現されるが、この対応関係はクラスラベルにも適用可能だと考える。この考えに基づき、ターゲットデータのクラスラベルを確率的に推定してターゲットデータについても計量学習を行い、識別性能の向上した識別部分空間を求めたい。

先述の通り、MLOT においてはソースデータ空間については計量学習を用いて  $\mathbf{L}_s$  を学習し、ターゲットデータ空間については主成分分析を用いて  $\mathbf{L}_t$  を一度学習する。ここで、ターゲットデータ空間についてもソースデータ空間と同様に計量学習を用いて射影行列  $\mathbf{L}_t$  を学習する手法を提案する。そのために、ターゲットデータのクラスラベルを推定して、その推定クラスラベルに基づいた確率的な計量学習を行う。なお、以後、2 クラス問題を想定して説明する。

### 3.1 ターゲットデータの確率的クラスラベル推定

輸送行列で表現されるソースデータとターゲットデータの対応関係はそれぞれのクラスラベルにも適用可能であるとする。そこで、輸送行列  $\mathbf{T}$  の列ベクトルを重み係数としてクラスラベル  $\hat{y}_j^t$  を推定する。その推定クラスラベルは次に定義される。

$$\hat{y}_j^t = \frac{\sum_{i=1}^n \mathbf{T}_{i,j} y_i^s}{\sum_{i=1}^n \mathbf{T}_{i,j}} = \frac{\sum_{i=1}^n \mathbf{T}_{i,j} y_i^s}{b_j}.$$

ここで、 $\hat{y}_j^t \in [0, 1]$  である。この推定ラベル  $\hat{y}_j^t$  を利用して、ターゲットデータ空間についても計量学習を行う。

### 3.2 確率的推定ラベルを用いた部分空間学習

MLOT の時と同様に、推定クラスラベルに基づきターゲット空間におけるラージマージン最近傍法を提案する。推定ラベルは重み付きラベルのため、ソースデータに対するラージマージン最近傍法を直接利用することはできない。そこで、新たに重み付きラベルに対するラージマージン

---

**Algorithm 1** MLOT with target domain

---

**Require:** stepsize  $\eta$ , max iteration #  $K$ ,  $\mathbf{X}^s, \mathbf{X}^t, \mathbf{y}^s$ ,  
regularization parameters  $\lambda_e, \lambda_c, \lambda_{l_s}, \lambda_{l_t}$ .

- 1:  $\mathbf{V}_s = \text{PCA}(\mathbf{X}_s), \mathbf{V}_t = \text{PCA}(\mathbf{X}_t)$
- 2: Initialize  $\mathbf{L}_s = \mathbf{V}_s^T \mathbf{V}_s, \mathbf{L}_t = \mathbf{V}_t^T \mathbf{V}_t$
- 3: **for**  $k = 0 \dots K$  **do**
- 4:  $\tilde{\mathbf{y}}_i^t = \mathbf{t}_i^T \mathbf{y}^s / b_i$
- 5:  $\arg \min_{\mathbf{T} \in \mathcal{U}(\boldsymbol{\mu}, \boldsymbol{\nu})} \langle \mathbf{T}, \mathbf{C}(\mathbf{L}_s, \mathbf{L}_t) \rangle + \lambda_e \Omega_e(\mathbf{T}) + \lambda_c \Omega_c(\mathbf{T}) + \lambda_{l_t} \Omega_{l_t}(\mathbf{L}_t, \tilde{\mathbf{y}}^t)$
- 6:  $\mathbf{L}_s = \mathbf{L}_s - \eta \nabla_{\mathbf{L}_s} (\langle \mathbf{T}, \mathbf{C}(\mathbf{L}_s, \mathbf{L}_t) \rangle + \lambda_{l_s} \Omega_{l_s}(\mathbf{L}_s))$
- 7:  $\mathbf{L}_t = \mathbf{L}_t - \eta \nabla_{\mathbf{L}_t} (\langle \mathbf{T}, \mathbf{C}(\mathbf{L}_s, \mathbf{L}_t) \rangle + \lambda_{l_t} \Omega_{l_t}(\mathbf{L}_t))$
- 8: **end for**
- 9:  $\tilde{\mathbf{X}}_s = \mathbf{T} \mathbf{L}_t \mathbf{X}_t$
- 10: classifier is produced
- 11:  $\mathbf{y}_{new}^t$  is produced by classifier

---

最近傍法の正則化項  $\Omega_{l_t}(\mathbf{L}_t, \tilde{\mathbf{y}}^t)$  を次に定義する。

$$\Omega_{l_t}(\mathbf{L}_t, \tilde{\mathbf{y}}^t) = (1 - \mu) \cdot \epsilon_{\text{pull}}(\mathbf{L}_t, \tilde{\mathbf{y}}^t) + \mu \cdot \epsilon_{\text{push}}(\mathbf{L}_t, \tilde{\mathbf{y}}^t). \quad (4)$$

ここで、 $\epsilon_{\text{pull}}(\mathbf{L}_t, \tilde{\mathbf{y}}^t), \epsilon_{\text{push}}(\mathbf{L}_t, \tilde{\mathbf{y}}^t)$  はマージンの最大化とインポスターの個数制限に寄与する項である。従来ラージマージン最近傍法のマージン最大化項 (2) はあるデータとそのターゲットネイバーの距離を最小化している。これを受けて、重み付きラベルの差分が小さいほどそのマハラノビス距離を最小化することを考慮し、 $\epsilon_{\text{pull}}(\mathbf{L}_t, \tilde{\mathbf{y}}^t)$  を次に定める。

$$\epsilon_{\text{pull}}(\mathbf{L}_t, \tilde{\mathbf{y}}^t) = \sum_{i,j} (1 - |\tilde{y}_i^t - \tilde{y}_j^t|) \|\mathbf{L}_t(\mathbf{x}_i^t - \mathbf{x}_j^t)\|_2^2. \quad (5)$$

同様に重み付きインポスターを次に定義する。

$$\epsilon_{\text{push}}(\mathbf{L}_t, \tilde{\mathbf{y}}^t) = \begin{cases} \sum_{i,j,l} f(\mathbf{L}_t, \tilde{y}_i^t, \tilde{y}_j^t, \tilde{y}_l^t), & \text{if } |\tilde{y}_i^t - \tilde{y}_j^t| > |\tilde{y}_i^t - \tilde{y}_l^t|, \\ 0, & \text{otherwise.} \end{cases}$$

ここで  $f(\mathbf{L}_t, \tilde{y}_i^t, \tilde{y}_j^t, \tilde{y}_l^t)$  を以下に定義する。

$$f(\mathbf{L}_t, \tilde{y}_i^t, \tilde{y}_j^t, \tilde{y}_l^t) = (1 - |\tilde{y}_i^t - \tilde{y}_j^t|) \|\mathbf{L}_t(\mathbf{x}_i^t - \mathbf{x}_j^t)\|_2^2 - \{\xi^t - |\tilde{y}_i^t - \tilde{y}_l^t|\} \|\mathbf{L}_t(\mathbf{x}_i^t - \mathbf{x}_l^t)\|_2^2.$$

以上を考慮して重み付きラージマージン最近傍項を損失に加え、最小化問題を次に定義する。

$$\min_{\substack{\mathbf{T} \in \mathcal{U}(\boldsymbol{\mu}, \boldsymbol{\nu}) \\ \mathbf{L}_s \in \mathbb{R}^{m \times m} \\ \mathbf{L}_t \in \mathbb{R}^{n \times n}} \langle \mathbf{T}, \mathbf{C}(\mathbf{L}_s, \mathbf{L}_t) \rangle + \lambda_e \Omega_e(\mathbf{T}) + \lambda_c \Omega_c(\mathbf{T}) + \lambda_{l_s} \Omega_{l_s}(\mathbf{L}_s) + \lambda_{l_t} \Omega_{l_t}(\mathbf{L}_t, \tilde{\mathbf{y}}^t). \quad (6)$$

#### 4. まとめと今後の課題

本稿では、ターゲットデータのクラスラベルを推定し、そのラベルを用いて計量学習を行う手法を提案した。特に、重み付き推定クラスラベルに対する計量学習の正則化

項を提案した。今後の課題は二つある。一つ目は、重み付きラベルに対してターゲットネイバーとインポスターを定める方法である。通常ラベルは整数値であるため、分類することが可能であるが、重み付きラベルはそれに当てはまらない。二つ目はマルチクラスラベルの拡張である。現状の推定方法では、クラスラベルの種類が多くなるほどソースクラスラベルの値が推定ラベルの値に影響を与える。

#### 参考文献

- [1] Fernando, B. and et al., A. H.: Unsupervised Visual Domain Adaptation Using Subspace Alignment, *ICCV* (2013).
- [2] Sun, B. and et al., J. F.: Return of Frustratingly Easy Domain Adaptation, *AAAI* (2015).
- [3] Wang, M. and Deng, W.: Deep visual domain adaptation: A survey, *Neurocomputing*, Vol. 312, pp. 135–153 (2018).
- [4] Courty, N. and Flamary, R. e. a.: Joint distribution optimal transportation for domain adaptation, *NIPS*, Vol. 30 (2017).
- [5] Courty, N. and Flamary, R. e. a.: Optimal Transport for Domain Adaptation, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 39, No. 9, pp. 1853–1865 (2017).
- [6] Kerdoncuff, T. and Emonet, R. e. a.: Metric Learning in Optimal Transport for Domain Adaptation, *IJCAI* (2020).
- [7] Villani, C.: *Optimal transport: Old and new*, Springer (2008).
- [8] Peyré, G. and Cuturi, M.: Computational Optimal Transport, *arXiv preprint: arXiv:1803.00567* (2020).
- [9] Chen, L. and Gan, Z. e. a.: Graph Optimal Transport for Cross-Domain Alignment, *ICML* (2020).
- [10] Su, B. and Hua, G.: Order-Preserving Wasserstein Distance for Sequence Matching, *CVPR* (2017).
- [11] Arjovsky, M. and et al., S. C.: Wasserstein Generative Adversarial Networks, *ICML* (2017).
- [12] Chakrabarti, S. and Huang, Y. e. a.: Quantum Wasserstein generative adversarial networks, *NIPS* (2019).
- [13] Cuturi, M.: Sinkhorn Distances: Lightspeed Computation of Optimal Transport, *NIPS* (2013).
- [14] Courty, N. and Flamary, R. e. a.: Domain Adaptation with Regularized Optimal Transport, *ECML PKDD* (2014).
- [15] Bredies, K. and et al., D. L.: A generalized conditional gradient method and its connection to an iterative shrinkage method, *Computational Optimization and Applications*, Vol. 42, No. 2, pp. 173–193 (2009).
- [16] Goldberger, J. and Hinton, G. E. e. a.: Neighbourhood components analysis, *NIPS*, Vol. 17, Citeseer (2004).
- [17] He, X. and Niyogi, P.: Locality preserving projections, *NIPS*, Vol. 16, No. 16 (2004).
- [18] Sugiyama, M.: Dimensionality reduction of multi-modal labeled data by local fisher discriminant analysis., *JMLR*, Vol. 8, pp. 1027–1061 (2007).
- [19] Weinberger, K. Q. and Tesauro, G.: Metric learning for kernel regression, *AISTATS*, PMLR (2007).
- [20] Weinberger, K. Q. and Saul, L. K.: Distance Metric Learning for Large Margin Nearest Neighbor Classification, *JMLR*, pp. 207–244 (2009).