進捗報告

田渕 智久

早稲田大学 基幹理工学研究科 情報理工・情報通信専攻

卒業研究の続き

Branched Optimal Transport

今後の研究分野

卒業研究のおさらい

最適輸送の外れ値に対する頑健性の欠如

- 最適輸送は外れ値のサンプルであっても必ず輸送される (左図)
- 不均衡最適輸送は外れ値のサンプルを輸送しない (中央)

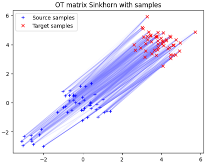


Figure: 最適輸送

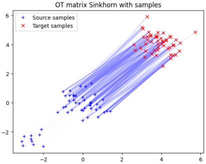


Figure: 不均衡最適輸送

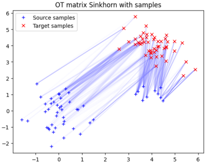


Figure: 別の外れ値

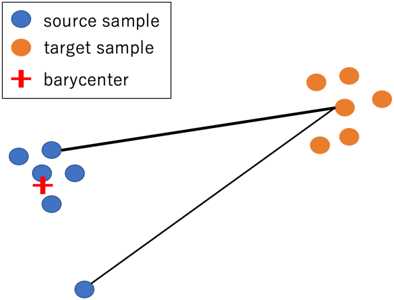
- しかし，不均衡最適輸送では対処できない外れ値も存在する (右図)
  ドメイン適応では外れ値への輸送が分類精度を劣化させる
  [Fatras et al., 2021]
- この外れ値に対しても輸送しないようにするための方法を提案した

卒業研究のおさらい

クラスタから離れているサンプルの輸送コストを大きくする

- 着眼点：外れ値は輸送されるべきでない
- アイデア: クラスタから離れているサンプルの輸送コストを大きくする
- 提案手法: コスト行列にクラスタ中心からの乖離尺度 $\mathbf{D}_{i,j}$ を加算

$$\mathbf{C}'_{i,j} = \mathbf{C}_{i,j} + \lambda \mathbf{D}_{i,j}$$



卒業研究のおさらい

実験内容

- 合成データ
  - (a) 外れ値を含む合成データを用いた最適輸送
    輸送計画の様子と転移後のサンプルの様子が適切かを評価
- 実データを用いた 2 つの実験
  - (b) 実データを用いた分類性能の評価
  - (c) 外れ値を付加した実データによる分類性能の評価
- 使用するデータセット
  - Office+Caltech データセット (物体検出タスク)

Table: データセットの詳細

ドメイン	サンプル数	次元	クラス数	略
Caltech	1123	800	10	C
Amazon	958	800	10	A
Webcam	295	800	10	W
DSLR	157	800	10	D

卒業研究のおさらい

実験の目的

- 実データの外れ値の加え方
  - クラス 2 の重心の付近にクラス 1 の外れ値をサンプル数の 10 分の 1 の数だけ加えた
- 期待する実験結果
  (a) 外れ値のサンプルへ輸送されていない
  (b) 通常の UOT と精度が変わらない
  (c) 外れ値への輸送をしないことによって提案した手法の精度が向上
- 各ハイパーパラメータは実験した中で最も精度が高かった値を採用
  - エントロピー正則化項
  - KL ダイバージェンス項
  - 提案したコスト行列の項

実験結果 (a)：輸送計画の可視化による評価

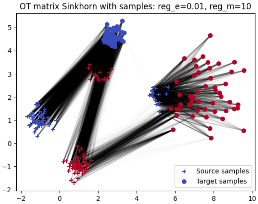


Figure: 輸送の様子 (既存手法)

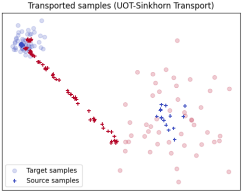


Figure: 転移後の様子 (既存手法)

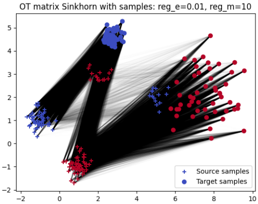


Figure: 輸送の様子 (提案手法)

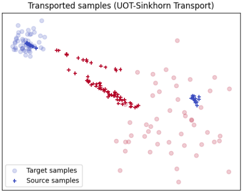


Figure: 転移後の様子 (提案手法)

実験結果 (b)(c)：外れ値を加えた実データの分類精度

Table: 実データの正解率 (%)

ドメイン	既存	提案
C→A	46.03	45.82
C→W	38.98	38.98
C→D	45.22	45.22
A→C	31.70	32.32
A→W	33.22	34.91
A→D	34.39	35.66
W→C	32.85	33.21
W→A	39.24	40.18
W→D	89.80	90.44
D→C	30.54	29.91
D→A	30.27	30.48
D→W	89.15	88.47
平均	58.81	59.05

Table: 外れ値を加えたデータの正解率 (%)

ドメイン	既存	提案
C→A	44.99	45.09
C→W	32.88	33.56
C→D	38.85	38.85
A→C	31.79	32.32
A→W	28.14	27.8
A→D	31.21	31.85
W→C	33.21	33.13
W→A	30.58	30.38
W→D	82.8	82.8
D→C	29.3	29.39
D→A	20.77	20.56
D→W	76.95	76.61
平均	54.82	54.72

卒業研究のおさらい

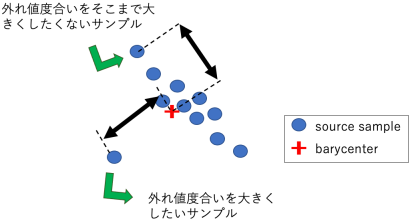
考察

- 実験結果 (a)
  - 外れ値への輸送量が小さくなり, 転移後のサンプルの位置が既存手法よりも少し改善された
- 実験結果 (b)
  - 既存手法と大きな違いがなかった
  - 元のデータセットには外れ値が多く存在していないため，期待通りの結果となった
- 実験結果 (c)
  - 既存手法と大きな違いがなかった
  - 既存手法は精度が大きく劣化するはずだが...
    - 外れ値の加え方が良くなかった?
    - 何か間違っていると思われる
  - 実験方法が正しいと仮定すると，提案した手法に問題がある (次ページ)

追加実験

提案した手法の問題点

- データが楕円形に広がる場合，「重心から離れているほど外れ値度合いが大きい」という考え方は不適切



- 点の近傍を考えるアイデアを追加する
  - 今までのアイデア：重心から離れているほど外れ値度合いが大きい
  - 追加するアイデア：その点から最も近いサンプルとの距離が大きいほど外れ値度合いが大きい

追加実験

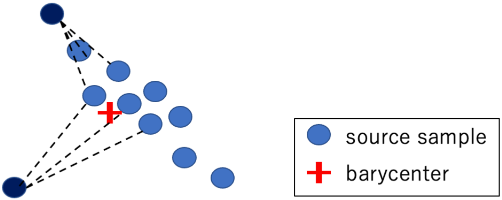
提案手法

- 最も近い 3 つのサンプルの距離の和を求める

$$\mathbf{Z}_{i,j} = z_1 + z_2 + z_3$$

- 求めた距離の和が大きいほど輸送コストを大きくする

$$\mathbf{C}'_{i,j} = \mathbf{C}_{i,j} + \lambda_1 \mathbf{D}_{i,j} + \lambda_2 \mathbf{Z}_{i,j}$$



新しく提案した手法 (UOTn) の実験結果

Table: 外れ値を加えたデータにおける実験結果

	1NN	OT	OTc	UOT	UOTc	UOTn
C→A	23.7	44.99	44.78	44.99	45.09	44.99
C→W	25.76	32.88	32.88	32.88	33.56	33.56
C→D	25.48	38.85	38.85	38.85	38.85	38.85
A→C	26.0	32.06	32.41	31.79	32.32	32.32
A→W	29.83	27.8	28.14	28.14	27.8	27.12
A→D	25.48	30.57	31.85	31.21	31.85	31.21
W→C	19.86	33.57	32.95	33.21	33.13	33.13
W→A	22.96	29.96	29.65	30.58	30.38	30.38
W→D	59.24	82.17	80.25	82.8	82.8	80.89
D→C	26.27	29.03	29.3	29.3	29.39	29.39
D→A	28.5	19.94	20.35	20.77	20.56	20.67
D→W	63.39	76.61	76.61	76.95	76.61	76.61
avg	48.53	54.54	54.35	54.82	54.72	54.52

この研究の思うところ

- 筋が悪い
  - コスト行列に加算する考え方は別に新しいわけではない
  - (精度が向上したとしても，他の外れ値を含むデータセットに対して精度が良くなるとは言いにくい)
  - そもそも最適輸送の外れ値に対する頑健性が重要視されていない

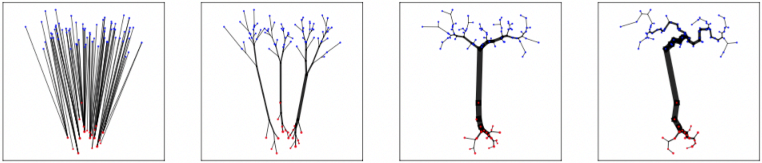
卒業研究の続き

Branched Optimal Transport

今後の研究分野

Branched Optimal Transport

- 合流した方が輸送コストが小さく済む ← 自然界に多く存在
  - トラックによる物資の輸送
  - 葉っぱの葉脈
- 輸送計画に枝分かれ構造が現れる



$\alpha = 1$ Optimal Transport | $\alpha = 0.95$ | $\alpha = 0.5$ | $\alpha = 0$ Euclidean Steiner Tree

- Branched OT が機械学習で利用できることを示唆 [Lippmann et al., 2022]
  - Branched OT の効率的な組合せ最適化ソルバーを提案
  - 具体的な応用は示していない

Branched Optimal Transport の定式化

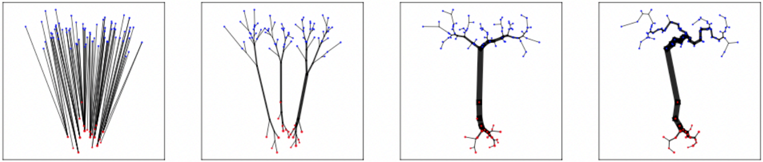
- Branched Optimal Transport
  - $m_{i,j}$ はエッジ i からエッジ j への輸送量
  - 合流することで，亜加法性より，全体の輸送コストが小さくなる

$$\underset{B,E,x_B,m_E}{\operatorname{argmin}} \sum_{(i,j)\in E} m_{i,j}^{\alpha} \|x_i - x_j\|_2$$

- 亜加法性 (subadditivity)

$$m_1^{\alpha} + m_2^{\alpha} > (m_1 + m_2)^{\alpha}$$

where $\alpha \in [0, 1]$



$\alpha = 1$ Optimal Transport | $\alpha = 0.95$ | $\alpha = 0.5$ | $\alpha = 0$ Euclidean Steiner Tree

Branched Optimal Transport の特徴

- 枝分かれした輸送計画が得られる
  - 通常の最適輸送と異なり，1 対 1 の輸送量は得られない
- 最適輸送問題とユークリッドシュタイナー木問題の中間として表現される
- 数学的性質はよく調べられている [Santambrogio, 2015]
- 1 点から複数点への輸送を考えると輸送の様子が綺麗

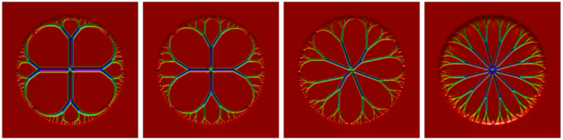


Figure: 1 点から複数点への輸送 [Santambrogio, 2015]

Branched Optimal Transport の研究の方向性

- 輸送計画の構造が有用な情報となる機械学習の応用先があるか
  - NLP のドキュメント間の最適輸送において，分岐点は何を表しているのだろうか
  - 機械学習においても輸送を合流させた方が人間の感覚に近い距離の測り方になっているのではないか
- 計算量が大きい
  - 組合せ最適化で解かれており，データ数が大きくなると厳しそう
  - 機械学習の性質を取り入れることで連続最適化に緩和できないか
    - 厳密に最適な分岐点を求める必要はないかも
    - スパース性?
- 現時点では何のタスクのために Branched OT を考える必要があるのか不明なので研究しにくい

卒業研究の続き

Branched Optimal Transport

今後の研究分野

今後の研究分野

- 最適輸送の新しいモデルの提案とその解析
  - Branched Optimal Transport を何かに応用できないか
  - 何をもって良い距離尺度とするのか
- Domain Adaptation/Domain Generalization
  - 最適輸送に限らず，広い視野で調査
- 生成モデル (GAN・Diffusion Model)
  - 拡散モデルの欠点の克服
  - 最適輸送に関する解析

References I

- Courty, N., Flamary, R., Habrard, A., and Rakotomamonjy, A. (2017).
  Joint distribution optimal transportation for domain adaptation.
  Advances in Neural Information Processing Systems, 30.
- Fatras, K., Séjourné, T., Courty, N., and Flamary, R. (2021).
  Unbalanced minibatch optimal transport; applications to domain adaptation.
  In Proceedings of the 38th International Conference on Machine Learning.
- Foret, P., Kleiner, A., Mobahi, H., and Neyshabur, B. (2021).
  Sharpness-aware minimization for efficiently improving generalization.
  In International Conference on Learning Representations.
- Keskar, N. S., Mudigere, D., Nocedal, J., Smelyanskiy, M., and Tang, P. T. P. (2017).
  On large-batch training for deep learning: Generalization gap and sharp minima.
  In International Conference on Learning Representations.

References II

- Li, H., Xu, Z., Taylor, G., Studer, C., and Goldstein, T. (2018).
  Visualizing the loss landscape of neural nets.
  In Advances in Neural Information Processing Systems.
- Lippmann, P., Sanmartín, E. F., and Hamprecht, F. A. (2022).
  Theory and approximate solvers for branched optimal transport with multiple sources.
  In Oh, A. H., Agarwal, A., Belgrave, D., and Cho, K., editors, Advances in Neural Information Processing Systems.
- Santambrogio, F. (2015).
  Optimal transport for applied mathematicians.
  Birkäuser, NY.
- Zhao, Y., Zhang, H., and Hu, X. (2022).
  Penalizing gradient norm for efficiently improving generalization in deep learning.
  In Proceedings of the 39th International Conference on Machine Learning.

References III