Robust computation of optimal transport by β-potential regularization (β-ポテンシャル正則化による最適輸送問題の頑健な計算方法について)

杉山・横矢・石田研究室
中村紳太郎 (47-206096)

概要

機械学習とは, データから必要な情報を抽出し, コンピュータに知的な作業（回帰分析, 画像分類, 画像生成, 非定常環境での意思決定など）をさせるデータ処理技術である. 重要なアプローチとして, データを生成した確率分布を我々が考えているモデルで近似することが挙げられる. そのためには, 確率分布間の「近さ」を定義することが必要である. 本論文では, 機械学習の分野において, 確率分布の近さの尺度として注目されている最適輸送問題を扱う. 最適輸送問題は計算量が多いため, もとの最適輸送問題をエントロピー罰則項で正則化し, シンクホルンアルゴリズムを用いて最適輸送問題の近似解を高速に計算するのが一般的なアプローチの１つである. しかし, シンクホルンアルゴリズムはカルバックライブラーダイバージェンスに関連した射影を実行するため, しばしば外れ値に弱くなる. この問題を解決するために本研究では, ロバスト統計学で発展した β-ダイバージェンスの下での射影を考える. 理論解析により我々の提案手法は確率質量が外れ値に輸送されるのを防ぐことができることを明らかにした. 本アルゴリズムの応用例として, 外れ値検知が可能であることを示し, 既存手法に比べて高い性能を示すことを紹介する.

1　背景

最適輸送問題の数学定式化及び, その凸正則化について述べる.

1.1　最適輸送問題

本研究では主に離散版の最適輸送問題について扱う。最適輸送問題は供給量 $\frac{1}{m}$ を持つ m 個の供給地から需要量 $\frac{1}{n}$ の n 個の需要地に最も輸送コストが小さくなるように運ぶ輸送方法を考える線形計画問題の一種である. 確率分布 P_x と P_y から得られた二つのサンプルセット $\{\boldsymbol{x}_i\}_{i=1}^m$ と $\{\boldsymbol{y}_j\}_{j=1}^n$ を考える. 対応する経験分布をそれぞれ $\hat{P}_x := \frac{1}{m}\sum_{i=1}^m \boldsymbol{x}_i\delta_{\boldsymbol{x}_i}$,$\hat{P}_y := \frac{1}{n}\sum_{i=1}^n \boldsymbol{y}_i\delta_{\boldsymbol{y}_i}$ と書くことにする. ここで, $\delta_{\boldsymbol{x}}$ は $\boldsymbol{x}$ におけるディラック関数である. $\boldsymbol{\gamma} \in \mathbb{R}_+^{m\times n}$ を γ_{ij} が $\boldsymbol{x}_i$ と $\boldsymbol{y}_j$ の間の距離を表すような距離行列と定義する. さらに, 輸送行列 $\boldsymbol{\pi} \in \{\boldsymbol{\Pi} \in \mathbb{R}_+^{m\times n} | \boldsymbol{\Pi}\mathbf{1}_n = \frac{\mathbf{1}_m}{m}, \boldsymbol{\Pi}^\top\mathbf{1}_m = \frac{\mathbf{1}_n}{n}\} =: \mathcal{G}(\frac{\mathbf{1}_m}{m}, \frac{\mathbf{1}_n}{n})$ を定義する。$\mathcal{G}(\frac{\mathbf{1}_m}{m}, \frac{\mathbf{1}_n}{n})$ をカップリング制約と呼ぶ。二つの経験分布 $\hat{P}_x$ と $\hat{P}_y$ の間の最適輸送問題は以下のとおりである [1].

$$\mathrm{OT}(\hat{P}_x \| \hat{P}_y) \quad := \quad \min_{\boldsymbol{\pi}\in\mathcal{G}(\frac{\mathbf{1}_m}{m}, \frac{\mathbf{1}_n}{n})} \sum_{i,j} \pi_{ij}\gamma_{ij}. \tag{1}$$

表記を簡潔にするために,

$$\langle \boldsymbol{\pi}, \boldsymbol{\gamma} \rangle := \sum_{i,j} \pi_{ij}\gamma_{ij} \tag{2}$$

と書くことにする.

1.2　最適輸送問題の凸正則化

ここではまず凸正則化最適輸送問題の定式化を行う. 次に、その解を得ることはカップリング制約を無視した凸正則化最適輸送問題の解をカップリング制約を満たす行列集合上へブレグマンダイバージェンスを最小化するような射影演算を行うことに対応することを示す.

凸正則化最適輸送問題は以下のように (1) をルジャンドル型 [2] の関数で正則化したものである:

$$L_\phi(\boldsymbol{\pi}) := \min_{\boldsymbol{\pi}\in\mathcal{G}(\frac{\mathbf{1}_m}{m}, \frac{\mathbf{1}_n}{n})} \langle \boldsymbol{\pi}, \boldsymbol{\gamma} \rangle + \lambda\phi(\boldsymbol{\pi}). \tag{3}$$

ここで, $\lambda > 0$ は正則化パラメーターであり, ϕ は行列の要素ごとに適用される.

ここで一度, カップリング制約を無視した場合の (3) を考える:

$$\min_{\boldsymbol{\pi}\in\mathbb{R}^{m\times n}} \langle \boldsymbol{\pi}, \boldsymbol{\gamma} \rangle + \lambda\phi(\boldsymbol{\pi}). \tag{4}$$

$\langle \boldsymbol{\pi}, \boldsymbol{\gamma} \rangle$ は π に関して線形で, ϕ は狭義凸なので (4) には大域最適解が存在し, 一次最適性条件

$$\boldsymbol{\gamma} + \lambda\nabla\phi(\boldsymbol{\xi}) = 0 \tag{5}$$

を解くことで、大域最適解である

$$\boldsymbol{\xi} = \nabla\psi(-\boldsymbol{\gamma}/\lambda) \tag{6}$$

が得られる. ここで, ψ は $\nabla\phi = (\nabla\psi)^{-1}$ を満たす ϕ のフェンシェル双対である. すると,

$$\begin{aligned} \boldsymbol{\pi}_\lambda^* &:= \operatorname*{argmin}_{\boldsymbol{\pi}\in\mathcal{G}(\frac{\mathbf{1}_m}{m}, \frac{\mathbf{1}_n}{n})} L_\phi(\boldsymbol{\pi}) && (7) \\ &= \operatorname*{argmin}_{\boldsymbol{\pi}\in\mathcal{G}(\frac{\mathbf{1}_m}{m}, \frac{\mathbf{1}_n}{n})} B_\phi(\boldsymbol{\pi}\|\boldsymbol{\xi}) && (8) \end{aligned}$$

であり, 最後の等式は

$$\langle \boldsymbol{\pi}, \boldsymbol{\gamma} \rangle + \lambda\phi(\boldsymbol{\pi}) - \lambda\phi(\boldsymbol{\xi}) - \langle \boldsymbol{\xi}, \boldsymbol{\gamma} \rangle = \lambda B_\phi(\boldsymbol{\pi}\|\boldsymbol{\xi}) \tag{9}$$

に依っていて,

$$B_\phi(\boldsymbol{x}\|\boldsymbol{y}) := \phi(\boldsymbol{x}) - \phi(\boldsymbol{y}) - \langle \boldsymbol{x} - \boldsymbol{y}, \nabla\phi(\boldsymbol{y}) \rangle, \tag{10}$$

をブレグマンダイバージェンスと呼ぶ. シンクホルンアルゴリズム [4] は正則化項として負のボルツマンシャノンエントロピー [5] を用い, 射影をカルバックライブラーダイバージェンスの下で行う.

2　外れ値に頑健な凸正則化最適輸送問題

ここでは提案法のアルゴリズムの概要とその理論解析について述べる.

2.1　β-ポテンシャル正則化及びアルゴリズムの概要

本研究では (3) の正則化項として, β-ポテンシャル ($\beta > 1$) $\phi(\pi) = \frac{1}{\beta(\beta-1)}(\pi^\beta - \beta\pi + \beta - 1)$ を使用し, 射影をロバスト統計学で発展した β-ダイバージェンス [3] の下で射影を行うことを考える, こうすることでデータに外れ値が含まれていても頑健に凸正則化最適輸送問題を解く. アルゴリズムを Algorithm 1 に示した. カップリング制約を満たす凸集合を $\mathcal{C}_0 = \mathbb{R}_+^{m\times n}$, $\mathcal{C}_1 = \left\{\boldsymbol{\pi} \in \mathbb{R}^{m\times n} | \boldsymbol{\pi}\mathbf{1}_n = \frac{\mathbf{1}_m}{m}\right\}$, $\mathcal{C}_2 = \left\{\boldsymbol{\pi} \in \mathbb{R}^{m\times n} | \boldsymbol{\pi}^\top\mathbf{1}_m = \frac{\mathbf{1}_n}{n}\right\}$ に分解し, それぞれへの射影を交互に繰り返すことで $\boldsymbol{\pi}_\lambda^*$ を得る. 行 2, 7, 11 が $\mathcal{C}_0$ への射影, 行 4–6 が $\mathcal{C}_1$ への射影, 行 8–10 が $\mathcal{C}_2$ への射影に対応している.

Algorithm 1 β-ダイバージェンス ($\beta > 1$) を用いた交互スケーリングアルゴリズム

1: $\tilde{\boldsymbol{\theta}} \leftarrow -\boldsymbol{\gamma}/\lambda$
2: $\boldsymbol{\theta}^* \leftarrow \max\{\nabla\phi(\mathbf{0}_{m\times n}), \tilde{\boldsymbol{\theta}}\}$
3: **repeat**
4: $\quad \boldsymbol{\tau} = \frac{\nabla\psi(\boldsymbol{\theta}^*)\mathbf{1}_n - \frac{\mathbf{1}_m}{m}}{\nabla^2\psi(\boldsymbol{\theta}^*)\mathbf{1}_n}$
5: $\quad \boldsymbol{\tau} \leftarrow \max(\boldsymbol{\tau},\ \hat{\boldsymbol{\theta}}^* - \nabla\phi(\frac{\mathbf{1}_m}{m}))$
6: $\quad \tilde{\boldsymbol{\theta}} \leftarrow \tilde{\boldsymbol{\theta}} - \boldsymbol{\tau}\mathbf{1}_n{}^\top$
7: $\quad \boldsymbol{\theta}^* \leftarrow \max\{\nabla\phi(\mathbf{0}), \tilde{\boldsymbol{\theta}}\}$
8: $\quad \boldsymbol{\sigma} = \frac{\mathbf{1}_m{}^\top\nabla^2\psi(\boldsymbol{\theta}^*) - (\frac{\mathbf{1}_n}{n})^\top}{\mathbf{1}_m{}^\top\nabla^2\psi(\boldsymbol{\theta}^*)}$
9: $\quad \boldsymbol{\sigma} \leftarrow \max(\boldsymbol{\sigma}, \hat{\boldsymbol{\theta}}^* - \nabla\phi(\frac{\mathbf{1}_n}{n}))$
10: $\quad \tilde{\boldsymbol{\theta}} \leftarrow \tilde{\boldsymbol{\theta}} - \mathbf{1}_m\boldsymbol{\sigma}^\top$
11: $\quad \boldsymbol{\theta}^* \leftarrow \max\{\nabla\phi(\mathbf{0}_{m\times n}), \tilde{\boldsymbol{\theta}}\}$
12: **until** convergence
13: $\boldsymbol{\pi}^* \leftarrow \boldsymbol{\nabla\psi(\theta^*)}$

表 1: 正しく正常データ/外れ値データを検出できた割合. 50 回の実験結果.

	Outliers	Inliers
One-class SVM [8]	49.8 ± 1.8 %	50.0 ± 0.1 %
Local outlier factor [9]	49.2 ± 3.8 %	**99.2 ± 0.1 %**
Isolation forest [10]	50.7 ± 10.2 %	65.5 ± 4.4 %
Elliptical envelope [11]	79.9 ± 7.0 %	80.0 ± 4.6 %
MoM-based[12]	79.7 ± 12.6 %	**98.9 ± 0.7 %**
Baseline technique	99.0 ± 0.6 %	84.8 ± 0.5 %
ROBOT [13]	**99.5 ± 0.3 %**	84.8 ± 0.5 %
提案法	99.1 ± 0.7 %	87.3 ± 0.5 %

2.2 理論解析

本アルゴリズムの理論解析を以下に述べる. Agorithm 1 の行 3–12 を「大ループ」と呼ぶことにする. すると, 以下の命題が成り立つ.

命題 1 任意の z $(> \frac{\lambda}{\beta-1})$ について, $J \subseteq \{1, \ldots, n\}$ を以下を満たすような集合とする.

$$\forall j \in J,\ \forall i \in \{1, \ldots, m\}, \gamma_{ij} \geq z. \tag{11}$$

大ループを t 回繰り返して得られた輸送行列を $\boldsymbol{\pi}^{\text{output}}$ が条件

$$\frac{(\frac{1}{m})^{\beta-1} + (\frac{1}{n})^{\beta-1}}{\beta - 1} t < \frac{1}{1-\beta} + \frac{z}{\lambda} \tag{12}$$

を満たすと,

$$\forall i, \pi_{ij} = 0 \quad \text{if} \quad j \in J \tag{13}$$

が満たされる.

命題 1 は直観的には, 条件 (12) が満たされれば, 距離が z 以上のデータ点には確率質量が送られないことを意味している.

3 実験

提案手法は外れ値検知に応用できる。Fashion-MNIST [6] の画像データ 9500 個（正常データ）と MNIST [7] 画像データ 500 個（異常値データ）からなるデータセットから外れ値データを検出する実験を行う. 本実験では, これとは別に Fashion-MNIST10000 個からなるデータセットがあると仮定し、それを二つに分割して距離行列を計算し, $z = \max_i \min_j \gamma_{ij}$ を求める. これは直観的には, 分割された二つのうち, 片方のデータセットの「各データ点の最近接点」の最大値に対応する. この z を基準とし, z 以上離れているデータを外れ値とみなす. この z を用いて, 命題 1 が満たされるように種々のハイパーパラメータ t, β, λ を設定した. 我々の提案法は既存の外れ値検出アルゴリズムや最適輸送問題を頑健に計算するアルゴリズムと比べても性能が高いことがわかる (表 1).

4 結論

本研究では,（離散）最適輸送問題を β-ポテンシャルで正則化することでシンクホルンアルゴリズムの行う射影演算を頑健にする手法を提案した. 外れ値に確率質量を送らないことができ、そのための理論保証を与えた. 提案手法の応用例として外れ値検知を紹介し, 既存手法に比べて高い性能を示すことが分かった.

参考文献

[1] G. Peyré and M. Cuturi, Computational Optimal Transport, ArXiv:1803.00567, 2018.

[2] H.H. Bauschke and J.M. Borwein. Legendre functions and the method of random bregman projections. Journal of Convex Analysis, 1997.

[3] A. Basu, I.R. Harris, N.L. Hjort, and M.C. Jones. Robust and efficient estimation by minimising a density power divergence. Biometrika, 85(3):549–559, 1998.

[4] M. Cuturi, O. Teboul, and J. Vert. Differentiable ranking and sorting using optimal transport. In NeurIPS, pp. 6858–6868, 2019.

[5] A. Dessein, N. Papadakis, and J. Rouas. Regularized optimal transport and the ROT mover's distance. Journal of Machine Learning Research, 19(15):1–53, 2018.

[6] H. Xiao, K. Rasul, and R. Vollgraf. Fashion-mnist: A novel image dataset for benchmarking machine learning algorithms. CoRR,abs/1708.07747, 2017.

[7] L. Deng. The MNIST database of handwrittendigit images for machine learning research. IEEE Signal Processing Magazine, 29(6):141–142, 2012.

[8] B. Schölkopf, R.C. Williamson, A.J. Smola, J.Shawe-Taylor, and J.C. Platt. Support vector method for novelty detection. In NIPS, pp. 582–588. 1999.

[9] M.M. Breunig, H. Kriegel, R.T. Ng, and J. Sander. LOF: Identifying density-based local outliers. In Proceedings of the 2000 ACM SIGMOD International Conference on Management of data, pp. 93-104, 2000.

[10] F.T. Liu, K.M. Ting, and Z. Zhou. Isolation forest. In 2008 Eighth IEEE International Conference on Data Mining, pp. 413–422, 2008.

[11] P.J. Rousseeuw and K.V. Driessen. A fast algorithm for the minimum covariance determinant estimator. Technometrics ,41(3):212-23 pp. 212–223, 1999.

[12] G. Staerman, P. Laforgue, P. Mozharovskyi, and F. d'AlcheBuc. When OT meets MoM: Robust estimation of Wasserstein distance. In AISTATS, pp.136-144, 2021.

[13] Debarghya Mukherjee, Aritra Guha, and Justin Solomon. Outlier-robust optimal transport. In International Conference of Machine Learning, pp. 7850–7860 2020.