機械学習における最適輸送のための緩和最適化とグラフ処理への応用の研究

Relaxed optimization methods and graph applications for optimal transport in machine learning

笠井 裕之(Hiroyuki KASAI, Dr. Eng.)
早稲田大学 基幹理工学部 情報通信学科 教授
(Professor, Waseda University, Department of Communications and Computer Engineering, School of Fundamental Science and Engineering)
IEEE 電子情報通信学会 情報処理学会 応用数理学会 他
受賞 : 安藤博記念学術奨励賞 (2000 年). 丹羽記念賞 (2001 年). エリクソン・ヤング・サイエンティスト・アワード (2001 年). 山下記念研究賞 (2003 年) 他
研究専門分野 : 機械学習、最適化理論、信号処理

あらまし

確率分布間の距離を表現可能な最適輸送理論は、近年、機械学習や統計的学習、信号処理等の分野で大きな注目を集めている。最適輸送問題は数理計画問題として定義されるが、その解を求めるには高い計算量が必要なことから、高速かつ柔軟な最適化手法への期待は高い。また、最適輸送問題の応用は広範囲に及ぶが、特に非構造データへの適用とその応用は極めて重要である。そこで本研究では、最適輸送問題の高速な求解手法の確立に向けて、緩和最適輸送問題のための制約緩和最適化手法、および複数最適輸送問題のための部分空間基底に基づく最適化手法を研究した。また、最適輸送問題のグラフ問題への応用について研究した。本稿では、それらの概要について報告する。

1. 研究の背景と目的

近年、機械学習の分野で精力的に研究が進められている最適輸送理論は、確率分布間の距離を計算することで、様々な種類のデータ間の「違い」を定量的に表現することができる。直感的には、最適輸送問題は、確率分布の土塊を地点 A から地点 B へ運ぶ際の最適な輸送計画 T を求める問題として捉えられ、ここで得られる最適な輸送計画 T によりデータ間の距離を定義することができる。この問題を定式化する上で注目すべきは、地点 A、B ともに条件として与えられた土塊量を満たす必要があるだけでなく、輸送に際して土塊の一部が地点 A に残る、また地点 A で存在していた以上の、あるいは以下の土塊が地点 B に輸送されることは許容されない、といった条件を考慮する点にある(図 1 左)。本研究では、これらの条件を質量保存条件と呼ぶことにする。よって、この最適化問題は質量保存条件を多面体制約条件で記述した「構造制約付き最適化問題」として定式化される。このように定義される最適輸送問題は、既存の線形計画問題*1 用の最適化手法により解くことが可能であるが、その計算量はデータ次元の三乗に比例して増加する。そこで、高速な最適化手法が多数提案されており、例えば、エントロピー*2 条件を追加したシンクホーン・アルゴリズムが広く注目を集めている。

本研究では、さらなる高速化と適用問題への柔軟性の向上を目指し、最適化問題における制約条件を緩和した緩和最適輸送問題に着目し、高速な制約緩和最適化手法を研究した。一方、機械学習の問題では、複数の最適輸送問題を同時に解く場合が頻繁に生じる。一般に、輸送対象となるデータの確率分布サイズは異なることが想定されるため、このような複数の最適輸送問題を効率的に計算する最適化手法は存在しない。そこで本研究では、部分空間の基底学習と、学習基底を用いたバッチ処理による高速最適化手法を研究した。最後に、非構造データであるグラフデータを対象として、グラフ分類問題やグラフ表現問題に対する最適輸送問題の応用について研究した。

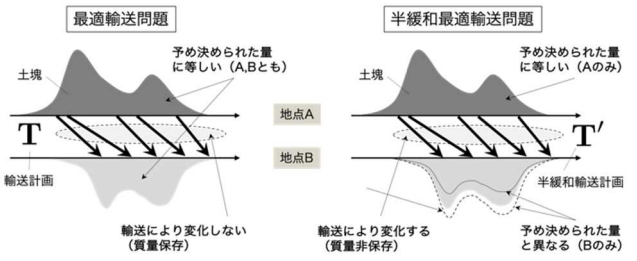

図 1 最適輸送問題と半緩和最適輸送問題

機械学習における最適輸送のための緩和最適化とグラフ処理への応用の研究

Relaxed optimization methods and graph applications for optimal transport in machine learning

３．研究の方法

3.1 制約緩和による最適化手法

最適輸送問題を実問題へ応用する際、質量保存条件が厳密に満たされる必要が無い場合では、かえって性能劣化を引き起こすことが報告されている。このため、質量保存条件を緩めた「緩和最適輸送問題」が研究されている。

本研究では、質量保存条件の内、片方（地点 B 側）の条件のみ緩めた「半緩和最適輸送問題」（図 1 右）に対して、座標毎に勾配法*3 を適用する高速化手法を研究し、その理論的収束レートを示した［1]。さらに、前述のシンクホーンアルゴリズムによる半緩和問題についても研究した［2]。後者については、目的関数値の誤差および最適解に対する誤差に関する理論的収束レート解析*4 を行い、最適輸送問題と比較して大幅な理論的性能向上を達成した。特に注目すべきは、質量保存条件の緩和問題において重要となる「質量保存誤差」についての収束レートを導出した点にある。この理論的成果により、アルゴリズムの停止時に、保存条件から質量がどれほど乖離しているかの最悪値を見積もることが可能となる。そして、応用問題への適用時に質量保存条件の緩和の程度を調節することも可能となる。

3.2 部分空間基底に基づく最適化手法

複数の最適輸送問題を同時に解く問題を考える。例えば、多数のグラフデータから似たようなグラフの集合を抽出する、観測された複数の時系列データと過去に観測されたデータを対象として類似のデータを検索しグループ化する、などがその例となる。このような問題において、複数のデータの全ての組み合わせに対する相互間距離を最適輸送により求めることを考える。

一般に、最適輸送問題では、与えられた特徴空間における確率分布は未知である場合が多い。そのため、計算においては、これらの確率分布を各点に対する一様分布として計算することが一般的である。これにより、二つのデータ間の比較を二つの既知の確率分布に関する問題とすることができる。一方で、離散分布のサイズが異なる場合は、GPU による並列化処理の適用は制限される。具体的には、異なるサイズの分布に起因してコスト行列も異なるサイズを有しているため、バッチデータとして GPU に入力することが困難となる。

本研究では、複数データに共通する部分空間上の基底を用いた最適化手法について研究した[3]。具体的には、最初に、輸送対象の元データの分布から部分空間の基底ベクトルを学習する。次に、求めた基底ベクトルを用いて元データを表現し、元々異なるサイズであった確率分布を固定サイズの分布に変換する。最後に、変換された分布間の最適輸送問題に対して GPU 並列計算処理を行う。本手法では、部分空間の基底ベクトルのサイズは常に一定であるため、前述したシンクホーン・アルゴリズムにより GPU 並列化を導入することが可能となり、距離計算をバッチ処理により高速化できる。表 1 は、ベンチマークであるグラフ・データセットである NCI 1 と ZINC を用いた距離計算における従来手法（詳細は省略）と提案手法の処理時間の比較を示している。表 1 から、提案手法は CPU 利用時においても高速化を実現しており、さらに GPU による著しい高速化を実現していることが分かる。以上から、提案手法の有効性は明らかである。

表 1 平均輸送処理時間（単位は[sec]）。提案手法は基底学習時間と輸送処理時間をそれぞれ表している。

手法	データセット	
	NCI1	ZINC
OT-EMD (CPU)	1717	2566
SW (CPU)	4512	5190
eOT (CPU)	18574	29090
BDS-eOT (GPU)	3970	5703
提案手法 (CPU)	7.09+1259	0.41+1458
提案手法 (GPU)	7.09+18.30	0.41+16.60

3.3 グラフ問題への適用

最適輸送問題は様々な機械学習問題へ応用されている。本研究では、グラフ学習問題への応用研究を行っ

機械学習における最適輸送のための緩和最適化とグラフ処理への応用の研究

Relaxed optimization methods and graph applications for optimal transport in machine learning

た。グラフ構造データは、ノード情報とエッジ情報から構成され、SNS 上のユーザー間のつながり、化学物質等を表現することが可能である。例えば、化学物質をグラフ構造データで表現し、そのデータ間の距離を定義することで、構造的に近い化学物質を発見でき創薬開発に寄与できる。さて、深層学習モデルはグラフ構造データには直接適用できないことから、グラフニューラルネットワーク（Graph Neural Network: GNN）*5 が注目されている。グラフ学習分野には、ノード分類、エッジ予測、グラフ分類の代表的なタスクがあるが、GNN は前者 2 つのタスクについて優れた性能を示すことが知られている。一方で、グラフ分類タスクについては、従来法よりも不安定で精度が低いことが報告されている。

このような背景を受けて、本研究ではグラフ構造データ間の距離を最適輸送問題により定義する研究を行った。最初に、グラフ上の探索手法から得られる走査パス比較手法の改良に基づくグラフ間距離を提案した[4]。具体的には、ノードとエッジから構成される系列の最長共通シーケンス集合を用いて最短パスを比較する手法を定義し、従来手法では不可能であった柔軟なパス間比較を実現した。そして、このように得られるパスの集合を離散分布と見なし、最長共通シーケンス要素間の非類似性を基底距離とする最適輸送問題を定義した。グラフ分類精度の数値実験から、提案手法が従来のカーネル法*6 と同等以上の性能を持つことを示した。

次に、微小な構造変化（ノードやエッジの挿入・削除・置換等）で経路が大きく変動する従来のグラフ探索手法の問題に着目し、安定した方法でグラフ構造を記述する手法を提案した 。具体的には、グラフカーネル分野で最も代表的な Weisfeiler-Lehman（WL）*7 に着目し、その解釈に用いる WL の部分木の概念をノード周辺の構造情報に適用した。WL 部分木間の木編集距離*8 を計算することでノード間の距離を定義した。また、高速に計算可能な近似距離を提案し、その高速アルゴリズムと理論的保証を与えた。さらに、提案した木編集距離を基底距離とする最適輸送問題を提案した。数値実験では、二つの同一グラフの一方だけにノイズを加えた際のグラフ間距離の挙動を評価した。その結果、従来手法では微小な構造変化に対して急激な増加が見られるのに対して、提案手法では全体的に滑らかな増加曲線を示すことを確認し、提案方式がグラフ構造の微小変化をグラフ間距離として効率的に表現可能であることを示した（図 2）。またグラフ分類精度の数値実験からも、提案手法が最先端のグラフカーネル手法と同等以上であることを示した。

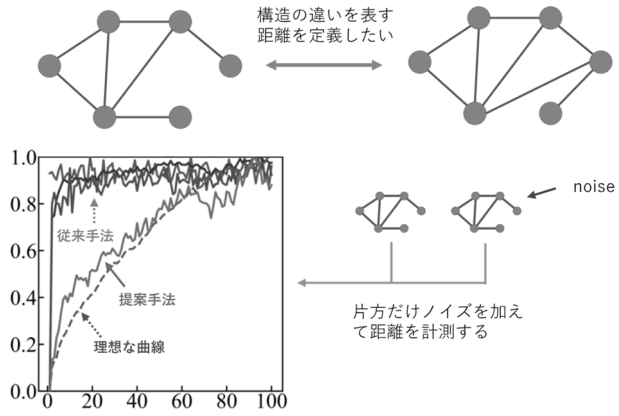

図 2　WL 部分木を用いた手法の距離識別性能

４．将来展望

応用数学分野に端を発する最適輸送理論は、現在、機械学習分野の重要な研究テーマの一つであり極めて広範な応用分野を有する。一方で、厳密に計算することは、データ次元の拡大に伴い困難を伴う。従って、より優れた最適化手法を開発するとともに、その特性を理論的また数値的に評価していくことは意義がある。今後は、その発展の一端を担えるよう研究を進めていく。

用語解説

*1 線形計画問題
目的関数と制約条件が 1 次式で表現される最適化問題。

*2 エントロピー
情報の乱雑さや不確実さを表す量。

*3 勾配法
ある初期点から開始し、現在の点から勾配情報により定まる移動方向に伸びる半直線に沿って、点を反復的に更新していくことで最適解を求める最適化

手法。

*4 収束レート解析
ある誤差を許容する最適解へ到達するための収束回数やそのための計算複雑度を推定する解析。

*5 グラフ・ニューラルネットワーク
グラフデータを扱うためのニューラルネットワークの一種。

*6 カーネル法
パターン認識において使われる手法の一つで、カーネル関数を用いて判別などのアルゴリズムに組み合わせて利用する手法。

*7 Weisfeiler-Lehman（WL）法
異なるグラフ間の構造の一致性を判定する代表的な手法の一つ。

*8 木編集距離
木構造を有するデータのノードとエッジを挿入・削除・置換の 3 つの操作で木 B に変換するための操作数。

参考文献

[1] T. Fukunaga and H. Kasai, “Block-coordinate Frank-Wolfe algorithm and convergence analysis for semi-relaxed OT problem,” IEEE 47th International Conference on Acoustics, Speech, and Signal Processing (ICASSP), 2022.

[2] T. Fukunaga and H. Kasai, “On the Convergence of Semi-Relaxed Sinkhorn with Marginal Constraint and OT Distance Gaps,” arXiv preprint:arXiv:2205.13846, 2022.

[3] J. Huang, X. Sun, Z. Fang, and H. Kasai, “Anchor Space Optimal Transport: Accelerating Batch Processing of Multiple OT Problems,” arXiv preprint:arXiv: 2310.16123, 2023.

[4] J. Huang, Z. Fang, and H. Kasai, “LCS graph kernel based on Wasserstein distance in longest common subsequence metric space,” International Journal of Signal Processing, Elsevier, Vol.189, 2021.

[5] Z. Fang, J. Huang, X. Sun, and H. Kasai, “Wasserstein Graph Distance based on L_1-Approximated Tree Edit Distance between Weisfeiler-Lehman Subtrees,” Thirty-Seventh AAAI Conference on Artificial Intelligence (AAAI-23), 2023.

関連文献

C. Villani, “Optimal Transport: Old and New,” Springer, New York, 2008.

G. Peyré and M. Cuturi, “Computational Optimal Transport: With Applications to Data Science,” Foundations and Trends® in Machine Learning: Vol. 11: No. 5-6, pp 355-607.

この研究は、令和元年度ＳＣＡＴ研究助成の対象として採用され、令和２～４年度に実施されたものです。