

巨大グラフとビッグデータ解析の基礎基盤：理論研究と高速アルゴリズム開発

Large Graphs: Theory and Algorithms

課題番号：18H05291

河原林 健一（KAWARABAYASHI KEN-ICHI）

国立情報学研究所・プリンシプル研究系・教授



研究の概要（4行以内）

研究提案は、提案者グループが主体となった「ERATO 河原林巨大グラフプロジェクト」の活動を引継ぎ、理論分野のさらなる強化と、理論分野の道具を利用によるアルゴリズムの高速化・スケール化に挑み、かつアルゴリズム研究における世界的拠点の構築も目指すものである。

研究分野：理論計算機科学、組合せ最適化

キーワード：グラフ、アルゴリズム

1. 研究開始当初の背景

現代の情報化社会が抱える大部分の問題は、センサー、画像、文書、音声などによって収集された多種類の大量のデータの解析、そして情報処理技術によって解決されることが期待されている。しかしながら、データ量が膨大であるため、HPCを使用しても解決が容易でないものばかりである。このような問題を解決するためには、アルゴリズムの革新が必要不可欠であり、計算モデルと数理の探求に基盤をおく革新的アルゴリズム設計技法の構築や体系化は、科学の共通基盤として最優先の意義を持つ。

実際、数学的理論に基づくアルゴリズムは、現代社会の進歩を加速度的に後押ししてきた。特に現在の情報検索、ゲノム情報処理などのアルゴリズム革新は国家規模のビジネス創成につながっている。ITビジネスの発展の歴史の中でも、Microsoft、IBM、Google、Yahoo、AT&T、Facebook、Amazonなどの巨大IT企業で、著名な理論研究者が斬新なソフト開発とさまざまな問題解決に貢献してきた。

2. 研究の目的

本研究では、以上の背景のもと、数学的理論を駆使することにより、アルゴリズムの理論分野（おもにグラフアルゴリズム）の強化および、理論分野の道具を利用によるアルゴリズムの高速化・スケール化に挑む。

3. 研究の方法

以下の3点の研究課題を中心にする予定である。

A. 劣モジュラ関数とその応用

劣モジュラ性は普遍的な概念であり、これまでに扱われてきた機械学習、人工知能分野だけではなく、自然言語、コンピュータビジョンにも応用されている。本提案では、近似アルゴリズム設計手法や代数的手法などさまざまな組合せ最適化手法を取り入れることでロバスト最適化（最悪時の解の質を担保する）などの、実社会に出現する最適化問題の解決に取り組む。

B. 基礎数理理論の探求：有向グラフマイナー理論

グラフ理論における最も重要な理論体系「グラフマイナー理論」を用いた効率的なアルゴリズムの設計は、従来のアルゴリズム理論をはるかに深化させることが過去20年で明らかになってきた。しかしながら従来のグラフマイナー理論およびそれに基づくアルゴリズムは、すべて無向グラフにおけるものであり、有向グラフで同様の議論を展開することは困難であると考えられてきた。本研究では、有向グラフ版グラフマイナー理論の構築をめざす。

C. 基礎理論分野の自然言語、深層学習、データマイニング等への応用

本研究は、理論研究中心であるものの、理論研究の応用も視野にいれている。

4. これまでの成果

グラフカット、連結問題における貢献

グラフの連結度を求める問題は、東西冷戦時代（つまり1950年代）より組合せ最適化における中心的課題の一つである。以下の論文

では、辺連結度に関して、初の「決定的」な「ほぼ」線形アルゴリズムを与えた。辺連結度に関する「非決定的」な「ほぼ」線形アルゴリズムは、Karger により2000 年に開発されていたが、「決定的」アルゴリズムは長年未解決であった。本論文はその最終的な解決を与えている。本論文は、コンピューターサイエンス分野の最高峰の国際学術雑誌 Journal of the ACM (J. ACM) に掲載されている。

向き付きグラフマイナー理論

RobertsonとSeymourは「Graph Minors」という共通のタイトルを持つ一連の論文（以下、グラフマイナー論文）で（向きがない）グラフの「グラフマイナー理論」を構築した。（向きがない）グラフ上で定義された「木幅」に関する研究は、80 年代の離散数学で最も深淵とされているこの「グラフマイナー理論」の発展とともに、近年のアルゴリズム分野・離散数学分野での中心的課題となってきた。しかしながら、グラフマイナー理論の向き付きグラフへの展開は、長年未解決であった。その第一歩として、Kreutzer氏と河原林は、1990 年代中盤にReed, Robertson, Seymour, Thomas などの著名な数学者によって予想された「向き付きグラフの木幅とグリッドマイナーのMin-Max 予想」を2015年に完全解決した。この結果をさらに拡張し「グラフマイナー理論」の向き付きグラフへの展開を以下で行った。

- A. 平面グラフ上での結果
- B. 弱「構造」定理を証明

劣モジュラ関数の解析

疎なモデルを得るための正則化の多くは劣モジュラ関数の Lovasz 拡張とみなすことができる。本研究ではそのような正則化項を持つ線形回帰に対する階層的確率モデルを考え、事後予測に対する変分推論手法を与えた。具体的には分配関数に対する理論的な上限を与え、この上限を計算する多項式時間アルゴリズムを与えた。我々の手法はMAP 推定のためのベイジアン Lasso モデルの拡張とみなすことができる。

5. 今後の計画

現時点で得られている研究成果は、3 つに大きく分けられる。1. 離散アルゴリズム、2. 機械学習分野における組合せ最適化、3.

AI、自然言語、データマイニング、深層学習への応用。

今後は、1 に関しては、分散計算においても、当該分野の最大の研究課題である「マッチング問題」「独立点問題」に貢献したいと考えている。

2 に関しては、引き続きオンライン最適化問題やバンディット組合せ最適化問題に対して、様々な組合せ的制約を取り入れ、最適なリグレットを達成する効率的なアルゴリズムの開発を行なう。

3 に関しては、引き続き、グラフニューラルネットワークで正確に学習ができる条件、あるいは入力データの条件などを解析していく予定である。

6. これまでの発表論文等（受賞等も含む）

Ken-ichi Kawarabayashi, Mikkel Thorup, Deterministic Edge Connectivity in Near-Linear Time, Journal of the ACM, 66(1), 4:1-4:50 (2019)

Ken-ichi Kawarabayashi, Anastasios Sidiropoulos, Polylogarithmic approximation for Euler genus on bounded degree graphs, Proceedings of the 51st Annual ACM Symposium on Theory of Computing(STOC), 164-175 (2019)

Bingkai Lin, A Simple Gap-Producing Reduction for the Parameterized Set Cover Problem, Proceedings of the 46th International Colloquium on Automata, Languages, and Programming (ICALP 2019), 81:1-81:15 (2019) (Best Paper Awards)

Ken-ichi Kawarabayashi, Bingkai Lin, A nearly $5/3$ -approximation FPT Algorithm for Min-k-Cut, Proceedings of the 2020 ACM-SIAM Symposium on Discrete Algorithms(SODA), 990-999 (2020)

Shinji Ito, Daisuke Hatano, Hanna Sumita, Kei Takemura, Takuro Fukunaga, Naonori Kakimura, Ken-ichi Kawarabayashi, Oracle-Efficient Algorithms for Online Linear Optimization with Bandit Feedback, The 33rd Conference on Neural Information Processing Systems (NeurIPS2019), 10589-10598 (2019)

7. ホームページ等

https://bigdata.nii.ac.jp/wp/k_keniti@nii.ac.jp