

加藤 直樹

関西学院大学理工学部
教授

ビッグデータ時代に向けた革新的アルゴリズム基盤

§ 1. 研究成果の概要

今世紀に新たに注目を浴びているビッグデータは、そのデータ量の膨大さ故に、その基礎となるアルゴリズムのモデルは根本的な変革を迫られている。例えば、これまで速いと考えられてきたアルゴリズムが、ペタスケールのビッグデータに対して計算資源や実行時間などの点で大きな困難が伴い、少なくとも線形、場合によっては劣線形時間や定数時間アルゴリズムが求められる。

そのために、本研究では、ビッグデータ時代に向けた新しい計算パラダイムとして「**劣線形時間パラダイム**」を提唱し、その中で、ビッグデータ用のアルゴリズムとデータ構造、およびモデリング技法を提案する。本研究は、劣線形時間アルゴリズムグループ、劣線形データ構造グループ、劣線形モデリングの3つの研究グループから構成されている。以下では、本年度の成果として学術的、および社会的にも注目を浴びた研究の概要を紹介する。

劣線形時間アルゴリズムグループは、組合せ剛性理論によるたんぱく質機能解析について重要な成果を得た。Sljoka, 土村らは、A2A アデノシンと呼ばれる GPCR (G たんぱく質共役受容体) におけるアロステリー信号伝達の分析をおこなった。現代の創薬の 50% 近くは、GPCR が関与していることから、アロステリー信号伝達の役割を解明することはとりわけ重要である。具体的には、アロステリーによる剛性伝達を調べるアルゴリズムを用いて、カルシウムやマグネシウムがアロステリー信号伝達を加速化させる作用があり、GPCR の一つである A2A 受容体の活性化を促進させる効果があることを明らかにした。さらに、ナトリウムは逆に A2A 受容体の活性化を低減させることも明らかにした。これらの成果は、これまで実証できなかった、GPCR におけるアロステリー信号伝達の存在が組合せ剛性理論に基づくアロステリー伝達分析アルゴリズムによってはじめて明らかになった重要な発見である。また、本プロジェクトの中心的課題の一つであるビッグデータ向け定数時間アルゴリズムの開発について、理論と実用の両面において重要な成果を得た。特に、伊藤は、疎グラフの代表的モデルである次数制限モデルに対し、単調な性質と遺伝的な性質について、片側誤りで定数時間検査可能な性質の特徴付けに成功した。さらに、伊藤らは、任意の定数時間アルゴリズム

ムと全データを読み込む線形以上の計算時間のアルゴリズムが存在するとき、その両者の計算時間をオーダーの意味で悪化させずに漸進型アルゴリズムを構築できることを証明した。

劣線形データ構造グループでは、劣線形データ構造研究の成果として、2つの新たな圧縮技術の開発に成功した。1つは、喜田らによる繰り返し構造が頻出するビッグデータに対応した超高压縮技術である。ビッグデータとされる様々なデータには、Wikipedia 履歴データなど大量の過去からの履歴を含むデータや、繰り返しの多いセンサーデータなど、通常よりもはるかに繰り返しの多いデータが含まれる。そのような繰り返しの多いデータはいかなる手法を用いても通常のデータと比べると高压縮率で圧縮することができるが、喜田らは、それらのデータの持つ性質を活用する新たな文法圧縮技術を開発することに成功し、当該手法が従来手法より必ず同等以上の性能を持つことを理論的に示すとともに、現実にも繰り返しの多いデータに対して圧縮性能が顕著であることを実験を通し実証した。もう1つの成果は、坂本・井・後藤らによる作業領域の小さな圧縮技術の開発である。坂本・井・後藤らは、RePair とよばれる高压縮率アルゴリズムを二段階圧縮を行うことで圧縮のための作業領域を大幅に縮小することに成功した。これによって、このような高压縮率アルゴリズムを、潤沢に資源の利用が不可能であるハードウェアでの実装をターゲットとした場合にも、容易に実装できる新しいデータ圧縮アルゴリズムとして、ビッグデータ・ストリーム圧縮処理などへ応用する道筋を新たに開くことに成功した。

劣線形モデリンググループの 2018 年度の関連する主な研究成果は、①周波数空間繰り返し込み理論による情報粗視化モデルの提案と②2016～2017 年度にかけて提案してきた深層学習モデルの高速なノイズ耐性検査手法の更なる高速化、③マルコフ確率場モデルに対する高速経験ベイズ推定アルゴリズムの提案である。①は当グループの初期段階の成果を更に実用的な状況に適用可能な形へと拡張したものである。②は深層学習モデルのノイズ頑健性を測る検査アルゴリズムに関する成果であり、前年度までの成果にランダムサンプリングの概念を輸入して更なる高速化(30,000 倍以上の高速化)に成功した。③は、統計的機械学習の一手法である、経験ベイズ法に基づく学習アルゴリズムを統計力学的情報粗視化法に基づき定式化したものである。通常の機械学習では、膨大な数のデータを必要とするが、一方、本手法は入力次元数 n に対してデータ数のオーダーが線形以下で学習可能であり、状況によっては $O(1)$ のデータ数(定数オーダー)からの学習が可能となっている。

【代表的な原著論文】

1. Ye L., C. Neale, Sljoka A., Pichugin D., Tsuchimura N., Sunahara R., Prosser S. et al, Bidirectional Regulation of the A2A Adenosine G Protein-Coupled Receptor by Physiological Cations, Nature Communications, 1 9:1372, 2018.
2. I. Furuya, T. Takagi, Y. Nakashima, S. Inenaga, H. Bannai, and T. Kida, "MR-RePair: Grammar Compression Based on Maximal Repeats", In Proceedings of Data Compression Conference (DCC2019), IEEE Computer Society Press, pp.508-517, March, 2019.
3. Muneki Yasuda, Hironori Sakata, Seung-Il Cho, Tomochika Harada, Atushi Tanaka, and Michio Yokoyama: An efficient test method for noise robustness of deep neural networks, Nonlinear Theory and its Applications, IEICE (in press)

§ 2. 研究実施体制

(A)「劣線形時間アルゴリズム」グループ

- ① 研究代表者:加藤 直樹 (関西学院大学 教授)
- ② 研究項目
 - ・ビッグデータ向け定数時間アルゴリズムの実用化と効率化
 - ・漸進型アルゴリズムの開発
 - ・線形時間アルゴリズムの開発
 - ・組合せ剛性理論によるタンパク質の機能解明
 - ・避難計画問題
 - ・革新的アルゴリズム基盤の構築

(B)「劣線形データ構造」グループ

- ① 主たる共同研究者:渋谷 哲朗 (東京大学医科学研究所 准教授)
- ② 研究項目
 - ・情報論的アプローチによる劣線形データ構造に関する研究
 - ・列挙論的アプローチによる劣線形データ構造に関する研究
 - ・実応用アプローチによる劣線形データ構造に関する研究

(C)「劣線形モデリング」グループ

- ① 主たる共同研究者:田中 和之 (東北大学大学院情報学研究科 教授)
- ② 研究項目
 - ・統計力学的粗視化アプローチによる劣線形モデリング
 - ・計算理論と統計的近似理論の融合による大規模システム上での高効率計算アルゴリズムの開発