

「ビッグデータ統合利活用のための次世代基盤技術の創出・体系化」  
平成26年度採択研究代表者

H29 年度  
実績報告書

加藤 直樹

学校法人関西学院関西学院大学  
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. 研究実施体制

### (A)「劣線形アルゴリズム」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

## § 2. 研究実施の概要

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、本年度、組合せ剛性理論によるたんぱく質機能解析について、2つの重要な成果を得た。一つは、Sljoka, 土村らによるもので、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。

もう一つは、Sljoka, 土村, 加藤らによって、抗体たんぱく質の親和性成熟の背景にある生物物理学的なメカニズムに新たな知見を与える研究成果を得た(代表的な原著論文2)。抗体は抗原の出現により突然変異を繰り返しながら進化する、これまで、少数の抗体を用いて、親和性成熟の進展が抗体の構造的柔軟性を減少させるという仮説が立てられていた。我々は、数十万の抗体構造を対象に、高速な剛性判定アルゴリズムを用いて、その仮説を検証した。ナイーブな抗体と成熟した抗体における柔軟性の差異は見られず、従来の仮説を反証するものである。この結果から、新たな研究課題が生じ、現在調査中である。

劣線形データ構造グループの新たな展開として、小野寺・渋谷は、Succinct Oblivious RAMとよぶ、劣線形パラダイムをプライバシー保護技術へと展開させたデータ構造の開発に成功した(代表的な原著論文1)。ORAM(Oblivious Random Access Memory)は、データベースのアクセス時に、どのデータにアクセスを行ったかどうかを秘匿するデータ構造として、近年脚光を浴びているプライバシー保護技術である。しかしながら、これまでの既存のいずれのORAM技術を用いたとしても、実用的なアクセス速度と高い秘匿性を両立するためには、最低でも実際のデータ容量の10倍を超えるデータ格納領域が必要であり、実際にビッグデータに対して活用することは非常に困難であった。これに対し小野寺・渋谷が開発したSuccinct Oblivious RAMは、劣線形パラダイムを活用することで、通常のORAMの設定においてアクセス速度が最速だとされるPath ORAMのアクセス速度を理論的にも実用的にもほぼ落とすことなく、実際に格納に必要な容量の実データ容量比が漸

近的に1に収束する, すなわち, 十分大きなデータに対しては, データ格納オーバーヘッドは無視できる初めての ORAM である. 実際の実データでの検証においても, 既存の ORAM と比較して, 多くとも 1/3 以下のデータ格納容量しか必要としないことも実証されており, この成果は今後のプライバシー保護技術のビッグデータへの応用の大きな第一歩となる成果であるといえる.

劣線形モデリンググループの主な研究成果は, ①深層学習に対する劣線形アルゴリズム開発(ノイズ耐性の高速テストアルゴリズムの開発)と②平均場理論を用いた高速アルゴリズム開発(画像修復に対する線形時間アルゴリズム)の2つである. ①は, 深層学習を含む一般の確率的パターン認識システムに対するノイズ耐性検査に関する新しいアルゴリズムである. 通常の検査では, 人工的に発生させた複数のノイズを実際にシステムに入力して耐性を測るのだが, 提案法は一回の入出力操作(つまり, “定数時間”の操作)で耐性を測ることができるアルゴリズムとなっている. 本成果は平成28年度から引き続き継続している成果である. ②は, マルコフ確率場と呼ばれるグラフィカルモデルを基礎とした画像ノイズ除去システムに対する新しい高速アルゴリズムである. これまで得られている最速アルゴリズムはピクセル総数  $n$  に対して  $O(n \log n)$  であった. また, その従来法は高速化のためにある種近似(グラフ構造に対する近似)を行っており, 厳密な(モデルに厳密に従った)処理結果を得ることができないという問題点がある. 対して, 当成果では“厳密”な処理結果を線形時間(つまり,  $O(n)$ )で得ることができるアルゴリズムを開発した(代表的な原著論文3). これは, 厳密計算を得ることのできるアルゴリズムの中で計算量において現時点での state-of-art となっている.

代表的な原著論文

1. T. Onodera and T. Shibuya, "Succinct Oblivious RAM", Proceedings of the 35th Symposium on Theoretical Aspects of Computer Science (STACS 2018), LIPIcs 96, pp. 52:1–52:16, 2018.
2. J. R. Jeliazkov, A. Sljoka, D. Kuroda, N. Tsuchimura, N. Katoh, K. Tsumoto and J. J. Gray, "Repertoire Analysis of Antibody CDR-H3 Loops Suggests Affinity Maturation Does Not Typically Result in Rigidification", Frontiers in Immunology, vol. 9, Article 413, pp. 1-16, 2018.
3. M. Yasuda, J. Watanabe, S. Kataoka, and K. Tanaka, Linear-Time Algorithm in Bayesian Image Denoising based on Gaussian Markov Random Field, IEICE Transactions on Information and Systems, in press.