

「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

H28 年度  
実績報告書

加藤 直樹

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

### (1)「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

### (2)「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

### (3)「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
- ・計合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

### 研究実施概要

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。

一般に酵素触媒による化学反応促進は、生化学分野の大きな謎であるが、劣線形アルゴリズムグループの Sljoka は、トロント大学の生化学研究グループと共同で、その解明に迫る重要な発見をおこない、その成果が、自然科学分野で最も影響力の高い「Science」に掲載された。フルオロ酢酸デハロギナーゼにおける酵素触媒による化学反応促進において、触媒反応時における酵素の動的変化が果たす役割を明らかにした(図1)。本論文では、その解明に迫る重要な発見をおこなった。CREST メンバーの Sljoka は、この触媒反応における、たんぱく質動力学とたんぱく質内の遠隔情報伝達の役割を組合せ剛性理論とそれに基づくアルゴリズムを用いて解明した。これにより、従来、分子動力学シミュレーションでは計算量の膨大さゆえ困難であった酵素の動的変化の解明が可能となった。この論文において、剛性理論に基づく数理的アルゴリズムが、たんぱく質の機能に関する生化学的謎の理解を助ける強力な道具であることが示されたが、さらに原子レベルで生命の謎を理解するための必須な道具であることも明らかにした。

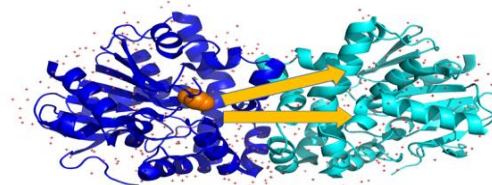


図1:フルオロ酢酸デハロギナーゼ(左の部分にあるオレンジ色の丸は酵素と結合したフルオロ酢酸を表している)とアロステリ信号伝達

劣線形アルゴリズムグループの伊藤は、複雑ネットワークのモデル化である多重グラフのクラス HSF (Hierarchical-Scale-Free)は超有限性を持つことを証明した。この結果から、多くの複雑ネットワークに対し、少なくとも理論的には「すべての性質の検査がグラフのサイズに無関係な定数時間でできる」ことになり、ビッググラフデータに対する大きな進展である。この成果は平成 27 年度に得られていたが、さらに 28 年度に、HSF を定義する際に用いられる孤立クリークを 2-孤立クリークまで拡張しても同様の定理が成り立つ目途が得られた。

避難計画の今年度の主な成果は、①梅田地下街の垂直避難のための収容人数制約付きの避難場所の領域分割問題を進展させたこと、②ZDD を用いた避難所割り当て問題においてパレート解の高速抽出アルゴリズムを開発したこと、③徒歩と自動車利用による緊急避難計画手法を開発したことである。

有用な組み合わせ集合の圧縮表現である ZDD はきわめて有用であり、列挙や論理値の高速検索にも用いられ、ビッグデータのための劣線形データ構造研究の鍵となるデータ構造の一つである。劣線形データ構造グループの伝住・定兼はこれまで、この ZDD を簡潔データ構造を用いてさらに圧縮して保持する超圧縮データ構造 DenseZDD を開発、さらにそれに対し、要素の追加を可能とした動的データ構造 Hybrid DenseZDD を開発することに成功するとともに、その圧縮の際に必要な作業メモリの空間計算量を極限まで改善することに成功した。これにより、ZDD と比べて 1/3 程度のスペースで ZDD と同様の計算が可能となり、より大規模なビッグデータをより小規模な計算機で扱うことが可能となった。

劣線形モデリンググループの安田は、深層学習の理論部分の深化に関する重要な研究を行った。深層学習は、劣線形モデリングに大きく寄与する基礎技術である。特に、深層学習における事前学習と呼ばれる計算アルゴリズムに対して統計力学的な数理解釈を与えたという理論研究の成果が顕著である。事前学習は現在の深層学習ブームの火付け役になったものであり、広く利用されている計算アルゴリズムであるが、非常に発見的なアルゴリズムとなっており、その数理的意味については提案以来未知のままであった。当成果は、統計力学的なアプローチを用いて事前学習の数理的背景を明らかとしたものであり、世界初の理論成果である。その理論の有意義性を認められ、当成果は機械学習分野でトップレベルの一つとして数えられている国際会議 (AISTATS: 採択率は毎年 2 割～3 割程度) の論文として採択された。当成果は深層学習分野のこれからの発展の礎を強固にするものであるが、それと同時に、当グループが目指す劣線形モデリング技法に深層学習のパワーを取り入れるための重要な方針を与えた理論成果となっている。

### 代表的な原著論文

1. T. Kim, P. Mehrabi, A. Sljoka, C. Ing, A. Bezginov, R. Pomes, S. Prosser and E. Pai, The Role of Dimer Asymmetry and Protomer Dynamics in Enzyme Catalysis, *Science* 355, 262, 2017.
2. H. Ito: Every property is testable on a natural class of scale-free multigraphs, *Proceedings of the 24th European Symposium of Algorithms (ESA 2016)*, Aug. 22-26, LIPICS, Vol. 57 (ISBN: 978-3-95977-015-6), pp. 51:1--51:12, 2016.
3. Muneki Yasuda:  
Relationship between PreTraining and Maximum Likelihood Estimation in Deep Boltzmann Machines, *Proceedings of the 19th International Conference on Artificial Intelligence and Statistics (AISTATS2016)*, pp.582-590, 2016.