

戦略的創造研究推進事業 CREST
研究領域「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」
研究課題「ビッグデータ時代に向けた
革新的アルゴリズム基盤」

研究終了報告書

研究期間 2014年10月～2021年3月
(新型コロナウイルス感染症の影響を受け2021年9月まで延長)

研究代表者:加藤直樹
(兵庫県立大学大学院情報科学研究科、
教授)

§ 1 研究実施の概要

(1) 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、劣線形モデリンググループの3つの研究グループから構成されている。以下では、本研究課題の成果として学術的、および社会的に注目を浴びた研究の概要を紹介する。

なお、本課題は新型コロナウイルス感染症の影響を受け、6か月の研究期間の延長を行った。

劣線形時間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となる多重グラフのクラスを提案し、このクラスは超有限性を持つことを証明した。この結果により、多くの複雑ネットワークに対し、少なくとも理論的には「すべての性質がグラフのサイズに無関係な定数時間で検査できる」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。さらに伊藤は Ilan Newman との共同研究によって、出次数制限有向グラフモデルにおいて、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1年の追加支援が得られたことにより、この特徴付けは単調な性質や遺伝的な性質に限らず一般の性質に関して成立するという予想を得た。また、たんぱく質機能解析に関して、従来は、分子動力学シミュレーションが用いられてきたが、時間スケールでは数ミリまでがシミュレーションの限界であった。Sljoka は、たんぱく質に小物質を付加することによって生じる自由度変化とその変化の範囲を高速に計算する数理的アルゴリズムを剛性理論に基づいて開発し、そのアルゴリズムが、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をいくつかの場合において明らかにした。その代表的な成果は、トロント大学の生化学研究グループと共同で、生化学分野の大きな謎である酵素触媒による化学反応促進の解明に迫る重要な発見をしたことである。この成果は、自然科学分野で最も影響力の高い国際学術誌「Science」に掲載された。さらに1年の追加支援が得られたことにより、新たな重要な成果を3つ得ることができた。

1. Mike Williamsonと共同で、核磁気共鳴装置(NMR)によるたんぱく質3次元構造データの正確性を定量的に評価する手法を開発した。これまでは、NMRによって得られたたんぱく質構造の正しさを定量的に確認する手法が存在しなかったが、本研究において世界で初めて、定量的な正確性評価手法の開発に成功した。その手法に関する論文は、「Nature Communication」に掲載された。
2. 人体で最もよく見られる受容体の一つである G-たんぱく質共役受容体(GPCR)のなかで、A2A アデノシン受容体に着目し、それが外部刺激をアロステリーによりその受容体と結合している G たんぱく質に伝達し、機能発現を促進・阻害させる相互作用の詳細を、剛性理論アルゴリズムと NMR による測定によって明らかにした。これも大きな発見であり、インパクトの高い専門誌「Cell」に掲載された。
3. グルタミン酸デヒドロゲナーゼ(GDH)は、炭素や窒素の代謝に重要な役割を果たす酵素である。Sljoka は、剛性理論によるアロステリー検出アルゴリズムと他の計算手法を組み合わせて、GDH がアロステリー調節により制御されていることを明らかにした。また、避難計画問題に関して、従来は、都市部の大規模な避難計画立案に対して、シミュレーションで検証するだけであったが、瀧澤は劣線形モデリンググループで開発された情報粗視化技術と数理最適化技術を巧みに組み合わせることで、従来の最適化手法が適用出来ないような大規模道路ネットワークに対しても、実行時間内に高精度の避難計画を策定することを可能とした。さらに1年の追加支援により、瀧澤は、大阪市全域の建物単位での大規模な被害シミュレーションを行い、2次避難者の発生数を推定した。さらに推定された2次避難者を、避難所に割り当てる大規模な数理計画問題を定式化し、避難所の地域割りあて計画を最適化で求め、大阪市の防災計画へ結果を提供した。

劣線形データ構造グループの小野寺・渋谷は、Succinct ORAM(Oblivious Random Access

Memory)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。ORAM は、データベースアクセスやメモリアccessを秘匿する近年脚光を浴びているデータ構造であるが、これまでの既存技術では、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だった。これに対し、小野寺・渋谷が開発した Succinct ORAM は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最適な初の実用的 ORAM である。その性能は実際の実データにおいても実証しており、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。さらに小野寺・渋谷は 1 年の追加支援により、これを大規模メモリのアクセス平準化によるメモリ寿命延長という新たな応用に展開することにも成功した。また、IoT の通信環境において、ハードウェア化による無遅延なロスレス圧縮技術はビッグデータ時代の通信環境を大幅に改善するが、これまではメモリ制限の厳しい FPGA 等のハードウェア上でそれを実用化できるアルゴリズムがなかった。これに対し、劣線形データ構造グループの山際・坂本は、そのような制限下でも実現可能なアルゴリズムの実現をはじめとする圧縮技術・理論の構築を多面的に行い、データの種別を問わずナノ秒オーダーの途切れない高速ロスレス圧縮・復号を低コストで可能とする世界初の新技術を実現させ、ビッグデータ時代における通信環境改善への道を大きく切り開いた。さらに山際は 1 年の追加支援により、この FPGA 実装技術を統計量の時系列変化にアダプティブに追従できる実装へと拡張させることにも成功し、さらに現実データへの適用性を高めることに成功した。

劣線形モデリンググループは統計力学における平均場理論と繰り込み群の方法に基づく確率的グラフィカルモデルの情報粗視化技術を体系的に開発することで劣線形モデリングの実現を目指した研究推進を行ってきた。そのポイントは超高次元の状態空間の部分空間の積分を解析的に実行することにより確率的グラフィカルモデルがどのように縮約された確率的グラフィカルモデルとして変換されていくかという縮約過程の流れをとらえながら、そこに統計的機械学習理論を組み込むという点にある。また、これらの知見をもとに深層学習における統計力学的なアプローチを通してその事前学習の数理的背景を明らかとし、機械学習分野でトップコンファレンスの一つである国際会議(AISTATS)で採択された。また、本研究課題を推進する過程で、その成果の一部を量子力学的に拡張された確率的グラフィカルモデルにおける統計的機械学習に展開するという当初は想定していなかった研究成果も得られている。これらの成果は統計的機械学習理論の計算時間における劣線形モデリングを実現するのみならず、深層学習をはじめとしてこれまで実問題に対する発見的手法にとどまっていた機械学習アルゴリズムの数理的背景を明らかとした点で革新的アルゴリズムの基盤構築に大きく貢献したものと位置付けられる。さらにこれらの成果はデータ数に対する劣線形モデリングへと展開され、量子計算技術にも波及しつつあり、1 年の追加支援が得られたことにより、量子アニーリングによる「CDMA 復調方式」、「ポートフォリオ最適化」、「量子ボルツマンマシンによる画像生成」への実践的展開が得られている。上述したように革新的アルゴリズム基盤に資する様々な成果が得られており、これらの成果の実応用分野への展開を推進するために、公開用ウェブサイト

<https://crest-sublinear.jimdofree.com/>

を構築した。具体的な成果として、(1)災害時避難シミュレーションの動画、(2)ストリームデータ高速圧縮の動画、(3)画像ノイズ除去、(4)グラフ列挙に関するプログラムを公開している。

さらに最終年度の新型コロナ研究への追加支援により、新型コロナウィルス感染者の予後予測の研究を実施した。兵庫県立尼崎総合医療センターとの共同で、国内の新型コロナウィルス入院対応をしている 11 施設のデータを収集し、収集した胸部の CT 画像と臨床情報を元に、入院患者が重症化するかどうかを予測する機械学習モデルの開発を行った。コロナ影響による延長期間において画像のテクスチャ特徴を重視した、よりパラメータ数の少ないモデルでも、同程度の予測が行えるという結果が得られた。また、新型コロナウィルス(SARS-CoV-2 たんぱく質)のスパイクたんぱく質が人間の受容体である ACE2 にどのように結合し、感染を引き起こすのかについてその動的振舞いについて研究を行い、コロナ影響による延長期間に剛性理論アルゴリズムや他の計算手法を用いて明らかにした。

(2) 顕著な成果

< 優れた基礎研究としての成果 >

1. 複雑ネットワークに対する定数時間アルゴリズムの開発

概要: インターネットやソーシャルネットワークなどの複雑ネットワークの「冪乗法則」と「クラスタ性」に着目し, ある種の階層的なクラスタ性を持つ多重グラフのクラスを提案し, そのモデル上では, すべての性質がグラフのサイズに無関係な定数時間で検査できることを証明した (**代表的な論文[1]**). さらに, 過去に提案されている複雑ネットワークのモデルにおいても, 階層性を利用してはいるものは, すべての性質が定数時間で検査可能であることを証明した.

2. 組合せ剛性理論によるたんぱく質機能解析

概要: 組合せ剛性理論を用いて, アロステリー現象が生じるメカニズムを説明するモデルを構築し, また, アロステリー信号伝達が生じる部位のペアを高速に同定するアルゴリズムを開発した (**代表的な論文[2]**). これにより, たんぱく質における互いに離れた位置にある二つの部分の間のアロステリー信号伝達に関する従来の仮説を立証することができた. また, 同アルゴリズムの適用によって, G たんぱく質共役受容体の一つである A2A 受容体の活性化促進・低減のメカニズムを明らかにするなど, たんぱく質の機能解明に大きな貢献を果たした (国際学術誌 Nature Communication に掲載).

3. 世界初の容量最適なプライバシー保護技術 Succinct ORAM の実現

概要: ORAM は, データベースアクセスを秘匿する技術として近年脚光を浴びている技術であるが, 必要なデータ容量のオーバーヘッドが極めて大きく超ビッグデータに適用するには難があった. これに対し小野寺, 渋谷が提案した Succinct ORAM は, 劣線形パラダイムを活用し, 従来の ORAM のセキュリティおよびアクセス速度を保ちつつ必要データ容量オーバーヘッドは漸近的に最適な容量に抑えた世界初の ORAM 技術である (**代表的な論文[3]**).

< 科学技術イノベーションに大きく寄与する成果 >

1. 革新的な省スペース圧縮技術の開発とそれによる超低コスト・超高速圧縮通信の実現

概要: ビッグデータ時代においては, 単純にデータをコンパクトに保持することだけでなく, それらの通信環境も問題となる. ハードウェア化による無遅延なロスレス圧縮技術はビッグデータ時代の通信環境を大幅に改善するが, これまではメモリ制限の厳しい FPGA 等のハードウェア上でそれを実用化できるアルゴリズムがなかった. 本研究では, そのような制限下でも実現可能なアルゴリズムの実現をはじめとする圧縮技術・理論の構築を多面的に行い, データの種類を問わずナノ秒オーダーの途切れない高速ロスレス圧縮・復号を低コストで可能とする世界初の新技术を実現させた. 本成果は ET/IoT アワード特別賞などを受賞した.

2. 量子アニーリング法を用いた実問題における実践的最適化手法の実用化の検討

概要: 量子アニーリング法を中心とする量子力学的効果を用いた高速アルゴリズムの実用化の検討を行った (学術誌 Nonlinear Theory and Its Applications, IEICE に掲載). 量子計算技術はカナダの D-Wave Systems 社による D-Wave マシンの出現により, 近年, 大きく加速している. 劣線形モデリンググループの大関は, 東北大学と東京工業大学の連携協定の締結および株式会社デンソーとの協力関係をもとに, 今後の量子計算技術の実践を通しての科学技術イノベーションを大きく牽引する拠点形成を行っている.

3. 大規模道路ネットワークに対する数理最適化に基づいた避難計画策定手法の開発

概要: 南海トラフに起因する大地震が発生した場合, 大阪市は通常の帰宅困難者問題に加えて津波による浸水が予想されており, 両者を考慮した避難計画の立案が求められている. 都

市部の大規模な避難計画はこれまでほとんど行われておらず、単純なシミュレーションを行う例がわずかにあった。本研究では、情報粗視化により数理最適化の技術を適用可能にし、従来の最適化手法が適用出来ないような大規模道路ネットワークに対しても、実用時間内に高精度の避難計画を策定することを可能とした。

< 代表的な論文 >

[1] Hiro Ito, Every property is testable on a natural class of scale-free multigraphs, Proceedings of the 24th European Symposium of Algorithms (ESA 2016), LIPIcs, Vol. 57, pp. 51:1-51:12, 2016.

概要: インターネットやソーシャルネットワークなどは複雑ネットワークと呼ばれ、典型的なビッグデータの一つであり、近年の研究により、「直径が小さい」「次数の冪乗法則」「クラスタ性」などの性質があることが分かっている。伊藤は、複雑ネットワークの「冪乗法則」と「クラスタ性」に着目し、「階層的孤立クリーク構造」というある種の階層的なクラスタ性を持つグラフは、任意の性質が検査可能であること(つまり、超有限であること)を証明した。この結果から、ほんの一部のデータのみを用いて、ネットワークのクラスタ性や中心性を定数時間で検証できる。

[2] 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.

概要: 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、本論文では、その解明に迫る重要な発見を行った。CRESTメンバーのSljokaは、アロステリー信号伝達と呼ばれる、たんぱく質の一つの部位の剛性変化が遠隔部位の剛性変化を引き起こす現象を予測する高速グラフアルゴリズム(Rigidity Transmission Algorithm)を開発して、この重要な発見に貢献した。この現象は数百ミリ秒から数秒を要することから、既存のMDシミュレーション技術は、計算負荷が高く、適用不可能であった。このアルゴリズムは、すでに、広範な種類のたんぱく質におけるアロステリー現象予測に適用実験中で、今後も多くの重要な発見を導くことが期待できる。

[3] T. Onodera and T. Shibuya, Succinct Oblivious RAM, Proceedings of the 35th Symposium on Theoretical Aspects of Computer Science (STACS 2018), LIPIcs, Vol. 96, pp. 52:1-52:16, 2018.

概要: ORAM (Oblivious Random Access Memory) は、データベースアクセスを秘匿する技術として近年脚光を浴びている技術であるが、実装に必要なデータ容量が実データ容量と比べて極めて大きいために超ビッグデータに適用するには難があった。小野寺・渋谷の提案した Succinct ORAM は、劣線形パラダイムを活用し、従来の ORAM のセキュリティおよびアクセス速度を保ちつつ、必要データ容量を漸近的に最適な容量にすることに成功した。これにより ORAM のさらなるビッグデータへの実応用が拓けていくことが考えられ、ひいては、プライバシー保護に限らない様々なより高度なビッグデータ処理技術の創出へとつながっていくことが期待できる。

§ 2 研究実施体制

(1) 研究チームの体制について

- ① 「劣線形時間アルゴリズム」グループ (研究代表者グループ)
 - ・研究代表者: 加藤直樹 (兵庫県立大学社会情報科学部, 教授)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たんぱく質の機能解明
 - ・避難計画問題
 - ・革新的アルゴリズム基盤の構築
 - ・【コロナ研究】臨床データと CT データに基づく COVID-19 の予後予測システムの研究開発
 - ・【コロナ研究】組合せ剛性理論に基づくたんぱく質構造の柔軟性, 動的振舞いの解析アルゴリズムによる SARS-CoV-2 (新型コロナウイルス) の分子機能の解明
- ② 「劣線形データ構造」グループ (渋谷グループ)
 - ・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)研究項目
 - ・情報論的アプローチによる劣線形データ構造
 - ・列挙論的アプローチによる劣線形データ構造
 - ・実応用アプローチによる劣線形データ構造
- ③ 「劣線形モデリング」グループ (田中グループ)
 - ・主たる共同研究者: 田中和之 (東北大学情報科学研究科, 教授)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高効率計算アルゴリズムの開発

(2) 国内外の研究者や産業界等との連携によるネットワーク形成の状況について

【国内外研究者との連携】

- ・たんぱく質機能解析においてはトロント大学の Scott Prosser 教授, Emil Pai 教授, Andrew Wooley 教授の研究グループ, York University の Derek Wilson 教授, Basel 大学の Stephan Grzesiek 教授, Timm Maier 教授等のグループと共同研究を開始している (Sljoka). また, Jeffrey Gray (John Hopkins University), Stuart Kauffmann (Institute for System Biology), Andreas Mueller (Johannes Kepler University), Daisuke Kuroda (University of Tokyo), Che Tao (University of North Carolina), Amit Chakraborty (Central University of Rajasthan), Alvaro Mallagaray (University of Luebeck) などの生化学者とも連携して研究を進めている (Sljoka).
- ・定数時間アルゴリズムに関しては, その分野の第一人者である Ilan Newman (Univ. of Haifa, イスラエル) と連携して研究を進めている (伊藤).
- ・離散アルゴリズムについて Erik Demaine (MIT), Stefan Langerman (ULB, ベルギー), John Iacono (NYU→ULB) らと連携して研究を進めている (伊藤).
- ・離散アルゴリズムについて Binhai Zhu (Montana State University), Guohui Lin (University of Alberta) と連携して研究を進めている (加藤, 東川, 玉置, 照山, 宮野).
- ・大阪市立大学都市防災教育研究センター (CERD) の研究員として, 避難計画問題のプロジェクト

クトを他のメンバーとともに進めており、九州大学マスコアインダストリ研究所の藤澤克樹教授らが進めるスマートシティ関連の研究プロジェクトに、避難計画で参加している(瀧澤)。

- 最速避難計画問題の理論研究においては、Mordecai J. Golin, Siu-Wing Cheng (Hong Kong University of Science Technology), Binay Bhattacharya(Simon Fraser University), Robert Benkoczi (University of Lethbridge), 神山直之氏 (九州大学) と連携して研究を進めている(加藤, 東川, Kameda)。
- 渋谷グループでは、山際・坂本らによるストリームデータ圧縮技術に関して、産業界との共同研究・事例展開を行っている。
- 渋谷グループにおける経営データに関する研究は、関西大学データサイエンス研究センターと連携して研究を行い、同センターが行っている多くの企業との共同研究を通じて、ビジネス領域への研究成果の社会還元が共同セミナーや一般公開の会議・ワークショップを通じて行われた。
- アムステルダム大学(University of Amsterdam)の Max Welling 教授との深層学習と状態空間の粗視化を組み合わせた革新的アルゴリズム創出に向けての連携関係を構築しつつある(田中)
- ローマ大学ラサピエンザ校(University of Roma, La Sapienza)の Federico Ricci-Tergenghi 准教授, Tommaso Rizzo 研究員と状態空間の粗視化のもとでのアルゴリズム動作解析, スパースモデリングの統計的性能解析に関して連携関係を継続している(大関, 田中)
- パリ高等師範校(Ecole Normale Supérieure)の Florent Krzakala 教授およびその研究グループとコミュニティ検出, ボルツマンマシンによる統計的機械学習のためのアルゴリズム設計に関して情報交換を行いつつ, 連携関係を構築しつつある(田中, 片岡)
- ユニバーシティ・カレッジ・ダブリン(University College Dublin)の Nial Friel 准教授とモンテカルロ法による統計的機械学習アルゴリズム設計に関する情報交換から連携関係を構築しつつある(大関, 安田, 田中)
- 台湾国立清華大学(National Tsing Hua University, Taiwan)の Candy Hsu 教授とマルチメディアにおけるビッグデータ活用と統計的クラスタリングアルゴリズム設計に関する共同研究を通して連携関係を継続している(片岡, 安田, 和泉, 田中)
- ウォータールー大学(University of Waterloo)の Naomi Nishimura 准教授と遷移問題に対するアルゴリズムについての共同研究を通して連携関係を継続している(鈴木)。
- パリ南大学(Université Paris-Sud)の Aurélien Decelle 研究員とパラメータ空間の間引き処理(Decimation)アルゴリズムの開発についての共同研究を行い, 連携関係を継続している(大関)
- フランス国立情報学自動制御研究所(Institut National de Recherche en Informatique et en Automatique; INRIA)のサクレイ(Saclay)研究拠点の Cyril Furtlehner 研究員と交通ネットワークにおける統計的機械学習アルゴリズム設計に関する共同研究を行い, 粗視化処理の転用に関する打ち合わせを行いつつ, 連携関係を継続している(安田, 片岡, 田中)

【産業界等との連携】

- 避難計画については、大阪市地下空間浸水対策協議会の専門委員として、梅田地下街の避難シミュレーションを担当。また、電気学会システム最適化と産業応用ベンチマーク問題調査専門委員会にて、避難計画のベンチマーク問題作成WGに参加(瀧澤)。
- 大阪市と大阪市立大との包括連携協定に基づいて、主に大阪市の湾岸5区を対象として、南海トラフに起因する大地震の津波に対する2次避難計画の策定業務を行い、新たな研究課題の発掘とフィードバックを行っている(瀧澤)。
- タウたんぱく質の構造解明の成果をもとにアルツハイマー等の神経性疾患の治療薬開発を目指している米国の製薬会社 Treventis に必要データを提供している(SIjokら)。
- 渋谷グループでは、ストリームデータ圧縮技術に関して、産業界との共同研究・事例展開を行っている。
- 渋谷グループにおける経営データに関する研究は、関西大学データサイエンス研究センター

と連携して研究を行い、同センターが行っている多くの企業との共同研究を通じて、ビジネス領域への研究成果の社会還元が速やかに行われる。こうしたビジネス領域への研究成果の社会還元は、非公開、公開の形態を柔軟に組み合わせ、様々な企業との共同研究はもちろん、共同セミナーや一般公開の会議・ワークショップを通じて行われる。本事業が主催する様々な会議に企業を積極的に招へいし、開発技術の普及を進める。2016年度は第二回として、秋にデータサイエンス研究センターとの共同会議を予定している(渋谷グループ)。

- 株式会社インフィニテックとの共同研究にて、テキストデータの情報圧縮によるデータ可視化システムの開発を行っている(安田)