

研究終了報告書

「離散幾何学が拓く計算系統学の新展開」

研究期間：2019年10月01日～2023年03月31日

研究者：早水 桃子

1. 研究のねらい

生物の進化的な系譜を解明する「系統樹」と呼ばれるツリー構造を推定する数理的な方法論は20世紀初頭からインテンシブに研究・体系化されて分子系統学の基礎理論となっているが、現代の生命科学にはそのような分子系統学の古典的な方法論では対処できない様々な課題がある。例えば、ウイルスや細菌(バクテリア)や植物の複雑な進化を解明する研究においては、系統樹を一般化した「系統ネットワーク」を活用したデータ解析手法を使いたいというニーズがある。また、再生医療にも関わる幹細胞生物学の研究においては、細胞分化の軌跡を推定するために遺伝子発現量を表す高次元ベクトルデータから「パス」や「ツリー」などの多様なグラフを構築することが必要とされている。

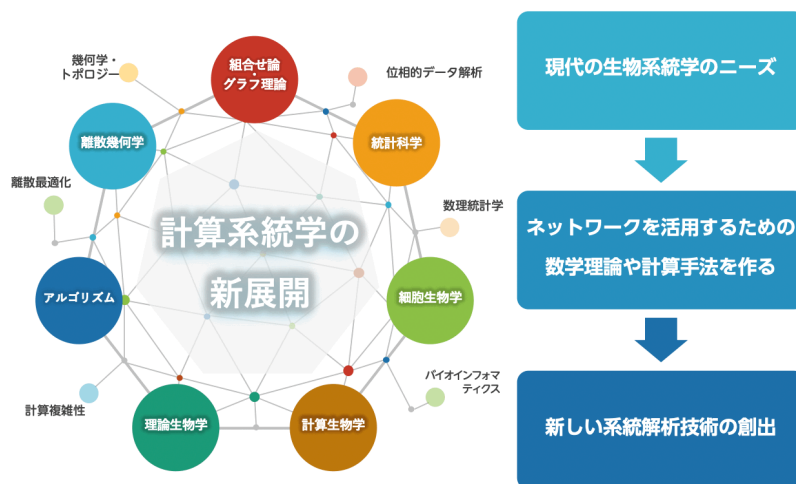
しかし、系統ネットワークは系統樹に比べて表現力が高い反面、網目のように複雑なネットワークも含む巨大なグラフのクラスなので、系統樹と違って数学的な考察やアルゴリズムの設計・解析において様々なハードルがあり、生物学者にとっても進化史を解釈することが容易でないという難点がある。また、細胞の分化の軌跡推定に用いられるシングルセルの遺伝子発現データは実験計測技術のブレークスルーにより10年ほど前に登場したばかりの新タイプのデータなので、現時点で多くの数理的・統計学的な課題があり、そのデータ解析手法を確立させることは計算生物学の大きなチャレンジの一つになっている。

早水はこれまでに細胞の分化や生物の進化といった生物学的現象の記述を「離散幾何学的モデリング」という包括的な枠組みで捉え、色んなグラフと距離の関係について考察を進めるとともに、関連する組合せ論の諸問題を研究することで、多様な生命現象のより良い理解に向けた方法論を整備してきた。本研究では、グラフ理論、最適化、数え上げ・列挙・幾何学、トポロジーなどの多様な数理・情報科学の発想を活かしてこれまでの研究をさらに発展させ、幅広い分野の研究者との交流・連携・協力を通じて計算系統学の新展開を拓く新しいデータ解析手法の創出を目指した。

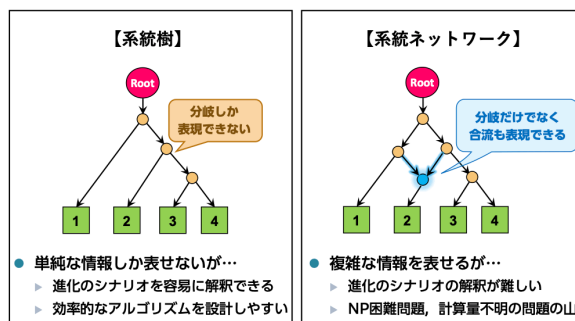
2. 研究成果

(1) 概要

採択時当初は(A)生物の進化に関する系統解析の方法論に絞った研究計画だったが、さきかけ数理構造活用領域内でのコラボレーションにより偶発的な進展が多くあり、当初の計画にない(B)細胞の分化に関するデータ解析手法についても共同研究を進めた。多様な数理・情報科学分野と関わりながら計算系統学の新展開を拓くという目標(下図)に沿った様々な成果を収めることができた。



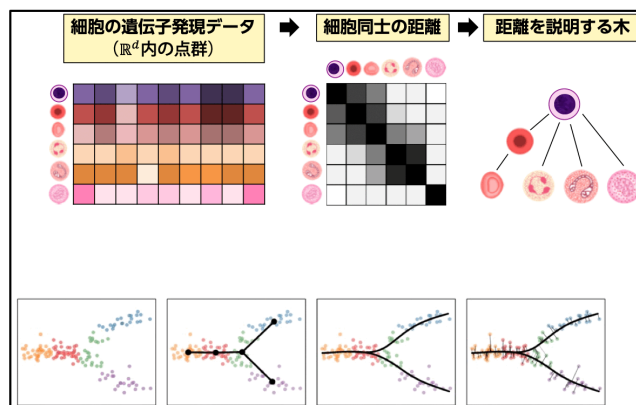
生物進化に関するテーマ(A)では、さきがけ研究期間中に単著・単著に近い貢献度の論文を3編出版し招待講演や特別講演も多数行った。3編のうち1編は、距離行列から系統カクタスという系統ネットワークを構築する方法に関する内容である(主な研究成果リスト(1):論文3)。



他の2編は根付き二分系統ネットワークの構造定理とそれを応用した全域系統樹に関する多数の高速アルゴリズムの開発に関する内容である(主な研究成果リスト(1):論文1・論文2)。論文1は組合せ論的系統学の複数の未解決問題を解決した単著論文で、離散数学の一流誌 SIAM Journal on Discrete Mathematics に掲載された。論文1を英語で解説した YouTube 動画は33万回以上再生され、日本語や英語だけでなく中国語に字幕翻訳された。論文2は系統樹をベクトルとしてエンコードする手法を開発するなどの新規アイデアによって論文1で与えた最適化や列挙のアルゴリズムを高度に一般化し上位ランキング問題を解く高速アルゴリズムを与えたもので IEEE/ACM Transactions on Computational Biology and Bioinformatics に掲載された。この他にも当初掲げていた研究項目の大半で部分的な結果や有望な予想などの進捗が得られた。

細胞分化に関するテーマ(B)でもさきがけ数理構造活用領域内の共同研究や実験研究者との連携により様々な研究結果を得た。多数の招待講演を行った。その他の共同研究では細胞の遺伝子発現データに限らず適用できる汎用的な手法の提案を目指す共同研究を進めた。

主な受賞として、令和4年度の文

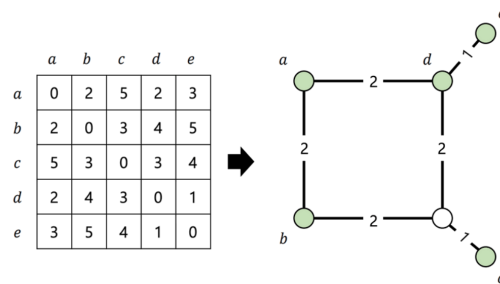


部科学大臣表彰若手科学者賞を受賞した。研究に直結する内容以外では YouTube を活用した大規模で国際的なオンラインでのアウトリーチに成功し WASEDA e-Teaching Award を 2 年連続で受賞した。YouTube 活動は産業界との連携、個人・法人からの寄付金、他大学の研究者や学生との研究教育交流など多様な展開につながった。さきがけ研究期間中に JST 創発的研究と早稲田大学 PI 飛躍プログラムに採択された。さきがけ研究者や関連する共同研究者と共同で数理・計算系統学に関する 4 セッションのミニシンポジウムを ICIAM2023 にて提案し採択された。

(2) 詳細

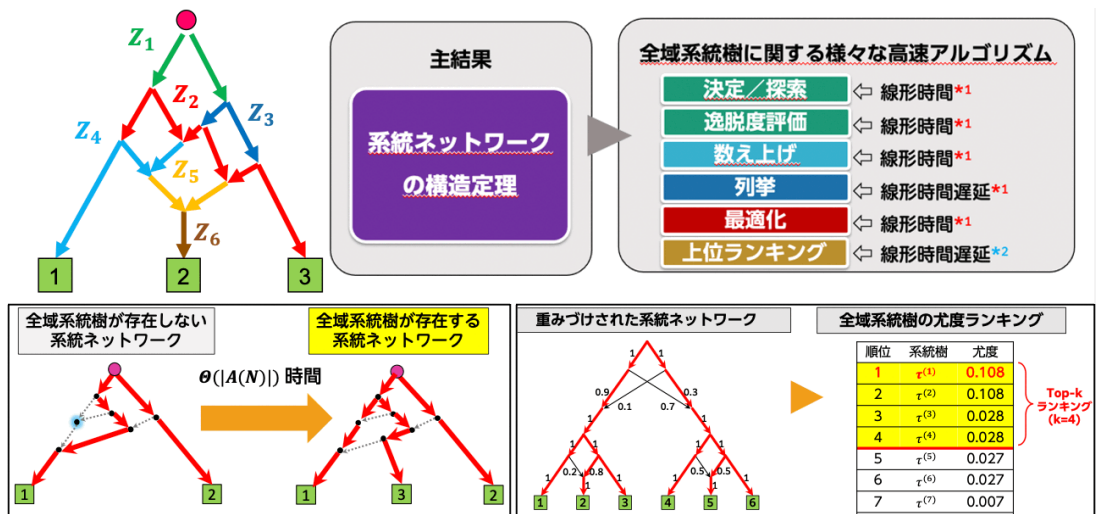
(A) 生物進化の系統解析

- ① 根つき二分系統ネットワークの構造定理のさらなる応用 その1: 全域系統樹が存在しない系統ネットワークを全域系統樹が存在する系統ネットワークに変換する高速アルゴリズムを設計した(論文1に含めて発表)。それまでは $O(|V|^{3/2})$ 時間アルゴリズム [Francis-Semple-Steel (2018), Adv. Appl. Math.] がベストだったが系統ネットワークの構造定理を活用した独自のアプローチで $\Theta(|V|)$ 時間アルゴリズムを与えた。
- ② 根つき二分系統ネットワークの構造定理のさらなる応用 その2: 牧野和久氏(京都大学数理解析研究所)との共同研究で、全域系統樹の上位ランキング問題を解く線形時間遅延アルゴリズムを発表した(論文2)。
- ③ 系統樹や木距離という概念を一般化した系統カクタスやカクタス距離を提案し、カクタス距離を実現する最適な系統カクタスが一意に定まり、 n 点距離空間を実現する最適な系統カクタスは $O(n^3)$ 時間で計算できることを示した(論文3)。



(B) 細胞分化の系統解析

- ① 以前のさきがけでシングルセル遺伝子発現データから推定された細胞分化の軌跡(ツリー)の信頼性を評価する手法を提案して Treefit というソフトウェアを発表した。こ



れはデータに摂動を加えたときの木構造の頑健性を測るものだが、データをどのように摂動するのが良いかについて数理構造活用領域のさきがけ研究者との共同研究を行い、新しい手法を提案した。

- ② さきがけ数理構造活用領域内の連携や実験研究者との協力により、多様な 1 細胞の遺伝子発現データを解析して様々な興味深いデータ解析結果を得た。その他、より汎用的な数理的手法の研究も行い、予備実験で有望な解析結果を得た。

(C) その他の成果

- ① 受賞((3)その他の成果 受賞 15, 16, 17)
 - ・令和4年度文部科学大臣表彰若手科学者賞 受賞
 - ・2021 年度 第 10 回 WASEDA e-Teaching Award (科目名:離散数学入門)
 - ・2020 年度 第 9 回 WASEDA e-Teaching Award 大賞(科目名:数学 B1(微分積分)) (受賞代表者:早水桃子)※中村憲史氏・さきがけ2期生の町田学氏との共同受賞
- ② さきがけ研究期間中に獲得したグラント:
 - ・JST 創発的研究支援事業
 - ・早稲田大学 PI 飛躍支援プログラム
- ③ 特記すべき講演:
 - ・日本数学会 特別講演((3)その他の成果 講演 1)
 - ・日本発生生物学会 招待講演((3)その他の成果 講演3)
- ④ アウトリーチの例:
 - ・YouTube での動画出版やライブ配信(チャンネル登録者 3.4 万人)
 - ・メタバース空間(Cluster)を活用したバーチャル学会 2022 でのポスター発表 (※ACT-X 研究者たちとの研究交流)

3. 今後の展開

(A)-①,②:根つき二分系統ネットワークの構造定理に基づく諸結果(アルゴリズム)は計算系統学における数学的な未解決問題を多数解決し、系統ネットワークの統計学的な応用も拓いた点で大きな成果だが、さらなる一般化が成功したり、データから有向系統ネットワークを構築するヒューリスティクスが発展したりすれば、本研究で開発したアルゴリズムを適用できる範囲も大幅に拡大し、様々な実データでの有効性を確かめることができるようになるだろう。さらにそれがソフトウェアとして利用可能になれば、例えば細菌やウイルスなどの病原体の変異解析、植物・動物の進化の解析など、多様な系統解析に役立つ新技術となるだろう。

(B)-②:今後は実験研究者との連携をさらに強化して、幹細胞生物学・幹細胞工学などの基礎医学の研究により直接的に貢献するデータ解析を行う。既にいくつかの予備的な結果を得ているので、近い将来、論文やソフトウェアを発表できる見込みである。

4. 自己評価

本研究では系統ネットワークの構造定理を応用した系統解析の様々な高速アルゴリズムを発表することで系統ネットワークの利活用を推進し、生物同士の距離(非類似度)データから解釈しやすい特殊な系統ネットワークを構築する効率的な手法についても論文を発表した。またさきがけ領域内・領域外の研究者や学生等と積極的に共同研究を行い、当初の計画に

ない細胞の系統解析というテーマ(1細胞の遺伝子発現データからの細胞分化の軌跡推定)についても様々な理論的考察や実データ解析を行い、色々な結果を得た。いずれのテーマについても、公開可能な成果については多数の招待講演を行った。外部評価としては、文部科学大臣表彰若手科学者賞を受賞し、98名の受賞者の代表者1名に選出された。さきがけ研究期間中にJST創発的研究支援事業や早稲田大学PI飛躍支援プログラムに研究代表者として採択された。産業界との連携・交流も行った。YouTubeでの教育活動・研究アウトリーチ活動は当初の想定をはるかに上回る社会的・経済的・教育的効果があり、今後も更に大きく成長することが見込める国際的な活動となった。以上より、総合的に考えて、本研究では当初の目標を様々な意味で上回り、さきがけ終了後も発展が見込める多様な成果を収めることができたと考える。

5. 主な研究成果リスト

(1) 代表的な論文(原著論文)発表

研究期間累積件数: 3件

<p>1. Momoko Hayamizu: A structure theorem for rooted binary phylogenetic networks and its implications for tree-based networks, <i>SIAM Journal on Discrete Mathematics</i>, 35 (4), 2490–2516, 2021. DOI: 10.1137/19M1297403</p>
<p>これまで多項式時間で解けるか不明だった全域系統樹の数え上げや列挙などの諸問題と、これまで研究されたことのない新規の問題(尤度最大化など最適な全域系統樹を求める問題)の全てに対して、線形時間/線形時間遅延アルゴリズム(存在し得る中で最も高速な解法)を与えた。配列データから最尤系統樹を計算するのは多項式時間では不可能なので、系統ネットワークを用いることで最尤系統樹を線形時間で推定できることは統計学的にも特筆すべき進展である。</p>
<p>2. Momoko Hayamizu and Kazuhisa Makino: Ranking top-k trees in tree-based phylogenetic networks, <i>IEEE/ACM Transactions on Computational Biology and Bioinformatics (TCBB)</i>, To appear. DOI: 10.1109/TCBB.2022.3229827</p>
<p>論文 1 では根付き二分木系統樹ネットワークの構造定理に基づき全域系統樹の最適化や全列挙などの線形時間/線形遅延アルゴリズムを与えたが、バイオインフォマティクス的には最適解を1つだけ求めるよりも最適解と最適解に近い解の両方を探索する方が実際に有用なことが多い。本論文では目的関数値を最大化する全域系統樹だけでなく上位 k 個の全域系統樹を上から順に列挙する(トップ k のランキングを作る)線形遅延アルゴリズムを与え、論文 1 の結果をさらに一般化して実用的なものにした。</p>
<p>3. Momoko Hayamizu, Katharina T. Huber, Vincent Moulton, Yukihiro Murakami: Recognizing and realizing cactus metrics, <i>Information Processing Letters</i> 157(105916) 1-5, 2020.</p>
<p>系統樹の拡張版となる新しい進化のモデルを見出すために系統樹で実現可能な距離を拡張した「カクタス距離」という新概念を導入し、最適な実現の一意性などの木距離の多くの良い性質がカクタス距離に継承されることを証明し、「系統カクタス」という進化のモデルを提案した。最適な系統カクタスは距離から系統樹を計算するのと同等の時間で計算できることも示した。</p>

(2) 特許出願

なし

(3) その他の成果(主要な学会発表、受賞、著作物、プレスリリース等)

1. 【特別講演】早水桃子 進化の系統樹と系統ネットワークに関する組合せ論 日本数学会2022年度秋季総合分科会 応用数学分科会(特別講演) 2022年9月13日 日本数学会
2. 【招待講演】 Momoko Hayamizu, Treefit: a method to measure the robustness of differentiation trajectories inferred from single-cell RNA-seq data, The 55th Annual Meeting of the Japanese Society of Developmental Biologists Symposium 4: Novel insights into cellular plasticity in development, regeneration, and aging, 2022年6月3日, 日本発生生物学会 (Japanese Society of Developmental Biologists)
3. 【招待講演】早水桃子, 進化の系統樹と系統ネットワークに関する組合せ論, 第17回 MACS コロキウム, 2021年11月19日 京都大学理学部
4. 【招待講演】 Momoko Hayamizu,, The structure of rooted binary phylogenetic networks and its applications to various computational problems in evolutionary analysis, RIMS Workshop: Toward an integration of fluids, ecology, and evolution in Biofluids 2021, 2021年7月6日
5. 【招待講演】 Momoko Hayamizu,, How to reconstruct a cell differentiation trajectory using scRNA-seq snapshot data, The 1st International Symposium on Human InformatiX X-Dimensional Human Informatics and Biology 2020年2月27日 Advanced Telecommunications Research Institute International (ATR)
6. 【受賞】2022年4月令和4年度 科学技術分野の文部科学大臣表彰 若手科学者賞(令和4年4月20日) 生命科学のデータ解析に関わる離散数学の諸問題の研究 文部科学省