

氏 名	やま ぐち あつ こ 山 口 敦 子
学位(専攻分野)	博 士 (情 報 学)
学 位 記 番 号	論 情 博 第 55 号
学位授与の日付	平 成 16 年 11 月 24 日
学位授与の要件	学 位 規 則 第 4 条 第 2 項 該 当
学位論文題目	Algorithms for Graph Theoretic Optimization Problems in Bioinformatics (バイオインフォマティクスにおけるグラフ理論的最適化問題に対するアル ゴリズム) (主 査)
論文調査委員	教 授 阿 久 津 達 也 教 授 後 藤 修 教 授 岡 部 寿 男

論 文 内 容 の 要 旨

本論文は、バイオインフォマティクス（生命情報学）に現れるグラフ理論的最適化問題に対する効率的なアルゴリズムについて述べたもので、6章から構成されている。

第1章は序論で、本論文で扱うバイオインフォマティクスにおけるグラフ理論的最適化問題について概観し、近似アルゴリズムや部分クラスに対する効率的アルゴリズムの開発の必要性について述べるとともに、本論文の構成の概要を示している。

第2章は、本論文における基本的な用語や概念について、特にグラフ理論、化合物のグラフ表現、NP 最適化問題、近似アルゴリズム、並列計算機モデル(PRAM)などに関する事項について、定義や説明を行っている。

第3章では、バイオインフォマティクスにおけるネットワーク（遺伝子ネットワークやタンパク質相互作用ネットワークなど）のユーザーフレンドリーな表現を主目的として、グラフ描画アルゴリズムを提案している。グラフ描画の審美基準として「辺の交差数の最小化」を用いて、二部グラフを対象にアルゴリズムを開発している。最初に、片側二部グラフ描画問題、すなわち、片方の層の頂点を固定して、もう一方の層の頂点のみを動かすことで辺の交差数を最小化する問題について考察している。この問題はNP 困難であることが知られ、また、重心法やメジアン法など様々な手法が提案されているが、本章では貪欲法に基づく近似アルゴリズムを提案し、このアルゴリズムの出力する描画の辺の交差数が最適解の $c(d)$ 倍以下であることを証明している。ここで、 d は入力グラフの最大次数であり、 $c(d)$ は d が小さければ2に近く、 d が大きくなるにつれて漸近的に3に近づく d の関数である。さらに、計算機実験によって、従来の重心法やメジアン法と比較して提案手法が平均的によい解を出力することを示している。次に、遺伝子ネットワークの描画をモデル化した、ペア付き二部グラフ描画問題を新規に定義し、そのNP 困難性を示すとともに、発見的アルゴリズムを提案し、計算機実験により有効性を確認している。

第4章では、グラフ理論においてグラフの複雑さの尺度として用いられる、木幅、局所木幅、全域木数などを用いて、生体内化合物の特徴の解析を行うとともに、生体内化合物に対する効率的な最大共通部分グラフ抽出アルゴリズムを提案している。まず、公開データベースより取得した生体内化合物データを用いて計算機実験を行い、生体内化合物の木幅は定数で抑えられること、全域木数が頂点数の多項式程度で抑えられることを示している。次に、局所木幅の概念を用いて木幅が比較的高い化合物について分類を行い、その分類の生物学的意味について論じている。さらに、これらの解析結果に基づいて、2種類の生体内化合物の最大共通部分グラフを求める多項式時間アルゴリズムを提案している。最大共通部分グラフ問題は一般にNP 困難であるが、このアルゴリズムにより、次数と木幅が定数、全域木数は多項式程度という制限をつけることで、多項式時間計算可能になることを示している。

第5章では、与えられた完全 k 分木の集合に対して、それらをすべて含む最小の木を求める問題について考察している。この問題は配列断片からのゲノム配列推定問題の一つを木に拡張したものとなっており、遺伝子配列における推定問題を進化系統樹や糖鎖構造の推定問題に拡張するための基礎問題として新規に提案されている。本章では、この問題がNP 困難で

あることを示すとともに、近似アルゴリズムを提案している。そして、このアルゴリズムが出力する木の大きさが最適解の $1+1/(2k-2)$ 以下であることを理論的に導いている。さらに、このアルゴリズムを PRAM 並列計算機モデル上で並列化し、 n^8 個のプロセッサを用いて $O(\log^4 n)$ 時間で実行できることを示している。

第 6 章は結論であり、本研究のまとめと今後の課題について述べている。

論文審査の結果の要旨

本論文は、バイオインフォマティクス（生命情報学）に現れるグラフ理論的最適化問題に対する効率的なアルゴリズムについて述べたもので、得られた成果は以下の通りである。

- (1) バイオインフォマティクスにおけるネットワークのユーザーフレンドリーな表現を主目的とし、グラフ描画の審美基準として「辺の交差数の最小化」を用いて、二部グラフを対象に 2 種類のアルゴリズムを開発した。一つは片側二部グラフ描画問題を対象としているが、この問題は NP 困難であることが知られ、また、重心法やメジアン法など様々な手法が提案されている。本論文では貪欲法に基づく近似アルゴリズムを提案し、それが出力する描画の辺の交差数が最適解の $c(d)$ 倍以下であることを証明した。ここで、 d は入力グラフの最大次数であり、 $c(d)$ は d が小さければ 2 に近く、 d が大きくなるにつれて漸近的に 3 に近づく d の関数である。さらに、計算機実験によって、従来の重心法やメジアン法と比較して提案手法が平均的によい解を出力することを示した。もう一つは、本論文において新規に提案したペア付き二部グラフ描画問題に対するもので、その NP 困難性を示すとともに、発見的アルゴリズムを提案し、計算機実験により有効性を示した。
- (2) グラフ理論における、木幅、局所木幅、全域木数などの概念を用いて、生体内化合物の特徴の解析を行うとともに、生体内化合物に対する効率的な最大共通部分グラフ抽出アルゴリズムを提案した。具体的には、計算機実験により、生体内化合物の木幅は定数で抑えられること、全域木数が頂点数の多項式程度で抑えられることを示した。また、この結果に基づき、次数と木幅が定数、全域木数は多項式程度という制限をつけた場合の最大共通部分グラフ問題に対する多項式時間アルゴリズムを開発した。
- (3) 与えられた完全 k 分木の集合に対して、それらをすべて含む最小の木を求める問題についての理論的解析を行った。この問題は配列断片からのゲノム配列推定問題の一つを木に拡張したものとなっているが、この問題が NP 困難であることを示すとともに、貪欲法に基づく近似アルゴリズムを提案した。アルゴリズムが出力する木の大きさが最適解の $1+1/(2k-2)$ 以下であることを理論的に導くとともに、PRAM 並列計算機モデル上での並列化を行い、 n^8 個のプロセッサを用いて $O(\log^4 n)$ 時間で実行できることを示した。

以上、本論文はバイオインフォマティクスにおいて重要な 3 種類のグラフ理論的最適化問題に対する効率的なアルゴリズムを示しており、当該分野の発展のために十分な寄与をしている。よって、本論文は博士（情報学）の学位論文として価値あるものと認める。

また、平成16年9月21日実施した論文内容とそれに関連した試問の結果合格と認めた。