

氏名	たむら たけ ゆき 田村 武 幸
学位(専攻分野)	博 士 (情 報 学)
学位記番号	情 博 第 214 号
学位授与の日付	平成 18 年 3 月 23 日
学位授与の要件	学位規則第 4 条第 1 項該当
研究科・専攻	情報学研究科通信情報システム専攻
学位論文題目	Graph Algorithmic Approaches for Structure Inferences in Bioinformatics (バイオインフォマティクスにおける構造推定のグラフアルゴリズムを用いたアプローチ)
論文調査委員	(主 査) 教授 岩間 一雄 教授 阿久津達也 助教授 伊藤大雄

### 論 文 内 容 の 要 旨

本論文では生命情報学における様々な構造推定問題に対して、理論計算機科学、特にグラフ理論やグラフアルゴリズムの立場から、ゲノムにおけるプローブ順序推定や、家系図の推定に関する結果が与えられている。

第 1 章では研究の背景および結果の概要が述べられている。

第 2 章では STS (配列タグ部位) に基づいた物理地図作成におけるデータの不完全さについての議論を行っている。物理地図作成問題の背景を踏まえ、現在までに得られたデータは、最終的な目標に対してどの程度不完全かという目安を明らかにすることの重要性が述べられている。その不完全さを今後追加するフラグメント数として定義し、情報の切れ目を意味するエッジという概念を導入することで、プローブの順序を一意に決定するために必要な最小数のフラグメント集合を求める多項式時間アルゴリズムを示している。さらに、ヒトゲノムの第 20 番染色体の遺伝子情報を用いたコンピュータ実験によって、その効果を検証している。また、エッジが密に存在する箇所に集中的に追加フラグメントを生成し、固定操作を高速化する技法が与えられている。

第 3 章では不完全さの尺度として追加フラグメントの長さの総和を用い、前章での追加フラグメント数を用いた場合との相違点について議論している。前者を用いた場合には、後者で有効であったエッジが機能しないことを指摘し、フラグメントの切れ目を PQ-木の部分木に対しスキャンさせるという技法を用い、動的計画法に基づいた多項式時間アルゴリズムを与えている。さらに、追加フラグメントを既存フラグメントの部分フラグメントに限定する場合や、追加フラグメントの長さを 2 に限定した場合についても効率的なアルゴリズムを与えている。

第 4 章では進化系統樹に関する議論を行っている。家系図が進化系統樹をさらに詳しくしたものと捉えられることを指摘し、遺伝的な距離から家系図を推定する問題を数理的にモデル化し、定式化している。家系図を入次数が 2 以下の非閉路的有向グラフとして扱い、入力行列に対応するトーナメントから枝の削除と、回数の制限された長さの変更を用いることによって、与えられた 2 節点間の遺伝的距離を満足する家系図を全て列挙する多項式時間アルゴリズムを与えている。また、所望の家系図の全てが、節点数を増やすことなく一つの重み付き有向グラフで表現できることを示している。また、より現実的なモデルに対しても、偽クリークを核家族と対応させる方法を紹介している。

最後に第 5 章では、以上の結果がまとめられている。

### 論 文 審 査 の 結 果 の 要 旨

生命情報学における様々な構造推定問題は、そのデータの膨大さゆえに、アルゴリズムの高速化が重要であり、その理論的保証が必要不可欠である。また、アルゴリズムの設計だけでなく、既存のデータをもとに現状の本質を深く理解することは、実験者が次の行動を正しく選択するための良い指針を与える。また新たな問題を発見し、それを数理的に定式化することは、既存の作業工程を変革へ導く可能性がある。これらの観点から、本論文の結果について特筆すべき点は以下の通りで

ある。

1. ゲノムにおけるプローブの順序を推定する問題において、現在得られている結果は最終的な目標に対してどの程度不完全かということを含め追加するフラグメントに関する関数として定義したこと。

2. 実験データに対応する PQ-木に対して、情報の切れ目を意味するエッジという概念を新たに導入し、その数がデータの不完全さに密接に関係することを示したこと。

3. ヒトゲノムの第20番染色体の遺伝子データを用いて、コンピュータ実験を行い、エッジが密に存在する箇所に集中的に追加フラグメントを生成することにより、実験時間を短縮できることを検証したこと。

4. 遺伝子の類似度から家系図を推定するという、新しい問題を定式化し、多項式時間アルゴリズムを与えていること。

論文の前半で議論されている STS に基づくマッピングは、DNA 配列の物理地図作成における最も重要な方法のひとつである。本研究は、データの不完全さを追加フラグメントの最小数と、追加フラグメントの長さの総和という2つの尺度で定義し、それぞれ解析を行ったことで、ゲノムにおけるプローブ順序を推定する問題の本質の理解を助けている。また、コンピュータ実験の際に用いた手法は、実験の高速化を実現するための指針を与えている。また、論文の後半で議論されている家系図は、遺伝学を扱ううえで欠かせないデータ構造である。生命情報学においては、家系図の存在を前提にした作業が多く、本研究で扱った家系図推定問題は、家系図が存在しない場合でもそれらの作業を可能にする場合がある。

以上、本研究は生命情報学における構造推定のためのアルゴリズム、特にグラフを用いた理論的なアプローチにおいて学術上意義深い結果を導いている。よって、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18年2月22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。