

氏 名	ふか 深 川 だい 大 路 じ 路
学位(専攻分野)	博 士 (情 報 学)
学 位 記 番 号	情 博 第 191 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 知 能 情 報 学 専 攻
学位論文題目	Discrete Algorithms for Optimization Problems for Structured Data in Bioinformatics (バイオインフォマティクスにおける構造データに対する離散最適化アルゴリズム)
論文調査委員	(主 査) 教 授 阿 久 津 達 也 教 授 岡 部 寿 男 教 授 永 持 仁

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）における構造データを扱ういくつかの離散最適化問題に対するアルゴリズムについて述べたもので、5章から構成されている。

第1章は序論で、本論文で取り扱う問題の背景、既存研究、論文の構成について述べるとともに、生物学的な背景、および、本論文で扱う構造データ、特に、遺伝子ネットワーク、RNA 二次構造、化学構造について説明している。

第2章では、遺伝子ネットワークの数理モデルの一つとして知られるブーリアンネットワークの推定問題について理論的解析を行っている。従来より、ブーリアンネットワークの推定問題は複数のブール関数を同時に推定する問題として定式化されており、また、個々のブール関数を推定するために貪欲算法に基づくアルゴリズムが提案されていた。既存の結果として、そのアルゴリズムは平均的には（入力変数の個数が最小となるという意味での）最適なブール関数を多くの場合に出力することが計算機実験により示唆され、また、非常に限定されたブール関数（リテラルの論理和もしくは論理積）については理論的証明がなされていた。第2章では、その性質が、unbalanced 関数と呼ばれる、より広いクラスのブール関数のクラスに対して成立することを理論的に証明し、さらに、入力変数の個数が増加するにつれ、ほとんどのブール関数が unbalanced 関数となることを示している。また、貪欲算法に基づくアルゴリズム、および、関連するアルゴリズムに対して、人工的に作成したデータを用いて計算機実験を行い、理論的結果について確認するとともに問題点などについて検討している。

第3章では、RNA や化合物の構造比較などと深い関連のある、木構造の比較問題に対するアルゴリズムについて記述している。特に、類似性の高い二個の順序無し木が入力された際に、最大の共通部分構造を検出する問題に対するアルゴリズムを提案している。その問題に対しては従来から動的計画法に基づくアルゴリズムが提案されていたが、それは入力データのサイズの少なくとも二乗のオーダーの計算時間を要するものであった。一方、第3章で提案しているアルゴリズムは、再帰的かつトップダウンな手続きと、部分木に対するラベルづけを組み合わせたもので、入力の木の違いが定数以下である場合に、線形時間で動作することが示されている。さらに、このアルゴリズムの木構造のアライメント問題に対する拡張も行っている。具体的には、類似度の高い次数限定の二個の順序無し木構造が入力された場合に、個々の編集操作が単位コストを要するというモデルのもとで、頂点間の最適対応関係（アライメント）を線形時間で計算するアルゴリズムを示している。

第4章では、ラベルつきパスの出現頻度に基づくグラフ構造の特徴ベクトル表現から、もとのグラフ構造を推定する問題について考察している。まず、この問題がサポートベクターマシンなどに代表されるカーネル法と関係があり、新規化合物の設計などに応用できる可能性があることを説明し、定式化を行っている。そして、グラフが木構造で、次数およびラベルの種類が限定され、かつ、特徴ベクトルに関与するパスの長さが限定されている場合に、この推定問題を解く動的計画法に基づくアルゴリズムを提案し、それが多項式時間で動作することを示している。一方、グラフを木構造に限定しても、次数

およびラベルの種類が限定されていない場合には、この推定問題が計算困難（強 NP 困難）であることを証明している。さらに、次数およびラベルの種類が限定されていても、パスの長さが限定されない場合には、木構造に対しても推定問題が計算困難となることも証明している。

第5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における構造データに対する離散最適化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) ブーリアンネットワークは遺伝子ネットワークの数理モデルの一つであり、その推定問題は複数のブール関数を同時に推定する問題として定式化されている。また、ブール関数およびブーリアンネットワーク推定のために従来から貪欲算法に基づくアルゴリズムが提案されており、計算機実験によって多くのブール関数に対し平均的に良い性能を示すことが示唆されていたが、理論的には非常に小さなクラスのブール関数に対してのみしか、そのことが示されていなかった。本論文では、unbalanced 関数と呼ばれるブール関数のクラスに対し、平均的に良い性能を示すことを理論的に証明し、かつ、そのクラスがほとんどのブール関数を含む広いクラスであることを示した。よって、これまで計算機実験により示唆されていたことを理論的に証明することに成功した。

(2) 順序無し木の比較問題は、RNA 二次構造や化学構造の比較と深い関連のある問題であるが、本論文では、二個の順序無し木が与えられた時に、最大共通部分木を計算するアルゴリズムを提案した。既存のアルゴリズムでは入力データのサイズの少なくとも二乗のオーダーの計算時間を要していたが、提案するアルゴリズムは、入力の木の違いが定数以内である場合に線形時間で動作することを証明した。また、このアルゴリズムを木構造のアライメント問題に対応するように拡張した。

(3) 新規化合物の設計を将来的な目標として、ラベルつきパスの出現頻度に基づくグラフ構造の特徴ベクトルから、もとのグラフ構造を推定する問題を定式化し、理論的考察を行った。具体的には、グラフが木構造で、次数およびラベルの種類が限定され、かつ、特徴ベクトルに関与するパスの長さが限定されている場合に、この推定問題を解く多項式時間アルゴリズムを示した。一方、グラフ構造を木構造に限定しても、次数およびラベルの種類が限定されていない場合には、この推定問題が計算困難（強 NP 困難）となることを証明し、さらに、次数およびラベルの種類が限定されていても、パスの長さが限定されない場合には、やはり木構造に対しても計算困難となることも証明した。

以上、本論文はバイオインフォマティクス（生命情報学）において重要な研究テーマである構造データに対する離散最適化問題に関して、独創的かつ多くの示唆を与えるアルゴリズムや理論的結果を示しており、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18年2月7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。