

氏 名	み うら てる ひさ 三 浦 輝 久
学位(専攻分野)	博 士 (情 報 学)
学 位 記 番 号	情 博 第 25 号
学位授与の日付	平 成 13 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 社 会 情 報 学 専 攻
学 位 論 文 題 目	ゲノム配列解析のためのアルゴリズムの研究

論文調査委員	(主 査) 教 授 石 田 亨	教 授 守 屋 和 幸	教 授 上 林 彌 彦
--------	--------------------	-------------	-------------

論 文 内 容 の 要 旨

本論文は、ゲノム配列解析手法である配列歩行問題と配列整列問題のためのアルゴリズムに関する研究をまとめたものである。両問題を配列類似度に基づく配列解析問題として定式化し、効果的に問題を解くためのアルゴリズムについて論じており、8章からなっている。

第1章は序論であり、本研究の背景、目的について概観している。

第2章では、ゲノム情報学の現状と本論文が扱う配列歩行問題と配列整列問題について述べている。ゲノム情報学は配列データの増加とともに生じた研究分野であり、配列データを情報科学の技術を用いて解析するものである。本論文で取り上げる問題は、配列の類似度に基づく解析問題である。

第3章では、配列歩行問題について説明し問題としての定式化を行っている。配列歩行問題は配列断片データベースを検索することを繰り返し、生物学的実験を行わずに遺伝子配列を推測し配列決定の際の時間と費用を削減する手法である。従来の配列歩行の作業は1日がかかりになる場合もあり、研究者にとり非常に負担となる。この原因はデータベースの検索に類似配列検索用ツールであるBLASTを用いていることに起因している。BLASTを配列歩行に用いることの問題点は、(1)類似検索ツールであるため不要な計算を実行し、配列歩行に適さない配列を出力する点、(2)計算量を落とすために多数のヒューリスティックを利用しているため配列歩行を制御できない点である。

第4章では、本研究で実装した配列歩行システムについて述べている。配列歩行問題は類似検索問題ではなく、誤りを許す文字列照合問題であることが示される。さらに従来のビット並列化による文字列照合アルゴリズムを配列歩行に適用するためのアルゴリズムの改良が提案される。このアルゴリズムを用いた実装システムはWWWを通して公開されている。実装システムは一致部分の最低一致長、誤り率、問い合わせ配列を与えられると自動的にデータベース検索を繰り返し配列歩行を行ない、従来のBLASTを用いるより高速な配列歩行を可能とする。

第5章は、配列整列問題についての定式化を行ない、従来研究について述べている。配列整列問題は与えられた複数の配列を整列させることにより、共通のパターンを抽出する問題である。できるかぎり多くの配列を同時に整列させることは、アライメントの信頼性を高める上で非常に重要である。しかし同時に多くの配列を整列させるためには非常に多くの記憶量が必要であり、従来のアルゴリズムにとって困難な問題である。配列整列問題を探索問題として定式化することで、探索アルゴリズムを適用することが可能になる。しかし最良優先探索や線形記憶量探索に基づく従来研究から、配列整列問題がどちらのアルゴリズムにとっても困難な問題であることが明らかになっている。配列歩行問題は探索問題として(1)非常に大きい分岐係数を持つ、(2)状態空間が束をなしているために同一節点を通る多数の経路が存在する、(3)辺のコストが多くの実数値をとるなどの特徴を持つ。この特徴が従来のアルゴリズムが配列を効果的に整列できない原因である。

第6章では、配列整列問題を効果的に解くために、新しい探索アルゴリズム、確率的節点記憶方式を提案し、配列整列問題に対する効果を検証している。さらに部分的節点展開方式が配列歩行問題の特徴を利用することで効果的に配列を整列さ

せることを示している。確率的節点記憶方式は従来の線形記憶量探索アルゴリズムの問題点である再訪を効果的に減らし、同様に節点を記憶し再訪を減らす MREC に比べ約 3 倍の速さで 7 配列の整列問題を解くことができることが示される。次に最良優先探索に基づく部分的節点展開方式を配列整列問題に適用することで、部分的節点展開方式が記憶量を効果的に抑え、従来最良優先探索が解くことができなかった 8 配列の配列整列問題を解くことができることが示される。

第 7 章では、探索分野における本研究の位置づけを行なっている。記憶制約下での探索アルゴリズムを最良優先探索 A* に基づくものと線形記憶量探索に基づくものとに分類し、それぞれのアルゴリズムで用いている探索の効率化手法に基づいて整理している。

第 8 章は結論であり、本論文で得られた成果について要約している。

論文審査の結果の要旨

本論文は、ゲノム配列解析の問題である配列歩行問題と配列整列問題を定式化し、問題を効率的に解くためのアルゴリズムに関して研究した成果をまとめたものであり、得られた主な成果は次のとおりである。

1. 従来、類似検索ツールを用いて実行されていた配列歩行の問題点を指摘し、配列歩行問題が文字列照合問題であることを明らかにした。従来の誤りを許す文字列照合アルゴリズムを改良し配列歩行システムを実装することにより、効果的な配列歩行を可能とした。

2. 配列整列問題を探索問題として定式化し、特徴を分析することで効果的な探索アルゴリズムを提案し有効性を示した。まず線形記憶量探索の再訪を減らすために確率的節点記憶方式を提案し、配列整列問題に対する有効性を示した。確率的な節点記憶という単純な機構が再訪を効果的に抑える節点の選択を可能とし、提案手法は従来の MREC に比べ 3 倍の速さで解を求めることができる。次に最良優先探索に基づく部分的節点展開方式を適用し、有効性を示した。本方式が問題の特徴を効果的に利用し、A* に比べ約 1120 の記憶量で探索を実行でき、A* が記憶量の制約により解けない 8 配列の配列整列問題を解けることを示した。

3. 従来の探索アルゴリズムを記憶量の利用法という観点から整理した。探索問題の対象となるデータが大規模になり、記憶量制約下で効果的に問題を解く探索アルゴリズムの重要性は高まりつつある。

以上、本論文はゲノム情報学における配列解析問題において、記憶量や計算量の制約下で効果的に問題を解決するアルゴリズムに関する研究をまとめたものであり、学術上、実際上寄与するところが少なくない。

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 13 年 1 月 24 日に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。