

論文審査の要旨および学識確認結果

報告番号	㊶／乙第 号	氏 名	秋山 真那斗
論文審査担当者：	主査	慶應義塾大学教授	博士（理学） 榊原 康文
	副査	慶應義塾大学教授	工学博士 岡 浩太郎
		慶應義塾大学教授	博士（地球環境科学） 土居 信英
		慶應義塾大学准教授	博士（工学） 舟橋 啓
<p>(論文審査の要旨)</p> <p>学士(理学)、修士(理学) 秋山真那斗君提出の学位請求論文は、「Development of RNA informatics for RNA sequence and structure analysis (RNA 配列と構造解析のための RNA インフォマティクスの構築)」と題し、4 章から構成されている。</p> <p>RNA はセントラルドグマを構成する三大生体高分子の一つである。その中でもタンパク質に翻訳されない非コード RNA (ncRNA) と呼ばれる RNA 分子は、ヒトゲノムの 70% の領域にコードされており、miRNA などのように生物活性を示すものもあることからポストゲノム時代において注目されている。このような膨大な量の非コード RNA を解析するためには、RNA インフォマティクスと呼ばれる情報解析技術が必須である。本論文は、RNA インフォマティクスの基本解析技術である二次構造予測、配列アライメント、クラスタリングという三つの課題に関して、最新の人工知能技術を応用した新しい手法を提案し、その有効性について報告している。</p> <p>本論文の第 1 章では、ncRNA とその情報解析分野である RNA インフォマティクスの重要性について述べるとともに、本論文で扱う三つの課題について概説している。</p> <p>第 2 章では、高精度な RNA 二次構造予測アルゴリズムの開発について報告している。ncRNA の機能は二次構造と密接に関連しているため、二次構造が決定されるとその生物学的機能を推測することができる。RNA 二次構造を予測するための一般的なアプローチは、熱力学的に最も安定した最小自由エネルギー構造を見つけるための熱力学的モデルである。一方で、予測精度の改善のために、より詳細な特徴量のモデリングが可能である機械学習に基づくアプローチが開発されている。機械学習ベースの詳細な特徴量を持つモデルは、熱力学的モデルと比較して予測精度において高いパフォーマンスを達成したが、学習データに過剰適合する問題点が報告されている。本論文では、熱力学的モデルと機械学習ベースのモデルを統合する RNA 二次構造予測のための新しいアルゴリズムを提案している。ベンチマークテストでは、提案アルゴリズムは大きな過剰適合を起こさず、既存の方法と比較して最高の予測精度を達成している。</p> <p>第 3 章では、RNA 配列および塩基のベクトル化技術の開発を行っている。DNA 配列やアミノ酸配列などの非構造データ（非数値データ）をベクトル化する埋め込みという技術が、DNA 配列モチーフの検出やタンパク質の機能予測の性能を向上させるために利用されている。一方で、RNA 配列の効果的な埋め込み技術はこれまで開発されていない。本論文では、RNA 配列を効果的に埋め込むために事前学習アルゴリズムを適用して、構造情報や配列の文脈情報を効果的に取り込んだ RNA 配列の埋め込みベクトルを計算する手法を開発している。事前学習によって得られた埋め込みベクトルの精度を検証するために、二つの基本的な RNA インフォマティクスの課題（構造配列アラインメントと遺伝子のクラスタリング）によるテストを実施し、既存の方法よりも優れた精度を達成している。</p> <p>第 4 章では、本研究を総括するとともに、開発した RNA 解析手法について今後の応用可能性について議論している。</p> <p>以上結論として、RNA 二次構造予測および RNA 塩基配列のベクトル化という二つのアプローチを用いて ncRNA の効果的な解析手法を開発することで、バイオインフォマティクス分野の発展に大きく資することができた。よって、本論文の著者は博士(理学)の学位を受ける資格があるものと認める。</p>			
学識確認結果	<p>学位請求論文を中心にして関連学術について上記審査会委員で試問を行い、当該学術に関し広く深い学識を有することを確認した。</p> <p>また、語学（英語）についても十分な学力を有することを確認した。</p>		