

Title	Identification and Analysis of Critical Sites in RNA/Protein Sequences and Biological Networks(Abstract_要旨)
Author(s)	Bao, Yu
Citation	Kyoto University (京都大学)
Issue Date	2018-09-25
URL	https://doi.org/10.14989/doctor.k21393
Right	
Type	Thesis or Dissertation
Textversion	ETD

(続紙 1)

京都大学	博士 (情報学)	氏名	鮑 雨 (BAO Yu)
論文題目	Identification and Analysis of Critical Sites in RNA/Protein Sequences and Biological Networks (RNA・タンパク質配列および生体ネットワークにおける重要部位の検出と解析)		
(論文内容の要旨)			
<p>本論文は、生物配列データおよび生物情報ネットワークデータにおける重要部位の予測・同定について述べられており、5章から構成されている。</p> <p>第1章では、研究の背景と動機、論文理解のために必要となるサポートベクターマシン (SVM)、カーネル関数、整数計画法などについての基礎事項、論文の構成について述べている。</p> <p>第2章では、RNA切断酵素によるマイクロRNAの切断部位をマイクロRNAの二次構造データ (配列データ+結合塩基対データ) から予測するための新規手法を提案している。この問題に対して従来から提案されていた、サポートベクターマシン (SVM) と二次構造の特徴を反映したカーネル関数の組み合わせについて説明した後、二次構造に関するより詳細な構造情報を反映した新たなカーネル関数を提案している。そして提案手法の有効性を示すために、既知のマイクロRNA切断部位データを用いた計算機実験により、既存手法との比較を行っている。比較結果は既存の2種類の二次構造予測プログラムのどちらを用いるか、どのような評価尺度を用いるかなどによって異なるが、多くの場合に提案手法により予測精度が数%上昇することを示している。</p> <p>第3章では、タンパク質切断酵素Caspaseによるタンパク質切断部位予測手法の網羅的な比較を行っている。まず既存の12種類の予測手法・ツールの特徴について詳細に説明した後、既知の切断部位データから複数のベンチマークデータを作成し、有用と考えられる6種類の予測手法の予測精度を比較している。その結果、予測精度はデータにより異なるが、配列パターンのスコア化に基づく手法、および、サポートベクター回帰を用いた手法が概して良い予測精度を持つことを示している。さらに計算機による予測の有用性を検証するために、切断されるかどうか未知のペプチド (短いアミノ酸の鎖) に対し、生物系研究者と共同で実際に切断されるかどうかの生物学実験を行い、その結果として切断される可能性が高いと予測されたペプチドの多くは実際にも切断されることを確認している。</p> <p>第4章では、生物情報ネットワークの制御に有用な頂点の新規選択手法を提案している。既存研究により、多くのダイナミクスに基づく有向ネットワークにおいて、帰還点集合 (Feedback Vertex Set) を制御すればネットワークを指定された定常状態に導けることが示されていた。しかしながら、最小のFVSは必ずしも一意に定まるとは限らないため、頂点の有用性を一意に評価できないという問題点があった。そこで、本論文では類似問題に対して定義されていた概念をFVSに適用し、必須頂点 (すべての最小FVSに含まれる頂点) と冗長頂点 (どの最小FVSにも含まれない頂点) という概念をFVSに対して定義している。そして、この必須頂点と冗長頂点を効率的に同定するために、ヒューリスティックな複数の規則に基づく前処理を提案し、それが厳密性を損なわないものであることを証明している。さらに、この前処理と整数計画法を組み合わせた計算手法を開発し、計算機実験によりその有効性を示している。また、実際の生物情報ネットワークデータの解析にも適用し、必須頂点の多くは生物学的に重要な遺伝子に対応することを示している。</p> <p>第5章は結論であり、本研究をまとめるとともに、今後の研究の方向性や課題について述べている。</p>			

注) 論文内容の要旨と論文審査の結果の要旨は1頁を38字×36行で作成し、合わせて、3,000字を標準とすること。

論文内容の要旨を英語で記入する場合は、400～1,100 wordsで作成し
審査結果の要旨は日本語500～2,000字程度で作成すること。

(論文審査の結果の要旨)

本論文は、RNA切断酵素によるマイクロRNAの切断部位予測、タンパク質分解酵素によるタンパク質切断部位予測についての機械学習手法、および、生物情報ネットワークにおける制御頂点同定についてのグラフ理論に基づく計算手法について述べたものであり、得られた成果は以下のとおりである。

(1) マイクロRNAの二次構造データからの切断部位予測については従来からサポートベクターマシン (SVM) を用いた予測手法が提案されていた。そこではRNA二次構造の特徴を反映したカーネル関数が提案・利用されていたが、本論文でそれを拡張した新たなカーネル関数を提案し、計算機実験により既存手法より優れた予測精度を持つことを示した。このカーネル関数ではループ領域の長さや位置に関する情報をカーネル関数に埋め込むことにより、二次構造情報をよりの確にSVM予測に反映することが可能となっており、新規性および有用性の高いものである。

(2) タンパク質切断部位予測は細胞の維持や分解のメカニズム解明に役立つ重要な研究課題であり、従来から数多くの予測手法が提案されてきた。しかしながら、どの手法が本当に有用かが明確ではなかった。そこで、本論文では複数のベンチマークデータを作成し、それを用いた計算機実験により、有用性の高いと考えられる6種類の予測法の比較・評価を行った。その結果、生物種やCaspase分解酵素の違いに応じて異なる手法の有効性が高いことなどが判明した。さらに、生物学実験による結果として、切断部位である可能性が高いと予測された部位は実際にも切断部位となっていることが判明した。これらの結果は、生物学者らが実際のデータ解析を行う際の指針の一つを与え、かつ、今後の研究の方向性に示唆を与える有用なものとなっている。

(3) 有向ネットワークで表現されるシステムの定常状態への制御のために、帰還点集合 (Feedback Vertex Set, FVS) を用いることが従来から提案されてきた。本論文では、最小サイズのどのFVSにも必ず含まれる頂点 (必須頂点)、どれにも含まれない頂点 (冗長頂点) を同定するための計算手法を提案した。計算手法の概要は、必須頂点や冗長頂点となる頂点の一部を前処理により同定した後、整数計画法を繰り返し適用するというものである。そして、この前処理のため新規にいくつかのヒューリスティックな規則を提案し、それが厳密性を失わない正当なものであることを証明した。さらに、計算機実験により、グラフが疎な場合に前処理が非常に有効であることを示すとともに、実際の生物情報ネットワークに適用して得られた必須頂点の多くが生物学的に重要なものであることを示した。

以上、本論文ではバイオインフォマティクスにおける重要な研究課題である生物配列データおよび生物情報ネットワークデータにおける重要部位の予測・同定について、機械学習およびグラフ理論に基づく新たな手法を提案し有用性を示すとともに、計算機実験による既存手法の比較・評価を行っている。提案手法のいずれもが新規性、有用性が高く、かつ、既存手法の評価結果は有用性の高いものであり、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30年7月18日に実施した論文内容とそれに関連した口頭試問を行った結果合格と認めた。

注) 論文審査の結果の要旨の結句には、学位論文の審査についての認定を明記すること。

更に、試問の結果の要旨（例えば「平成 年 月 日論文内容とそれに関連した
口頭試問を行った結果合格と認めた。」）を付け加えること。

Webでの即日公開を希望しない場合は、以下に公開可能とする日付を記入すること。
要旨公開可能日： 年 月 日以降