氏名	西郷浩人（さいごう ひろと）
学位(専攻分野)	博士（情報学）
学位記番号	情博第221号
学位授与の日付	平成18年7月24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	Local Alignment Kernels for Protein Homology Detection （タンパク質相同性検出のための局所アラインメントカーネル）
論文調査委員	（主査）教授 阿久津達也　教授 後藤　修　教授 山本章博

論文内容の要旨

本論文は，バイオインフォマティクスにおいて基本的かつ重要である生物学的配列の相同性検出を行う方法について述べており，7章から構成されている。

第1章は本論文の背景となる既存研究を概観し，動機，研究の独自性について述べている。

第2章ではタンパク質の構造，機能別クラスについて述べた後，代表的な相同性検出法であるSW（Smith-Waterman）アルゴリズムを説明している。次に，近年盛んに研究されているサポートベクターマシン（SVM）とそのカーネルについて触れた後，SWアルゴリズムをカーネルとして扱うことの可能性について述べている。

第3章ではまず，Convolutionカーネルという，部分領域に対するカーネルから全体に対するカーネルを導き出す一般的な手法を振り返っている。一方，SWアルゴリズムなどの局所アラインメントとよばれる手法においては，アラインメント後の領域を良く似たアミノ酸同士の一致領域，（その間に入りうる）ギャップ領域，アラインメントの前後の領域の3つに分けることができるが，SWアルゴリズムはカーネルとしての定義を満たさないことが知られている。そこで，本論文では，SWアルゴリズムとConvolutionカーネルを組み合わせ，それらの領域それぞれにカーネルを定義し，全ての可能な組み合わせを考慮することでLA（Local Alignment）カーネルを定義している。さらに，LAカーネルとSWアルゴリズムの間にある理論的な関係が成立することを示している。また，LAカーネルが動的計画法を用いて効率良く計算できることを示すとともに，実際のデータ適用時に発生する対角優位性問題への対処法も提案している。

第4章ではLAカーネルを実際の問題に適用した例を示している。最初の例は微弱なタンパク質の相同性検出とよばれる問題で，目標はタンパク質の構造，機能別のクラスを配列情報のみから学習することである。SVMとカーネルを組み合わせた既存手法はいくつかあるが，代表的な手法であるSVM-Pairwise，SVM-Fisher，SVM-Mismatchとの計算機実験による比較を行っている。その結果として，LAカーネルが最も良い性能を持つことが示されている。もう一つの例として，タンパク質のシステイン架橋構造予測という問題に適用している。この問題に対しても，計算機実験により既存のカーネルとの比較を行っている。その結果として，LAカーネルはシステイン架橋構造予測に対しても良好な性能が得られることを示している。

第5章ではLAカーネルのパラメータ最適化法について検討している。SWアルゴリズムおよびLAカーネルのいずれにおいてもアミノ酸置換行列というパラメータ集合を用いているが，4章までは既存の置換行列を用いていた。この章ではアミノ酸置換行列を微弱な相同性検出のために最適化する手法を提案している。具体的には，LAカーネルの微分の計算が動的計画法を用いて効率良く行えることを示し，その微分と最急降下法を組み合わせた方法を提案している。そして，SWアルゴリズムと最急降下法を組み合わせた既存手法と比べて，提案手法では最適化が滑らかに進展するという長所があることを示している。また，複数のデータセットを用いた計算機実験を通して，LAカーネルに最適化されたアミノ酸置換行列が，SWアルゴリズムに最適化されたアミノ酸置換行列よりも微弱な相同性検出において優れていることを示している。さ

らに，これらの実験結果から，LA カーネルは単体で用いても SW アルゴリズムよりも優れている場合があること，および，パラメータの最適化においても SW アルゴリズムより優れていることを指摘している。

第 6 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。また，付録として第 7 章において，第 3 章で示した定理の証明を記載している。

論 文 審 査 の 結 果 の 要 旨

本論文は，配列情報からのタンパク質の相同性検出手法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。

⑴ タンパク質の相同性検出はバイオインフォマティクスにおいて重要な問題であり，SW（Smith-Waterman）アルゴリズムなどの配列アラインメントに基づく方法が広く利用されてきた。近年はサポートベクターマシン（SVM）という分類アルゴリズムを利用する様々な方法が提案されている。本論文では，SVM と組み合わせて使う類似度関数「LA（Local Alignment）カーネル」を提案した。そして，LA カーネルがカーネルとしての性質，すなわち，正定値性を持つことを示し，さらに，カーネルの計算が動的計画法を用いて効率的に行えること，および，SW アルゴリズムとの間に成立する理論的な性質を示した。また，実データに適用する際に生じる対角優位性問題への対処法も示した。

⑵ LA カーネルと SVM との組み合わせの有効性を示すために，微弱な相同性検出問題，および，システイン架橋構造予測問題に対し，実際のタンパク質配列データを用いた計算機実験を行った。その結果として，提案手法は代表的な既存手法と比較して，十分に有効性を持つことが示された。

⑶ SW アルゴリズム，LA カーネルのいずれにおいても，相同性検出における性能はアミノ酸置換行列とよばれるパラメータ集合が影響している。そこで，置換行列を最適化する新たな手法を提案した。具体的には，LA カーネルが高階微分可能な滑らかな関数であることを利用し，最急降下法と組み合わせて置換行列を最適化する手法を提案した。また，LA カーネルの微分の計算が動的計画法を用いて効率的に行えることを示した。そして，微弱な相同性検出に適した目的関数を設定して，実際のタンパク質配列を用いた計算機実験を行った。その結果，LA カーネルを用いる提案手法は既存手法より優れていることが示された。また，LA カーネルを SVM との組み合わせでなく単体で用いても，SW アルゴリズムより優れている場合が数多くあることが示された。これらの結果より，本論文で提案した LA カーネルは，SVM との組み合わせ，配列類似性のスコア関数としての単体での利用，最急降下法との組み合わせによるアミノ酸置換行列の最適化のいずれにおいても有効であることが示された。

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タンパク質の相同性検出において，近年の機械学習の成果と組み合わせた独創的な方法を提案するとともに，多くの計算機実験を通じてその有効性を示しており，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 6 月12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。