

|  |  |    |                   |
|--|--|----|-------------------|
| 京都大学   | 博士 (情報学)   | 氏名 | LI Ruiming (李 瑞明) |
| 論文題目   | Analysis of Biological Networks by Graph Theory-based Methods<br>(生物情報ネットワークのグラフ理論に基づく解析法) |    |                   |
| (論文内容の要旨)  |  |    |                   |
| <p>本論文は、グラフ構造を持つ生物情報データ、具体的にはタンパク質相互作用ネットワークとタンパク質立体構造データに対する情報解析手法と評価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、バイオインフォマティクスにおけるグラフ構造データ解析における機械学習に基づく手法とグラフ理論に基づく手法の適用範囲の比較、関連研究、成果の概要、論文の構成について述べている。</p> <p>第2章では、まず、本論文で用いるグラフ理論や計算手法についての概念、具体的には、グラフデータ構造、帰還点集合、強連結成分、最大密度部分グラフ (densest subgraph)、線形計画法について説明し、次に、本論文で重要な役割を果たす生物学に関する概念、具体的には、がん遺伝子、タンパク質複合体におけるhot spot、タンパク質立体構造データベースについて説明している。</p> <p>第3章では、方向性を持つタンパク質相互作用ネットワーク・データと、正常細胞、および、がん細胞における遺伝子発現データの両者を用いることにより、がん遺伝子を予測するための計算手法を提案し、実際のデータを用いた計算機実験により有効性を評価している。具体的には、グラフ理論においてよく知られた概念である最小帰還点集合 (MFVS) をもとに、頂点に重みをつけ、重みの和が最大となる最小帰還点集合を求める問題を提案し、目的とする集合を重み付き最小帰還点集合 (WMFVS) と名付けている。次に、最小帰還点集合計算において最適性を損なわずにグラフデータを縮小するための既存の前処理法を説明するとともに、WMFVSに対する新たな前処理法を提案している。そして、MFVS問題に対する既存の整数線形計画法による計算法をもとに、WMFVS問題に対する整数線形計画法に基づく計算法を提案している。さらに、重みの和が最大となる帰還点集合 (WFVS) に対する整数線形計画法に基づく新規の計算法を提案している。そして、ヒトのタンパク質相互作用ネットワーク・データと遺伝子発現データを用いた計算機実験を行っている。具体的には、正常細胞とがん細胞の発現量の違いに基づき頂点の重みを定義し、ランダムに遺伝子を選択する手法、重みの大きな遺伝子を選択する手法、MFVS法、WMFVS法、WFVS法のがん遺伝子の予測能力についての比較を行っている。</p> <p>第4章では、タンパク質複合体の3次元構造データから、複数のタンパク質が結合するのに重要となる部位 (複数のアミノ酸)、すなわち、hot spotとよばれる部位をグラフ論的概念に基づき予測する方法を提案している。まず、最大密度部分グラフ、および、すべての極小最大密度部分グラフの組み合わせとそれらの線形計画法に基づく既存手法を詳細に説明し、次に、極大最大密度部分グラフの整数線形計画法、および、線形計画法に基づく新たな計算手法を提案し、局部極小最大密度部分グラフを組み合わせる計算手法も新たに提案している。一方、タンパク質複合体の3次元データをグラフ構造データに変換するために、アミノ酸を頂点とし、アミノ酸間の距離とファンデルワールス半径に基づき辺の有無を判定して無向グラフを構成する方法を提案している。そして、そのグラフに対して最大密度部分グラフなどを計算し、その部分グラフに含まれる頂点をhot spotと予測する。提案手法の有効性を示すために、Protein Data Bankという実際のタンパク質立体構造データを格納したデータベースから取得したデータを用いて、二部グラフにおけるクリーク (完全部分グラフ) に基づく既存手法、最小カット木 (minimum cut tree) に基づく既存手法などとの計算機実験</p> |  |    |                   |

による比較を行っている。さらに、辺に重みがついた場合の最大密度部分グラフ問題を新たに定義し、その線形計画法に基づく計算手法を示している。

第5章は結論であり、本研究をまとめるとともに、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、有向グラフに対する重み付き最小帰還点集合の計算手法とその生物情報ネットワーク解析への適用手法、および、無向グラフに対する最大密度部分グラフとその変形版の計算手法とタンパク質複合体データ解析への適用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) 複数種類の生物情報データを統合して解析し、がんの発生や進展などの原因となる遺伝子、すなわち、がん遺伝子を予測することはバイオインフォマティクスのみならず、生物学や医学においても重要な研究課題である。本研究では、タンパク質相互作用データと遺伝子発現データを統合して解析するために、重み付き最小帰還点集合 (WMFVS) という概念を提案し、前処理によるグラフの縮約と整数線形計画法を組み合わせたWMFVSの計算手法を開発した。そして、タンパク質相互作用データと正常細胞、および、がん細胞の遺伝子発現データから重み付き有向グラフを構成する方法を提案した。さらに、ヒトにおけるそれらのデータと既知のがん遺伝子データを用いて、単純な計算手法や通常の最小帰還点集合に基づく計算手法などとの計算機実験による比較を行った。その結果、提案手法やその変形版が、より高い精度でがん遺伝子を予測できることが示された。さらに、dark geneとよばれる遺伝子発現量の変化だけでは検出困難ながん遺伝子の予測においても提案手法が有効であることも示された。

(2) タンパク質複合体の立体構造から、hot spotとよばれる結合に関する重要部位を予測することは、タンパク質の機能を解析する上で重要な研究課題である。本研究では、タンパク質立体構造をグラフ構造として表現し、そのグラフ構造からhot spotを予測するために、最大密度部分グラフに基づく計算手法を提案した。具体的には、最大密度部分グラフ、および、すべての極小最大密度部分グラフの組み合わせに対する線形計画法に基づく既存手法を利用する方法、および、極大最大密度部分グラフに対する整数線形計画法および線形計画法に基づく新規手法、複数の局部極小最大密度部分グラフを組み合わせる計算手法 (Min-SDS法) を開発した。実際のタンパク質複合体の立体構造データを使用し、変異体とのエネルギー差から実験的に同定されたhot spotデータを正解データとして、既存手法と提案手法を計算機実験により比較した。その結果、提案手法、特にMin-SDS法がより良い予測精度を持つことが示された。さらに、将来の拡張性を考慮し、辺に重みを与えた場合の最大密度部分グラフの線形計画法に基づく計算手法を開発し、その正当性を示した。

以上、本論文ではグラフ構造データに対する実用的な計算手法という情報学における重要な研究課題に取り組み、開発した計算手法の生物情報ネットワーク解析によるがん遺伝子予測、および、タンパク質複合体における結合部位予測への適用法を提案し、それぞれの手法を実際の生物学データを用いた計算機実験により評価した。提案手法のいずれもが新規性、有用性が高く、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、令和4年12月26日、論文内容とそれに関連した事項について試問を行った結果、合格と認めた。なお、本論文のインターネットでの全文公表についても支障がないことを確認した。