

審査の結果の要旨

論文提出者氏名 葛 錫金

デジタル化の急激な進行の中で大規模な情報が容易に入手できるようになったが、そこから必要な情報、知識を抽出することは容易ではない。本研究では、知識抽出の問題を情報圧縮と補完とを含む逆問題と考え、冗長性圧縮による知識抽出を行うための多変量解析、非負条件下での単純な順結合階層型ニュートラルネットワーク及び複雑システム間の写像による可視化の3つの手法を開発した。知識抽出の対象としては、2D画像から生体情報に至る質的に異なる課題を選定して、本研究で開発した3手法の適用可能性について検討した。その結果、非負行列分解 (Non-negative Matrix Factorization 以下NMFに略) を使用して、局所的構造要素の重ねあわせから全体像をつかむことができるようになった。また、不十分なデータの補完についてもクラスタリングによる手法を提案し、多様なデータベース群からの知識抽出についての見通しを得て、その結果を8章からなる論文としてまとめている。

第1章は序論であり、知識抽出に関する研究の現状を要約し、本研究の目的および論文構成について述べている。

第2章は標準的多変量解析手法としての主成分解析 (Principal Component Analysis 以下PCAと略) について説明し、第3章では、本研究で開発したニュートラルネットワークすなわち非負行列分析 (Non-negative Matrix Factorization 以下NMFと略)、非負自動連想 (Non-Negative Auto association 以下NNAと略) について説明し、PCA、NMF、NNAをHintonらによる標準パターンに適用した結果を、精度、収束性、抽出内容、適用可能性等の観点から論じている。PCAは、全体像の特定の部分空間への線形写像であるが、本研究で使用したニュートラルネットワークは3層から成り、入力層と出力層のニューロンは同数とし、NNAモデルでは、すべてのシナプスが正または0に限定することにより、統計的相関ではなく、自己説明性のある解析結果を得ている。

第4章は、全体一部分関係の抽出に関する検討で、人の顔への適用例を示し、NNAとNMFの違いを内部層の非線形性、誤差逆伝播学習アルゴリズムの違いによるものとしている。第5章はテキストへの適用例で手書き、漢字、英単語の解析例を示し、全体一部分関係の抽出が効果的に行われていることを示し、次に文献データベースの分類についての検討結果を述べている。第6章は、実用的な課題として、遺伝子発現パターンによるガン分類への適用について述べている。部分群の認識も重要な課題であるが、与えられたパターンがどの集合に属するか正確に判断することも同様に重要なことであり、本論文では、材料シミュレーションで幅広く使用される原子間ポテンシャルの考えに基づいた新しい分

類アルゴリズムを提案している。この方法では、パターンを原子で表現し、個々のパターンの類似性により原子間力を定義し、類似のパターンが集合体を形成する場合、系全体が安定状態となるとして、大量のデータの中にあるパターンの抽出を試みている。つまり、クラスタリング問題を安定構造を見出す問題に変換することにより、分子動力学で活用されている計算手法、解析手法、例えばランダムな構造の場合の標準的な例であるシミュレートドアニーリング手法などを利用可能にしている。アヤメのデータのクラスタリングを試み、PCAで得られた結果と類似の結果を得て、本手法の妥当性を確認し、実用的問題として、急性白血病を例に採った遺伝子パターンの分類を試みている。急性白血病には、急性リンパ球性白血病 (Acute Lymphoblastic Leukemia 以下ALLに略) と急性骨髄性白血病 (Acute Myeloid Leukemia 以下AMLに略) 2つのタイプがあり、ALLとAMLを区別することは、適切な治療を施すために重要なことである。ここでは遺伝子発現パターンに72人の患者のそれぞれ7129種類の遺伝子群を使用し、遺伝子発現量を成分に持つ巨大なベクトルに本アルゴリズムを適用することにより、2次元マップを得て、ALLとAMLの2種類を明らかに区別することが可能であることを示している。

第7章は、以上の事例についての議論、第8章は結論である。本研究は、大規模なデータベースから知識を抽出するため、全体一部分関係を抽出するための新しいニュートラルネットワークの手法及びクラスタリングによる複雑な関係の可視化手法を開発し、その手法がデータ一般の知識抽出に有用であることを示したものであり、工学分野のみならず他分野の知識抽出に寄与するところ少なくない。

よって本論文は、博士 (工学) の学位請求論文として合格と認められる。