

論文の内容の要旨

論文題目 **Classified Clustering for Quantitative Biological Data Analysis**
(数量的な生命科学データの解析に向けた分類付きのクラスタリング)

氏 名 瀬 々 潤

マイクロアレイなどに見られる近年の生命工学の発展は、大量の数量的な生命科学データを生み出した。そして、それらのデータをどのように生命科学の知識として活かしていくかが問題となっている。数量データの解析には、クラスタリングが利用される。クラスタリングの結果は、生成したクラスタが生命科学の既知の概念と対応が取れたとき、クラスタ内のデータの解釈に利用することができる。しかし、クラスタリングを行うことと、その後のクラス分類、既知の概念との対応付け、が独立して行われているため、必ずしも対応が取れるとは限らず、対応が取れない場合は、数量データの解釈は著しく困難となる。このため、対応の取れるクラスタを発見できるよう、独立したクラスタリングとクラス分類を結びつけ、互いに矛盾の無いクラスタを見つける手法が渴望されている。

本博士論文では、この問題を解決するため、クラス分類付きのクラスタリング“**Classified Clustering**”を提案する。この手法は、クラスタリングのグループ分割を同一の特徴を持つものに制限し、その分割の中でグループ間の級間分散を最大にする分割を発見する手法である。しかし、級間分散を最大にする分割を発見することは NP 困難であるため、探索空間を効率よく枝刈りし、最適解を現実的な時間で回答するアルゴリズムを開発した。そして、酵母のマイクロアレーデータを利用した実験で、このアルゴリズムが高速に動作することを実証した。さらに、**Classified Clustering** の応用を 2 つの異なったデータに対し行い、有用性を確認した。

一つの応用は、酵母のゲノムワイドな cis 制御配列の情報と遺伝子発現量データを利用し、Classified Clustering が既知の cis 制御配列と遺伝子発現量制御の関係が導けることを確認した。さらに、cis 制御配列の未知の組み合わせと、条件特異的な遺伝子発現制御を確認した。もう一つの応用は、酵母より複雑と考えられている人の遺伝子発現解析への応用である。本論文では肝癌の遺伝子発現データと、患者の病理学的特徴を含むデータに、Classified Clustering を適用することで、“腫瘍があり”かつ“男性”や、“腫瘍が無く”かつ“肝機能が正常”である、といった病理学的に特徴のある遺伝子クラスタを発見することに成功した。これらのクラスタは、従来からある k-中心法では見逃されたクラスタである。