

論文審査の結果の要旨

氏名 ポール トポン クマル

本論文は進化論的計算を用いたマイクロアレイデータの分類とガン関与遺伝子群の検出と題し、9章からなり、遺伝子発現量の分類及びバイオマーカーの同定を主題とし、マイクロアレイなどのデータを進化論的計算によって解析する手法を提案し、提案手法である遺伝的プログラミングによるルール統合手法の有効性を実験的に検証している。

1章では本論文の導入が行われている。研究背景及び研究目的、これまでに提案されている癌分類手法の利点や欠点について説明している。

2章では遺伝子が発現するまでの流れ、遺伝子発現量測定手法、Affymetrix GeneChip ソフトによって生成されたマイクロアレイデータの様々な特徴について議論している。また、幅広く用いられているマイクロアレイデータの詳細及び事前処理手法について説明している。

3章では、マイクロアレイデータから重要な遺伝子を抽出する決定論的及び確率的計算手法の詳細について説明している。マイクロアレイデータは、サンプル数と比較して、非常に多くの遺伝子を含んでいるのが特徴である。

4章では、様々なクラス発見及び同定手法の説明をしている。クラス発見とはサンプルを似たサンプル同士に分類する作業を、クラス同定とは新しいサンプルを既知のクラスにラベル付けする作業を指している。クラス発見は教師なし学習であり、クラスタリング手法が幅広く用いられている。クラス同定は教師あり学習であり、様々な機械学習手法が用いられている。この章では様々なクラスタリング手法及びクラス同定手法の説明をしている。

5章では、ランダム確率モデル遺伝的アルゴリズム (RPMBGA) を提案している。RPMBGA は GA の拡張であり、マイクロアレイデータから有用な遺伝子を選択するのに用いる。RPMBGA では、遺伝子は確率分布にしたがって選択される。この確率分布は前の世代の分布と現在の世代の周辺分布についての、ランダム性を伴う重みつき平均によって更新される。はじめに、サポートベクターマシーン (SVM) 及び k-Nearest Neighbor (kNN) を分類器として用い、遺伝子を一つ一つ用いた場合と、すべてを同時に用いた場合で分類する。さらに、Signal to noise ratio を用い、一つの遺伝子のデータ同定能力にもとづいて、一度に一つの遺伝子を選択する手法で分類している。最後に、SVM 及び kNN を分類器として RPMBGA によって実験を行っている。RPMBGA によって選択された遺伝子セットの分類精度は、一つ一つの遺伝子、全ての遺伝子、Signal to noise ratio によって選択された遺伝子のサブセットより高いものとなっている。さらに、RPMBGA によって選択された遺伝子セットのサイズは、GA や PBIL などの従来の進化論的計算によって選択されるセットよりコンパクトなものとなっている。RPMBGA によって選択された、もっとも高い分類精度の遺伝子セットに含まれる遺伝子は、単独では高い精度で分類することができない。そのため、最良の分類を行う遺伝子セットの遺伝子間には相関関係があると考えられる。一方で、RPMBGA などの進化計算手法では、一度に複数の遺伝子を選択するサブセットを生成するため、遺伝子間の依存関係が保持され、高い分類が可能となっている。

6章では、GP のマイクロアレイデータの解析への適用について説明をする。様々な遺伝子選択手法や分類手法が提案されているが、選択される遺伝子の正確さ及び分類精度は、分類器の能力に大きく依存している。遺伝的プログラミングは、分類器であると同時に遺伝子選択も行うこ

とができるという利点を持つ。さらに、獲得されたルールは分かり易い関数で与えられるため、定量的な遺伝子間の関係を推定することができる。しかしながら、マイクロアレイデータに対する実験の結果、クロスバリデーションを用いることが出来ないため通常の GP により得られた単独のルールでは分類精度が必ずしも高くないことが示されている。

7章では、Majority Voting Genetic Programming Classifier (MVGPC) を提案してしている。MVGPC は独立した試行で得られた GP ルールを統合することで、分類精度を改善した手法である。この手法は信頼性が高く、ロバストな分類を行うことができる。この章では、最適 Ensemble Size の決定手法、ルール統合手法を用いたテストサンプルのラベル推定の方法、マイクロアレイデータからのバイオマーカの抽出方法について説明している。バイオマーカの同定手法として、はじめに分類器を生成し、その後で獲得されたルールの中で頻出の遺伝子を抽出するという手法を提案している。遺伝子の定常的な頻度分布を獲得するためには、MVGPC をマイクロアレイデータに対して複数回適用する必要がある。様々なマイクロアレイデータに対して MVGPC を適用することで、AdaBoost+GP などの他の手法と比較して優れた性能であることを示している。さらに、MVGPC によって選択された遺伝子のうちいくつかは、知られている癌関与遺伝子と非常に深い関係があることを確かめている。

8章では、MVGPC をマイクロアレイ以外のデータに適用し、金融データなどに対しても MVGPC が有効であることを検証している。このデータではマイクロアレイデータの場合とは異なり、非常に多くのサンプルを用いることができる。実験結果から、GP を単独で用いた場合と比較して、MVGPC は高い精度で分類できるという結果を得ている。

9章では、本論文の結論及び本研究の課題について述べている。

以上これを要するに本論文は、進化論的計算に基づいた分類学習手法である RPMBGA と MVGPC を提案し、遺伝子発現データの分類及び癌のバイオマーカの同定に適用することでその有用性を実証しており、情報学の基盤の発展に貢献するところが少なくない。

したがって、博士（科学）の学位を授与できると認める。