

遺伝子ネットワークの解析と可視化システムの開発

●篠原 歩 ◆竹田 正幸

九州大学大学院システム情報科学研究院情報理学部門

研究の目的と進め方

本研究は、遺伝子の相互作用を明らかにするために、破壊や過剰発現による実験結果を集めた遺伝子発現プロファイルデータから、そこに内在する相互作用を表現するための規則をユーザとの対話の中で発見するためのシステムを開発することを目的とする。我々の研究グループでは、重み付きネットワークモデルを導入し、発現プロファイルデータから遺伝子間の力関係を辺の重みとして取り出すアルゴリズムを設計し、予備実験を行ってきた。このネットワークを効率よく可視化するための方法論を確立し、それを実装することが目的である。

2001 年度の研究の当初計画

グラフィカルモデリングと呼ばれる統計的手法の中で用いられている、共分散選択の技法を応用する。また、遺伝子発現プロファイルデータのみならず、遺伝子の配列情報を合わせてネットワークの同定をより高精度に行うことを目指して、配列データからそこに内在する共通なパターンを抽出する問題について、特に高機能化と高速化に力点をおいて研究を展開する。

2001 年度の成果

パターン抽出に関して、部分文字列パターンおよびエピソードパターンに関して、実用的な時間で最も分類精度の高いパターンを見つけだすアルゴリズムの開発に成功し、それを機械発見システム BONSAI の中に

組み込んだ。このことにより、これまでの部分文字列パターンのみを用いた表現方法よりも簡潔でより分類精度のよい決定木が抽出できることを確認した。

国内外での研究の位置づけ

パターン発見の問題は、テキストマイニングやデータベースからの知識発見の研究分野で大変注目されており、さまざまな方式が提案されているが、我々の方法は、目的とするスコア関数による評価値を最大化する部分列パターンやエピソードパターンを見つける非常に強力なものであり、幅広い応用が期待できる。

達成できなかったこと、予想外の困難、およびその理由

現時点では、新たなデータ集合に対する実験を行っておらず、実データに対する評価をしていく必要がある。

今後の課題

引き続き、遺伝子の配列情報を合わせてネットワークの同定をより高精度に行うことを目指す。そのため、配列データからそこに内在する共通なパターンを抽出する問題について、さらなる高機能化と高速化に力点をおいて研究を展開する。

成果公表リスト

- ・ M. Hirao, S. Inenaga, A. Shinohara, M. Takeda, and S. Arikawa, "A Practical Algorithm to Find the Best Episode Patterns", Proc. of Discovery Science 2001, pp.435-440.