

氏名	西田有一郎	学籍番号	150054
主テーマ指導教員	佐藤賢二	副テーマ指導教員	伊藤孝行
主指導教員	小長谷明彦	副指導教員	佐藤賢二

<研究テーマ>

遺伝子発現情報からの遺伝子機能推定に関する研究

<研究の目的>

本研究はマイクロアレイデータより得られる遺伝子発現情報だけでなく、現在蓄積されている遺伝子の機能情報等をも解析に取り入れた機能ネットワーク推定的手法を取り入れ、更に遺伝子のグループ発見としての **clustering** と遺伝子のグループへの機能割り付けのための **classifier** を組み合わせた多段的機能推定を行う方法論の提案と、それをを用いた実際のシステムの構築を目的とする。

この方法によって、現在多く行われているマイクロアレイデータの発現情報から目的が明確な機能解析ができると期待できると考えている。

<研究の背景・特色>

ゲノムプロジェクトの進展よりヒトゲノム塩基配列の決定は完成に近づいてきた。現在、遺伝子が 3 万～4 万と予測されている。それらの遺伝子の発現を網羅的に解析する手法としてマイクロアレイを用いるのは、いまや常道の手段となり、さまざまなマイクロアレイによる遺伝子発現データが得られている。それらのマイクロアレイによる発現データに対して **SOM** などによる、さまざまなクラスター解析結果が報告されている[1]が、本質的な問題として遺伝子のグループ分けとしての **clustering** は個々の遺伝子について生物学的な知見から自由な分類が可能である分、逆にいうと結果の妥当性と何の意味を示すかを検証するのが困難である。一方で遺伝子のグループに対する機能割り付けのための **classifier** は最近 **SVM** などから良好な結果が報告されている[2]が、**classify** するための前提、即ちそれぞれの割り付ける機能が互いに排反であるとして **classify** を行う。しかしながら、必ずしも遺伝子は一つの機能を持つだけではなく、複数の機能をもつこともあるため、機能同士が背反であると言う前提で **classifier** を行うのは他の機能を無いものとして扱ってしまうという問題がある。

そこで、この二つを組み合わせることで、**classifier** で解析目的とする情報を明確にし、その明確となった目的とする情報を反映した **clustering** によるクラスター解析を行う。これにより、**clustering** によってグループ化された遺伝子群がグループの意味をはっきりとさせつつ、複数の機能をもつ場合においても機能推定が可能となる。

更にマイクロアレイによる発現情報に関しては技術的に新しいこともあり、現在のところ非常にノイズの多いデータであることが知られている。それに対して、当研究室では先行研究として出芽酵母の発現情報に加えて、YPD と **ENZYME** の二つのデータベースから得られる機能情報や分類情報を導入した解析手法が考案されている。このようなマイクロアレイからの発現情報以外に機能を推定するための情報を加えたデータに対して上記の手法を適用しようと考えている。

<研究計画・方法>

1. 研究のための準備

- (a): アレイデータだけではなく、ゲノムデータベースからの情報を用いるために、ゲノムデータベースについてのサーベイをおこなう。
- (b): classifier からの目的情報を clustering に生かすための情報の効果的なモデル化を行う。
- (c): 出芽酵母遺伝子の機能を理解する為に、生物学全般に対するの知識を深める。

2. システムの構築

現在の時点では、classifier にしても clustering にしてもソフトコンピューティングに属する手法を用いた構築することを考えている。これはソフトコンピューティングがノイズに強いという特徴をもつ為であるが、一般的に計算量が多いのが難点である。また複数の手法を段階的に用いるため、計算量が膨大になることが予想される。そのためにシステムの最適化と、高速化を行う必要がある。

3. 結果に対する検証

検証を行う際に、重点とするのは、結果が生物学的に有用であるかという点と、解析目的に柔軟にシステムが対応できるかという2点である。しかし、前者を詳しく検証するためには、実際に実験を行って解析を待つ研究者の方と協議する必要があると思われる。

4. 改良とアプリケーション化

3で得た結果をもとに、必要ならば改良を行い、アプリケーション化を行う。アレイデータだけではなく、ゲノムデータベースの情報を用いるため、Web 上でのサービスとなると思われる。

5. 論文執筆

参考文献

- [1] 実験医学増刊 “ゲノム医科学と基礎からのバイオインフォマティックス”, 羊土社
- [2] Knowledge-based analysis of microarray gene expression data by using support vector machines,
Michael P. S. Brown, William Noble Grundy, David Lin, Nello Cristianini, Charles Walsh Sugnet, Terrence S. Furey, Manuel Ares, Jr. , and David Haussler*, *Proceedings of the National Academy of Science*. 97(1):262-267.

<現在までに単位取得した専門教科>

概論：3科目、方法論：2科目、専門：2科目

K215 イノベーション概論 B	K216 知識科学概論 A	K217 物理科学概論 A
K212 知識ベース方法論 B	K214 知識処理方法論 A	
K416 物質知識論	K417 知識創発論	