2V-4

ベイジアンネットワークによる遺伝子ネットワークの推定

島田 公敬†

芝浦工業大学大学院 工学研究科‡

1 はじめに

ゲノム規模の実験生物学アプローチによって、効率的にこれまでの仕事をこなすことができるようになったばかりでなく、生理現象に潜む普遍的な設計原理を探索することも可能になってきた。その一例として近年マイクロアレイ技術の発展に伴い、数千から数万の遺伝子の発現状態を同時に観測することが可能となった。また大規模な遺伝子発現の構造だけでなく、細胞周期や環境変化に対する応答、体内時計、発生・組織特異的な発現制御などの個別具体的な動的な生命現象を支える複雑なネットワーク構造を明らかにしてきた[1]。

このように、生命科学全体で分子・要素レベルからシステムネットワークレベルへと理解の枠組みが大きく変化しつつある[2]。その結果機能未知の遺伝子も含んだ膨大な相互作用ネットワークのデータが蓄積している。

この大量かつ多種類の情報を用いた遺伝子ネットワークの推定問題は、現在バイオインフォマティクスにおいて最も精力的に研究が進められている分野の一つである[3]。

本研究はこの遺伝子ネットワーク推定において広く使用されているマイクロアレイのデータからベイジアンネットワークモデルにより遺伝子ネットワークの推定方法を提案する。さらに他の生物学的データを用いてネットワークの修正を行い、より最適なネットワークを生成することをねらう。

2 マイクロアレイ

「遺伝子が発現する」とは本来「遺伝子を基にタンパク質が生合成される」ことを意味する[3]。遺伝子を基にしてタンパク質が合成されるプロセスは簡略化すると図1のようになる。

マイクロアレイは主に以下の 2 つに大別される。

- Affymetrix 社の光リソグラフィー法に基づくマイクロアレイ（GeneChip）
- Stanford 大学のスポット法に基づくマイアレイ（cDNA マイクロアレイ）

GeneChip は、cDNA マイクロアレイの 20 倍近い密度を持ち、規格製品であるため入手も比較的容易である。しかし、研究目的に合わせた注文生産が困難となる。

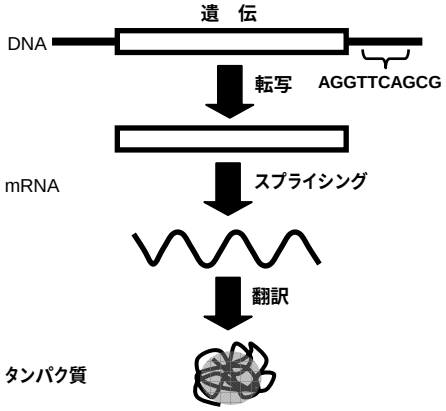

図1：遺伝子からタンパク合成のプロセス

一方 cDNA マイクロアレイは、研究目的にあわせてマイクロアレイをデザインでき、その簡便さから広く用いられている。実験方法は 2 種類の細胞を用意する(通常細胞・サンプル細胞)。正常細胞とサンプル細胞から全遺伝子に関して mRNA を抽出し、それを鋳型として cDNA を生成する。cDNA のコピーを稠密にスポットしたガラス表面に、細胞から抽出し蛍光標識(Cy3,Cy5)した遺伝子転写産物を結合させることで、遺伝子発現の変化を測定できる。測定サンプルにより発現の絶対量は異なるため、正常例と異常例のそれぞれを異なる色で蛍光標識し、標準となるスポットで正規化した後に、二色の相対強度を遺伝子発現量として測定する(GeneChip の場合ガラス表面への DNA の植え付け法が既知配列のスポッティングではなく、光リソグラフィーを用いた固相合成であることが多い)。

またこの手法は、発現パターンが似た遺伝子は機能的にも関連するという前提に基づいているが、これが生物学的に正しい保証はない。そのため、遺伝子の依存関係そのものをデータから学習する試みとしてブーリアンモデル、線形モデル、ニューラルネットワークが提案されてきた。しかしこれらのモデルは、実際に大量データを処理するには至っていない。初めて 800 遺伝子という大量データの解析に踏み込んだのがベイジアンネットワークモデルである[4]。

マイクロアレイ解析の問題点として常に挙げられるのは、以下の 3 つである。

1. 遺伝子発現は細胞内メカニズムの一面にすぎない
2. 数千の遺伝子に対し統計的に有意な結果を出せるほどデータが揃わない
3. データのノイズが非常に大きい

Estimation of Genetic Networks by Bayesian Netoworks
†Kmitaka Shimada
‡Gradeuate School of Engineering,Shibaura Institute of Techonology

3 ベイジアンネットワーク

ベイジアンネットワーク(以下 BN)とは 1988 年 Pearl によって最初に提案された確率推論のための枠組みであり、不確実性を含む対象領域における予測や意思決定を定量的に取り扱うことができる手法として注目を浴びた。

BN は設計変数間の依存関係を有向グラフにより表現する。BN は、以下のような性質を有するグラフ構造として提示される。

- 変数の集合がネットワークのノードを形成
- リンクまたは矢印の集合がノード対を結ぶ。ノード X からノード Y への矢印の直感的な意味は X が Y に直接的影響を与える(Y が X に依存する)
- グラフは矢印の方向にサイクルを持たない(有向非循環グラフ)

BN の例を図 2 に示す。

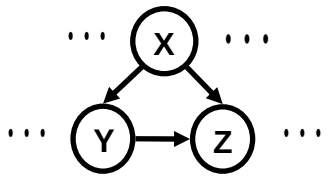

図 2：ベイジアンネットワークの例

図 2 において、ネットワークを構成する x, y, z という 3 つのノードがそれぞれ変数を表わしており、矢印は依存関係を示す。つまり z は x,y に依存し、y は x に依存することを示す。影響を与える変数(矢印の始点となるノード)を親ノードと呼び、影響を受ける(矢印の終点となるノード)を子ノードと呼ぶ。

4 本研究について

マイクロアレイの問題に対し、ベイジアンネットは以下のように対処できる。

1. マイクロアレイ解析は細胞集団の平均値を観察する技術であり、遺伝子発現の大局的な依存関係が観測される。遺伝子 X が Y を活性化する(または抑制する)という関係は、各遺伝子を変数とみなした条件付き確率モデルで表現できる
2. 観測する遺伝子数は数千あるが、データセットは多くても数十である。しかし、遺伝子ネットワークは非常に疎なグラフ構造を持つ。よって局所的な構造を学習することでデータの少なさを補うことできる
3. マイクロアレイで見る発現とは配列にも依存した物理化学反応(水素結合) であるため、実験の条件や計測法により発現量は大きく変化する。そのため、遺伝子発現のデータは発現レベルに応じて離散化する。マイクロアレイを用いた解析の多くで、データは-1,0,1 の 3 段階(非常に低い、同程度,非常に高い)に分けられる。通常のしきい値は、発現量が数倍変化する点に設定する

さらに本研究では、マイクロアレイの問題点の 2. と 3.に対して他の生物学的データを用いてベイジアンネットワークを学習させていくことで補う。この学習には遺伝子１つをエージェントにみたて、全体のネットワークを崩すことなく局所的にネットワークを修正することのできる、エージェント間交渉を用いる予定である。また修正に用いる生物学的データとして京都大学の KEGG(Kyoto Encyclopedia of Genes and Genomes)プロジェクト[5]ではこれまで蓄積されてきた代謝系やシグナル伝達系のデータなどを使用する。

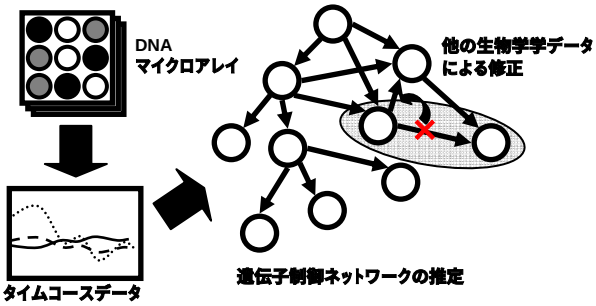

図 3：本システムの流れ

5 今後の展望

今回遺伝子ネットワークにおいて DNA マイクロアレイの問題点を補い、より最適なネットワークを構築するシステムの提案を示してきた。

今後システムを実装し、他の手法や既存の遺伝子ネットワークの推定結果などと比較検証を行っていく。

謝辞

本研究に関して、貴重なご意見を頂いた指導教員である相場亮教授に深く感謝いたします。

参考文献

[1] 上田 泰己:遺伝子発現ネットワークダイナミクス:人工知能学会誌: Vol.20,No.3,2005/5.
[2] 五斗 進:生体分子の相互作用ネットワークにおけるスケールフリー性,人工知能学会誌,Vol.20,No.3,2005/5.
[3] 井元 清哉:マイクロアレイデータ解析における統計的方法論の開発
[4] T.A.Brown,"Genoumes,"BIOS Scientific Publishers,1999.
[5] http://www.genome.ad.jp/kegg/