

グラフィカルモデルに基づく相互作用推定法の開発と適用

●堀本 勝久¹⁾ ◆藤 博幸²⁾ ◆油谷 幸代¹⁾ ◆吉田 寛³⁾

1) 産業技術総合研究所生命情報工学研究センター 2) 九州大学生体防御医学研究所 3) 九州大学大学院数理学研究院

<研究の目的と進め方>

本研究は、グラフィカルモデル (GM) に基づく統計因果推定法を生命情報データから相互作用を推定する方法の開発に適用し、遺伝子発現プロファイル、蛋白質配列・立体構造情報などの生物情報から生命現象におけるシステム間及び分子間の因果関係を推定することで、生命システムを構成する物質間の関係性の理解を深めることを目的とする。

具体的には、以下の通りである。

- 1) 近年開発したグラフィカル・ガウシアン・モデル (GGM) によるシステム間の関連構造推定法に基づいた、グラフィカル連鎖モデル (GCM) による因果推定法を確立する、
- 2) GM の一つである path consistency (PC) アルゴリズムの改良によって、分子間関連及び因果推定法を開発する、
- 3) 知識ベースに推定された関連・因果モデルに基づいて評価する方法を開発する、
- 4) GGM による関連構造推定法を、カテゴリカル・データが解析可能な混合法に拡張する、
- 5) 1) 2) 3) の推定法を統合することによって、遺伝子制御ネットワークの推定精度の向上を図る、
- 6) 2) 4) の統合により蛋白質間相互作用領域の推定を行う。

<2007年度の研究の当初計画>

2007年度は、2006年度開発した方法の精度向上と統合を行う。具体的には、

1. カテゴリカル・データを解析する混合法の開発 (研究目的 4)
2. 2006年度開発した推定法の統合による遺伝子制御ネットワークの推定精度の向上 (研究目的 5)
3. 2006年度の2と2007年度の1の推定法の統合による蛋白質間の相互作用推定 (研究目的 6)
4. 記号計算に基づくネットワーク解析手法の開発である。

1) によって、蛋白質構造の階層的な情報を相互作用領域推定に利用できる。

2) によって、遺伝子システム間及び遺伝子間のマクロとミクロなレベルの両方についての関連構造を統一的に推定する。

3) ではカテゴリカル・データと数量データの両方を同時に採用した推定法を構築する。2年間の方法開発によって、多様な生命情報データについて柔軟に対応できる関連及び因果推定法の開発を目指す。

4) では、上記研究目的に加え、ネットワーク解析について記号計算に基づく新規手法の開発を行う。近年のソフトウェア性能の向上で記号計算により大規模な変数消去が可能になっている。既知ネットワーク構造と同等な反応方程式を仮定することで、この特徴をネットワーク解析に導入する。2007年年度は、グラフィカルモデルで解析不能な巡回グラフを含むネットワークに関する解析の開発と計測不能な変数を含むネットワーク構造において計測不能変数を代数的に消去することによるネットワーク動態に関する手法の開発を行う。

<2007年度の成果>

グラフィカル連鎖モデル (GCM) の適用法については、細胞周期及び肝癌進展過程の解明のため、それぞれの遺伝子発現プロファイルに適用した。細胞周期については隣接する細胞状態の関係性に加え時間的に離れた細胞状態間の遺伝子発現の関連性を推定することができた。肝癌進展過程の解析では、臨床及び病理の知見と一致した関係性が確認された。

PC アルゴリズムの推定性能を確認するために、ノイズの多い生命情報への適用の前に、ノイズの少ない物性データで特徴づけられる薬理化合物の関連性解析に適用した。その結果、薬理化合物の物性的特徴を極めて忠実に反映した関連性グラフが推定され、それらが薬理効果をもよく反映していることが判明した。さらに、PC アルゴリズムのカテゴリデータ解析への適用手法を実装し、アルツハイマー遺伝子関連遺伝子に関する SNP データに適用した。その結果、既知の関連性遺伝子間の関係をネットワーク形式で表現することに成功した。

モデルと計測データとの整合性評価法については、ガウスモデルに基づく統計的方法を確立したので、シミュレーション及び計測データを用いて性能評価を行いその有効性を確認した。大腸菌の既知遺伝子制御ネットワーク 30 について、嫌気性下で計測された発現プロファイルとの整合性を推定した結果、嫌気呼吸関連ネットワークと maltose 代謝関連ネットワークの2つが整合性を示す結果を得た。共に嫌気性下に特徴的なネットワークであり、特に嫌気呼吸関連ネットワークは 150 以上の遺伝子から構成される複雑な構造を持つネットワークである。既知ネットワークデータについて、特異的条件下で活性化される制御ネットワークを厳格に検出する方法として有効であることがわかり、今後既知ネットワークの整備と様々な状況下で計測されたデータの適用により、新たな活性化ネットワークの発見を期待している。ただし、統計的手法では適用グラフが directed acyclic graph (DAG) に限定される。その問題を解決するために、代数的な手法を開発した。現在、シミュレーションによって cyclic loop を持つグラフについて有効であることを検証した。

GM によって推定されたネットワークは時間に応じて変化する多様なネットワーク構造のスナップショットに過ぎない。一方、数値的に推定され遺伝子間の関連性の強弱は、ネットワークの振る舞いを見積もるためのパラメータ最適化において初期値として利用できる。このような状況を鑑み、ネットワーク構造をより多面的に評価するために、動的な振る舞いから構造の吟味を開始した。また、動態解析の新規なアプローチとして、計算機代数に基づいた手法を考案し、その有効性を調査した。その結果、従来の数値解析アプローチでは解析解が非常に複雑になるためパラメータ推定不能である場合にも、ラプラス変換によって代数方程式に変換することで解析可能になる場合が発見された。

<国内外での成果の位置づけ>

グラフィカルモデルに基づくネットワーク解析法は、解析法の数学が理解が簡単である、利用に際して実装が容易である、パソコンレベル以上の計算機性能を要求しない、また、解析結果の直

感的理解が容易である、などの利点があり、実験研究との連携が容易であり、様々な共同研究が進行している。成果報告が部分的に遅れているが、今後その成果を発表することで多方面での研究が進行することが期待される。

また、グラフィカルモデルの一部は、代数統計学において近年盛んに研究されている。より広い視点から代数算法の生物学諸問題への有効的な適用法の開発を目指し、国際会議 Algebraic Biology を主催している。幸い国際的に注目を集めることができ、2007年度はヨハネス・ケプラー大学記号計算研究所（リンツ、オーストリア）で開催し、その要旨集は Springer 社から出版された（Anai, H., Horimoto, K. and Kutsia, T. (eds) *Algebraic Biology. Lecture Notes in Computer Science 4545*, Springer, Heidelberg）。2008年度も同じくヨハネス・ケプラー大学記号計算研究所で開催が決定し、2010年度は、米国で開催予定である。国際的な規模の共同研究により、代数算法によるネットワーク解析法の開発が期待される。

<達成できなかったこと、予想外の困難、その理由>

これまで開発した方法である、グラフィカル・ガウシアン・モデル、グラフィカル連鎖モデル、PC アルゴリズムの適用範囲が広く、グラフィカルモデルに基づく新規手法の開発が遅れている。また、グラフィカルモデルを含む統計的手法の欠点を克服するための代数手法の開発が順調であることも、グラフィカルモデルに基づく手法の開発遅延の原因と考える。

さらに、成果発表が遅れ気味である。手法開発の報告はできているが、新規生物学的知見の発見のために有効であることは各共同研究で確認済みであるにもかかわらず、その報告ができていない。既開発法の適用範囲が広いことが、各課題についての成果発表が遅れの原因となっている。

ネットワーク構造解析と動態解析の方法の統合、遺伝子発現情報とタンパク質相互作用情報の利用法に関する統合など、既開発法の統合に関する研究ができなかった。上記のように、各方法の有効性を示す研究に集中したためである。

<今後の課題>

課題は以下の通りである。

- I. 方法自体の開発に関する報告は済んでいるので、様々な生命情報データへの適用によって、生物学的知見発見のために有用であることを示す報告を行う。
- II. 新規方法の開発を行う。特にネットワーク構造解析と動態解析の橋渡しになる、ネットワーク構造変化を推定が可能な時系列データの解析手法を開発する。
- III. 開発方法の統合を行う。構造推定法による静的ネットワークについて動態解析法を有機的に連携し、データ入力により動的ネットワーク解析が可能なシステムを構築する。
- IV. 既開発法の公開を行う。現在、グラフィカル・ガウシアン・モデルの適用法のみ公開しているが、新規解析法を公開し、解析データの性質や解析目標など利用状況に応じた包括的なネットワーク解析サイトを構築する。
- V. 既開発法と生物情報データとの連携を行う。実験研究者が所有するデータの解析だけでなく、既知データの有効利用を実行できるように関連データベースとの連携が可能なシステムの構築を行う。

<成果公表リスト>

1) 論文/プロシーディング (査読付き)

1. 0801291226

Aburatani, S., Sun, F., Saito, S., Honda, M., Kaneko, S. and Horimoto, K., Gene systems network inferred from expression

profiles in hepatocellular carcinogenesis by graphical Gaussian model. EURASIP J. Bioinfo. Systems Biol., 47214, (2007).

2. 0702131228

Yoshida, H., Anai, H. and Horimoto, K., Derivation of rigorous conditions for high cell-type diversity by algebraic approach. BioSystems, 90, 486-495, (2007)

3. 0801291230

Yoshida, H., Horimoto, K., and Anai, H., Inference of probabilities over a stochastic IL-system by quantifier elimination. Math. Compu. Sci., (in press)

4. 0801291238

Sato, T., Yamanishi, Y., Horimoto, K., Kanehisa, M. and Toh, H., Inference of Protein-Protein Interactions by Using Co-evolutionary Information. In Anai, H., Horimoto, K. and Kutsia, T. (eds), Algebraic Biology 2007, Lecture Notes in Computer Science 4545, p.322-333, Springer, Heidelberg, (2007).

5. 0801291246

Aburatani, S., Inference of Complex Regulatory Network for Cell Cycle System in Saccharomyces Cerevisiae. In Anai, H., Horimoto, K. and Kutsia, T. (eds), Algebraic Biology 2007, Lecture Notes in Computer Science 4545, p.350-364, Springer, Heidelberg, (2007).

6. 0801291251

Yoshida, H., Nakagawa, K., Anai, H. and Horimoto, K. Exact parameter determination for Parkinson's disease diagnosis with PET using an algebraic approach. In Anai, H., Horimoto, K. and Kutsia, T. (eds), Algebraic Biology 2007, Lecture Notes in Computer Science 4545, p.110-124, Springer, Heidelberg, (2007).

7. 0801291257

Yoshida, H., Nakagawa, K., Anai, H. and Horimoto, K., An Algebraic-Numeric Algorithm for the Model Selection in Kinetic Networks. Proceedings of 10th CASC, Lecture Notes in Computer Science 4770, p.433-447, Springer, Heidelberg, (2007).

8. 0801291303

Hayashida, M., Sun, F., Aburatani, S., Horimoto, K. and Akutsu, T., Integer Programming-based Approach to Allocation of Reporter Genes for Cell Array Analysis. Proceedings of 10th OSB, Lecture Notes in Operation Research, p.288-301, World Publishing Corporation, Beijing, (2007).

9. 0801291306

Katoh, K., Toh, H., PartTree: an algorithm to build an approximate tree from a large number of unaligned sequences. Bioinformatics, 23, 372-374, (2007).

10.0801291308

Standley, D.M., Toh, H., Nakamura, H., ASH structural alignment package: Sensitivity and selectivity in domain classification. BMC Bioinformatics 8, 116, (2007).

(班員間での共同研究)

応用ゲノム班員の井ノ上逸朗教授（東海大学・医学部）と、SNP解析による疾患原因遺伝子の同定に関する研究において、本研究の課題の一つであるPCアルゴリズムの適用を行っている。