社団法人　情報処理学会　研究報告
IPSJ SIG Technical Report

2005－MPS－57（4）
2005／12／20

遺伝子発現プロファイルを用いた遺伝子制御ネットワーク推定のためのバイクラスタリングの利用

瀧　　浩平[†]　竹中要一[†]　松田秀雄[†]

遺伝子発現プロファイルの蓄積に伴い，遺伝子制御ネットワークの推定に，より多くの実験条件を含む発現プロファイルを用いることが可能になった．しかし，この様な発現プロファイルに対してモジュールネットワークの様な推定手法を適用すると，推定精度の低下を招く恐れがある．モジュールネットワークは，実験条件の大半で類似した発現を示す遺伝子が多数存在することを前提とするが，多くの実験条件を含む発現プロファイルほどその様な遺伝子は少ない．そこで本研究では，発現プロファイルの実験条件がバイクラスタに含まれる部分のみから推定を行うことで，推定精度の低下の軽減を図る．本手法を出芽酵母の発現プロファイルに対して適用し，有効性を検証した．

Inference of Gene Regulatory Networks from Gene-expression Profiles with Utilization of Biclustering Results

Kohei Taki[†] ,Yoichi Takenaka[†] and Hideo Matsuda[†]

The accumulation of gene-expression profiles can allow an inference of gene regulatory networks by using a profile measured under many experimental conditions. However, the conventional methods such as module network may perform not enough accurate inferences against such profiles, because of following two facts. 1) Module network can perform accurate inferences only for genes showing similar gene-expression. 2) In gene-expression profiles with more conditions, fewer genes show similar gene-expression. To alleviate the accuracy loss, we perform an inference by using gene-expression patterns only under experimental conditions included in biclusters. The performance of our method is demonstrated by applying to inferences using various gene-expression profiles of budding yeast.

1. はじめに

遺伝子の機能解析は，個々の遺伝子の機能解析から，複数の遺伝子が協調して働くことで果たされる機能の解析へと焦点が移りつつあり，遺伝子相互の機能的関連により形成される遺伝子ネットワークの構造を解明することが求められている．遺伝子ネットワークの構造を解明するための研究の１つとして，遺伝子発現プロファイルを用いて遺伝子制御ネットワークを推定する試みがなされている．遺伝子発現プロファイルとは，細胞を様々な条件下に置いた場合に各遺伝子が働いた量（遺伝子発現量）を示す，遺伝子×実験条件の行列データである．遺伝子制御ネットワークとは，遺伝子の間の発現量の制御関係を表したグラフである．

遺伝子制御ネットワークの推定は，ベイジアンネットワーク [1] などを用いて研究が進められてきた．また，推定における組合せ爆発の問題を軽減するために，推定の前段階として同じ制御関係に従う遺伝子をまとめる，遺伝子のクラスタリング手法 [2] の研究が進められてきた．そして，クラスタリングとネットワーク推定を同時に行う手法として，モジュールネットワークモデル [3] に基づいた推定の試みがなされており，従来の手法よりもロバストな推定が可能である．

遺伝子制御ネットワークの推定では，より多くの実験条件を含む遺伝子発現プロファイルを用いた方が，精度の高い結果が得られることが期待される．近年の遺伝子発現プロファイルの蓄積により，その様な遺伝子発現プロファイルの取得が可能になりつつある．本研究の目的はこの様な遺伝子発現プロファイルを用いて，より精度の高い推定を試みることである．

しかし，この様な遺伝子発現プロファイルに対してモジュールネットワークモデルを適用すると，十分な精度が得られない可能性がある．モジュールネットワークモデルは，大半の実験条件で類似した発現量を示す遺伝子の集合（モジュール）が多数存在することを前提とするが，より多くの実験条件を含む遺伝子発現プロファイルではその様な遺伝子は多くない．このため，同じモジュールにまとめられた遺伝子の間でも多数の実験条件で発現量が類似しないことが多くなり，推定精度が低下することが予測される．

そこで本研究では，バイクラスタリング [5] によって選別された実験条件を，モジュールネットワークの推

† 大阪大学 大学院情報科学研究科 バイオ情報工学専攻
Department of Bioinformatic Engineering, Graduate School of Information Science and Technology, Osaka University

定に利用する．遺伝子発現プロファイルのバイクラスタに含まれる実験条件のみを用いることで，十分に類似した発現パターンによる制御関係の推定を試みる．

2. 遺伝子の制御関係の推定

2.1 遺伝子発現プロファイルと制御関係の推定

遺伝子とは DNA 配列上の領域であり，生命活動に不可欠とされるタンパク質の設計図である．タンパク質の合成は，遺伝子がコードされた DNA 配列上の領域が mRNA に転写されることで開始される．これを遺伝子の発現と呼ぶ．遺伝子が転写された量は遺伝子発現量と呼ばれ，DNA マイクロアレイなどを用いて測定することが出来る．遺伝子発現プロファイルとは，細胞を様々な条件下に置いて各遺伝子の発現量を測定した，遺伝子×実験条件の発現量の行列データである．本論文では，遺伝子発現プロファイルの遺伝子 A の行ベクトルを A の発現パターンと呼び，発現パターンが類似する事を発現が類似すると言う．

発現量には遺伝子間で依存関係があることが観測されている．遺伝子 A の発現量の増減が遺伝子 B の発現量の増現をもたらすとき，A は B を制御するという．遺伝子の制御関係をグラフ構造で表現したのが遺伝子制御ネットワークであり，遺伝子をノードによって，その間の制御関係を有向辺によって表す．

遺伝子発現プロファイルによる遺伝子間の制御関係の推定は，遺伝子間の発現パターンの依存関係に基づいて行われる．遺伝子 A の発現パターンを遺伝子 B の発現パターンの関数として表すことが可能な場合，A は B に制御されると推定される．このモデル化に基づく制御関係の推定は，ベイジアンネットワーク [1] などを用いて研究が進められてきた．遺伝子制御ネットワークの推定は制御遺伝子の組合せ最適化問題としてモデル化されるため，遺伝子数に対する組合せ爆発が問題となる．この軽減のため，推定の前段階として同じ制御関係に従う遺伝子をまとめる，階層型と分割型のクラスタリング手法を用いた研究が進められてきた [2]．そして，分割型クラスタリングとベイジアンネットワークによる推定を組み合わせた，モジュールネットワークに基づく推定手法が提案されている [3]．

2.2 モジュールネットワーク

これまでに，同じ遺伝子によって制御される多数の遺伝子が類似した発現を示すことが，多くの遺伝子発現プロファイルで確認されてきた．モジュールネットワークモデルでは，同じ遺伝子によって制御され類似した発現を示す遺伝子の集合を，モジュールとしてまとめて扱う．同じモジュールに含まれる全ての遺伝子が，同じ遺伝子に制御される様にモデルを制限して，推定における組合せ爆発の問題を軽減している．

モジュールネットワークはモジュールの集合とその間を結ぶ有向辺により表される．モジュールは同じ遺伝子に制御される遺伝子の集合として定義される．有向辺は遺伝子からモジュールへ結ばれ，その遺伝子がモジュールが含むすべての遺伝子を制御することを表す．**図 1** は例えば，遺伝子 CLN1，CLN2 が同じモジュールに含まれ，SWI6 に制御されることを表す．

モジュールネットワークの推定は図 1 に示す様に，1) 各遺伝子が含まれるモジュールの推定，2) 各モジュールの間の制御関係の推定，の 2 段階に分かれる．ネットワークの評価値が収束するまで 2 つの段階を交互に繰り返される．推定の入力としては，遺伝子発現プロファイルと制御遺伝子候補となる遺伝子の集合が与えられる．各モジュールを制御する遺伝子を，制御遺伝子候補に含まれる遺伝子に限定した推定が行われる．

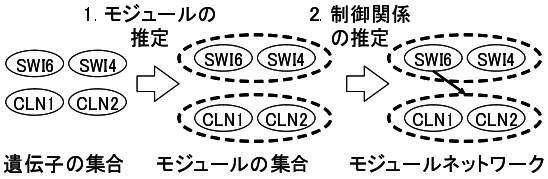

図 1 モジュールネットワークの推定
Fig. 1 Inference of a module network

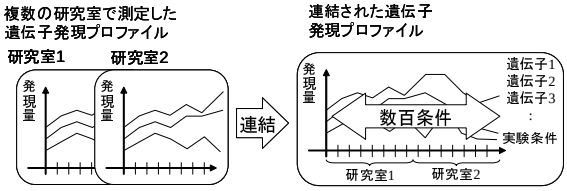

図 2 遺伝子発現プロファイルの実験条件を連結することによって作成した，多くの実験条件を含む発現プロファイルを
Fig. 2 To obtain a gene-expression profile that contains more experimental conditions, concatenating profiles measured by several laboratories.

2.3 遺伝子発現プロファイルの蓄積

制御関係の推定では，推定に用いた遺伝子発現プロファイルに含まれる実験条件の数が多いほど，精度の高い推定が可能になる．遺伝子発現プロファイルは，DNA マイクロアレイの技術発展に伴って膨大な数が蓄積されつつあり，より多くの実験条件を含む遺伝子発現プロファイルを用いた推定が可能になりつつある．

複数の遺伝子発現プロファイルを用いた解析は，既に試みられている [6]．多数の遺伝子発現プロファイルの実験条件を，**図 2** の様に繋げることで，多くの実験条件を含む 1 つの遺伝子発現プロファイルとして解析が行われた．この様にして得た多くの実験条件を含む遺伝子発現プロファイルを用いて，より精度の高い推定を行うことが本研究の目的である．

しかし，モジュールネットワークモデルの様な従来の推定手法を，この様な遺伝子発現プロファイルに対して適用する場合には，十分な推定精度を得られない可能性がある．2.2 節で述べた様に，モジュールネットワークモデルは，大半の実験条件で発現が類似した遺伝子から成るモジュールが多数存在することを前提とする手法である．このため，多くの実験条件で発現が類似しないモジュールの制御遺伝子推定で，精度が低下することが予測される．一方で，含まれる実験条件が多い遺伝子発現プロファイルほど，任意の 2 つの遺伝子の発現は類似しない．このため，多くの実験条件で発現が類似しない遺伝子が，同じモジュールに分類されることが多くなると予想される．従って，モ

ジュールネットワークを用いた場合には，十分な推定精度を得られない可能性がある．

この問題点を解決するため本研究では，バイクラスタリングを，モジュールネットワークの推定に利用することを提案する．

3. 遺伝子と実験条件のバイクラスタリング

遺伝子の集合が類似した発現を示すのは一部の実験条件に限られ，その他の実験条件ではほとんど独立に発現することが観測されている．この様に特定の実験条件でのみ類似した発現を示す遺伝子の集合を見つけるために開発されたのが，バイクラスタリング [5] である．遺伝子発現プロファイルを入力として与えると，発現パターンが類似した遺伝子と実験条件の部分集合の組から成るバイクラスタの集合を出力する．

バイクラスタは遺伝子の部分集合と実験条件の部分集合の組から成り，**図 3** の例の様に，行列データの遺伝子発現プロファイルの対応する列と行から成る部分行列を表す．バイクラスタに含まれる遺伝子の評価には，そのバイクラスタに含まれる実験条件のみから成る発現パターンが用いられる．

4. バイクラスタリングを利用した遺伝子制御ネットワークの推定

本研究ではモジュールネットワークを多くの実験条件を含む発現プロファイルに適用する際に，発現が類似しない実験条件も推定に用いられる問題点を解決するために，バイクラスタリングを制御関係の推定に利用する．図 1 で示した，モジュールネットワークの制御関係推定の段階で，バイクラスタに含まれる実験条件のみを用いることで，類似した発現パターンのみを用いた制御関係の推定を実現する．

提案手法には入力として，遺伝子発現プロファイルと制御遺伝子候補が与えられる．まず，入力として与えられた遺伝子発現プロファイルに対して，バイクラスタリングが適用される．検出されたバイクラスタからは，発現パターンが類似した実験条件の集合が遺伝子の集合ごとに得られる．そこで次に，バイクラスタに含まれる実験条件のみから成る発現パターンを用いて，バイクラスタに含まれる遺伝子を制御する遺伝子を推定する．これは図 3 に示す様に，行列形式の遺伝子発現プロファイルから，バイクラスタに含まれる実験条件の集合に対応する列のみから成る部分行列を取り出すことに相当する．その部分行列に含まれる発現パターンのみを用いて，制御関係の推定を行う．

制御関係の推定はバイクラスタごとに行う．バイクラスタごとに含まれる実験条件の集合が異なるため，バイクラスタごとに異なる列から成る遺伝子発現プロファイルの部分行列を用いて，制御関係を推定する．例えば図 3 では，バイクラスタ 1 に含まれる遺伝子 1,2,3 を制御する遺伝子の推定には，実験条件 1,2 の列のみから成る部分行列を用いるが，バイクラスタ 2 では実験条件 2,3,4 の列のみから成る部分行列を用いる．

バイクラスタに含まれる遺伝子の集合をモジュールとみなし，バイクラスタに含まれる遺伝子を制御する

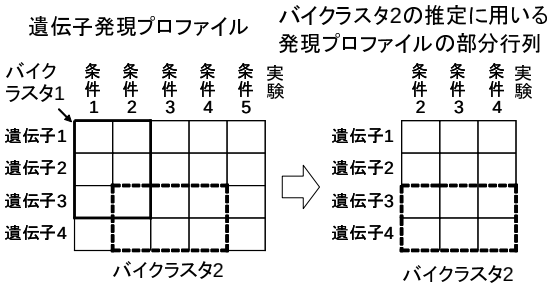

図 3 バイクラスタに含まれる実験条件の集合に対して，遺伝子発現プロファイルの対応する列の部分行列を取り出す
Fig. 3 Extracting column vectors corresponding to experimental conditions included in the bicluster 2.

遺伝子を推定する．図 3 の例では，バイクラスタ 1 に含まれる遺伝子 1,2,3 から成るモジュールと，バイクラスタ 2 に含まれる遺伝子 3,4 から成るモジュールそれぞれを，制御する遺伝子が推定される．制御関係の評価には，モジュールネットワークの評価関数をそのまま用いることができる．

5. 評 価 実 験

遺伝子発現プロファイルによる遺伝子制御ネットワークの推定を，以下の様な条件で行った．遺伝子発現プロファイルとしては，遺伝子発現データベース GEO[4] から得られた出芽酵母のデータを用いた．これらの遺伝子発現プロファイルの実験条件を繋げて，448 個の実験条件を含む遺伝子発現プロファイルを作成した．遺伝子間の転写制御関係のデータベース TRANSFAC[8] に記載されている制御関係を，既知の制御関係として用いた．既知の制御関係は，91 個の制御遺伝子と 182 個の被制御遺伝子の間の 276 個の制御関係を表している．推定で用いる制御遺伝子候補として，他の遺伝子を制御することが既知の 91 遺伝子を与えた．

以下では，本研究で提案する実験条件の選別に関して，妥当性を検証した結果を示し，提案手法による遺伝子制御ネットワークの推定結果を評価する．

5.1 実験条件選択の効果の検証

制御される遺伝子の発現の類似性によって，実験条件を選別することが，制御関係の推定精度を向上させるか検証した．

まず，個々の例に関して検証を行った．遺伝子 EUG1 と PDI1 は，遺伝子 HAC1 に制御されることが知られている．遺伝子 EUG1 と PDI1 から成るモジュールを制御する遺伝子を推定した．その結果，実験条件を選別しなかった場合には，既知の制御因子 HAC1 とは全く異なる発現を示す REB1 という遺伝子が推定された．一方で選別した場合には，HAC1 と類似した発現を示す遺伝子 CDC10 が推定された．

これは本手法では発現が類似した遺伝子同士を区別することは困難であることを示しており，既知の制御遺伝子と発現の類似した遺伝子が，制御遺伝子として推定される傾向がある．そこで本研究では，推定された制御遺伝子だけでなく，それと発現の類似した遺伝

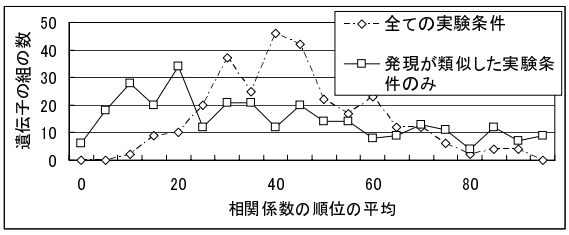

図 4 推定に利用する実験条件を選別した場合とすべての実験条件を用いた場合の間の相関係数の順位の比較
Fig. 4 Comparing ranks of correlation coefficients in case of selecting experimental conditions with ones in case of not selecting.

子も推定された制御遺伝子の候補として提示する．ここでは遺伝子の発現の類似性を測る指標として相関係数を用いた．実験条件を選別しなかった場合の REB1 と HAC1 の間の相関係数が 0.23 と無相関に近かったのに対して，選別した場合の CDC10 と HAC1 の間の相関係数は 0.75 と強い相関があった．

以上から，推定された制御遺伝子と既知の制御遺伝子の間の発現の類似性から，推定結果の妥当性を評価することにした．しかし，発現パターンに含まれる実験条件の数が多いほど，相関係数は低下する傾向にあるため，相関係数の値そのもので評価することは適切ではないと考えられる．そこで，相関係数の値の大きさではなく順位から評価を行った．まず，推定された遺伝子と制御遺伝子候補に含まれる 91 個の遺伝子の間で相関係数を求めた．そして，既知の制御遺伝子との相関係数が 91 個中の上位何番目に入るかによって，推定された制御関係の妥当性を評価した．実験条件を選別した場合に制御遺伝子として推定された CDC10 の順位は 3 位と上位に位置したのに対して，選別しなかった場合に制御遺伝子として推定された REB1 の順位は 36 位だった．

この様に，実験条件の選別によって，既知の制御遺伝子そのものは推定できなかったが，既知の制御遺伝子と発現の類似した遺伝子を制御遺伝子として推定することができた．以上から，既存の手法よりも精度の高い推定が可能になることが期待される．

そこで次に，全体についても上の例と同様の効果があるかどうか傾向を調べた．同じ遺伝子に制御されることが既知の任意の 2 つの遺伝子について，上の例と同様の検証を行った．各モジュールついて，実験条件の選別を行った場合と，行わなかった場合の結果を比較した．**図 4** は各モジュールの結果の相関係数の順位の分布をヒストグラムで示している．実験条件の選別を行った場合の相関係数の順位の分布は，選別を行わなかった場合に比べて順位の高い左方に偏っていることが分かる．従って図 4 は，発現の類似性から実験条件を選別する事で，任意の組でもより良い推定結果を示す傾向があることを示していると考えられる．

5.2 遺伝子制御ネットワークの推定

他の遺伝子に制御される事が既知の遺伝子 182 個に対して，これらの遺伝子を制御する遺伝子を推定した．提案手法とモジュールネットワークそれぞれを用いて，制御関係の推定を行いその結果を上と同様に相関係数の順位によって評価した．バイクラスタリングの手法としては，Plaid モデル [7] を用いた．提案手法の性能評価のため，2 つの手法の推定精度を比較した．

既知の制御遺伝子が上位 5 番目に入る様な制御関係を正解とする場合における，推定精度をまとめたのが **表 1** である．表 1 から，提案手法は従来手法よりも高い推定精度を示したことが分かる．推定された制御関係を正解とする閾値として，5 以外の値を選んだ場合について推定精度を比較した場合でも，提案手法の方がより高い推定精度を示す事が確認できた．

推定手法	推定精度
提案手法（Plaid モデル）	11.2%
モジュールネットワーク	6.3%

表 1 提案手法と従来手法の推定精度の比較
Table 1 Comparing of an accuracy of proposed method with the conventional 's one.

6. おわりに

遺伝子発現プロファイルに基づく遺伝子制御ネットワークの推定に，バイクラスタリングを利用する手法を提案した．多くの実験条件を含む遺伝子発現プロファイルを用いた遺伝子制御ネットワークの推定に有効である．バイクラスタリングによって実験条件を選別することで，制御関係の推定精度が向上することを検証した．今後は，制御関係の推定結果に応じたバイクラスタの最適化の方法を検討することが挙げられる．

謝　辞

本研究は一部，服部報公会工学研究奨励援助金によっている．

参 考 文 献

1) Friedman, N., Linial, M., Nachman, I. and Pe 'er, D.: Using Bayesian networks to analyze expression data, *J. Comput. Biol.*, Vol.7, pp.601-620 (2000).
2) Hartigan, J.A.: *Clustering Algorithms*, John Wiley & Sons, New York (1975).
3) Segal, E., et al.: Learning Module Networks, *Proc. 19th Conf. on Uncertainty in Artificial Intelligence*, Acapulco, Mexico (2003).
4) Barrett, T., et al.: NCBI GEO: mining millions of expression profiles–database and tools, *Nucl. Acids Res.*, Vol.33, Database Issue, pp.D562-D566 (2005).
5) Sara, C.M., Arlindo, L.O.: Biclustering Algorithms for Biological Data Analysis: A Survey, *IEEE/ACM Trans. Comput. Biol. Bioinfo.*, Vol.1, No.1, pp.24-45 (2004).
6) Luscombe, N.M., et al.: Genomic analysis of regulatory network dynamics reveals large topological changes, *Nature*, Vol.431, pp.308-12 (2004).
7) Lazzeroni, L. and Owen, A.: Plaid Models for Gene Expression Data, *Statistica Sinica*, Vol.12, No.1, pp.61-86 (2002).
8) Matys, V., et al.: TRANSFAC: transcriptional regulation, from patterns to profiles, *Nucl. Acids Res.*, Vol.31, pp.374-378 (2003).