Vol.2014-MPS-98 No.9
Vol.2014-BIO-38 No.9
2014/6/26

ベイジアンネットワークによる遺伝子制御ネットワーク推定結果の反復構築のための計算速度向上手法

津田 絢子　瀬尾 茂人　竹中 要一　松田 秀雄

概要：複数の遺伝子間で行われている転写制御関係をグラフにより可視化したものを遺伝子制御ネットワークと呼ぶ．遺伝子制御ネットワークの全容は未だ解明されておらず，計算機によって推定することができれば医学，創薬の分野において大幅に実験コストを削減できる．その為，遺伝子制御ネットワークの推定手法の研究はバイオインフォマティクスにおいて非常に重要なテーマである．
遺伝子制御ネットワークの解析手法の一つに，ベイジアンネットワークがある．このモデルは他のモデルに比ベノイズに強く，データ数が少ない場合にも推定が可能である．しかしこのモデルはデータ数が増えると爆発的に探索空間が大きくなるため，探索を行うには非常に大きな計算量が必要となる．この欠点を回避するため，近似法であるグリーディ法を用いることが多い．
遺伝子制御ネットワークを推定する研究では，ネットワークを推定した後に生物学実験や生物学者の議論の結果より得られた制御関係を考慮して反復的にネットワーク推定を行う．推定結果の反復構築とは，ネットワークを推定した後に新たな制御関係に応じてネットワークを修正するためにこれを考慮してネットワークを再度推定することである．しかし反復構築を行う場合，反復回数に比例して計算量が必要となる．本研究では，グリーディ法で反復的に推定結果を構築する場合における計算量の軽減のための手法を提案する．反復的な推定のために構築された結果のさらなる利用により計算量の軽減を図り，従来の反復推定の手法と比較することでその有効性を検証した．

キーワード：遺伝子制御ネットワーク，ベイジアンネットワーク， 反復構築， 計算量の軽減

1. はじめに

生物の重要な構成要素である遺伝子からタンパク質が生成される一連の過程は発現と呼ばれ，様々な条件下で計測されている．遺伝子の発現ではある遺伝子の転写量は別の遺伝子の転写産物やタンパク質によって促進や抑制されており，複数の遺伝子間で様々な転写制御が行われている．転写制御関係は遺伝子の働きを調整するために非常に重要であることから，解明が望まれている．

マイクロアレイ技術 [1] の開発により，多数の遺伝子の発現量を同時に計測することが可能になり、同時発現する遺伝子の探索とそれに基づく遺伝子機能の推定などの解明の可能性が出てきた [2]．ただし，実験計測される遺伝子発現量は膨大であるため，高精度かつ高速に情報を解析する方法論やアルゴリズムの開発が必要となっている．遺伝子制御ネットワークの解明もそのひとつであり，バイオインフォマティクスにおいて重要なテーマとなっている．

遺伝子制御ネットワークの推定手法は，数理モデルに基

Vol.2014-MPS-98 No.9
Vol.2014-BIO-38 No.9
2014/6/26

づいて行われている．使用される数理モデルは，ブーリアンネットワーク [3]，偏微分方程式モデル [4][5]，グラフィカルガウシアンモデル [6] といった様々なモデルが存在する．中でも統計的手法としてよく使われるものに，ベイジアンネットワーク [7][10] がある．確率的推論モデルの一種であるベイジアンネットワークは，数式での表現が困難な事象を確率で表現でき，これはマイクロアレイのようなノイズを含むデータから挙動を解析するのに適しているといえる．

ベイジアンネットワークは非常に有用なモデルではあるが，問題として全件探索を行うと計算量が膨大であり時間がかかるという点が挙げられる．この問題点の解消のために，近似法を用いることが多い．しかし近似法では全件探索に比べると精度が悪い．遺伝子制御ネットワークを推定する研究では，ネットワークを推定した後に生物学実験や生物学者の議論の結果により得られた遺伝子制御関係を考慮してネットワークを再度推定することがある．これを反復構築と呼ぶ．従来では，反復構築時に推定を新たに始めから行っており，これは修正という観点では無駄が多いと考えた．さらに，スムーズに修正を行うためには反復構築する前の結果を反復構築時に使用することができれば良いのではないかと考えた．

本研究では，ベイジアンネットワーク推定の近似法における結果の反復構築時に，推定結果を利用することで計算量を軽減させることを目的とする．今回は，反復的な推定のために構築された結果を利用したネットワークを初期状態として利用するという方法を使用した．

2. 遺伝子の発現と遺伝子制御ネットワーク

2.1 遺伝子の発現と転写制御，遺伝子制御ネットワーク

生物の重要な構成要素である遺伝子は DNA の配列であり，A(アデニン)，T(チミン)，G(グアニン)，C(シトシン) の 4 種類の塩基配列で構成されている．DNA は相対的な塩基対結合である A と T，G と C 塩基の間の水素結合により，二重螺旋構造を形成する．

遺伝子の主な機能は，タンパク質を生成することである．タンパク質の生成には，まず遺伝子の塩基配列が mRNA に転写される．次に mRNA がタンパク質に翻訳される段階を経て，タンパク質が生成され，酵素や生体構成部品として役割を果たす．このように遺伝子からタンパク質が生成される一連の流れのことを遺伝子の発現と呼ぶ．

遺伝子からタンパク質が生成される量は一定ではなく，そのときに必要な量にあわせて変化する．遺伝子の発現ではある遺伝子の転写量は別の遺伝子の転写産物やタンパク質によって促進や抑制されていることが知られており，何千，何万もの遺伝子が他の遺伝子を制御しあって生体内で全体として調和の取れたタンパク質生成を行っている．この制御関係を転写制御と呼び，転写因子と呼ばれるタンパク質が遺伝子領域の上流，あるいは下流にある特定の DNA 塩基配列に結合することで実現している．

複数の遺伝子間で行われている転写制御関係を遺伝子制御ネットワークと呼ぶ．遺伝子制御ネットワークは，遺伝子をノード，制御関係を有向辺とした有向グラフとして表される．生体内で行われている動的な制御関係をネットワークとして捉えることで，遺伝子間の制御関係だけでなく遺伝子間の間接的な制御関係も把握できるようになるため，遺伝子の働きに影響を与える遺伝子についてより広く理解することが容易になる．

2.2 遺伝子発現プロファイル

遺伝子の発現した量は発現量と呼ばれ，発現したタンパク質の量で表される．しかし，直接タンパク質の量を計測することは困難であるため，発現量はタンパク質の元となる mRNA の量として測定されている．

遺伝子の発現量を測定する方法として，マイクロアレイを用いた方法が存在する．この方法では，一回の実験によって数万個の遺伝子発現情報を得ることができるため，網羅的な発現解析に広範に利用されている．

このようにして集積した発現量のデータを遺伝子発現プロファイルと呼ぶ．遺伝子発現プロファイルは様々な条件下で測定された遺伝子の発現量を記録したものであり，遺伝子と実験条件が対になった表としてまとめられている．マイクロアレイ実験の計測データによる遺伝子発現プロファイルはデータ量が非常に大きいため，計算機を用いた解析が必要となっている．

マイクロアレイ実験の計測データによる遺伝子発現プロファイルは，公的なデータベースである GEO(Gene Expression Omnibus)[8][9] に保管されている．GEO は，実験・管理された遺伝子発現プロファイルを検索し，ダウンロードするのに有用なツールである．

2.3 遺伝子制御ネットワーク推定

近年，発現プロファイルの規模の増大に伴い，計算機による遺伝子ネットワーク推定が求められている．遺伝子ネットワークの推定とは，発現プロファイルなどの遺伝子の働きを示したデータから，複数の遺伝子間の制御関係を網羅的に推定することである．ここでは入力データとして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に限定して説明する．

発現プロファイルを用いた推定は，推定対象となる遺伝子群の発現プロファイルを入力として，そこから推定された遺伝子制御ネットワークを出力とする．制御関係にある遺伝子間には発現量の依存関係が存在する場合が多く，発現プロファイルにはその依存関係が測定されている可能性が高い．そのため発現プロファイルを用いた推定では，発現量に依存関係がある遺伝子の間に制御関係がある事を仮

Vol.2014-MPS-98 No.9
Vol.2014-BIO-38 No.9
2014/6/26

定して，発現プロファイルから制御関係を推定する．

本研究では，遺伝子制御ネットワークを推定する際の数理モデルとしてベイジアンネットワークを用いる．

3. ベイジアンネットワークによる遺伝子制御ネットワーク推定とネットワークの利用法

3.1 ベイジアンネットワークによる遺伝子制御ネットワーク推定

ベイジアンネットワークとは，データ間の関係を確率によって表現しようと試みる手法である．具体的には，非循環有向グラフと条件付確率分布の表によって構成され，変数間の依存関係を表現する．各変数とグラフのノードを 1 対 1 で対応付け，変数間に依存関係がある場合にグラフの対応するノードに有向線を引いて表現する．依存関係のある変数同士がどのような確率変数に従うかは，各変数に対応する条件付確率分布の表を用いて表す．

ベイジアンネットワークの考え方を応用して遺伝子制御ネットワークを推定することができる．制御関係を調べる遺伝子群の発現プロファイルを入力とし，ネットワーク全体としての評価関数が最大となるグラフ構造を組み合わせ最適化問題として探索することによって推定が行われる．以下では，ベイジアンネットワークの評価関数と，組み合わせ最適化問題についてそれぞれ説明する．

ベイジアンネットワークは，データセット D が与えられた場合のネットワーク B の事後確率 $p(B|D)$ によって評価される．事後関数 $p(B|D)$ は，ベイズの定理から式 1 のように分解される．それぞれの変数が親変数を除くと互いに独立であることを仮定しており，式 1 の事後確率は変数ごとに独立に求められる．

$$p(B|D) = \frac{p(D|B)p(B)}{p(D)} \tag{1}$$

ここで，$p(B)$ はベイジアンネットワーク B の事前確率を，$p(D)$ はデータセット D の事前確率を示している．ベイジアンネットワークスコアは，この式 1 により求められる．式 1 を用いて推定されるネットワークは，非循環有向グラフとして表現される．

ベイジアンネットワークによるネットワーク推定では，ネットワークを組み合わせ最適化問題として探索し，複数の遺伝子間の複雑な制御関係が推定可能である．組み合わせ最適化問題の制約条件として，ネットワーク構造が非循環有向グラフであることが挙げられる．したがって，循環構造が存在しないという制約条件の下で目的関数 $p(B|D)$ を最大化するように各変数について最適な親変数の組み合わせを探す問題となる．

最適解を求めるための全件探索のアルゴリズムを以下に示す．

入力：遺伝子発現プロファイル
出力：遺伝子制御ネットワーク
(1) ネットワークに対する事前確率分布を作成する
(2) 事前確率分布から，現在のネットワーク構造に対して入力データへの学習を行い，事後確率分布を得る
(3) 適当なネットワーク構造を仮定する
(4) ネットワーク構造を逐次変化させ，ネットワークスコアを計算する
(5) ネットワークの全組み合わせのスコアの計算後，最もスコアの高いネットワークを出力する

ベイジアンネットワークの組み合わせ最適化問題で最適解を求めるための全件探索を行うと，変数の個数に対して指数関数的に探索空間が増加し，組み合わせ爆発を起こすといった問題点がある．全件探索によるネットワーク推定は現実的でないため，近似アルゴリズムであるグリーディ法 (Greedy Search)[10] を用いることが多い．

ネットワーク構造を推定する場合に確実にネットワークに含まれるまたは確実に含まれないと分かっているエッジが存在することがある．ホワイトリストやブラックリストはそのようなエッジを格納しておくためのデータセットである．ホワイトリストに含まれたエッジは常にネットワークに含まれ，ブラックリストに含まれたエッジは常にネットワークから除外される．

3.1.1 グリーディ法

組み合わせ最適化問題におけるネットワーク構造の全件探索アルゴリズムは，入力遺伝子数によって探索空間が指数関数的に増加するため，近似法としてグリーディ法が用いられる．

ホワイトリストやブラックリストを加えてベイジアンネットワークを用いた遺伝子ネットワーク推定を行う場合，グリーディ法ではアルゴリズムが以下のようになる．

入力：遺伝子発現プロファイル
　　リスト (ホワイトリスト，ブラックリスト)
出力：遺伝子制御ネットワーク
(1) ネットワークに対する事前確率分布を作成する
(2) 事前確率分布から，現在のネットワーク構造に対して入力データへの学習を行い，事後確率分布を得る
(3) リストにホワイトリストがある場合，初期状態にホワイトリストを追加し，そのネットワークスコアを計算する
(4) リストがブラックリストのみの場合，初期状態のネットワークスコアを計算する
(5) 枝をランダムに 1 本選び，リストとして示された制約条件の中で制御関係の付加，削除，方向変換を行い，スコアが最良のものを残す
(6) スコアが改善されれば再び 5. を行う
(7) スコアが改善されなければ，その時点で最もスコアの高い遺伝子制御ネットワークを出力する

グリーディ法では全件探索では不可能であった遺伝子数に対しての推定の可能であるが，推定精度が悪い．その理由として，局所的最適解に陥る可能性があることが挙げられる．グリーディ法ではスコアが改善されなくなった時点で終了するため，初期状態によってはすぐに終了してしまう可能性がある．精度向上のための手法はあるが，グリーディ法において局所解に陥ることを完全に防ぐことはできない．

3.2 推定した遺伝子制御ネットワークと反復構築

反復的な推定結果の構築とは，ネットワークが推定された後に生物学実験により遺伝子制御ネットワークが判明するために，これをホワイトリストやブラックリストとしてリストに追加し，ネットワークを再度推定することである．追加されるリストはネットワークが一度推定されないと判明しないため，最初の推定時に利用することができず，反復構築が必要となる．グリーディ法で反復構築を行う場合のアルゴリズムを以下に示す．ここでリスト 1 は事前知識として与えられているリストを，リスト 2 は反復構築のために途中で加えられるリストを指す．

- ステージ 1
  入力：遺伝子発現プロファイル
  　　リスト 1(ホワイトリスト，ブラックリスト)
  出力：遺伝子制御ネットワーク
  　　リスト 2(ホワイトリスト，ブラックリスト)
  (1) リスト 1 を制約条件としてグリーディ法を実行する
  (2) 実行結果より，ネットワークに含まれるべきである，または含まれるべきではないと判断される部分を探し出す
  (3) 2. で探し出した部分を，反復構築のためのリスト (リスト 2) に格納する
- ステージ 2 (反復構築)
  入力：遺伝子発現プロファイル
  　　リスト 1(ホワイトリスト，ブラックリスト)
  　　リスト 2(ホワイトリスト，ブラックリスト)
  出力：遺伝子制御ネットワーク
  (1) リスト 1 とリスト 2 の両方を制約条件としてグリーディ法を実行する

構築を重ねる場合の問題点として，推定を再度行うために以前の結果を作成したときと同程度の時間がかかるということがある．これは，グリーディ法の初期状態が空であるため，ステージ 1 とステージ 2 では同じ初期状態から似たようなネットワークを作成しており，グリーディ法でスコアが改善されなくなり推定結果が出力されるまでに作成されるネットワークの数が同程度となるからであると考えられる．この問題は，ネットワークが大きくなるほど一度の試行に時間がかかるため，より大きな問題となる．さらに試行錯誤のために一度のみならず何度も反復して構築され，推定を行う毎に反復される回数に比例した多量の時間を必要とする．

このような現状のもとで，グリーディ法を用いて推定結果の反復構築を行う場合に計算量の軽減が必要であると考え，これを本研究の目的とした．

4. 推定結果の反復構築のための計算速度向上手法

4.1 提案手法の目的

本研究の目的は，ベイジアンネットワークによる遺伝子制御ネットワーク推定においてグリーディ法を用いて推定結果の反復構築を行う場合に計算量の軽減を図ることである．ベイジアンネットワークは発現プロファイルから遺伝子制御ネットワークを推定する場合において，利点が多数がある．しかし全件探索は推定遺伝子数が増加すると計算量が爆発的に増加する．これを防ぐためのグリーディ法を用いても，精度を向上させるために推定結果の反復構築を重ねる場合は反復回数に比例した多量の計算時間がかかる．ここで，グリーディ法の利点を生かしつつ反復構築時における計算時間を減らすための手法を提案する．

提案手法では，従来手法ではベイジアンネットワーク推定における初期状態が空ネットワークであるという点に着目する．初期状態としてスコアの良いネットワークを与えてグリーディ法における推定を行うと，従来手法の途中で作成されるネットワークから推定を始めることと同様になり，スコアが改善されなくなるまでに作成されるネットワークが減るのではないかと考えた．また，初期状態として与えるスコアの良いネットワークとして推定結果を利用することにした．これは推定結果が局所解であるものの反復的に構築される結果に近いネットワークであると考えたからである．これらの観点より，提案手法ではネットワーク推定の初期状態を推定結果を利用したネットワークに変更し，推定を従来のグリーディ法の途中から行うことで，計算時間の軽減を図る．

4.2 提案手法の概要とアルゴリズム

提案手法を用いて反復構築を行う場合，まず事前に与えられたリスト (リスト 1) を制約条件としてグリーディ法を用いた推定を行い，推定結果より反復構築のためのリスト (リスト 2) を作成する．これをステージ 1 と呼ぶ．次に反復構築の実行 (ステージ 2) として，ステージ 1 で推定された結果を基に初期状態を作成し，その初期状態を指定してリスト 1，リスト 2 を制約条件としたグリーディ法による推定を行う．ベイジアンネットワーク推定における初期状態はデフォルトでは空のネットワークであるが，提案手法では反復構築時に初期状態を指定することで，計算時間の削減を図っている．手法の全体図と従来・提案手法の差異

Vol.2014-MPS-98 No.9
Vol.2014-BIO-38 No.9
2014/6/26

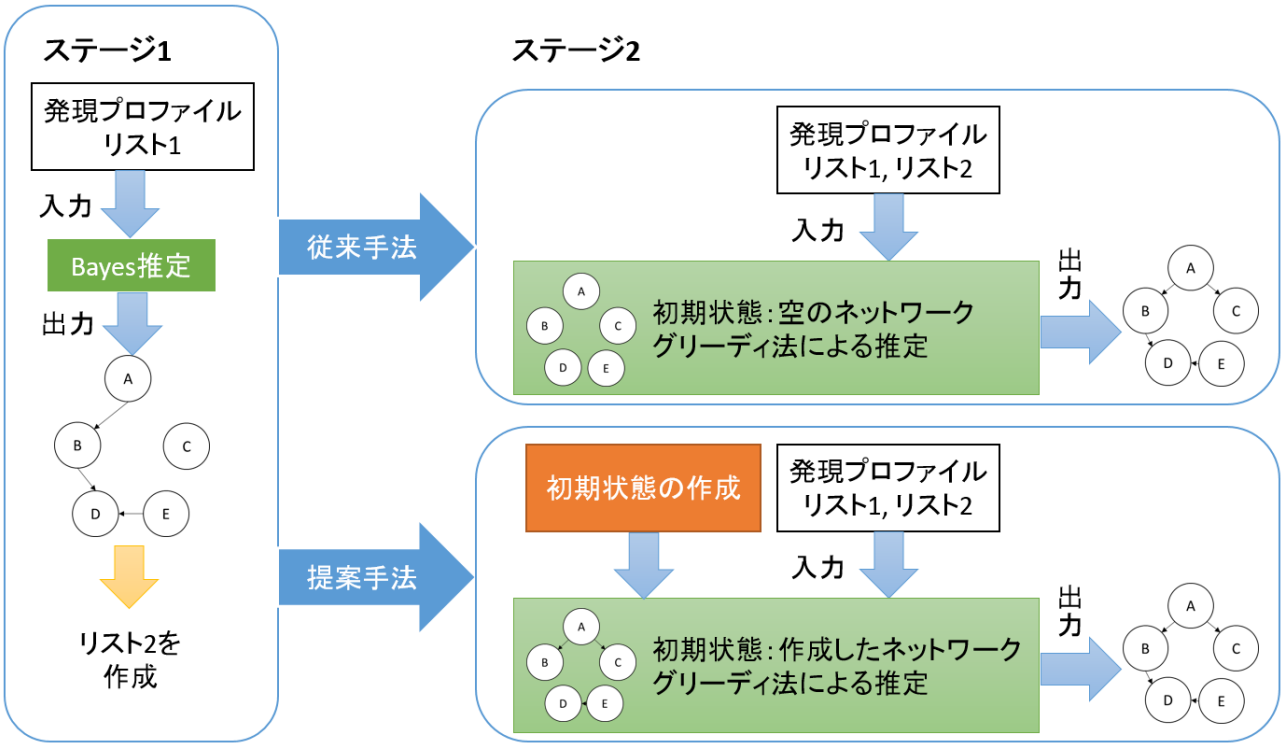

図 1 手法の概要と従来・提案手法の差異

を図 1 に示す．

ステージ 1 は従来のグリーディ法と提案手法において共通である．提案手法のステージ 2 のアルゴリズムは以下の様である．

入力：遺伝子発現プロファイル
　　リスト 1，リスト 2
　　ステージ 1 で作成された推定結果
出力：遺伝子制御ネットワーク
(1) リスト 1，リスト 2 とステージ 1 で作成された推定結果を用いて 4.2.1 章のアルゴリズムに従い初期状態を作成する
(2) 作成した初期状態を指定して，リスト 1，リスト 2 を制約条件とし，遺伝子発現プロファイルを入力としたグリーディ法を実行する

4.2.1 初期状態の作成

提案手法における初期状態として，ステージ 1 の推定結果を全体的に活用することを考える．初期状態としてステージ 1 の推定結果をそのまま利用しようとすると，その推定結果にリスト内のホワイトリストを加えたネットワークが循環構造を持つことがあり，ベイジアンネットワークによるネットワーク推定が不可能となる．

そこで，ステージ 1 の推定結果を活用し，さらにリスト内のホワイトリストを加えても循環構造を持たない初期状態を作成する．リスト 1，リスト 2 とステージ 1 で作成された推定結果から初期状態を作成するアルゴリズムは以下のようになる．

(1) リスト 1，リスト 2 の少なくとも片方に存在するエッジの始点または終点のノードを取り出す
(2) ステージ 1 で作成された推定結果から，1. で取り出したノードに関わるエッジを消去する
(3) 2. で作成されたネットワークにリスト 1，リスト 2 の中のホワイトリストに含まれるエッジを追加する

以上のアルゴリズムに従い作成されたネットワークを初期状態として利用する．リスト 1，リスト 2 とステージ 1 で作成された推定結果によっては初期状態が空グラフとなることも考えられる．

5. 実験と考察

提案手法が従来の初期状態を指定しないグリーディ法に比べ計算時間が軽減されているかを調査するために実験を行った．この章では実験内容，実験結果及び考察を述べる．

5.1 実験条件

ここでは実験で用いるデータと実行環境について述べる．

入力となる発現プロファイルとして GEO(Gene Expression Omnibus) に登録されているデータのうち，platform が”Affymetrix Mouse Gene 1.0 ST”であり，GPL6246 と記載されているものを使用する．サンプル数は 497 である．

提案手法と比較する従来手法として，ベイジアンネットワークの近似手法であるグリーディ法で，初期状態が空グラフであるものを用いた．また，従来手法・提案手法で共

Vol.2014-MPS-98 No.9
Vol.2014-BIO-38 No.9
2014/6/26

表 1　提案手法と従来手法の計算時間

遺伝子数	従来手法 (s)	提案手法 (s)	計算時間率
20	0.95	0.73	76.84%
50	23.27	12.56	53.97%
80	179.15	68.27	38.10%
100	457.50	139.80	30.55%
150	3952.50	985.17	24.92%

に用いるグリーディ法として，R の bnlearn パッケージ [11] に搭載されている関数”hc”を用いた．

計算の実行に用いた計算機環境は以下の様である．

CPU Intel®$Core^{TM}i5-2540MvPro^{TM}$ 2.60GHz

5.2　実験 1

5.2.1　実験内容

提案手法が従来の初期状態を指定しないグリーディ法に比べ計算量が軽減されていることを，計算時間の比較により実験した．

入力となる発現プロファイルに対して抽出を行い，規定数の遺伝子を取り出す．取り出した遺伝子に対してグリーディ法で推定を行い，結果を得る (ステージ 1)．ステージ 2 で追加されるホワイトリストとして取り出された遺伝子の中の 10 遺伝子からなるネットワークを使用し，この結果に対して提案手法と従来手法によって推定を行い，結果を得るまでの計算時間を比較する．

以上の実験を，取り出す遺伝子数を 20,50,80,100,150 とした場合の比較を行った．

なお，ステージ 1 における計算は提案手法，従来手法ともに共通であるため，比較は行っていない．

5.2.2　実験結果

表 1 は提案手法と従来手法の結果を得るまでの計算時間の比較である．この表での計算時間は CPU 時間 (秒) で構成されている．また，計算時間率は従来手法の計算時間を 100% としたときの提案手法の計算時間を示している．

表 1 より，遺伝子数が多くなるほど計算時間率が下がっていることが分かる．これより，遺伝子数が少ないときは従来・提案手法間に大差はないが，遺伝子数が増大するほど提案手法の従来手法における計算時間率が小さくなり，提案手法は従来手法に対しより優れた効果を発揮すると分かる．

5.3　実験 2

5.3.1　実験内容

グリーディ法による計算量は結果を得るまでに作成されるネットワークの数と比例することを利用して，提案手法が従来の初期状態を指定しないグリーディ法に比べ計算量が軽減されていることを，グリーディ法で結果を得るまでに作成されるネットワーク数の比較により実験した．

入力となる発現プロファイルに対して無作為抽出を行い，50 個の遺伝子を取り出す．取り出した遺伝子に対してグリーディ法で推定を行い，結果を得る (ステージ 1)．ステージ 1 において，リスト 1 は空である．ステージ 2 で追加されるリスト 2 として 50 遺伝子の中の 10 遺伝子からなるネットワークのエッジをホワイトリストで使用し，この結果に対して提案手法と従来手法によって推定を行い，結果を得るまでに作成されるネットワークの数と初期状態でのスコア，最終状態でのスコアを比較する．

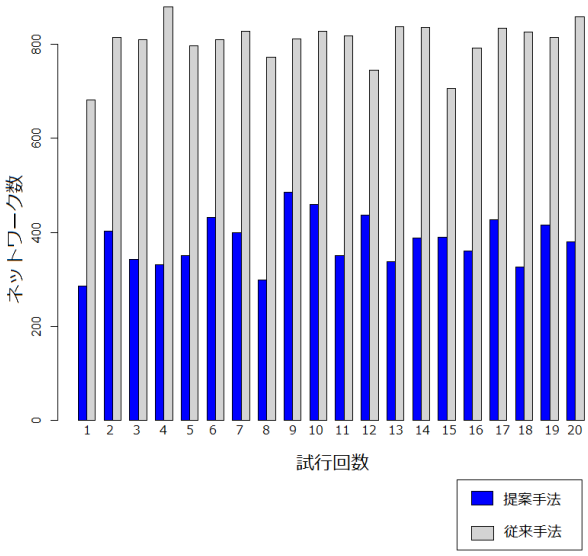

図 2　遺伝子数 50 における作成されたネットワーク数

以上の実験を，20 回ずつ行った．

なお，ステージ 1 における計算は提案手法，従来手法ともに共通であるため，比較は行っていない．

5.3.2　実験結果

図 2 は提案手法と従来手法の結果を得るまでに作成されたネットワークの数の比較である．横軸は実験の実行回数を，縦軸は結果を得るまでに作成されたネットワーク数を示している．また，右側に示された灰色の棒グラフが従来手法を，左側に示された青の棒グラフが提案手法を示している．常に提案手法は従来手法より少ないネットワーク数を示しているため，データに特異的ではなく提案手法は従来手法より計算量を減らすことに成功していると分かる．

実験結果の一例として，ある遺伝子群におけるステージ 2 での実験結果を図 3 に示す．この実験では，ホワイトリストとして有向辺が 29 本のものを使用した．この図において，黒の三角で示されたグラフが従来手法を，青の丸で示されたグラフが提案手法を示している．また，横軸は初期状態から結果を得るまでの作成されるネットワークの順を，縦軸はネットワークスコアを表している．従来手法と提案手法の横軸は終点をそろえてあり，横軸に書かれているネットワーク番号は従来手法のものである．従来手法の結果を得るまでに作成されるネットワークの数は 681，初

Vol.2014-MPS-98 No.9
Vol.2014-BIO-38 No.9
2014/6/26

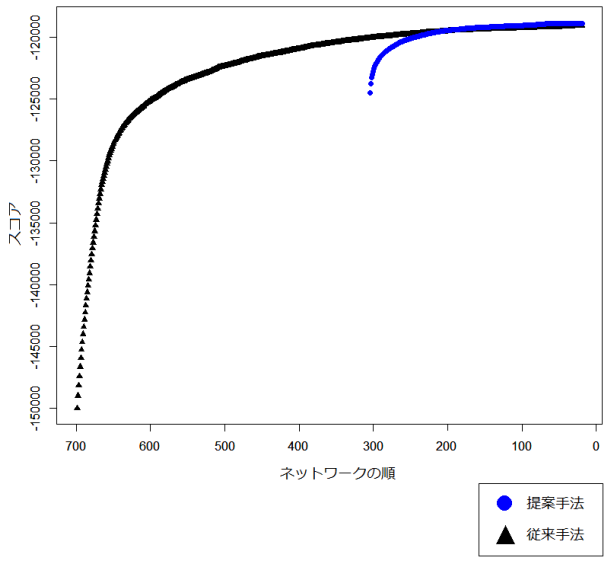

図 3 提案手法・従来手法比較の一例

期状態 (空グラフ) でのスコアは-150003.8，最終状態でのスコアは-119077.6 である．提案手法の結果を得るまでに作成されるネットワークの数は 286，初期状態でのスコアは-124506.1，最終状態でのスコアは-118870.3 である．図 3 より，提案手法では従来手法より初期状態が良いスコアであるために，計算量を従来手法の 50% 以下に減らすことに成功していると分かる．

この実験におけるステージ 2 での結果を得るまでに作成されたネットワークの数の 20 回の平均は，従来手法では 804.75 個，提案手法では 380 個となった．提案手法の計算量は従来手法の 47% になっている．

5.4 考察

実験 1，実験 2 では共に提案手法が従来のグリーディ法に比べ計算量が軽減されていることが示されている．

実験 1 では，遺伝子数と従来・提案手法における計算時間の差を確認した．表 1 により，遺伝子数が増大するほど提案手法は従来手法よりも優れた効果を発揮すると確認できる．遺伝子数が少ないときは従来・提案手法間に大差はないが，実行時間そのものが短いためこれは大きな問題にならないと考えられる．

実験 2 では遺伝子数を固定し，提案手法が従来手法と比較してどのような振る舞いをするかということと，それがデータ依存でないことを確認した．実験によって得られた図 2 より，データに関わらず提案手法は結果を得るまでに必要なネットワーク数が従来手法より少ないことを確認した．また，図 3 より，提案手法は従来手法よりもスコアの良い状態から探索を始めることで結果を得るまでに作成されるネットワークの数の削減に成功していることを確認できる．計算量は結果を得るまでに作成されるネットワークの数と比例するため，この実験ではデータに関わらず計算量の削減できるということと，スコアの良い状態から推定を始めると計算量が軽減できるということが確認された．

提案手法ではスコアの良い初期状態を設定しても推定結果が空ネットワークからグリーディ法を用いて推定する場合と同様の結果になり，精度に差がない状態で計算時間の向上を行うことができた．

6. おわりに

本研究では，ベイジアンネットワークによる遺伝子制御ネットワーク推定の推定結果に対し反復構築を行う場合にグリーディ法よりも少ない計算量で推定を行うことを目的として，グリーディ法の初期状態として推定結果を利用した状態を使用する手法を提案した．初期状態が空グラフの状態でのネットワーク推定と比較して，マウスの遺伝子の発現プロファイルを用いて評価実験を行った．その結果，初期状態が空グラフとして推定を行う場合より，推定結果を利用して初期状態を作成する提案手法は反復構築時において計算時間を削減することができた．

参考文献

[1] J. DeRisi , P. Brown, *Exploring the Metabolic and Genetic Control of Gene Expression on a Genomic Scale*, Science, 1997,Vol.278,Num.5338,Pages 680–686.
[2] 藤渕 航，堀本 勝久, マイクロアレイ データ統計解析プロトコール, 羊土社, 2008.
[3] 阿久津 達也, 遺伝子発現制御ネットワークの論理的解析, 生物物理,1999, Vol.39, Num.6, Pages 381–385.
[4] M.A Savageau, *Biochemical Systems Analysis: A Study of Function and Design in Molecular Biology*, Addison-Wesley, 1976.
[5] T. Chen , HL He and GM Church, *Modeling gene expression with differential equations*, Pacific Symposium of Biocomputing, 1999.
[6] H. Toh , K. Horimoto, *Inference of a genetic network by a combined approach of cluster analysis and graphical Gaussian modeling*, Bioinformatics, 2002, Vol=18, Pages 287–297.
[7] N. Friedman, M. Linial, I. Nachman and Dana Pe'er, *Using Bayesian Networks to Analyze Expression Data*, Journal of Computational Biology, 2000,Vol.7,Pages 601–620.
[8] T. Barrett, T. O. Suzek, D. B. Troupa, S. E. Wilhite, WC Ngau, P. Ledoux, D. Rudnev, A. E. Lash, W. Fujibuchi, and R. Edgar. *NCBI GEO: mining tens of millions of expression profiles - database and tools*, Nucleic Acids Research, 2005, Vol.33, Pages 562–566.
[9] Gene Expression Omnibus *http://www.ncbi.nlm.nih.gov/geo/*
[10] P. Spirtes, C. Meek, *Learning Bayesian networks with discrete variables from data*, KDD,1995.
[11] Marco Scutari, *Learning Bayesian Networks with the bnlearn R Package*, Journal of Statistical Software, 2010,Vol.35.