情報処理学会論文誌　数理モデル化と応用　Vol.6　No.3　151–162 (Dec. 2013)

複数時系列遺伝子発現プロファイルを利用した遺伝子制御ネットワーク推定の精度向上手法

渡邉 之人[1,†1]　瀬尾 茂人[1]　竹中 要一[1]　松田 秀雄[1,a)]

受付日 2013年1月30日，再受付日 2013年3月7日，
採録日 2013年4月18日

概要：遺伝子は生命の遺伝情報を記録しており，互いに制御し合っていることが知られている．この複雑な遺伝子間の制御関係を解明するために，個々の遺伝子制御関係を統合した遺伝子制御ネットワークの推定が行われている．しかし遺伝子 × 実験条件の行列データである遺伝子発現プロファイルは遺伝子数に比べ 1 実験あたりの測定回数が少ない．数千，数万の遺伝子に対して測定は数十回程度しか行うことができず，測定回数が 10 以下の遺伝子発現プロファイルも多い．そのため遺伝子発現プロファイルの情報量不足を原因とした遺伝子制御ネットワーク推定の精度低下が問題点となる．本研究では情報量不足を軽減するために複数の時系列遺伝子発現プロファイルを利用し，全プロファイルで一貫して存在する共通の部分ネットワークを全プロファイルを用いて推定する．その後各実験条件で特徴的な部分ネットワークを共通する部分ネットワークをもとに各プロファイルを用いて推定することで推定精度の向上を目指す．本手法を実際の遺伝子制御ネットワークを基にしたシミュレーションデータと，網膜視細胞の桿体，錐体分化時の遺伝子発現プロファイルに対して適用し，有効性を明らかにした．

キーワード：遺伝子制御ネットワーク，遺伝子発現プロファイル，ベイジアンネットワーク，複数時系列

A Robust Method for Estimating Gene Regulatory Networks Using Multiple Time Series Gene Expression Profiles

Yukito Watanabe[1,†1]　Shigeto Seno[1]　Yoichi Takenaka[1]　Hideo Matsuda[1,a)]

Received: January 30, 2013, Revised: March 7, 2013,
Accepted: April 18, 2013

Abstract: Genes contain genetic information and regulate each other. Estimating gene regulatory networks reveals complicated regulations. However, the number of conditions or time points of gene expression profiles is fewer than that of genes. It causes degrading the estimation accuracy. In this study, we propose a robust method for estimating gene regulatory networks using multiple time series gene expression profiles. First, the proposed method estimates a common network under multiple conditions. Second, the common network is extended by adding characteristic regulations of each condition. We demonstrate the effectiveness of our method by applying it to *in silico* datasets and differentiation processes of mouse retina to rod and cone photoreceptors.

Keywords: gene regulatory network, gene expression profile, Bayesian network, multiple time series

1. はじめに

ヒトゲノムプロジェクトによって 2003 年にヒトゲノム解読が完了し，個々の遺伝子の機能解析に加えて複数の遺伝子間における協調関係の解析がさかんに行われている．この協調関係は遺伝子をノード，遺伝子間の制御関係を辺

[1] 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, Suita, Osaka 565–0871, Japan
[†1] 現在，日本電信電話株式会社
Presently with Nippon Telegraph and Telephone Corporation
[a)] matsuda@ist.osaka-u.ac.jp

とした有向グラフとして表現することができる．有向辺の始点にあたるノードに対応する遺伝子が，有向辺の終点にあたるノードに対応する遺伝子を制御していることを表している．有向グラフは遺伝子制御ネットワーク [16] と呼ばれ，その構造を解明するために数理モデルに基づいたネットワーク構造の推定が行われている．

遺伝子制御ネットワーク推定のための研究の 1 つとして，遺伝子発現プロファイルを用いる試みがなされている [6], [23]．遺伝子発現プロファイルとは，様々な細胞組織，時期，実験条件下における遺伝子発現を測定した，遺伝子 × 実験条件の行列データである．一般的にすべての遺伝子が同時期に発現することはなく，発生の異なる段階や環境の変化によって様々な遺伝子が異なるレベル，時期で発現する．そのため遺伝子発現プロファイルは様々な実験条件下で変化する遺伝子の発現量から構成されていたが，近年需要が高まっている遺伝子制御ネットワークの動的変化をとらえるためには，対象組織・実験条件による環境変化を固定し時間経過に沿って発現量を測定した，遺伝子 × 時間の行列データである時系列遺伝子発現プロファイルが必要となる [10]．

遺伝子には細胞，時期，実験条件の違いによって異なる発現量を示すものがある．たとえば細胞分化と呼ばれる現象においては，ある細胞が異なる細胞になることでそれまでに存在しなかったタンパク質が出現する．新しいタンパク質が出現するためには，それまでに発現していた遺伝子が発現しなくなる，またはそれまでに発現していなかった遺伝子が発現する必要がある．他の例として，乾燥，高温などの様々なストレスに対する応答として，個々の細胞における遺伝子発現のパターンが変化するものなどがある．これらの遺伝子は細胞，実験条件に対して特徴的に発現しており，それぞれ異なる発現量を示す．そのため細胞や実験条件が変わると，観測される遺伝子制御関係が変化する可能性があり，遺伝子制御ネットワーク推定をより複雑にしている．

数理モデルを利用した遺伝子制御ネットワーク推定手法として，ブーリアンネットワーク [14]，グラフィカルガウシアンモデリング [25]，微分方程式モデル [4], [12], [21]，ベイジアンネットワーク [3], [8], [11], [19] などが用いられる．手法によって推定できるネットワークの構造や計算量などの特性が異なるため [13]，問題に合致した手法を選ぶことが必要となる．これらの手法の中でもベイジアンネットワークは遺伝子制御ネットワークのモデルとして広く用いられ，その有効性が報告されている [2], [7], [15]．様々な有効性の中でも遺伝子発現の依存関係を制御の方向の情報を含めて推定できる点，条件付き確率分布群によって対象をモデル化するためノイズに比較的強い点に着目し，本研究ではベイジアンネットワークに基づいて遺伝子制御ネットワークを推定する．

遺伝子発現プロファイルにおける実験条件数は遺伝子数と比較して非常に少なく，十分な推定精度を得られない可能性があり問題点となる．遺伝子の発現量計測には時間的，金銭的コストがかかるため，数千，数万の遺伝子に対して実験は数十回程度しか行うことができず，実験条件数が 10 以下の遺伝子発現プロファイルも多い．一方，遺伝子発現データベースの発展と整備により数多くの遺伝子発現プロファイルが蓄積されており，様々な実験条件下での遺伝子の発現量を利用することが可能となりつつある [1]．異なる実験条件下では発現する遺伝子，遺伝子間制御関係が変化する可能性があるが [5]，複数の実験条件下において共通する制御関係については，複数の遺伝子発現プロファイルを利用することでノイズの影響を軽減し推定精度の向上が期待できる．

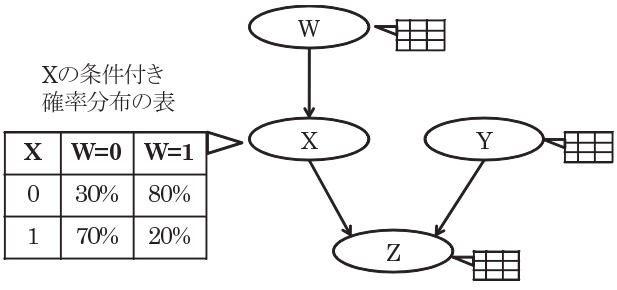
そこで本研究では，まず複数の実験条件別の時系列遺伝子発現プロファイルからそれぞれの遺伝子制御ネットワークを独立に推定し，その共通部分のネットワークを決定する．その後共通ネットワークを基に，各実験条件へ特徴的な形状へ共通ネットワークを拡張することで，各時系列遺伝子発現プロファイルに対応した遺伝子制御ネットワークを得る．これによりベイジアンネットワークを用いた遺伝子制御ネットワークの推定において，複数の遺伝子発現プロファイルを利用することができ，推定精度の向上につながると考えられる．

以下ではまず，2 章で遺伝子制御ネットワーク推定の従来手法とその問題点について説明する．3 章で提案手法について説明し，4 章では本手法の有効性を評価するために行った実験の結果を示す．

2. 遺伝子制御ネットワーク推定

2.1 ベイジアンネットワーク

ベイジアンネットワークは，**図 1** の例に示すような閉路なし有向グラフ（Directed Acyclic Graph: DAG）と条件付き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 [8], [19]．各変数を DAG のノードと 1 対 1 で対応付け，変数間に依存関係がある場合 DAG の対応するノード間に有向辺を引いて表現する．依存関係のある変数



X	W=0	W=1
0	30%	80%
1	70%	20%

図 1　ベイジアンネットワーク
Fig. 1　Bayesian network.

どうしがどのような確率関数に従って依存するかは，各変数に対応する条件付き確率分布の表を用いて表す．

ベイジアンネットワークは，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になるネットワークのグラフ構造を，組合せ最適化問題として探索することによって行われる [2], [11]．以下では，ベイジアンネットワークの評価関数と，組合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．事後確率 $p(B|D)$ 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，親変数を除く変数とは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．

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

$p(B)$ はベイジアンネットワーク B の事前確率を表しており，$p(D)$ はデータセット D の事前確率を表している．$p(D)$ は計算することが困難であり，データセット D は推定中につねに一定であるため，実際に $p(B|D)$ を推定に利用する場合には，$p(D)$ の計算は省略される場合が多い．

組合せ最適化問題の制約条件として，ネットワーク構造が DAG であることがあげられる．したがって，閉路なしの制約条件の下で，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合せを探す問題となる．以下では，ベイジアンネットワークのアルゴリズムを 3 種類説明する．

2.1.1　全件探索のアルゴリズム

ベイジアンネットワークの全件探索では，可能なすべてのネットワーク構造に対して目的関数 $p(B|D)$ を最大化する最適な親変数の組合せを探索する．以下で全件探索のアルゴリズムを説明する．

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) 適当なネットワーク構造を仮定する．
(3) 事前確率分布から，ネットワーク構造に対して入力データの計算を行い，事後確率分布を得る．
(4) ネットワーク構造を逐次変化させ，ネットワークのスコアを求める．
(5) ネットワークの全組合せのスコアの計算後，最もスコアの高い遺伝子制御ネットワークを出力する．

ベイジアンネットワークの全件探索ではネットワーク空間は $\mathrm{O}(3^{n^2})$ となり現実的ではないため，組合せ最適化問題はヒューリスティックに解くことが多い．

2.1.2　グリーディ法のアルゴリズム

ネットワーク構造の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的に増加するため，上記の全件探索ではなく近似アルゴリズムであるグリーディ法を用いることが多い [3]．以下でグリーディ法のアルゴリズムを説明する．

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適当に作成する．
(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) 枝をランダムに選び，制御関係の削除，付加，方向逆転を行い，スコアが最良のものを残す．
(4) スコアが改善されれば，再び (3) を行いスコアが改善されなくなるまで繰り返す．
(5) スコアが改善されなくなれば，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．

2.1.3　部分問題結合法

グリーディ法はベイジアンネットワークの近似アルゴリズムとして非常に高速に探索を行うことができるが，初期ネットワークによって推定結果が大きく変わる点，局所最適解に陥りやすい点が問題点となる．これらの問題点を解決するために，入力に対して一意に推定結果を決定し，推定精度がより高い部分問題結合法が開発された [26]．部分問題結合法では，一度に推定する遺伝子数を 3 とすることでネットワーク空間を $\mathrm{O}(n^3)$ としている．以下で部分問題結合法のアルゴリズムを説明する．

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) 推定対象の遺伝子から，すべての組合せの 3 遺伝子からなる 3 つ組を作成する．
(2) 各 3 つ組について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を推定する．
(3) 3 つ組の各制御関係に推定結果のスコアを重みとして付ける．
(4) 全ノード間の制御関係について，制御の向きの重みの和が最も高いものを選択する．

2.2　従来手法の問題点

1 章で述べたとおり，ベイジアンネットワークを用いて遺伝子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際に，遺伝子発現プロファイルにおける実験条件数が非常に少ないことが問題点となる．そのため従来の手法である，1 つの遺伝子発現プロファイルから 1 つの遺伝子制御ネットワークを推定する手法は精度が悪い，不安定なネットワークを推定する原因となる．この問題に対して同一実験条件下で測定されたデータを複数用いるアプローチが考えられるが，遺伝子発現データベースに蓄積されている遺伝子発現プロファイルには完全に同一の実験条件下で測定されたものは少ない．一方，同一の細胞組織に対して複数の実験条件下で測定された遺伝子発現プロファイルは

多く存在する．そのため遺伝子発現プロファイルにおける実験条件数が不足する問題点に対してのアプローチの１つとして，複数の実験条件下の時系列遺伝子発現プロファイルを利用することが考えられる．

しかし従来手法は１つの時系列遺伝子発現プロファイルから１つの遺伝子制御ネットワークを推定するものであり，複数の入力データを想定していない．また，

(1) 遺伝子は細胞の種類や実験条件によって発現量や制御関係が変動する可能性がある点
(2) 異なる環境で作成された遺伝子発現プロファイル間には，偶然誤差だけでなく系統誤差が存在する点

の２点の理由ため，複数の時系列遺伝子発現プロファイルを既存の遺伝子制御ネットワーク推定手法で扱うことができる形へ変換することは，推定精度低下の原因となる．そのため複数の時系列遺伝子発現プロファイルを利用して遺伝子制御ネットワークを高精度で推定する新たな手法が必要だと考えられる．

3. 提案手法

3.1 提案手法の目的と概要

本研究では，複数の時系列遺伝子発現プロファイルを利用することで遺伝子制御ネットワークの推定精度を向上させることが目的である．複数の時系列遺伝子発現プロファイルを利用するために，前節であげた２点の問題点を解決した新しい手法を提案する．

１つ目の問題点は，遺伝子は細胞の種類や実験条件によって発現量や制御関係が変動する可能性がある点である．複数実験条件下それぞれの遺伝子制御ネットワークは，各実験を通して一貫して観測することができる共通ネットワークと，各実験条件について特徴的なネットワークに分けることができる．この共通ネットワークにおいては，複数の時系列遺伝子発現プロファイルを利用することでノイズの影響を軽減させ推定精度の向上が期待できる．一方特徴的な制御関係はある限られた実験条件下によってのみ観測できるため，複数の実験条件下の時系列遺伝子発現プロファイルを利用することで，より観測しにくくなると考えられる．そのため提案手法では各実験条件に共通するネットワークと，各実験条件について特徴的なネットワークに分けて推定を行うことで推定精度の向上を狙う．

２つ目の問題点は，異なる環境で作成された遺伝子発現プロファイル間には，偶然誤差だけでなく系統誤差が存在する点である．マイクロアレイによる遺伝子発現の計測には多くのノイズが含まれており，推定精度を低下させる原因の１つとなっている．誤差は測定値から真の値を引いた値であるが，真の値を正確に知ることはできないため誤差を決定することは不可能である．またマイクロアレイ実験に含まれる誤差は，系統誤差と偶然誤差の２種類に分けることができる．系統誤差は蛍光色素の違い，アレイの違い，バックグラウンドの輝度などの実験環境の違いによってもたらされる偏りである．一方偶然誤差は環境要因には依存せず，つねに系に内在するバラツキである．異なる実験環境によって計測されたプロファイルを等しく扱うことは系統誤差の影響を受けやすいと考えられるため，複数の時系列遺伝子発現プロファイルを結合して１つの入力データセットとして遺伝子制御ネットワークを推定することは，推定精度低下の原因となる．提案手法では複数の時系列遺伝子発現プロファイルから各遺伝子制御ネットワークを独立に推定した後に結合することで，各実験を通して観測できる共通ネットワークを推定する．

一方各実験条件に特徴的な制御関係は共通ネットワークには現れにくく，各時系列遺伝子発現プロファイルから独立に推定する必要がある．そのため実験条件に特徴的なネットワークを推定するために，一時的に推定した共通ネットワークを利用し，各実験条件に特徴的な制御関係を付与することで共通ネットワークを拡張して各実験に対応した遺伝子制御ネットワークを推定する．しかし１つの時系列遺伝子発現プロファイルのみから推定をすることはノイズの影響を受けやすく，誤った制御関係を推定する可能性が高い．そこで提案手法では，共通ネットワークへ付与する制御関係に制約条件を設け，それを満たす制御関係のみを付与することで共通ネットワークを拡張する．

以上より，提案手法はフェイズ 1，2 の 2 フェイズに分けることができる．フェイズ１では複数の時系列遺伝子発現プロファイルから各遺伝子制御ネットワークをそれぞれ独立に推定し，推定結果を結合することで複数の実験条件下で一貫して存在する共通ネットワークを決定する．フェイズ２では，フェイズ１で決定した共通ネットワークへ各実験条件に特徴的な制御関係を付与し，特徴的な遺伝子制御ネットワークへ拡張することで，時系列遺伝子発現プロファイルの数だけ遺伝子制御ネットワークを決定する．提案手法全体のイメージ図を**図 2** に示す．

3.2 フェイズ 1

フェイズ１では複数の時系列遺伝子発現プロファイルを用いて，各実験条件下で一貫して存在する共通ネットワークを推定することを目的とする．

提案手法ではベイジアンネットワークによって複数プロファイルから各遺伝子制御ネットワークを独立に推定し，その推定結果を結合して１つの遺伝子制御ネットワークを決定する．しかしベイジアンネットワークによって推定されたネットワークは，式 (1) より導き出されたそのネットワークの尤もらしさを表すネットワークスコアのみを持ち，ネットワーク内の制御関係ごとの重みを判断できない．同様に，制御関係がないと推定された遺伝子間では弱い制御関係が切り捨てられた可能性があり，それらをすべて等価に扱うことは制御関係のとりこぼしにつながると考えられ

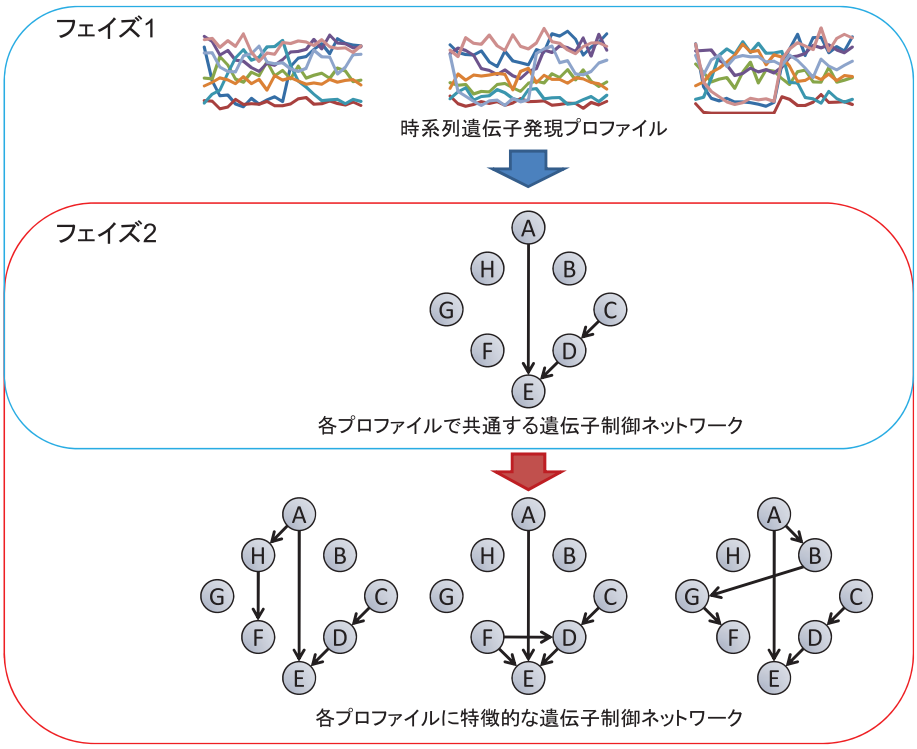
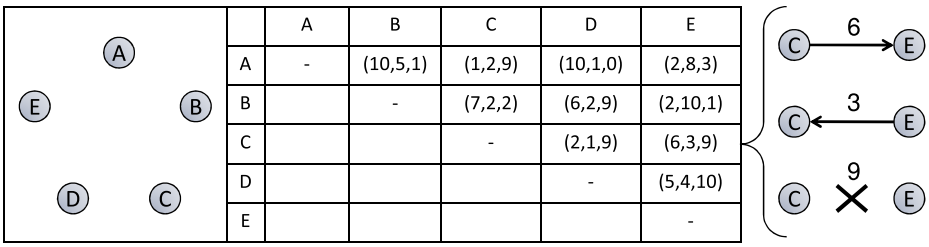


図 2　提案手法全体のイメージ図

Fig. 2　Conceptual representation of the proposed method.



	A	B	C	D	E
A	-	(10,5,1)	(1,2,9)	(10,1,0)	(2,8,3)
B		-	(7,2,2)	(6,2,9)	(2,10,1)
C			-	(2,1,9)	(6,3,9)
D				-	(5,4,10)
E					-

図 3　重み付き遺伝子制御ネットワークの例

Fig. 3　A sample of the weighted gene regulatory network.

る．これらの問題点を解決するために，本研究では部分問題結合法を用いて時系列遺伝子発現プロファイルから重み付き遺伝子制御ネットワークを推定し，結合することで各実験条件下で一貫して存在する遺伝子制御ネットワークを推定する．ここでは重み付き遺伝子制御ネットワークを，すべてのノード間に重みが付くグラフとする．また遺伝子 v_i，v_j の制御関係は，v_i から v_j への有向辺，v_j から v_i への有向辺，制御関係なしの 3 種類についての重みを同時に持つ．以下では，これらの制御関係をエッジタイプとする．また，本論文で用いる重み付き遺伝子制御ネットワークの例を図 **3** に示す．

フェイズ 1 についての概略図を図 **4** に示す．

3.2.1　部分問題結合法

提案手法では重み付き遺伝子制御ネットワークを推定するために，ベイジアンネットワークの近似法である部分問題結合法を利用する [26]．部分問題結合法ではベイジアンネットワークでの推定時の遺伝子数を 3 とすることで指数関数的に増加する探索空間を軽減するとともに，全組合せの 3 つ組に対するネットワークを適切に結合することで推定精度の低下を抑えている．部分問題結合法の 3 つ組結合時では，各 3 つ組のネットワークが持つネットワークスコアをその 3 つ組に存在する制御関係のスコアとして利用し，エッジタイプごとの重みの和が最も高い制御関係を選択している．提案手法では 3 つ組の結合時にはエッジタイプを決定せず，各遺伝子間はそれぞれエッジタイプごとの重みの和を保存する．

3.2.2　遺伝子制御ネットワーク組合せ手法

提案手法では，P（= 実験条件数）個の重み付き遺伝子制御ネットワークを組み合わせて，各遺伝子制御ネットワークで一貫して存在する制御関係を持つ共通ネットワークを作成する．組合せ手法は，ステップ 1，2 の 2 ステップから構成される．

ステップ 1 では全ノード間の制御関係の有無を決定する．ここで s_{ijpe} を時系列遺伝子発現プロファイル E_p

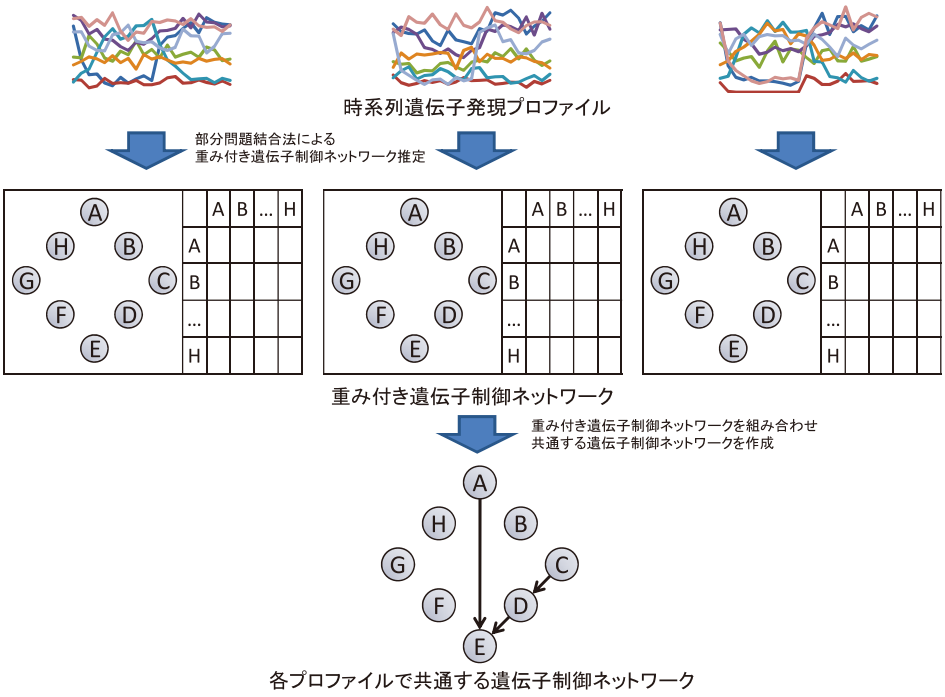


図 4　フェイズ 1 についての概略図
Fig. 4　A schematic representation of phase 1.

$(1 \leq p \leq P)$ から推定した重み付き遺伝子制御ネットワークにおける遺伝子 v_i，v_j 間の重みとする．ただし e は 3 種類のエッジタイプである（$e = 1$ は v_i から v_j への有向辺，$e = 2$ は v_j から v_i への有向辺，$e = 3$ は制御関係なし）．また時系列遺伝子発現プロファイル E_p から推定した重み付き遺伝子制御ネットワークにおける全遺伝子間の重みの集合を s_p とする．遺伝子 v_i，v_j 間の制御関係の有無を，式 (2)，(3) に従って決定する．

$$\sum_{p=1}^{P} s_{ijp1} + \sum_{p=1}^{P} s_{ijp2} > \sum_{p=1}^{P} s_{ijp3} \tag{2}$$

$$\sum_{p=1}^{P} s_{ijp1} + \sum_{p=1}^{P} s_{ijp2} > t \tag{3}$$

式 (2)，(3) を同時に満たすとき，遺伝子 v_i，v_j 間には制御関係があると決定してステップ 2 へ進む．それ以外の場合は制御関係はないと決定して終了する．ただし t は閾値とする．

ステップ 2 では，ステップ 1 において制御関係があると決定したノード間に関して，式 (4) に従ってその制御関係の向きを決定する．

$$\sum_{p=1}^{P} s_{ijp1} > \sum_{p=1}^{P} s_{ijp2} \tag{4}$$

式 (4) を満たすとき，有向辺の向きは遺伝子 v_i から v_j だと決定する．式 (4) を満たさない場合，有向辺の向きは逆となる．

以上のステップをすべての遺伝子間に対して行った結果として得られた遺伝子制ネットワークを共通ネットワークとして出力する．

3.3　フェイズ 2

フェイズ 2 では，フェイズ 1 で決定した共通ネットワークをもとに，各実験条件へ特徴的な制御関係を付与することで各実験条件へ特徴的な遺伝子制御ネットワークへ拡張することを目的とする．

各実験条件へ特徴的な制御関係は，その実験条件下における時系列遺伝子発現プロファイルからのみ観測できる．しかし 3.1 節で述べたように，1 つの時系列遺伝子発現プロファイルのみから制御関係を推定することでノイズの影響を受けやすくなり，誤った推定結果を導き出す可能性がある．そこで提案手法では 2 つの制約条件を設け，それを満たす制御関係のみを共通ネットワークへ付与することで各実験条件へ特徴的な遺伝子制御ネットワークを推定する．

1 つ目の制約条件は，ノイズの判定である．遺伝子発現プロファイルはノイズを多く含むデータであり，発現変動をしていない遺伝子であってもノイズによる変動が発生する．誤った制御関係を共通ネットワークへ付与する可能性を軽減するために，遺伝子の発現量の変動がノイズによるものかをランダムシャッフルサロゲート法（RS サロゲート法）[22], [24] によって検定し，ノイズと判定したものについては共通ネットワークへ付与しない．

2 つ目の制約条件は，共通ネットワークの構造保持である．3.2 節で述べたように共通ネットワークは複数の時系列遺伝子発現プロファイルを利用して推定されており，1

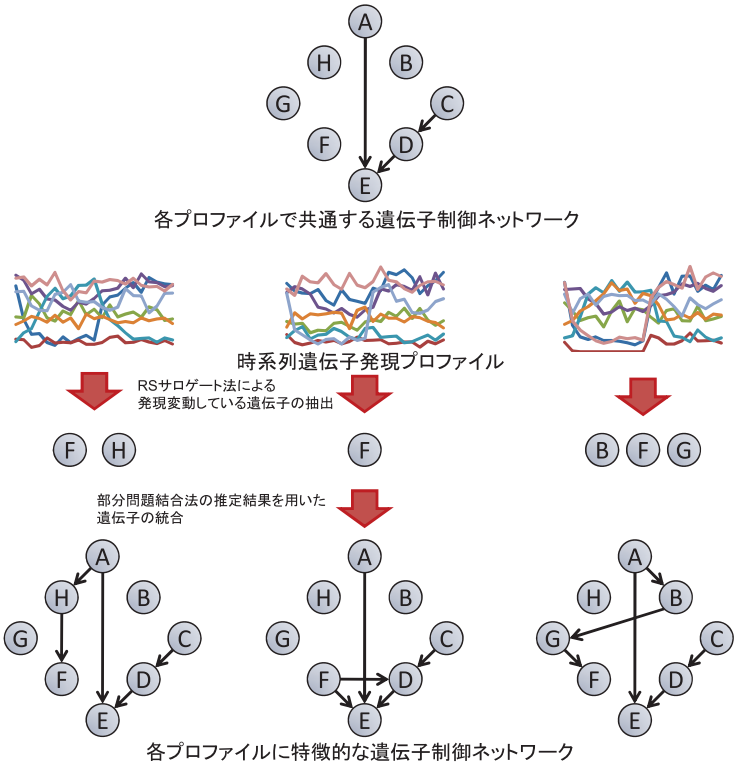


図 5　フェイズ 2 についての概略図

Fig. 5　A schematic representation of phase 2.

つの時系列遺伝子発現プロファイルから推定したものと比べて誤った推定結果が少ないことが期待できる．したがって実験条件へ特徴的な制御関係を付与する際，共通ネットワークの制御関係に反するものについては付与しない．

フェイズ 2 についての概略図を図 5 に示す．

3.3.1　RS サロゲート法による検定

提案手法では，RS サロゲート法による検定を用いてある遺伝子の時系列発現変動が不規則なノイズであるかどうかを検定する [22], [24]．サロゲート法はカオス時系列解析で用いられる手法の 1 つであり，解析対象となる時系列の統計的性質の一部を保存しその他の性質を破壊することにより，両者の統計的性質に有意差があることを示すことで破壊した性質の重要性を主張するものである．

提案手法では，サロゲート法の 1 つである RS サロゲート法による検定を行う．また破壊する統計的性質を，1 次マルコフ過程とする．そのため帰無仮説を「観測された時系列は時間的に無相関である」とする．この仮説に従う場合，対象遺伝子の時系列変動は時間的相関がないため，各点をランダムに入れ替えても統計的な有意差が算出されないと考えられる．帰無仮説が棄却された場合，オリジナルデータとランダムシャッフルデータとの間に有意な差があることになり，オリジナルデータがランダムノイズではないことがいえる．提案手法ではランダムシャッフルデータを 5 つ作成し，オリジナルデータの時点間の相関がランダムシャッフルデータの時点間の相関群より有意に大きい場合に帰無仮説を棄却する．検定は片側 5%の t 検定によって検定を行っている．

3.3.2　遺伝子制御ネットワークへの遺伝子の統合

提案手法では，RS サロゲート法によって有意に発現変動していると決定された遺伝子を，共通ネットワークへ統合する．その際に統合する遺伝子を含む 3 つ組の推定結果であるネットワークを用いることで，新たな制御関係を追加する．

共通ネットワークを構成する遺伝子群と統合する遺伝子を用い，統合する遺伝子が必ず含まれるすべての組合せの 3 つ組みを作成する．作成した 3 つ組の推定結果であるネットワークにおいて，共通ネットワークに存在する制御関係に反しない中で最もネットワークスコアが高いものを選び，そのネットワークの制御関係を共通ネットワークへ追加する．以上のステップを統合する遺伝子すべてに行い，特徴的な遺伝子制御ネットワークを決定する．

3.4　アルゴリズム

各フェイズのアルゴリズムを以下に示す．

3.4.1　フェイズ 1

入力：$E_1, \ldots, E_P$：時系列遺伝子発現プロファイル，t：閾値

出力：C：共通ネットワーク，$Z = (X_1, Y_1), \ldots, (X_P, Y_P)$：3 つ組の各（ネットワーク構造，ネットワークスコア）の 2 つ組の集合

変数：n：遺伝子数，P：入力時系列遺伝子発現プロファイル数，s_p：E_p から推定された重み付き遺伝子制御ネットワークの全遺伝子間の重みの集合

(1) $i \leftarrow 1$
(2) E_i から，すべての組合せの 3 つ組を作成する．
(3) すべての組合せの 3 つ組について，それぞれベイジアンネットワークの全件探索によってすべてのネットワークの構造 X_i とそのネットワークスコア Y_i を計算する．
(4) X_i と Y_i を用いた重み付き遺伝子制御ネットワークを作成し，全遺伝子間の重みの集合を s_i とする．
(5) $i \leftarrow i+1$
(6) $i \leq P$ を満たす限り (2) から (5) を繰り返す．
(7) $i \leftarrow 1$
(8) $j \leftarrow i+1$
(9) 遺伝子 v_i，v_j 間の制御関係の有無を，式 (2)，(3) に従って決定する．
(10) 遺伝子 v_i，v_j 間に制御関係があると決定された場合，式 (4) に従って制御関係の向きを決定する．
(11) $j \leftarrow j+1$
(12) $j \leq n$ を満たす限り (9) から (11) を繰り返す．
(13) $i \leftarrow i+1$
(14) $i < n$ を満たす限り (8) から (13) を繰り返す．
(15) すべての遺伝子間の制御関係が決定された遺伝子制御ネットワークを，共通ネットワーク C として出力する．

3.4.2　フェイズ 2

入力：$E_1, \ldots, E_P$：時系列遺伝子発現プロファイル，C：共通ネットワーク，Z：3 つ組の各（ネットワーク構造，ネットワークスコア）の 2 つ組の集合

出力：$N_1, \ldots, N_P$：遺伝子制御ネットワーク

(1) $i \leftarrow 1$
(2) E_i から，C で用いられていない遺伝子をすべて抽出し，追加候補遺伝子群とする．
(3) 追加候補遺伝子群の遺伝子すべてに対して RS サロゲート法による検定を行い，ノイズと判定された遺伝子をすべて追加候補遺伝子群から削除する．
(4) 追加候補遺伝子群の遺伝子について，その遺伝子と C に属する遺伝子 2 つによる 3 つ組のネットワーク構造の中から，C の制御関係に反せず最もネットワークスコアが高いものを Z を用いて決定し，そのネットワーク構造の制御関係を C へ付与する．
(5) 追加候補遺伝子群のすべての遺伝子を C へ付与したものを遺伝子制御ネットワーク N_i とする．
(6) $i \leftarrow i+1$
(7) $i \leq P$ を満たす限り (2) から (6) を繰り返す．
(8) $N_1, \ldots, N_P$ を出力する．

4. 検証実験

本手法の評価を検証するため，2 つの実験を行った．実験 1 では提案手法のフェイズ 1 についての実験を行った．実験 2 では提案手法のフェイズ 1，2 を通した全体での実験を行った．その方法，結果および考察を述べる．

実験にあたっては，提案手法を R 言語により実装した．提案手法中のベイジアンネットワークによるネットワーク推定部については，R のパッケージである deal [3] を使用した．また入力データとして連続値である遺伝子発現プロファイルを利用し，グラフ構造である遺伝子制御ネットワークを出力する．

4.1　実験 1

提案手法のフェイズ 1 についての遺伝子制御ネットワーク推定精度の比較を行うため，グリーディ法，部分問題結合法，提案手法の 3 手法を用いた実験を行った．従来の遺伝子制御ネットワーク推定手法の比較対象として，グリーディ法を用いる．また，複数の遺伝子制御ネットワークを結合する際に重み付き遺伝子制御ネットワークを用いる点の比較対象として，部分問題結合法を用いる．

本実験では，遺伝子制御ネットワークの推定精度を比較するために receiver operating characteristic（ROC）曲線を用いる．ROC 曲線は specificity(TN/TN+FP) と sensitivity(TP/TP+FN) の対比をプロットしてグラフとして表現する．ここで TN，FP，TP，FN はそれぞれ真陰性，偽陽性，真陽性，偽陰性の数である．

本実験では入力データとして時系列遺伝子発現プロファイルを 5 種類与える．そのためグリーディ法，部分問題結合法ではそれぞれ 5 つの遺伝子制御ネットワークが推定できる．推定した 5 つの遺伝子制御ネットワークを 1 つに統合する際に，5 つの中のいくつ以上のネットワークに共通する制御関係を用いるかという点を変えながら統合する．つまり，推定した 5 つの遺伝子制御ネットワークのすべてのネットワークに共通して存在する制御関係を統合したネットワーク，4 つ以上のネットワークに共通して存在する制御関係を統合したネットワーク，といった例のように 5 つの統合ネットワークを作成する．作成した 5 つの遺伝子制御ネットワークについて sensitivity と specificity を算出して ROC 曲線を作成する．提案手法では，閾値 t を変動させることで ROC 曲線を作成する．

実験データとして，DREAM4 *in silico* データセットを用いる [17], [18], [20]．データセットの遺伝子数は 10，時点数は 21 である．本実験では，これらの条件を持つ時系列 5 本を複数時系列として用いる．5 本の時系列はすべて同じ遺伝子制御ネットワークから生成されており，またそれぞれ異なる実験条件下を想定した発現変動をしている．そのため 1 本の時系列において 10 遺伝子すべてが発現して

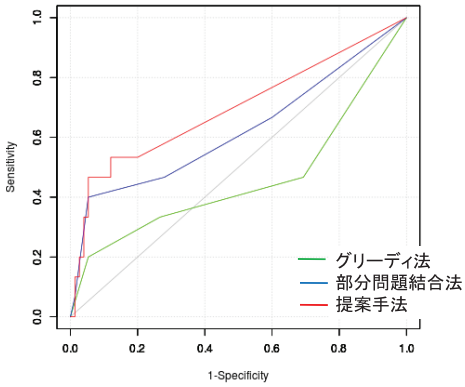


図 6　グリーディ法，部分問題結合法，提案手法の ROC 曲線
Fig. 6　ROC curves of the greedy method, the uniting of partial problems, and the proposed method.

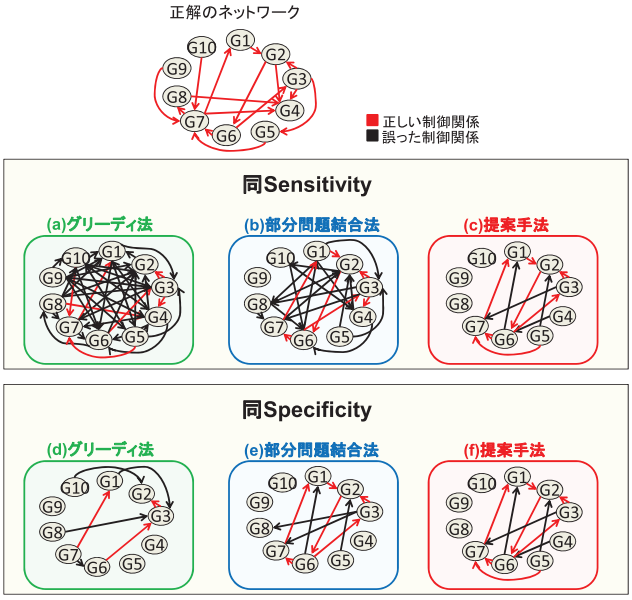


図 7　ネットワーク比較図
Fig. 7　Comparison of the networks.

いるわけではなく，発現しているのはその一部である．正解の遺伝子制御ネットワークは，**図 7** 上部のものである．

グリーディ法，部分問題結合法，提案手法の ROC 曲線を**図 6** に示す．

図 6 から，提案手法が他の 2 手法と比べ sensitivity，specificity ともに著しく改善していることが分かる．特に，同 sensitivity 帯での specificity の改善が著しい．さらなる比較のために，3 手法において同じ sensitivity を持つ遺伝子制御ネットワーク，同じ specificity を持つ遺伝子制御ネットワークをそれぞれ 3 つ取り出したものを図 7 として示す．

図 7 は，ネットワーク比較図である．また図 7 の各ネットワークの推定精度を**表 1** にまとめる．上部のネットワークを正解のネットワークとする．ネットワーク (a)，(b)，(c) は sensitivity がすべて 46.7%のネットワーク，(d)，(e)，(f) は specificity が 94.7%のネットワークである．同じ specificity を持つネットワーク (d)，(e)，(f) を見ると，グリーディ法は部分問題結合法，提案手法と比較して正しい制御関係をあまり推定できていないことが分かる．また部分問題結合法と提案手法の比較では，提案手法のほうが正しい制御関係は多いが，大きな差はない．一方同じ sensitivity を持つネットワーク (a)，(b)，(c) を見ると，非常に大きな差があることが分かる．ネットワーク (a)，(b)，(c) の specificity はそれぞれ 30.7%，72.0%，94.7%であり，提案手法は正しい制御関係の数を減らすことなく，誤った制御関係を減らすことに成功している．

表 1　推定精度の比較 1
Table 1　Comparison of the estimation accuracy 1.

	sensitivity	specificity
(a) グリーディ法	46.7%	30.7%
(b) 部分問題結合法	46.7%	72.0%
(c) 提案手法	46.7%	94.7%
(d) グリーディ法	20.0%	94.7%
(e) 部分問題結合法	40.0%	94.7%
(f) 提案手法	46.7%	94.7%

4.2　実験 2

提案手法のフェイズ 1，2 を通した全体での遺伝子制御ネットワーク推定精度の比較を行うため，グリーディ法，提案手法による実データを用いた実験を行った．

実験データとして，網膜視細胞のデータセットを用いる．データセットの遺伝子数は 16，時点数は 5 時点である．本実験では細胞視細胞の桿体分化時，錐体分化時の時系列 2 本を複数時系列として用いる．本論文では遺伝子発現プロファイルとして GEO（アクセッション番号：GSE4051）のものを，正解とする遺伝子制御ネットワークとして Hao らの論文 [9] のものを用いている．

網膜視細胞は網膜を構成する細胞の 1 つであり，他には神経節細胞，アマクリン細胞，双極細胞，水平細胞，ミュラーグリア細胞が存在する．網膜視細胞には暗所での視覚をつかさどる桿体と，明所や色覚をつかさどる錐体の 2 種類が存在する．網膜の 6 種類の細胞の中で網膜視細胞へ運命決定された細胞について，桿体か錐体のいずれに分化するかを誘導する転写因子として，NRL（neural retina leucine zipper）があげられる．NRL は視細胞に特徴的に発現し，桿体への分化誘導を行い錐体への分化を抑制する転写因子である．そのため NRL 遺伝子のノックアウトマウスはすべての桿体の錐体への形質転換が起こる．本実験では，NRL 遺伝子ノックアウトマウスの時系列遺伝子発現プロファイルを錐体分化，ワイルドタイプコントロールマウスの時系列遺伝子発現プロファイルを桿体分化のものとして用いる．また正解とする遺伝子制御ネットワークにおいては，ノックアウトする NRL 遺伝子と NRL と制御関係にある RORB を除いたものとしている．

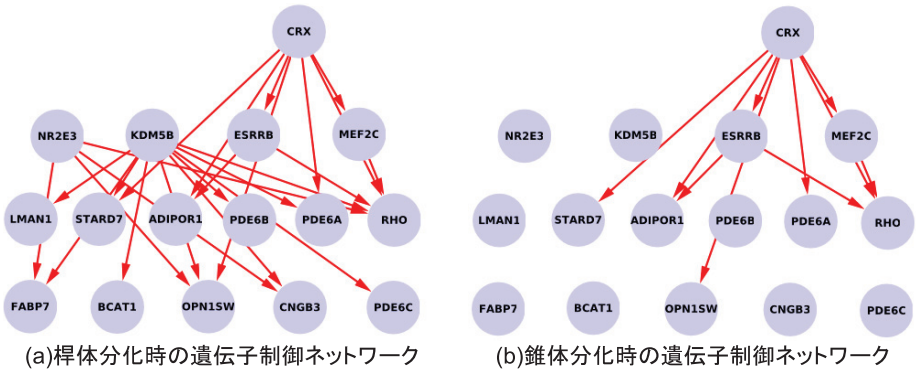


図 8　網膜視細胞の桿体分化，錐体分化時の正解のネットワーク
Fig. 8　True networks of rod and cone photoreceptors.

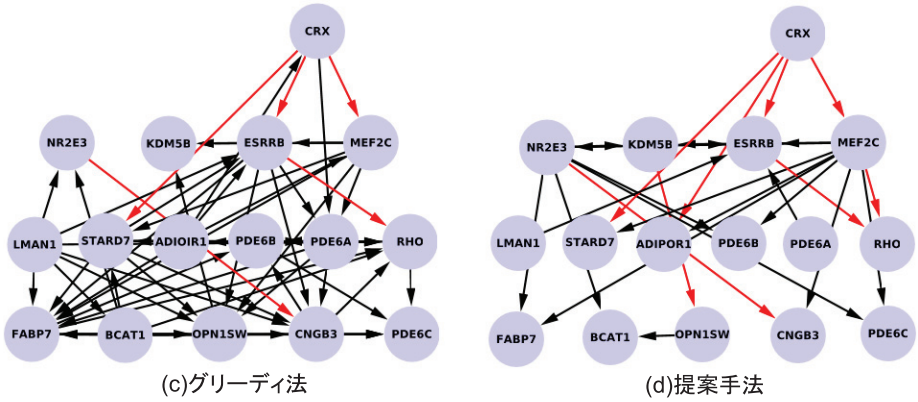


図 9　グリーディ法，提案手法による網膜視細胞の桿体分化時の推定ネットワーク
Fig. 9　Estimated networks of rod photoreceptor by the greedy method and the proposed method.

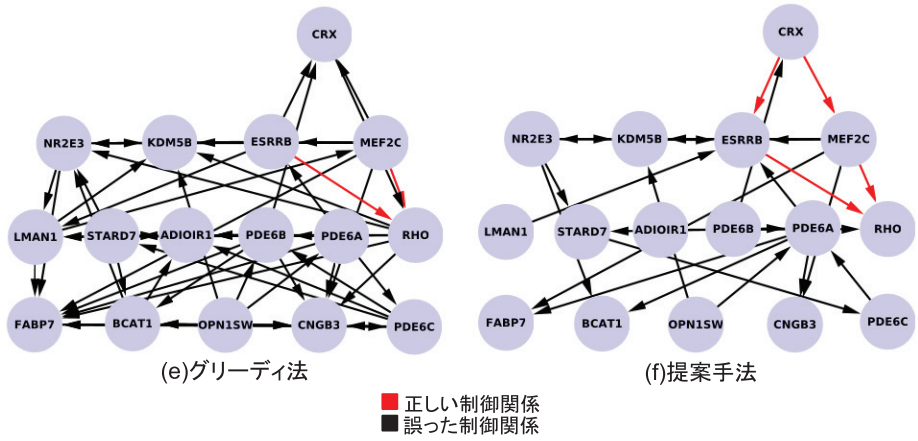


図 10　グリーディ法，提案手法による網膜視細胞の錐体分化時の推定ネットワーク
Fig. 10　Estimated networks of cone photoreceptor by the greedy method and the proposed method.

本実験で網膜視細胞の桿体分化，錐体分化それぞれ正解とするネットワークを**図 8** に示す．

またグリーディ法，提案手法による桿体分化時の推定ネットワークを**図 9** に，錐体分化時の推定ネットワークを**図 10** に，推定精度のまとめを**表 2** に示す．

赤矢印が正しく推定した制御関係，黒矢印が誤って推定した制御関係である．桿体分化時のグリーディ法と提案手法の sensitivity はそれぞれ 20.8%と 37.5%，specificity は

表 2　推定精度の比較 2
Table 2　Comparison of the estimation accuracy 2.

		sensitivity	specificity
桿体分化	グリーディ法	20.8%	37.5%
	提案手法	43.8%	82.3%
錐体分化	グリーディ法	20.0%	40.0%
	提案手法	50.9%	82.7%

43.8%と 82.3%であり，ともに提案手法が上回っていることが分かる．また図 9 より，グリーディ法によって推定できた正しい制御関係はすべて提案手法によっても推定できていることが分かる．

錐体分化時のグリーディ法と提案手法の sensitivity はそれぞれ 20.0%と 40.0%，specificity は 50.9%と 82.7%である．錐体分化時の正解の遺伝子制御ネットワークは桿体分化時のものと比べ制御関係が少ないため，グリーディ法・提案手法ともに正しく推定した制御関係が減っているにもかかわらず sensitivity，specificity は大きく低下していない．錐体分化時においても，グリーディ法によって推定できた正しい制御関係はすべて提案手法によっても推定できている．

4.3　考察

実験 1 の図 7 を見ると，G5 から G7 への制御関係を正しく推定できているのはグリーディ法の (a) と提案手法のみである．5 つの時系列から推定した 5 つの遺伝子制御ネットワークの中で，この制御関係が現れていたものはグリーディ法，部分問題結合法ともに 1 つのみであった．そのためネットワーク (a) のように，すべてのネットワークに存在する制御関係を用いて統合を行わない限りこの制御関係は現れない．しかしこの制御関係が現れなかった他の 4 つのネットワークにおいては弱い制御関係が存在しており，それをネットワーク単位で切り捨てることによって最終的に統合結果には現れなかった．提案手法ではネットワーク単位で制御関係を切り捨てず，制御関係がある場合・ない場合の重みをすべて残した重み付きグラフを用いることによって，この制御関係を推定することができている．

実験 1，2 において，提案手法は既存手法と類似したネットワークを推定している．特に正しい制御関係においては，既存手法によって推定できたが提案手法によって推定できていないものはほとんどない．また実験 1 における G9 から G7 への制御関係，G3 から G5 への制御関係や，実験 2 における CRX から OPN1SW，PDE6A への制御関係など，どの手法によっても推定することができない制御関係も多い．そのため提案手法は正しく推定できる制御関係の数を劇的に増やすものではなく，また従来の手法ではどうしてもとらえられなかった制御関係を必ずしもとらえられるものではないと考えられる．一方誤った制御関係を推定する数は，実験 1，2 を通して大きく減らすことができている．これはフェイズ 1 において 3 つのエッジタイプそれぞれの重みを持つ遺伝子制御ネットワークを組み合わせることで一定以下の数の実験条件下でのみ現れる制御関係やノイズによる誤った制御関係を適切に排除し，またフェイズ 2 においてフェイズ 1 で決定した共通部分ネットワークの構造を維持しつつ各実験条件へ特徴的なネットワークへ拡張することで無駄な制御関係を増やすことを避けることができていると考えられる．

実験 2 において，網膜視細胞の錐体分化時の正解のネットワークの制御関係は，桿体分化時のネットワーク中にすべて現れるものとなっている．したがって桿体分化時，錐体分化時の共通ネットワークは錐体分化時のネットワークと同じものとなる．すなわちこの 2 つのネットワークは共通部分が多いネットワークであり，提案手法の有効性を十分に発揮できるデータであったといえる．しかしより共通部分が少ないものやまったく共通部分がないものに対しては，提案手法は有効に作用しない可能性がある．提案手法による推定を行う前に，対象の複数時系列遺伝子発現プロファイルがどの程度共通部分を持つのか判断する必要があり，今後の課題である．

5.　おわりに

本研究では時系列発現プロファイルに基づく遺伝子制御ネットワーク推定時の精度向上を目的として，複数時系列を利用した遺伝子制御ネットワーク推定手法を提案した．

複数の遺伝子制御ネットワークを通して共通する部分ネットワークにおいては複数の遺伝子発現プロファイルを用いて共通ネットワークを推定することで推定精度の向上が認められた．非共通部分ネットワークにおいては，共通する部分ネットワークの構造を保持したまま各実験条件下へ特徴的なネットワークへ拡張させることで推定精度の低下を防ぐことで sensitivity，specificity の両者の上昇が認められた．特に specificity については大きく上昇している．

提案手法は複数時系列遺伝子発現プロファイルすべてを通して共通する部分ネットワークを推定するため，共通部分が少ない遺伝子発現プロファイルが入力されると推定精度が下がる．そのため提案手法による推定前に複数時系列遺伝子発現プロファイルを共通する部分ネットワークが多い組合せに分割するなどの前処理が必要となる点が今後の課題である．

謝辞　本研究は，JSPS 科研費 22310125，22680023 および文部科学省 HPCI 戦略プログラム分野 1 の助成を受けている．

参考文献

[1] Barrett, T., Wilhite, S.E., Ledoux, P., et al.: NCBI GEO: Archive for Functional Genomics Data Sets — Update, *Nucleic acids research*, Vol.41, pp.D991–D995 (2013).
[2] Bøttcher, S.G.: *Learning Bayesian Networks with Mixed Variables*, Department of Mathematical Sciences (2004).
[3] Bøttcher, S.G. and Dethlefsen, C.: deal: A Package for Learning Bayesian Networks, *Journal of Statistical Software*, Vol.8, pp.1–40 (2003).
[4] Chen, T., He, H.L. and Church, G.M.: Modeling Gene Expression with Differential Equations, *Pacific Symposium On Biocomputing*, Vol.4, No.5, pp.29–40 (1999).

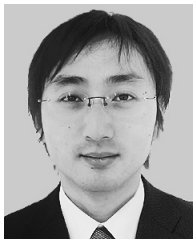
[5] Davidson, E.H.: Emerging Properties of Animal Gene Regulatory Networks, *Nature*, Vol.468, No.7326, pp.911–920 (2010).
[6] DeRisi, J.L., Iyer, V.R. and Brown, P.O.: Exploring the Metabolic and Genetic Control of Gene Expression on a Genomic Scale, *Science*, Vol.278, No.5338, pp.680–686 (1997).
[7] Dondelinger, F., Husmeier, D. and Lebre, S.: Dynamic Bayesian Networks in Molecular Plant Science Inferring Gene Regulatory Networks from Multiple Gene Expression Time Series, *Euphytica*, Vol.183, pp.361–377 (2012).
[8] Friedman, N., Linial, M., Nachman, I. and Pe'er, D.: Using Bayesian Network to Analyze Expression Data, *Journal of Computational Biology*, Vol.7, pp.601–620 (2000).
[9] Hao, H., Kim, D.S., Klocke, B., et al.: Transcriptional Regulation of Rod Photoreceptor Homeostasis Revealed by in Vivo NRL Targetome Analysis, *PLoS Genetics*, Vol.8, No.4, p.e1002649 (2012).
[10] Hecker, M., Lambeck, S., Toepfer, S., et al.: Gene Regulatory Network Inference: Data Integration in Dynamic Models-a Review, *Bio Systems*, Vol.96, No.1, pp.86–103 (2009).
[11] Heckerman, D.: *A Tutorial on Learning with Bayesian Networks*, Microsoft Research (1996).
[12] Iba, H. and Mimura, A.: Inference of a Gene Regulatory Network by Means of Interactive Evolutionary Computing, *Information Sciences*, Vol.145, pp.225–236 (2002).
[13] Karlebach, G. and Shamir, R.: Modelling and Analysis of Gene Regulatory Networks, *Nature Reviews Molecular Cell Biology*, Vol.9, No.10, pp.770–780 (2008).
[14] Kim, H., Lee, J.K. and Park, T.: Boolean Networks Using the Chi-square Test for Inferring Large-scale Gene Regulatory Networks, *BMC Bioinformatics*, Vol.8, p.37 (2007).
[15] Kim, S., Imoto, S. and Miyano, S.: Dynamic Bayesian Network and Nonparametric Regression for Nonlinear Modeling of Gene Networks from Time Series Gene Expression Data, *Biosystems*, Vol.75, No.1-3, pp.57–65 (2004).
[16] Kitano, H.: Systems Biology: A Brief Overview, *Science*, Vol.295, pp.1662–1664 (2002).
[17] Marbach, D., Prill, R.J., Schaffter, T., et al.: Revealing Strengths and Weaknesses of Methods for Gene Network Inference, *PNAS*, Vol.183, pp.6286–6291 (2010).
[18] Marbach, D., Schaffter, T., Mattiussi, C. and Floreano, D.: Generating Realistic in Silico Gene Networks for Performance Assessment of Reverse Engineering Methods, *Journal of Computational Biology*, Vol.16, pp.229–239 (2009).
[19] Pe'er, D., Regev, A., Elidan, G. and Friedman, N.: Inferring Subnetworks from Perturbed Expression Profiles, *Bioinformatics*, Vol.17, No.2-3, pp.S215–S224 (2001).
[20] Prill, R.J., Marbach, D., Saez-Rodriguez, J., et al.: Towards a Rigorous Assessment of Systems Biology Models: The DREAM3 Challenges, *PLoS One*, Vol.5, p.18 (2010).
[21] Savageau, M.A. and Rosen, R.: *Biochemical Systems Analysis: a Study of Function and Design in Molecular Biology*, Addison-Wesley Educational Publishers Inc. (1976).
[22] Schreiber, T. and Schmitz, A.: Improved Surrogate Data for Nonlinearity Tests, *Physical Review Letters*, Vol.77, p.635 (1996).
[23] Spellman, P.T., Sherlock, G., Zhang, M.Q., et al.: Comprehensive Identification of Cell Cycle–regulated Genes of the Yeast Saccharomyces Cerevisiae by Microarray Hybridization, *Molecular Biology of the Cell*, Vol.9, No.12, pp.3273–3297 (1998).
[24] Theiler, J., Eubank, S., Longtin, A., et al.: Testing for Nonlinearity in Time Series: The Method of Surrogate Data, *Physica D: Nonlinear Phenomena*, Vol.58, No.1-4, pp.77–94 (1992).
[25] Toh, H. and Horimoto, K.: Inference of a Genetic Network by a Combined Approach of Cluster Analysis and Graphical Gaussian Modeling, *Bioinformatics*, Vol.18, No.2, pp.287–297 (2002).
[26] Watanabe, Y., Seno, S., Takenaka, Y. and Matsuda, H.: An Estimation Method for Inference of Gene Regulatory Network Using Bayesian Network with Uniting of Partial Problems, *BMC Genomics*, Vol.13, p.S12 (2012).

渡邉 之人

2011 年大阪大学基礎工学部情報科学科卒業．2013 年大阪大学大学院情報科学研究科バイオ情報工学専攻博士前期課程修了．同年日本電信電話株式会社入社．

瀬尾 茂人　（正会員）

大阪大学大学院情報科学研究科助教．2006 年大阪大学大学院情報科学研究科修了．博士（情報科学）．同年同研究科助手．2007 年より現職．JSBi，MBSJ，IEEE 各会員．



竹中 要一　（正会員）

大阪大学大学院情報科学研究科准教授．2000 年大阪大学大学院基礎工学研究科修了．博士（工学）．同年大阪大学大学院基礎工学研究科助手．2002 年大阪大学大学院情報科学研究科助教授．2007 年より現職．電子情報通信学会，言語処理学会，JSBi，ISCB，IEEE 各会員．

松田 秀雄　（正会員）

大阪大学大学院情報科学研究科教授．1987 年神戸大学大学院自然科学研究科修了 (学術博士)．同年同大学工学部助手．1994 年大阪大学基礎工学部助教授．2002 年より現職．JSBi, ISCB, IEEE CS，ACM 各会員．