

# ダイナミック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推定の部分問題化による近似解法

上木 怜<sup>1</sup> 瀬尾 茂人<sup>1</sup> 竹中 要一<sup>1</sup> 松田 秀雄<sup>1</sup>

**概要:** 本研究ではダイナミック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推定法を提案する。提案手法では任意の3遺伝子についての部分問題を作成し、その解を情報量基準に基づいて結合することで近似解を得る。遺伝子制御ネットワークの高速かつ高精度な推定手法の確立は生物学分野だけでなく医学・創薬分野への発展にもつながる重要課題である。ネットワーク推定のモデルとして、時系列を考慮する確率モデルであるダイナミックベイジアンネットワークがあるが、この問題はNP困難であり、近似解法が求められている。従来の代表的な近似法である Greedy Hill-Climbing 法と提案手法の精度・計算時間を比較し、今後の課題を述べる。

## 1. はじめに

すべての生物は遺伝子をもっており、遺伝子はDNAにコードされている。DNAがRNAを介してタンパク質を合成する分子生物学の基本原則をセントラルドグマと呼び、セントラルドグマに従って遺伝情報をタンパク質に変換する過程を遺伝子発現と呼ぶ。遺伝子発現は個々の遺伝子が独立して行うのではなく、それぞれが互いに抑制・促進といった相互作用をもち発現量の制御を行っている。この相互作用はネットワークで表すことができ、これを遺伝子制御ネットワークという。遺伝子制御ネットワークの解明は細胞機能の解析、創薬ターゲットの選定などを可能にし [1]、医学・創薬分野で重要な課題とされる。

近年では、DNAマイクロアレイや次世代シーケンサに代表される生物学的実験技術の発達により、膨大な量のデータが蓄積されるようになった。しかし、生命現象の複雑さや情報量の多さにより、それら大量のデータから生物学的手法のみで実際の生物情報を解析するのは困難を極める。このため人工知能をはじめとする情報学的なアプローチによる、高速で高精度な解析手法が必要とされている。遺伝子制御ネットワークの推定はそのひとつであり、バイオインフォマティクスにおける重要なテーマである [2]。

遺伝子制御ネットワークの推定に用いられるモデルとしてブーリアンネットワーク [3], [4]、微分方程式モデル [5] などが存在し、中でも多く用いられているモデルのひとつ

として、ベイジアンネットワーク [6], [7], [8] がある。推定に用いるデータは多くの誤差を含むが、ベイジアンネットワークは確率的推論モデルの一種であり、誤差を含むデータから確率的に挙動を解析するのに有用なモデルである。

さらに、時系列データを用いることでベイジアンネットワークを拡張したダイナミックベイジアンネットワーク [9], [10] による推定手法も盛んに研究されている。時系列を考慮することで、ベイジアンネットワークでは表現ができなかったタンパク質生合成による遺伝子制御の遅延を表すことができる。

ダイナミックベイジアンネットワークは遺伝子制御ネットワーク推定に有用なモデルであるが、最適解を求める問題はNP困難として知られる。近似解法として Greedy Hill-Climbing 法 (GHC) による推定方法 [11], [12] が存在するが、求めた局所解が大域最適解に近い保証はない。本研究では、ダイナミック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推定において、問題を部分問題化することによる計算量の軽減とGHCより高精度な推定を目的とする。

## 2. 遺伝子制御ネットワーク

### 2.1 遺伝子発現制御と遺伝子制御ネットワーク

遺伝子発現は大きく分けて転写、翻訳と呼ばれる二つの段階からなる。まず、転写によりDNA上の遺伝子領域の塩基配列から相補的な塩基配列を持つRNAが合成される。このRNAはmRNA (messenger RNA) と呼ばれる。次に翻訳により、mRNAを元にタンパク質が合成される。

遺伝子から合成されるタンパク質の量は一定ではなく、

<sup>1</sup> 大阪大学大学院情報科学研究科  
Graduate School of Information Science and Technology  
Osaka University  
Yamadaoka, Suita City, Osaka 565-0062, Japan

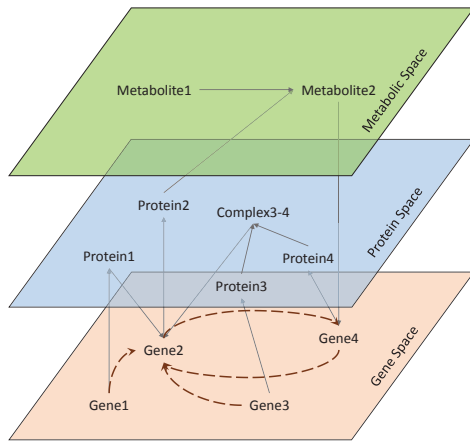


図 1 3層からなる遺伝子制御ネットワーク  
(Brazhnik et al. [13] より引用)

Fig. 1 Gene Regulatory Network

表 1 時系列遺伝子発現プロファイル

Table 1 Time Series Gene Expression Profile

	時刻 1	時刻 2	...	時刻 $n$
遺伝子 1	$x_1^1$	$x_1^2$	...	$x_1^n$
遺伝子 2	$x_2^1$	$x_2^2$	...	$x_2^n$
⋮	⋮	⋮	⋮	⋮
遺伝子 $p$	$x_p^1$	$x_p^2$	...	$x_p^n$

他の遺伝子から合成されたタンパク質や代謝産物によってその量が抑制・促進される。このように遺伝子，タンパク質，代謝産物は複雑な相互関係をもっており，これらの関係はネットワークとして捉えられる。

遺伝子制御ネットワークは遺伝子発現における遺伝子間の相互作用をネットワークとして表したものである。これを代謝産物間の化学反応を表す代謝ネットワーク，タンパク質の複合体形成などタンパク質間の相互作用を表すタンパク質ネットワークとともに層構造として表すと図 1 のように表される。このことから遺伝子制御ネットワークは代謝産物間やタンパク質間の相互作用を遺伝子間の相互作用として射影したものであると言える。

### 2.1.1 遺伝子制御ネットワーク推定

遺伝子制御ネットワークの推定では，ある時点での遺伝子発現量を記録した，遺伝子発現プロファイルと呼ばれるデータを学習データとする。時系列遺伝子発現プロファイルは表 1 のような表形式で表される。

遺伝子制御ネットワークの推定アルゴリズムの構築とは，遺伝子発現プロファイルを GEP，遺伝子を  $V$ ，制御関係を表す辺を  $E$  として，

$$f : \text{GEP} \rightarrow G = (V, E) \quad (\text{最適化: } \text{score}(G, \text{GEP}))$$

となる  $f$  を求めることである。ここで  $\text{score}$  とは GEP に対する  $G$  の評価関数で，式 (1) などで表される。以下にこの問題を解くにあたって克服する必要がある課題を説明

する。

### 組み合わせ爆発

遺伝子数が増加するにつれ，それらの遺伝子が構成しうる遺伝子制御ネットワークの数は幾何級数的に増加する。ヒトの遺伝子数は約 2 万と推定されており，全遺伝子についてのネットワーク推定を行うには考えられる全ネットワークから効率的に解を求めなければならない。

### 誤差を含むデータからの推定

遺伝子発現プロファイルは誤差を多く含む特徴がある。したがって誤差に強いネットワーク推定モデルを選択しなければならない。

**発現制御の時間差** 実際の発現制御は，タンパク質，代謝産物を介する。それらの合成時間によって発現制御には時間遅延が生じる。したがって時間差を考慮できるモデルとアルゴリズムが必要である。

このことから，誤差に強く時系列を考慮できるモデルを利用し，考えられる遺伝子制御ネットワークから効率的に解を求めるアルゴリズムを構築することが研究対象である。

## 3. 遺伝子制御ネットワーク推定に用いるモデルと手法

本研究では遺伝子制御ネットワークの推定にダイナミックベイジアンネットワークを用いる。ダイナミックベイジアンネットワークはベイジアンネットワークを拡張したものであり，まずベイジアンネットワークについて解説したあと，ダイナミックベイジアンネットワークを解説する。

### 3.1 ベイジアンネットワーク

ベイジアンネットワークはグラフィカルモデルの一種であり，辺に向きを持たせることで変数間の因果関係を表した非循環有向グラフ (Directed Acyclic Graph: DAG) によるモデルである [6]。

ベイジアンネットワークは，確率変数間の因果関係をその変数間に定義される条件付き確率によって表す。確率変数  $X_i, X_j$  間に  $X_i$  を原因， $X_j$  をそれに対する結果とした因果関係があるとき，ベイジアンネットワーク上で  $X_i \rightarrow X_j$  と有向辺を引く。このとき  $X_i$  を  $X_j$  の親ノード， $X_j$  を  $X_i$  の子ノードと呼ぶ。 $X_j$  の親ノードの集合を  $\pi(X_j) = \{X_{j1}, X_{j2}, \dots, X_{ji}\}$  と表すと変数  $X_j$  と親ノード間の関係は条件付き確率  $\Pr[X_j | \pi(X_j)] = \Pr[X_j | X_{j1}, X_{j2}, \dots, X_{ji}]$  で定義される。さらに  $n$  個の確率変数  $X_1, X_2, \dots, X_n$  が存在するとき，すべての確率変数の同時確率は  $\Pr[X_1, X_2, \dots, X_n] = \prod_{j=1}^n \Pr[X_j | \pi(X_j)]$  と表される。

ベイジアンネットワークを用いて遺伝子制御ネットワーク推定を行う場合，確率変数  $X_i$  は遺伝子  $i$  の発現量を表す。ネットワークの推定には評価関数を用いてスコアの最

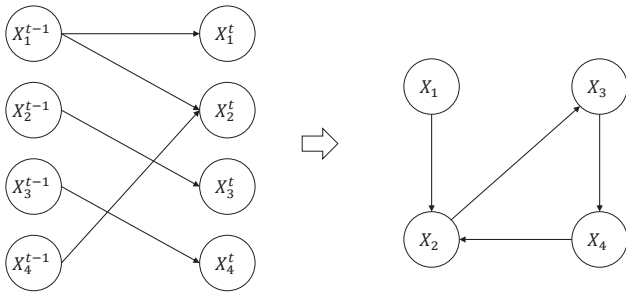


図 2 ダイナミックベイジアンネットワークによる  
遺伝子制御ネットワークの例

Fig. 2 Gene Regulatory Network by  
Dynamic Bayesian Network

も良いものを組み合わせ最適化問題として探索する。ベイジアンネットワークの評価関数として用いられるものにベイズ情報量規準 (Bayesian Information Criteria: BIC) があり、次式で表される。

$$\text{BIC} = \log L - \frac{d}{2} \log n \quad (1)$$

ここで  $L$  はモデルの尤度、 $d$  はモデルのパラメータ数、 $n$  はデータ数である。式 (1) を最大にするモデルが最適解である。

### 3.2 ダイナミックベイジアンネットワーク

ダイナミックベイジアンネットワークはベイジアンネットワークを拡張したもので、データの時系列を考慮し、時間を表す変数  $t$  を用いることで離散時間における確率過程を表現できるモデルである。

遺伝子制御ネットワークで用いられるダイナミックベイジアンネットワークは図 2 のような形になる。図の左はダイナミックベイジアンネットワークそのものであり、 $X_i^t$  は時刻  $t$  における遺伝子  $i$  の発現量を表している。右は同一遺伝子をまとめ簡略化したもので  $X_i$  は遺伝子  $i$  およびその発現量を表す。ただし自己ループは省略してある。遺伝子発現制御は時間的に遅れてはたらくため、1 次マルコフ性を仮定し時刻  $t-1$  から  $t$  の辺のみを考える [14]。時刻  $t$  での確率変数の集合を  $Z_t = \{X_k^t \mid k = 1, 2, \dots, p\}$  とすると、 $Z_t$  はそれぞれの遺伝子の時刻  $t$  における発現量を表す。

ダイナミックベイジアンネットワークにおいて各時刻の確率変数の個数を  $p$ 、時点数を  $n$  とすると、各変数の同時確率は

$$\Pr[Z_1, Z_2, \dots, Z_n] = \prod_{t=2}^n \Pr[Z_t \mid Z_{t-1}] \quad (2)$$

と表される。また、依存関係のない変数間は独立であるため  $\Pr[Z_t \mid Z_{t-1}]$  は

$$\Pr[Z_t \mid Z_{t-1}] = \prod_{i=1}^p \Pr[X_i^t \mid \pi(X_i^t)] \quad (3)$$

と分解できる。

ダイナミックベイジアンネットワークの推定にはベイジアンネットワークと同様に式 (1) などの情報基準量を評価に利用できる。

### 3.3 ネットワーク推定のアルゴリズム

#### 3.3.1 GHC によるアルゴリズム

ベイジアンネットワーク、ダイナミックベイジアンネットワーク推定に全探索を用いるのは実用的でない。そこで近似解法が存在し、代表的な近似アルゴリズムである GHC は以下ようになる [11], [12]。

- (1) 適当なネットワークを作成する
- (2) 各辺について、追加・削除・反転を行いネットワークを作成する
- (3) 作成したネットワークのスコアを計算する
- (4) 最もスコアを高くするネットワークに更新する
- (5) スコアが更新されなくなるまで 2-4 を繰り返し、その時点でのネットワークを出力する

GHC では局所最適解に陥る欠点があるが、初期解の選び方によって異なる局所解に到達する。このことを利用し、Restart 回数を指定し試行回数を増やすことで大域最適解により近い解を推定する手法がある。

#### 3.3.2 部分問題化による近似アルゴリズム

渡邊ら [15] はベイジアンネットワーク推定の部分問題化による近似解法を示した。3 遺伝子からなるネットワークを全探索で推定し、推定結果を結合するものである。本研究ではこの推定方法をダイナミックベイジアンネットワークに適用している。

### 3.4 モデルの利点と問題点

ベイジアンネットワークは確率的に解析を行うため、発現プロファイルのもつ誤差の多さやデータ数の少なさという性質に対応できる。しかし最適解を求める問題は NP 困難である。GHC では大幅な時間削減が可能だが、局所最適解に陥る可能性がある。また、ベイジアンネットワークは DAG であるためループ構造を表せないことや、同時刻で観測されたデータについて推定を行うため発現制御の遅延を表現できないことが問題として挙げられる。

ダイナミックベイジアンネットワークは時系列を考慮し、ベイジアンネットワークでは表せなかったループ構造や時間的に遅れた制御を表現できる。しかしベイジアンネットワークと同様に最適解を求めるのは難しい。そこで、ダイナミック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の推定において、全探索より時間計算量が小さく、GHC より精度のよいアルゴリズムを構築できれば、優れた手法であると考えられる。本研究では遺伝子制御ネットワーク推定において部分問題化による推定時間の短縮を図る。

## 4. 提案手法

### 4.1 提案手法の概要

提案手法ではダイナミック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推定を部分問題に分割することで計算量の軽減を図る。

ダイナミック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の推定を全探索で行う場合、推定対象の遺伝子数を少なくすれば計算量の増加を抑えることができる。提案手法では推定対象の遺伝子から異なる3遺伝子を取り出した部分問題を作成し、その部分問題に対して全探索による推定を行い組み合わせ爆発を抑制する。そして、その推定結果を結合することで全体の遺伝子制御ネットワークを得る。

提案手法は、部分問題化と解の結合の2段階から構成される。次節では提案手法全体と部分問題への分割、解の結合方法について説明する。

### 4.2 提案手法のアルゴリズム

提案手法では調べる辺を減らすため、以下の仮定をする。これらの仮定は遺伝子制御ネットワークが疎なネットワークであることに基づいている。

仮定1 自己ループは存在しない

仮定2 異なる2遺伝子が互いに直接制御する関係、すなわち  $X_i \leftrightarrow X_j$  は存在しない

全体の流れは以下ようになる。

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異なる3遺伝子を選び、全組み合わせを求める
  - (2) 遺伝子発現プロファイルを推定対象ごとに分割し、全組み合わせの部分問題を作成する
  - (3) 各部分問題に対して全探索によるダイナミックベイジアンネットワーク推定を行う
  - (4) 全体のネットワークの各辺に対し、部分問題の解のスコアを用いて辺の向きを決定する
- 1-3, 4の詳細を4.2.1項, 4.2.2項で説明する。

#### 4.2.1 部分問題への分割

部分問題への分割の流れは以下のとおりである。擬似コードを **Algorithm 1** に載せる。

- (1)  $n$  個の遺伝子から異なる3個の遺伝子を取り出す。
- (2) 取り出した遺伝子に変数を割り当て、考えられるネットワークすべてに式(1)によるスコアをつける
- (3) 各ネットワークのスコアと「辺が全く引かれていない解(図3の形)」の平均スコアの差をとり、それを新たなスコアとする
- (4) スコアで順位付けする
- (5) 3個の遺伝子の組み合わせすべてについて2-4を行う

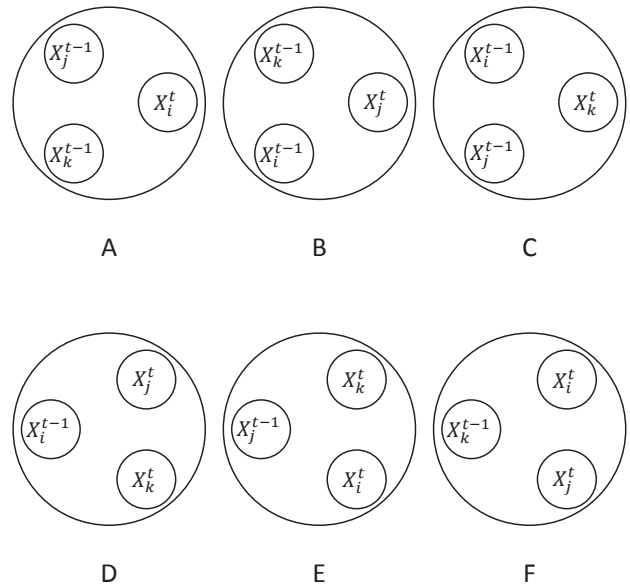


図3 3変数に対する  $t-1, t$  の割り当て方

Fig. 3 Assignment of  $t-1$  and  $t$  to 3 variables

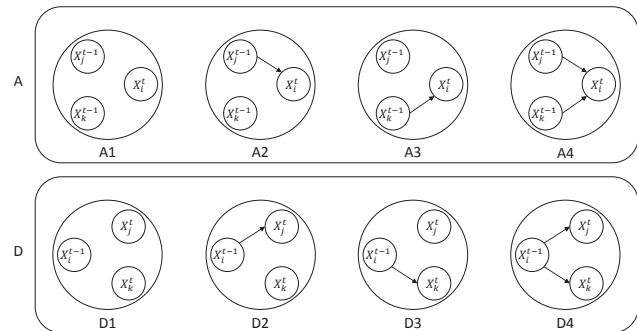


図4 考えられるネットワークの形

Fig. 4 All the Possible Networks of A and D

遺伝子の組み合わせは全部で  $nC_3$  通りになる。各遺伝子が時刻  $t-1$  と  $t$  のどちらであるか考慮し変数を割り当てるが、3変数がすべて  $t-1$ 、またはすべて  $t$  であるときは辺を引くことができないのでその場合は除く。このとき遺伝子を  $i, j, k$  とすると割り当て方は図3のように6通りある。さらに A, D に対して考えられるネットワークが図4のように4通り、B, C, E, F も同様に4通りあるので、3個の遺伝子からなるネットワークは合計24通りになる。

#### 4.2.2 部分問題の解の結合

部分問題で得られた解を結合し全体のネットワークを推定する方法は以下ようになる。擬似コードを **Algorithm 2** に示す。

- (1) 異なる2つの遺伝子  $\{X_i, X_j\}$  を選び、その頂点間の辺の向きを決定する
  - (a)  $\{X_i, X_j\}$  についての各部分問題に対し  $\{X_i^{t-1}, X_j^t\}$  または  $\{X_i^t, X_j^{t-1}\}$  を含むネットワークの中で1位のものを探す
  - (b)  $X_i^{t-1} \rightarrow X_j^t$  であれば  $S_{\rightarrow}$  にスコアを足し、



---

**Algorithm 1** 部分問題への分割

---

**Require:**  $V = \{1, 2, \dots, n\}$ :  $n$  個の遺伝子からなる集合, GEP:  $V$  の遺伝子発現プロファイル  
**Ensure:**  $Z$ : 部分問題の解を格納したリスト  
 $V$  の異なる 3 遺伝子からなる部分集合  $U$  の集合族  $\mathbf{V}$  を作成  
**for**  $U$  **in**  $\mathbf{V}$  **do**  
 $U$  で考えられる 24 個の DAG からなる集合  $\mathbf{D}_u$  を作成  
**for**  $D \in \mathbf{D}_u$  **do**  
 GEP からスコア  $S_D$  を計算  
**end for**  
 空グラフの平均スコア  $S_{empty}$  を計算  
**for**  $D \in \mathbf{D}_u$  **do**  
 $S_D \leftarrow S_D - S_{empty}$   
 $(D, S_D)$  をリスト  $L$  に追加  
**end for**  
 $L$  中の  $D$  を  $S_D$  で順位付け  
 $(L, U)$  を  $Z$  に追加  
**end for**

---



---

**Algorithm 2** 部分問題の解の結合

---

**Require:**  $Z$ : 部分問題の解を格納したリスト  
**Ensure:**  $G$ : 遺伝子  $V$  の遺伝子制御ネットワーク  
 $V$  の異なる 2 遺伝子からなる部分集合の集合  $\mathbf{E}$  を作成  
**for**  $\{i, j\} \in \mathbf{E}$  **do**  
 $Z$  中で  $\{i, j\} \in U$  である  $L$  の集合を  $\mathbf{L}_{ij}$  とする  
 $S_{\rightarrow} \leftarrow 0$ ;  $S_{\leftarrow} \leftarrow 0$ ;  $S_0 \leftarrow 0$   
 $k \leftarrow 1$   
**for**  $L \in \mathbf{L}_{ij}$  **do**  
**for**  $(D, S_D) \in L$  **do**  
**if**  $D$  が  $\{X_i^{t-1}, X_j^t\}$  または  $\{X_i^t, X_j^{t-1}\}$  を頂点にもつ中で順位が  $k$  **then**  
 辺の種類に従って  $S_D$  を  $S_{\rightarrow}$ ,  $S_{\leftarrow}$ ,  $S_0$  に加算  
**end if**  
**end for**  
**if** 最も高いスコアをもつ辺が一意に決まる **then**  
 $G$  に辺を追加  
**break**  
**else**  
 $k \leftarrow k + 1$   
**end if**  
**end for**  
**end for**

---

$X_i^t \leftarrow X_j^{t-1}$  ならば  $S_{\leftarrow}$ , 辺が引かれていないならば  $S_0$  にスコアを足す

- (c)  $S_{\rightarrow}$ ,  $S_{\leftarrow}$ ,  $S_0$  の内最大のものに従って辺を決定する
- 最大のものが一意に決まらない場合は 2 位, 3 位, ... と次点のネットワークのスコアを足していく

(2) すべての頂点の組み合わせについて 1 を行う

**4.2.3 時間計算量**

部分問題の総数は遺伝子数を  $n$  とすると  ${}_n C_3$  なので部分問題に対するネットワーク推定の時間計算量は  $O(n^3)$  に従う。また部分問題の解の結合時に調べる頂点の組み合わせが  ${}_n C_2$ , 各々の組み合わせに対して調べなければならない部分問題が  $n - 2$  個あるため, 結合時の時間計算量は  $O(n^3)$  に従う。したがって全体の時間計算量は  $O(n^3)$  と

表 2 TP, FP, TN, FN の定義

**Table 2** Definition of TP, FP, TN, and FN

		真のネットワークに含まれる	
		含まれる	含まれない
推定されたネットワークに	含まれる	TP	FP
	含まれない	FN	TN

なる。

**5. 実験と考察**

提案手法と GHC による推定結果を真のネットワークと比較することで, 提案手法による推定の精度を確認した。

**5.1 実験条件**

遺伝子数を 10 個から 70 個まで 10 個おきに変化させ, それぞれの遺伝子数について 20 種類のネットワークを生成, 各々のネットワークに対して 1 個のシミュレーションデータを生成した。20 種類のネットワークは辺の割合をランダムに 10%–50% から選び生成し, データの時点数は 75 とした。シミュレーションデータは式 (4) で表される線形重回帰モデルによって得られる。

$$X_i(t) = \sum_{j=1}^n \alpha_{i,j} X_j(t-1) + \epsilon_i(t) \quad (4)$$

まず, ノードの個数それぞれに対し提案手法と GHC による推定をし, 計算時間, 適合率, 再現率を比較した。次に, GHC において初期ネットワークを変更して推定し直す Restart 回数を指定することで提案手法より計算時間が大きくなるようにし, その適合率, 再現率を提案手法のものと比較した。ネットワークの辺について TP, FP, TN, FN を表 2 のように定義すると, 適合率と再現率は以下の式で表される。ここでは辺の向きも考慮している。例えば真のネットワークに  $X_i \rightarrow X_j$  があり, 推定されたネットワークに  $X_j \rightarrow X_i$  がある場合は FP, FN となる。

$$\text{適合率 (Precision)} = \frac{TP}{TP + FP}$$

$$\text{再現率 (Recall)} = \frac{TP}{TP + FN}$$

**5.2 実験結果**

実行時間, 適合率, 再現率を図 5–図 7 に示す。GHC の Restart 回数は表 3 のとおりである。各図は平均値をプロットしたものであり, 適合率, 再現率では標準偏差をエラーバーで表している。

Restart 指定のない GHC と提案手法について, 実行時間を比べると提案手法は GHC より時間の増え方が大きくなっている。適合率を比較するとどちらも似た挙動を示しており, 遺伝子数が少ない場合はわずかに GHC が勝っている。一方再現率については遺伝子数 10 の場合 GHC が

表 3 GHC の Restart 回数

Table 3 The Number of Times of GHC Restart							
遺伝子数	10	20	30	40	50	60	70
Restart 回数	1283	815	1027	1620	1800	2197	2977

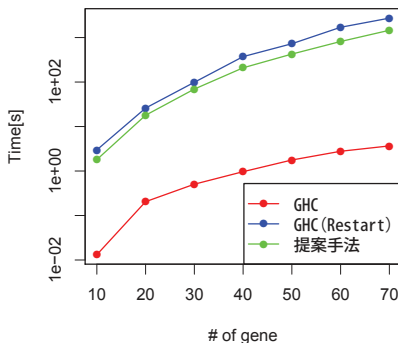


図 5 実行時間  
Fig. 5 Execution Time

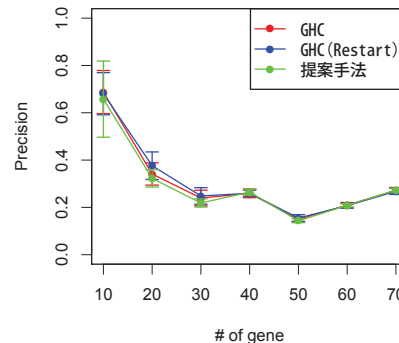


図 6 適合率  
Fig. 6 Precision

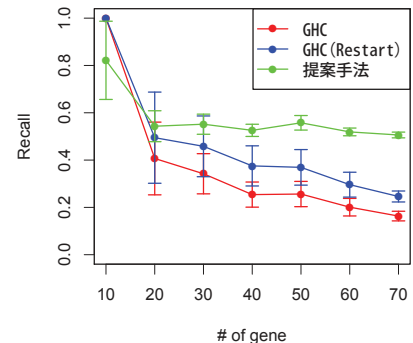


図 7 再現率  
Fig. 7 Recall

1.0 を示しており、真のネットワークに含まれる辺がすべて推定されていることがわかる。しかし遺伝子数 20 以上では GHC が提案手法を下回り、提案手法は 0.5 以上で安定している。

### 5.3 考察

#### 5.3.1 推定精度に関して

適合率について、遺伝子数が GHC の Restart なし・あり、提案手法のどれも似た挙動を示しており、提案手法は他の 2 つに比べ僅かに低い。

再現率について、GHC は遺伝子が 10 個の場合 1.0 となっており、すべての辺を推定できている。提案手法は遺伝子が少ない場合 GHC より低いが、遺伝子数が増えても 0.5 以上と安定している。

以上のことから推定精度に関しては、遺伝子が多い場合には GHC より提案手法のほうが真のネットワークに近い推定を行えていると言える。

#### 5.3.2 時間効率に関して

遺伝子数が増えると提案手法は GHC に比べ急激に推定時間が増えている、提案手法と GHC では GHC が非常に短い時間で推定ができる。Restart 指定により推定時間を大きくした結果、提案手法と適合率はほぼ変わらず再現率は提案手法より低くなった。このことより、提案手法は GHC より推定時間効率が良いと言える。

## 6. おわりに

本研究では、ダイナミック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、計算量の軽減を目的に部分問題化による近似解放を提案した。他の近似法である GHC と比較するためシミュレーションデータ

を用いて評価実験を行った。その結果、遺伝子数が多ければ提案手法の再現率が GHC を上回った。適合率に関しては GHC とほぼ等しくなる結果となった。計算時間は GHC より加速度的に増えるが、全探索の時間計算量が  $O(2^{n^2})$  であるのに対し提案手法は  $O(n^3)$  なので、全探索と比べると大幅に軽減できたといえる。

提案手法では適合率が低く本来存在しない辺を多く引いていることが分かる。存在しない辺のより正確な推定と計算時間の軽減のため、結合方法の見直しが必要である。

謝辞 本研究は一部 JSPS 科研費 15K00402 および 26280106 の助成を受けたものです。

### 参考文献

- [1] Csermelyemail, P., Agoston, V. and Pongor, S.: The efficiency of multi-target drugs: the network approach might help drug design, *Trends Pharmacological Sciences*, Vol. 26, pp. 178–182 (2005).
- [2] Zhang, Y., Pu, Y., Zhang, H., Su, Y., Zhang, L. and Zhou, J.: Using gene expression programming to infer gene regulatory networks from time-series data, *Computational Biology and Chemistry*, Vol. 47, pp. 198–206 (2013).
- [3] Somogyi, R. and Sniegowski, C. A.: Modeling the complexity of genetic networks: Understanding multigenic and pleiotropic regulation, *Complexity*, Vol. 6, pp. 45–63 (1996).
- [4] Akutsu, T., Miyano, S. and Kuhara, S.: Identification of genetic networks from a small number of gene expression patterns under the boolean network model, *Proc. Pacific Symposium on Biocomputing*, pp. 17–28 (1999).
- [5] Savageau, M. A.: *Biochemical Systems Analysis: A Study of Function and Design in Molecular Biology*, Addison Wesley (1976).
- [6] Friedman, N., Linial, M., Nachman, I. and Pe'er, D.: Using Bayesian Networks to Analyze Expression Data, *Journal of Computational Biology*, Vol. 7, pp. 601–620

- (2000).
- [7] Su, C., Andrew, A., Karagas, M. R. and Borsuk, M. E.: Using Bayesian networks to discover relations between genes, environment, and disease, *BioData Mining*, Vol. 6 (2013).
  - [8] Spirtes, P., Glymour, C., Scheines, R., Kauffman, S., Aimala, V. and Wimberly, F.: Constructing Bayesian Network Models of Gene Expression Networks from Microarray Data (2000).
  - [9] Murphy, K. P.: Dynamic Bayesian Networks: Representation, Inference and Learning (2002).
  - [10] 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*, pp. 104–113 (2003).
  - [11] Gamez, J. A., Mateo, J. L. and Puerta, J. M.: Learning Bayesian networks by hill climbing: efficient methods based on progressive restriction of the neighborhood, *Data Mining and Knowledge Discovery*, Vol. 22, pp. 106–148 (2011).
  - [12] Chickering, D. M. and Boutilier, C.: Optimal structure identification with greedy search, *Journal of Machine Learning Research*, Vol. 3, pp. 507–554 (2002).
  - [13] Brazhnik, P., de la Fuente, A. and Mendes, P.: Gene networks: how to put the function in genomics, *TRENDS in Biotechnology*, Vol. 20 (2002).
  - [14] Husmeier, D.: Sensitivity and specificity of inferring genetic regulatory interactions from microarray experiments with dynamic Bayesian networks, *Bioinformatics*, Vol. 19, pp. 2271–2282 (2003).
  - [15] 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 (2012).