

ノンパラメトリックベイジアンT過程アルゴリズムによる 時間的構造変化を考慮した遺伝子発現ネットワーク推定

宮下 弘樹^{1,a)} 中村拓磨^{1,b)} 井田安俊^{1,c)} 鈴木知彦¹ 松本 隆^{1,d)} 錦木崇史^{2,e)}

概要: 本稿では, ノンパラメトリック・ベイズ的枠組みの一つである T 過程を回帰モデルに用いることにより, 時間的に構造が変化するダイナミックベイジアンネットワークを推定することを目的とする. 特に, 実データとして *Drosophila melanogaster* の遺伝子発現データを用い, 生命現象に内在すると考えられる非線形性の依存関係および各成長段階における依存関係の変化を考慮した上での遺伝子発現ネットワーク推定を主題とする. 全体の推定アルゴリズムを構築するにあたり, リバーシブルジャンプマルコフ連鎖モンテカルロ法 (RJMCMC) の枠組みによる実装を採用している.

キーワード: バイオインフォマティクス, 機械学習, ノンパラメトリックベイジアンモデル, ベイジアンネットワーク

Nonparametric Bayesian T-Process Algorithm for Heterogeneous Gene Regulatory Network

HIROKI MIYASHITA^{1,a)} TAKUMA NAKAMURA^{1,b)} YASUTOSHI IDA^{1,c)}
TOMOHIKO SUZUKI¹ TAKASHI MATSUMOTO^{1,d)} TAKASHI KABURAGI^{2,e)}

Abstract: A nonparametric Bayesian model is employed to estimate gene regulatory networks of *Drosophila melanogaster*. A T-process-based algorithm is expected to capture nonlinear dynamics in the life phenomenon and reconstruct the gene regulatory interactions with consideration of the actual timing of morphogenic transitions. The whole algorithm is implemented by a reversible jump Markov Chain Monte Carlo.

Keywords: Bioinformatics, Machine Learning, Nonparametric Bayes, Bayesian Network

1. はじめに

時間的に構造が変化する遺伝子発現ネットワークの推定には, 以下の2つの対象が含まれていると見なせる.

- 遺伝子発現ネットワーク構造推定
- 構造変化が起きる時刻の推定

これら2つの対象を独立な問題として扱うことは難しく, 従来の時間依存性を考慮しない遺伝子発現ネットワーク推定問題へのアプローチと比較して, 更なる工夫を必要とする. この問題に対していくつかの先行研究が存在する. Robinson et al. は離散化した時系列データをベイジアン-ディリクレメトリックを用いて推定するアルゴリズムを提案した [1]. より成功しているアルゴリズムとして, Dondelinger et al. は連続値データに対してリバーシブルジャンプマルコフ連鎖モンテカルロ法を適用した [2]. この先行研究では, データの離散化は必要なく, また, 遺伝子間の依存関係には線形性を仮定している. 依存関係の線形性という仮定は, モデルの構築を比較的容易にする上, まずまずの成果を上げることに成功している. そのため先

¹ 早稲田大学 先進理工学研究所 電気・情報生命専攻
Waseda University, Tokyo, Japan

² 学習院大学 計算機センター
Gakushuin University, Tokyo, Japan

a) miyashita10@matsumoto.eb.waseda.ac.jp

b) nakamura10@matsumoto.eb.waseda.ac.jp

c) ida11@matsumoto.eb.waseda.ac.jp

d) takashi@matsumoto.elec.waseda.ac.jp

e) takashi.kaburagi@gakushuin.ac.jp

行研究では良く用いられる傾向がある．しかし，生命現象に内在することが指摘される非線形性のダイナミクスを的確に捉えることが出来るか疑問が残る．本稿では，ノンパラメトリックベイジアン手法を用いることで非線形性を考慮したより柔軟なモデルを提案する．なお，本稿における「ネットワーク」とはすべてベイジアンネットワークを意味することを先に断っておきたい．

2. 定式化

2.1 構造変化点の定義

Dondelinger et al. [2] の定式化に従い，構造変化点は観測値列の中から選ばれるものとする．1つの構造変化点により観測値列は2つのセグメントに区分され，それぞれのセグメントごとにネットワークが存在するとする．あるノード i と j の依存関係が， j から i に流れるリンクで表現される時，ノード i は j を親ノードに持つと見なす．ノード i を子ノードとする h 番目のセグメントにおけるベイジアンネットワークを $M_i^h = \{j_1, j_2, \dots\}$ と表現する．

2.2 事前分布

c_i をノード i の観測値列上の構造変化点数， p をノード数とする．更に， s_i^h を M_i^h におけるノード i の親ノード数とする．すなわち， $s_i^h = |M_i^h|$ が成立する．提案手法では， c_i 及び s_i^h の事前分布には切断ポアソン分布を用いる．すなわち， c_i の事前分布は平均値 λ と最大値 \bar{c} を決めることで

$$P(c_i|\lambda) \propto \frac{\lambda^{c_i}}{c_i!} \frac{1}{\sum_{m=0}^{\bar{c}} \frac{\lambda^m}{m!}} \quad (1)$$

と表現される． c_i が与えられた上での構造変化点位置 ζ_i の条件付き確率は，観測値列の長さ N を用いて

$$P(\zeta_i|\lambda) = 1 / \binom{N - c_i - 2}{c_i} \quad (2)$$

s_i^h の事前分布も平均値 Λ 及び最大値 \bar{s} を用いて同様に記述される． λ, Λ の事前分布には，ガンマ分布を採用している．

2.3 ノンパラメトリックベイジアン回帰モデル

Kikuchi et al. は遺伝子発現ネットワーク推定問題にノンパラメトリックベイジアン手法を用いている [3]．マウスの時計遺伝子間の依存関係が非線形性の性質を保持しているとの考察から，回帰モデルにガウス過程を用いた．

ガウス過程をベイジアンネットワーク推定に適用する手順は以下の通りである．ベイジアンネットワーク M_i^h におけるノード i の観測値数を T_i^h ，観測値列を \mathbf{y}_i^h と定義する． \mathbf{y}_i^h は親ノード $\{j_1, j_2, \dots, j_{s_i^h}\}$ の観測値 $\mathbf{x}_{i,j}^h = (x_j^1, \dots, x_j^{T_i^h})$ によって決まる．ガウス過程においては， $\mathbf{x}_{i,j}^h$ から共分散行列を以下の様に計算する．

$$K_i^h = \begin{pmatrix} k_{1,1} & \cdots & k_{1,T_i^h} \\ \vdots & \ddots & \vdots \\ k_{T_i^h,1} & \cdots & k_{T_i^h,T_i^h} \end{pmatrix} \quad (3)$$

ここで， $k_{p,q}$ は共分散関数であり，例として以下の式で表わされる．

$$k_{p,q} = \left(1 + \frac{\epsilon}{2\delta} \sum_{j \in M_i^h} |x_j^p - x_j^q|^2\right)^{-\delta} + \exp\left(\frac{-2s \sin^2\left(\frac{\sum_{j \in M_i^h} |x_j^p - x_j^q|}{l^2}\right)}{l^2}\right) \quad (4)$$

共分散関数の選択には任意性があり，共分散行列 (3) が半正定値性を満たすことを条件に，様々な候補が考えられる．式 (4) では，右辺第1項から RQ 共分散関数とサイン関数を用いた共分散関数を採用している．学習アルゴリズムの複雑化を避けるため，ハイパーパラメータ ϵ, δ, l の値は固定とする．ガウス過程は確率過程の1つであるため，共分散行列にはノイズの存在が仮定される．従って，単位行列 I とスケールに関するハイパーパラメータ α, β を用いて，共分散関数は

$$C_i^h = \alpha^{-1} K_i^h + \beta^{-1} I \quad (5)$$

と再定義される．ガウス過程における観測値 \mathbf{y}_i^h の尤度はガウス分布の形状を有しており，以下の様に記述される．

$$f(\mathbf{y}_i^h) = \frac{1}{|C_i^h|^{\frac{1}{2}}} \exp\left(-\frac{1}{2} \mathbf{y}_i^h C_i^h^{-1} \mathbf{y}_i^h\right) \quad (6)$$

\mathbf{y}_i^h と C_i^h のベクトルには1時刻のズレが生じることに注意されたい．これはグラフィカルモデルとしてダイナミックベイジアンネットワークを仮定しているからである．子ノード i の時刻 t における観測値は，時刻 $t-1$ における親ノード j の観測値を入力にすることで求められる．尤度の式 (6) を t 分布に書き直すことで，ガウス過程は T 過程に変更される．この場合の尤度は自由度 ν を用いて以下の様に表現される．

$$f(\mathbf{y}_i^h) = \frac{\Gamma\left(\frac{T_i^h}{2} + \frac{\nu}{2}\right) |C_i^h|^{-\frac{1}{2}}}{\Gamma\left(\frac{\nu}{2}\right) (\pi\nu)^{\frac{T_i^h}{2}}} \left(1 + \frac{\mathbf{y}_i^h C_i^h^{-1} \mathbf{y}_i^h}{\nu}\right)^{-\left(\frac{T_i^h}{2} + \frac{\nu}{2}\right)} \quad (7)$$

2.4 事後分布

2.2 及び 2.3 より事後分布はベイズの定理を用いて以下の様に導出される．

$$P(c_i, \zeta_i, \mathbf{s}_i, \mathbf{M}_i | \mathbf{y}_i) \propto \left(\sum_{m=0}^{\bar{s}} \frac{\Lambda^m}{m!}\right)^{-(c_i+1)} \lambda^{c_i} \frac{(N - 2c_i - 2)!}{(N - c_i - 2)!} \prod_{h=1}^{c_i+1} \Lambda^{s_i^h} \frac{(p - s_i^h)!}{p!} \times \frac{\Gamma\left(\frac{T_i^h}{2} + \frac{\nu}{2}\right) |C_i^h|^{-\frac{1}{2}}}{\Gamma\left(\frac{\nu}{2}\right) (\pi\nu)^{\frac{T_i^h}{2}}} \left(1 + \frac{\mathbf{y}_i^h C_i^h^{-1} \mathbf{y}_i^h}{\nu}\right)^{-\left(\frac{T_i^h}{2} + \frac{\nu}{2}\right)} \quad (8)$$

2.5 RJMCMC によるダイナミックベイジアンネットワーク推定

リバーシブルジャンプマルコフ連鎖モンテカルロ法を用いることで、構造変化点推定とダイナミックベイジアンネットワーク推定を同時に行うことができる [2]。このアルゴリズムでは、(i) 構造変化点追加操作、(ii) 構造変化点削除操作、(iii) 構造変化点位置移動操作、そして (iv) 各ダイナミックベイジアンネットワークの回帰操作の 4 つをイテレーションごとに繰り返す。それぞれの操作は現在の構造変化点数 c_i を基に、イテレーションごとに確率的に選択される。それぞれの選択確率は以下の通り定義する。

構造変化点追加操作の選択確率

$$b_{c_i} = \frac{1}{4} \min \left(1, \frac{P_c(c_i + 1)}{P_c(c_i)} \right) \quad (9)$$

構造変化点削除操作の選択確率

$$d_{c_i} = \frac{1}{4} \min \left(1, \frac{P_c(c_i - 1)}{P_c(c_i)} \right) \quad (10)$$

構造変化点位置移動操作の選択確率

$$p_{c_i} = \frac{1}{2} (b_{c_i} + d_{c_i}) \quad (11)$$

各ダイナミックベイジアンネットワークの回帰操作の選択確率

$$\pi_{c_i} = 1 - (b_{c_i} + d_{c_i} + p_{c_i}) \quad (12)$$

2.5.1 構造変化点追加操作

$\zeta_i = (\zeta_i^1, \dots, \zeta_i^{c_i})$ をノード i の観測値列上の構造変化点位置とする。 ζ 以外の観測値列から新しい構造変化点候補 ζ_i^{new} をランダムに 1 つ選ぶ。 ζ_i^+ を $\zeta_i^+ = \{\zeta_i^{new}\} \cup \zeta_i$ と定義する。すると ζ_i^+ の条件付き確率は

$$q(\zeta_i^+ | \zeta_i) = \frac{1}{N - 3 - c_i - x_i} \quad (13)$$

と書き表わすことができる。ここで x_i は構造変化点を置くことのできない観測点の数である。同様に、 ζ_i の条件付き確率は以下の様になる。

$$q(\zeta_i | \zeta_i^+) = \frac{1}{c_i + 1} \quad (14)$$

ζ_i^{new} は既存のセグメント h を 2 つの新しいセグメントに分割することになる。このとき、1 つのセグメントにおけるダイナミックベイジアンネットワークはそのまま継承されるが、もう一方のセグメントのダイナミックベイジアンネットワークは新規に生成される。構造変化点追加操作における受容確率 $\alpha_{c_i, c_{i+1}}$ は、式 (8), (9), (10), (13), (14) を用いて以下の様に計算される。

$$\alpha_{c_i, c_{i+1}}(\zeta_i, \zeta_i^+) = \min \left(1, r_{c_i, c_{i+1}}(\zeta_i, \zeta_i^+) = \frac{P(c_i+1, \zeta_i^+, s_i^+, M_i^+ | y_i)}{P(c_i, \zeta_i, s_i, M_i | y_i)} \frac{d_{c_i+1}}{b_{c_i}} \frac{q(\zeta_i | \zeta_i^+)}{q(\zeta_i^+ | \zeta_i)} \right) \quad (15)$$

アルゴリズムの流れを以下に示す。

1. ζ_i^{new} をサンプリング。 ζ_i^{new} が所属するセグメントは h_L, h_R に区分される。
2. 既存のセグメント h における s_i^h を用いて Λ を更新。
3. 更新された Λ を用いて新規ネットワークを生成。
4. 3 で生成されたネットワークを h_L, h_R のいずれかに所属される。
5. 4 で選択されなかったセグメントに既存のセグメント h のネットワークを所属される。
6. 式 (15) を用いて $\alpha_{c_i, c_{i+1}}(\zeta_i, \zeta_i^+)$ を計算。
7. $u \sim U_{[0,1]}$ をサンプリング。
8. $u \leq \alpha_{c_i, c_{i+1}}(\zeta_i, \zeta_i^+)$ なら ζ_i^{new} を受容。そうでなければ ζ_i^{new} を棄却、ネットワークを元に戻す。
9. 終了。

2.5.2 構造変化点削除操作

この操作では、既存の構造変化点の中から 1 点を選び、削除する。すなわち、 ζ_i からランダムに ζ_i^h をサンプリングする。

$$\zeta_i^h \sim U_{\{\zeta_i\}} \quad (16)$$

次に、 ζ_i^h に接触する 2 つのセグメントの内、ランダムに 1 つを選び、これを新しいセグメントとしダイナミックベイジアンネットワークを引き継ぐ。もう一方のセグメントは消去する。この操作の受容確率は、以下の式で表わされる。

$$\alpha_{c_i+1, c_i}(\zeta_i^+, \zeta_i) = \min(1, r_{c_i+1, c_i}(\zeta_i, \zeta_i^+)) \quad (17)$$

注意すべきは、式 (17) において $c_i + 1$ は、操作前の既存の構造変化点数を意味していることである。

2.5.3 構造変化点位置移動操作

この操作では、既存の構造変化点を 1 つ選び、その位置を移動させる。まず、 ζ_i からランダムに ζ_i^h をサンプリングする。

$$\zeta_i^h \sim U_{\{\zeta_i\}} \quad (18)$$

次に、 ζ_i^h の新しい位置を以下の範囲内でランダムに決める。

$$\left[\zeta_i^h - \frac{W}{2}, \zeta_i^h + \frac{W}{2} \right] \cap [\zeta_i^{h-1} + 2, \zeta_i^{h+1} - 2] \quad (19)$$

ただし、 W は ζ_i^h の行動範囲を決めるパラメータである。ここで、 ζ_i^{h*} を ζ_i^h の新しい位置、 ζ_i^* を ζ_i^{h*} を含む新しい構造変化点集合とする。 ζ_i^* 及び ζ_i の条件付き確率は以下の通り計算される。

$$q(\zeta_i^* | \zeta_i) = \frac{1}{c_i W - e} \quad (20)$$

$$q(\zeta_i | \zeta_i^*) = \frac{1}{c_i W - e^*} \quad (21)$$

ここで e, e^* は、隣り合う構造変化点との距離が $W/2$ 以下であるために、新しい構造変化点位置として選択すること

ができない観測点の数である．この操作における受容確率は以下の通りになる．

$$\alpha_{c_i}(\zeta_i, \zeta_i^*) = \min\left(1, r_{c_i, c_i+1}(\zeta_i, \zeta_i^*) = \frac{P(c_i, \zeta_i^*, s_i, M_i | y_i) q(\zeta_i | \zeta_i^*)}{P(c_i, \zeta_i, s_i, M_i | y_i) q(\zeta_i^* | \zeta_i)}\right) \quad (22)$$

2.5.4 各ダイナミックベイジアンネットワークの回帰操作
この操作では，さらに細かい2つの操作が分解される．対象のダイナミックベイジアンネットワーク M_i^h に新しく親ノードを追加する操作（親ノード追加操作）と既存の親ノードを削除する操作（親ノード削除操作）である．各イテレーションではこれらのどちらかのみ実行する．それぞれの操作が選択される確率 $b_{s_i^h}, d_{s_i^h}$ は以下の通りである．

$$b_{s_i^h} = \frac{1}{2} \min\left(1, \frac{P_s(s_i^h + 1)}{P_s(s_i^h)}\right) \quad (23)$$

$$d_{s_i^h} = \frac{1}{2} \min\left(1, \frac{P_s(s_i^h - 1)}{P_s(s_i^h)}\right) \quad (24)$$

$P(\cdot)$ には切断ポアソン分布を用いる．いずれかの操作によって $s_i^h \rightarrow s_i^{h*}, M_i^h \rightarrow M_i^{h*}$ がなされた場合の受容確率は以下の通りである．

$$\alpha_{s_i^h, s_i^{h*}}(M_i^h, M_i^{h*}) = \min\left(1, r_{s_i^h, s_i^{h*}}(M_i^h, M_i^{h*}) = \frac{P(c_i, \zeta_i, s_i^{h*}, M_i^{h*} | y_i^h)}{P(c_i, \zeta_i, s_i^h, M_i^h | y_i^h)}\right)$$

この操作は多少複雑であるため，アルゴリズムを以下にまとめる．これを全てのセグメントのダイナミックベイジアンネットワークに対して行う．

1. 式 (23), (24) を用いて $b_{s_i^h}, d_{s_i^h}$ を計算する．
2. $u \sim U_{[0,1]}$ をサンプリング．
3. $u \leq b_{s_i^h}$ ならば5へ．
そうでなければ4へ
4. $u > b_{s_i^h}$ かつ $u \leq b_{s_i^h} + d_{s_i^h}$ ならば6へ．
そうでなければ10へ
5. 新規に親ノードを1つ追加．
6. 既存の親ノードを1つ削除．
7. 式 (25) を用いて $\alpha_{s_i^h, s_i^{h*}}(M_i^h, M_i^{h*})$ を計算する．
8. $u \sim U_{[0,1]}$ をサンプリング．
9. $u \leq \alpha_{s_i^h, s_i^{h*}}(M_i^h, M_i^{h*})$ ならば s_i^{h*} 及び M_i^{h*} を受容．そうでなければ棄却，ネットワークを元に戻す．
10. 終了．

3. 実験

3.1 ノイズ付加人工データによる実験

提案手法を人工データに対して適用する．特に, T 過程及びガウス過程ベースの2つのアルゴリズム (TP-TVDBN, GP-TVDBN) を用意し, その性能差の検証を目的とする．

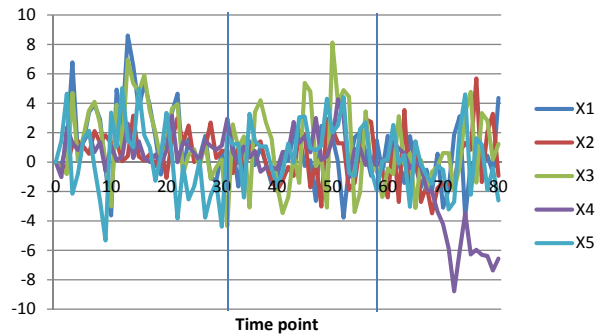


図 1 生成した人工データ．実際の構造変化点を青線で示している．
Fig. 1 Synthetic data. Blue lines indicate the actual positions of change points.

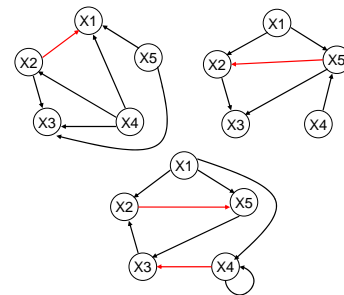


図 2 ノード間の依存関係を構造変化点によって区分されるダイナミックベイジアンネットワークごとに示した．黒矢印は線形，赤矢印は非線形の依存関係を示す．上段左：1 時刻目から 30 時刻目までのダイナミックベイジアンネットワーク．上段右：31 時刻目から 56 時刻目まで．下段：57 時刻目から 80 時刻目まで．

Fig. 2 Interactions between five synthetic data through 80 time points. Black arrows indicate linear interactions, whereas red arrows correspond to nonlinear interactions.

評価指標には, ROC (area under the curve) 及び BER (balanced error rate) を用いる．

本実験では, ノード数は5個, それぞれのノードが80時刻の観測点を持つ人工データを生成した (図 1) . それぞれのノードを子ノードとするダイナミックベイジアンネットワークは少なくとも1つの構造変化点を持ち, それは31時刻目および57時刻目の観測点に限られる．ノード間の依存関係は線形および非線形のいずれかを含む．本実験における非線形依存関係は, 自然対数, 平方根, 三角関数が該当する．また, 推定問題をより難解にするために, 各ノードの各観測値には分散0.2の正規ノイズを付加している．各ダイナミックベイジアンネットワークにおける依存関係を図 2 に示す．

本実験では以下の共分散関数を使用する．

$$k_{p,q} = \xi \left(1 + \frac{\epsilon}{2\delta} \sum_{j \in M_i^h} |x_j^p - x_j^q|^2\right)^{-\delta} + \eta \sum_{j \in M_i^h} x_j^p x_j^q \quad (26)$$

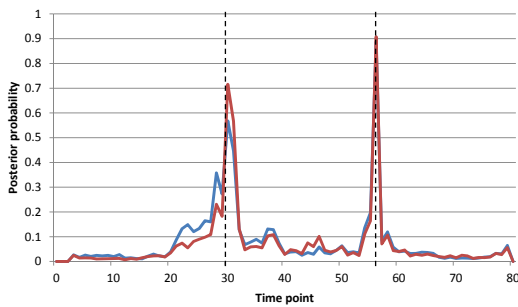


図 3 構造変化点位置推定結果．TP-TVDBN(赤線) 及び GP-TVDBN(青線) アルゴリズムにより得られた事後分布を示す．実際の構造変化点を点線で示している．

Fig. 3 Change points in the synthetic data. Posterior probabilities inferred by TP-TVDBN (red line) and GP-TVDBN (blue line). Dashed lines indicate the actual positions of change points.

表 1 それぞれのダイナミックベイジアンネットワークにおける AUC の比較．

Table 1 AUC values of each dynamic Bayesian network.

	NET1	NET2	NET3	allINT
TP-TVDBN	0.952	0.868	0.919	0.909
GP-TVDBN	0.889	0.886	0.882	0.900

表 2 それぞれのダイナミックベイジアンネットワークにおける BER の比較．

Table 2 BER values of each dynamic Bayesian network.

	NET1	NET2	NET3	allINT
TP-TVDBN	0.099	0.083	0.092	0.108
GP-TVDBN	0.183	0.083	0.063	0.132

共分散関数のスケールに関するハイパーパラメータ ξ, η はそれぞれ 1.0, 0.5 に固定する．

構造変化点位置推定の結果を図 3 に示す．それぞれの観測点において、少なくとも 1 つのノードが構造変化点を持つ確率を表わしている．31 時刻目の観測点において TP-TVDBN による推定結果が GP-TVDBN を上回っていることが見て取れる．ただ、いずれのアルゴリズムも正しい構造変化点を正確に捉えつつ、その他の観測点における誤検出を免れている．従って、ノンパラメトリックベイズモデルをベースとしたネットワーク構造変化点推定アルゴリズムの有効性が示されたと言える．

表 1 に各ダイナミックベイジアンネットワーク (NET1, 2, 3) 及び全時刻を通してすべての依存関係 (allINT) について計算した AUC の値を示す．同様に、表 2 には BER の値を示す．表 2 において、閾値は BER が最小になるように自動的に選択される．表 1 及び表 2 を見ると、全体的に TP-TVDBN の結果が GP-TVDBN のそれを上回る傾向が分かる．

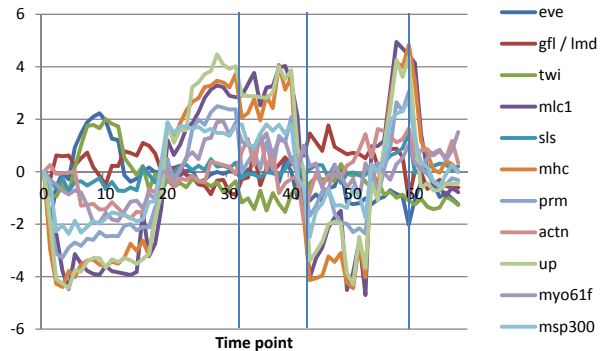


図 4 *Drosophila melanogaster* 遺伝子発現データ．青線は生物学的な形態変化が生じた時点を示している．

Fig. 4 *Drosophila melanogaster* gene expression data. Blue lines indicate the morphogenesis transitions.

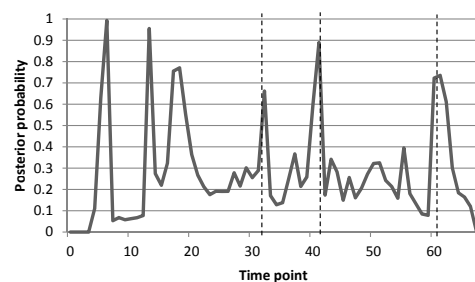
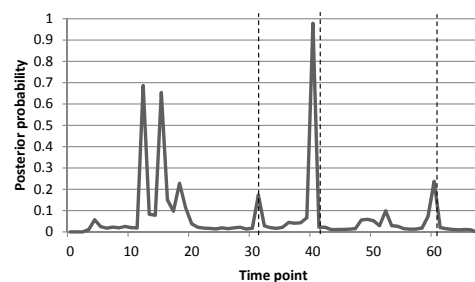


図 5 構造変化点位置推定結果．提案手法 (上段) 及び先行研究 [2](下段) のアルゴリズムにより得られた事後分布を示す．点線は生物学的な形態変化が生じた時点を示している．

Fig. 5 Change points during morphogenesis in *Drosophila melanogaster*. Top panel : posterior probability inferred by TP-TVDBN. Bottom panel: the result by HetDBN. Dashed lines indicate the morphogenesis transitions.

3.2 *Drosophila melanogaster* 遺伝子発現データによる実験

本項では、提案手法を *Drosophila melanogaster* 遺伝子発現データ [6] に対して適用する．このデータセットは 4028 個の遺伝子発現データが各 67 時刻観測されているが、本実験では特に筋肉の発達を司ると言われている 11 個の遺伝子を対象にする (図 4)．比較対象として、ダイナミックベイジアンネットワークの依存関係の回帰問題に線形モデルを用いたものを採用する (HetDBN)[2]．

この実験には式 (4) の共分散関数を用いる．構造変化点推定の結果を図 5 に示す．3 回の形態変化に伴う、遺伝子

表 3 全 67 時刻の事後分布の実現値の平均値. 及び中央値
Table 3 Average and median value of the posterior probability among 67 time points.

	平均値	中央値
TP-TVDBN	0.0713	0.0211
HetDBN	0.285	0.215

表 4 適合率, 再現率, F 値の比較 (閾値=表 3 の平均値).
Table 4 Precision, Recall and F-measure. Their thresholds were determined according to their averages.

	適合率	再現率	F 値
TP-TVDBN	0.231	1.0	0.375
HetDBN	0.136	1.0	0.240

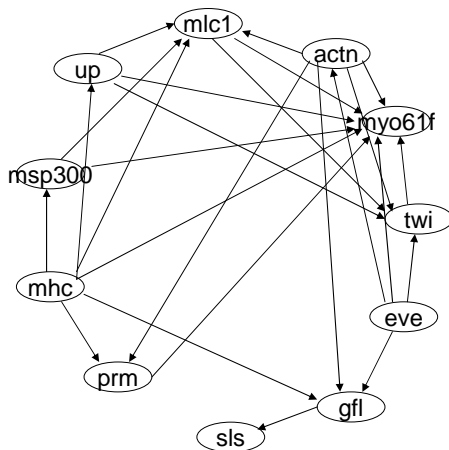


図 6 提案手法により推定された, 成長初期段階における遺伝子発現ネットワーク (閾値=0.25)

Fig. 6 Reconstructed network for the morphological phase of an embryo using TP-TVDBN method (thr = 0.25).

間の依存関係の変化が捉えられている事が分かる. 特に, 提案手法は先行研究と比較しても True-Negative(本来の構造変化点以外の誤検出を避けること)の性能が良いことが分かる. これを定量的に確認するため, F 値による比較を実施する. その際に用いる閾値として, 全 67 時刻における事後分布の実現値の平均を取る (表 3). 恣意的に決める場合に比べて, 公平性を欠かないと思われる. また, 実際の構造変化点から 2 点隣の観測値に 0.15 以上の事後分布の実現値が存在した場合, それは正しい構造変化点として認めるものとする.

表 4 を見ると, 適合率及び F 値の比較において先行研究 [2] を大きく上回っていることが分かる. 表 3 において, 事後分布の実現値の平均値に約 4 倍の差があることを考慮すれば, 提案アルゴリズムがいかに True-Negative の性能が良いかが定量的にはっきり分かる.

4. まとめ

時間的に構造が変化する遺伝子発現ネットワークの推定に, T 過程を適用することで, 遺伝子間の非線形性の依存

関係を捉え得る新しい手法を提案した. 人工データを用いた実験では, ネットワーク構造推定結果の評価を AUC 及び BER の 2 つの指標を用いて示し, 軒並み良好な結果を得ることができた. 同様に, *Drosophila melanogaster* 遺伝子発現データを用いた実験では, 先行研究 [2] の実験結果との比較を定量的に行い, 提案手法の優位性を示した. ただ, T 過程の問題点として, 共分散関数の設定の仕方によってアルゴリズムの性能が大きく左右されることが挙げられる. 従って今後は, データセットの性質ごとに適切な共分散関数を選択する方法の確立を目指したい.

参考文献

- [1] J. W. Robinson and A. J. Hartemink, Learning Non-Stationary Dynamic Bayesian networks.: *Journal of Machine Learning Research* (2010).
- [2] F. Dondelinger, S. Lebre, and D. Husmeier.: Non-homogeneous dynamic Bayesian networks with Bayesian regularization for inferring gene regulatory networks with gradually time-varying structure, *Machine Learning* (2012).
- [3] T. Kikuchi, T. Suzuki, Y. Nakada, T. Kaburagi, T. Matsumoto, T. Kimiwada and K. Wada.: Gene Regulatory Network Prediction using a Dynamic Gaussian Process: MCMC approach, *IEICE Technical Report Neurocomputing* (2011).
- [4] M. Grzegorzczak and D. Husmeier.: Bayesian regularization of non-homogeneous dynamic Bayesian networks by globally coupling interaction parameters, *Proceedings of the 15th International Conference on Artificial Intelligence and Statistics (AISTATS), US* (2012)
- [5] Yi. Jia and J. Huan.: Constructing non-stationary Dynamic Bayesian Networks with a flexible lag choosing mechanism, *BMC bioinformatics* (2010)
- [6] M. N. Arbeitman., E. E. M. Furlong, F. Imam, E. Johnson, B. H. Null, B. S. Baker, M. A. Krasnow, M. P. Scott, R. W. Davis and K. P. White.: Gene Expression During the Life Cycle of *Drosophila melanogaster*, *Science* (2002).
- [7] J.M. Wang, D. J. Fleet and A. Hertzmann.: Gaussian Process Dynamical Models, *Advances in Neural Information Processing Systems (NIPS)*(2005).
- [8] C. Andrieu, P. Djuri'c, and A. Doucet.: Joint Bayesian model selection and estimation of noisy sinusoids via reversible jump MCMC, *IEEE Trans. On Signal Processing*(1999)
- [9] S. Lébre.: Stochastic process analysis for Genomics and Dynamic Bayesian Networks inference, PhD Thesis, Université d' Evry-Val-d' Essonne, France (2007).
- [10] P. Green.: Reversible jump Markov chain Monte Carlo computation and Bayesian model determination, *Biometrika*(1995).
- [11] C. E. Rasmussen and C. K. I. Williams.: *Gaussian Process for Machine Learning*, Adaptive Computation and Machine Learning, MIT Press (2005).
- [12] D. J. C. MacKay, *Introduction to Gaussian Processes*, In C.M. Bishop, *Neural Networks and Machine Learning* (1998).