Vol.2011-BIO-25 No.6
2011/6/23

ダイナミック・ガウス過程遺伝子発現ネットワーク予測：MCMC 実装

菊地貴彰[†1]　鈴木知彦[†1]　中田洋平[†2]
鏑木崇史[†3]　松本隆[†4]
君和田友美[†5]　和田圭司[†6]

本研究では，ダイナミック・ガウス過程によるノンパラメトリック・ベイズ的枠組みにより遺伝子発現データから遺伝子発現ネットワーク構造を推定することを目的とする．提案手法は，ガウス過程により遺伝子発現生成メカニズムに内在しているといわれる非線形性を捉える一方，ノンパラメトリックな枠組みにより，柔軟なモデルを構築しようとするものである．提案手法をマルコフ連鎖モンテカルロ法により実装し，本研究グループが採取した時計遺伝子の発現データに適用する．

Gene Regulatory Network Prediction using a Dynamic Gaussian Process: MCMC approach

Takaaki Kikuchi,[†1] Tomohiko Suzuki,[†1]
Yohei Nakada,[†2] Takashi Kaburagi,[†3]
Takashi Matsumoto,[†4] Tomomi Kimiwada[†5]
and Keiji Wada[†6]

A dynamic Gaussian process-based algorithm is proposed for predicting a gene regulatory network structure of mouse clock genes, which regulate circadian rhythm. The proposed algorithm attempts to capture nonlinear dynamics associated with gene expression values with Bayesian non-parametrics. Implementation was performed by a Markov Chain Monte Carlo (MCMC).

1. はじめに

本研究で用いるデータを含め，一般に遺伝子発現データにはいくつかの注意すべき点が内在していると考えられる：

(i) マイクロ・アレイであっても，本研究で用いる定量 RT-PCR であっても，観測雑音が内在する．

(ii) 遺伝子間には非線形な依存関係が内在していると考えられている．

(iii) 一般に遺伝子発現データは静的なものではなく時系列データであることが多い．

(iv) データが豊富にあるとは限らない．例えば本研究では，遺伝子発現データは 19 点からなる時系列が 3 本のみである．

本研究の目的は，ダイナミック・ガウス過程[1)–4),12),13),17)] を用いたノンパラメトリック・ベイズ的枠組みにより，遺伝子発現データに内在していると考えられる非線形性とダイナミクス，そして不確定性を捉える新たな手法を提案し，その性能を人工データと実データを用いて評価することである．

この研究では遺伝子発現ネットワーク推定問題をグラフ構造の推定問題として捉え，ベイズ的枠組みからグラフ構造の事後期待値を，マルコフ連鎖モンテカルロにより近似計算する．この問題に限らないが，一般に推定問題では，

(a) 対象とするデータの構造を極力正確に捉えること，

(b) 仮定するモデル構造の柔軟性，そして

(c) 推定を実装する手法の性能，

が良いアルゴリズム構築のポイントと考えられる．この研究では，(a) に対して遺伝子発現

†1 早稲田大学大学院 先進理工学研究科
Waseda University, Graduate School of Advanced Science and Engineering
†2 青山学院大学 経営システム工学科
Aoyama Gakuin University, Department of Industrial and Systems Engineering
†3 学習院大学 計算機センター
Gakushuin University, Computer Center
†4 早稲田大学 理工学術院
Waseda University, Faculty of Science and Engineering
†5 宮城県立こども病院 脳神経外科
Miyagi Children's Hospital, Neurosurgery
†6 国立精神・神経医療研究センター 神経研究所
National Center of Neurology and Psychiatry

データを各時刻で独立な変数ではなく時系列として捉え，(b) に対してノンパラメトリック・ベイスとしてのガウス過程を，そして (c) に対してマルコフ連鎖モンテカルロ法を考えることでアプローチする．ノンパラメトリックな枠組みにより，極めて限られたデータからの推定問題に取り組むこともポイントの一つである．

1.1 先行研究

遺伝子発現ネットワーク推定では，これまでに様々な手法が考案されてきた．代表的なものでは，ブーリアンネットワークによるアプローチ[21)–23)]，離散値によるベイジアンネットワーク[15),16),23)] や S-system などの特別な微分方程式系によるアプローチ[18),20)] などが知られている．

本研究グループでは，定量 RT-PCR による発現量を時系列として捉え，マルコフ・ダイナミカルシステムによる定式化を行い，遺伝子ネットワーク推定を行ってきた[7)–10),25)]．25) では連続な発現量を離散化しており，離散化による情報量の欠落が考えられる．また，7)–10) では連続量として遺伝子発現データを捕らえたが，遺伝子間の依存関係に線形性を仮定していた．また，これらのモデルでは，その柔軟さにおいても改善の余地があると考えられる．少なくとも本研究グループが実験で得ることが可能なデータ数は極めて限られており，今回も例外ではなく，そのような厳しい条件の下でいかに推定問題に挑戦するかも重要課題の一つである．

1.2 定式化

時刻 t における遺伝子 1 から遺伝子 n の発現量を $x_{1:n,t} := (x_{1,t}, x_{2,t}, \cdots, x_{n,t})^\top$，それを全ての時刻について考えたものを $x := (x_{1:n,1}, x_{1:n,2}, \cdots, x_{1:n,T})$ と表すものとする．ベイズ的な枠組みでは，未知パラメータ θ を介して尤度関数 $P(x|\theta, G)$ が定義されることが多い．ここで，G は遺伝子間の依存関係を定義するグラフ構造である．この時，遺伝子発現データ x が与えられたもとでの遺伝子発現ネットワークのグラフ構造 G の事後確率は，

$$P(G|x) = \frac{P(x|G)P(G)}{\sum_{G' \in \mathcal{G}} P(x|G')P(G')} \tag{1}$$

と表せる．ただし，$\mathcal{G}$ は n 個の遺伝子が作る全てのグラフ構造の集合であり，$P(G)$ はグラフ構造の事前分布である．分子の第 1 項 $P(x|G)$ はパラメータ (あるいはハイパーパラメータ)θ を周辺化した周辺尤度であり，θ の事前分布 $P(\theta)$ を用いて次のように表される：

$$P(x|G) = \int P(x|\theta, G)P(\theta)d\theta \tag{2}$$

一般に，遺伝子発現ネットワーク推定を含むネットワーク推定問題では，式 (1) の計算の複雑さが問題となる．式 (1) の計算をするには，まず式 (2) を計算する必要がある．式 (1) の分母はグラフ構造 G の取り得る全ての組み合わせを考慮しなければならないため，組み合わせ的爆発が起きる．例えば，20 ノードの場合，有向非循環グラフであっても，その取り得る組み合わせが約 10^{72} 通りにもなる．[6)] 本研究の定式化では，自己ループも含めループ構造も許容するため，取り得る組み合わせは更に膨大になる．提案手法ではマルコフ連鎖モンテカルロを用いてグラフ構造事後確率からサンプルを採取し，その平均を推定値とする．

$$\sum_{G \in \mathcal{G}} GP(G|x) \approx \frac{1}{S} \sum_{k=1}^{S} G^{(k)} \tag{3}$$

ただし，$G^{(k)}$ は式 (1) からサンプリングした k 番目のグラフ構造のサンプルとする．

2. 提案手法

2.1 定式化

この研究では，グラフ G のノードが一つの遺伝子に対応し，ある遺伝子の他の遺伝子への依存性がグラフのリンク（枝）で表現されると仮定する．すなわち，ある遺伝子が他の遺伝子から影響を受ける場合リンクがあり，そうでない場合リンクはないと考える．実験により得られる発現量データは，対象とするグラフの各ノードから各時刻ごとに得られる値と考える．

時刻 t における遺伝子 m の発現量を $x_{m,t}$ とする．遺伝子 m は自分または他の遺伝子が発現してできたタンパク質を介して機能が制御される．本研究ではこのメカニズムを静的 (スタティック) なものでなく，動的 (ダイナミック) なものとして捉える．この違いが予測アルゴリズムの枠組みに与える影響は比較的大きい．後者で扱うことができないループ構造を，前者ででは考慮することが可能となるからである．自己ループも含まれる．

遺伝子発現ネットワーク予測問題には少なくとも 2 つの不確定性が内在していると考えられる．ひとつは生物実験のデータ観測に付随するもの，そしてもうひとつは遺伝子発現現象そのものに内在する不確定性である．

図 1 はグラフ上のダイナミカルシステムを模式的に表したものである．ここではノード数 $n = 7$ であり，方向を含めたノード依存性は矢印で示されている．グラフ構造が視覚的に見やすいのは，例えば図 2 であり，ここではノードが 2 次元空間に適当に配置してあり，図 1 を時間方向に“つぶして”ある．矢印の先頭で時刻が τ であれば，根元の部分では時刻は $\tau - 1$ を意味する．

Vol.2011-BIO-25 No.6
2011/6/23

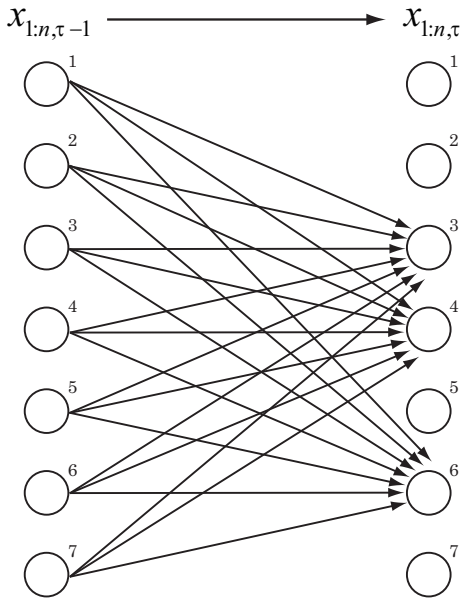


図 1 グラフ上の時系列データ模式図.
Fig. 1 A schematic picture of time series data on a graph.

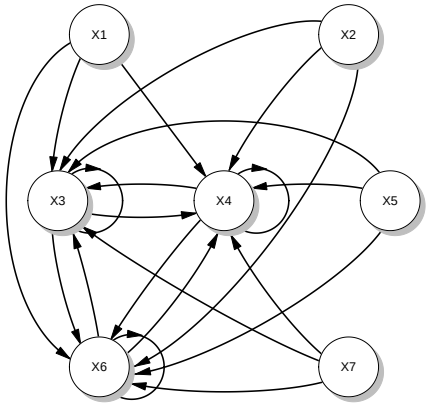


図 2 時間方向を圧縮したグラフ上の時系列データ模式図.
Fig. 2 A schematic picture of time series data on a graph with time collapsed.

2.2 尤　　度

$x_{1:n,t} = (x_{1,t}, \cdots, x_{n,t})^\top$ を n 個の遺伝子の時刻 t における発現量とし，時系列が採取される時間を $1, \cdots, T$ とする．遺伝子 m の発現量の軌跡 $x_{m,1:T}$ は，その親ノードの軌跡 $pa_{m,1:T}$ に依存すると仮定し，尤度を次のように定義する：

$$P(x_{m,1:T}|pa_{m,1:T}, \theta_m, G) = \prod_{m=1}^{n} P(x_{m,1:T}|pa_{m,1:T}, \theta_m, G) \tag{4}$$

ここに，θ_m は後述するハイパーパラメータである．

尤度関数右辺の時系列データを次のようなガウス過程により定義する．

$$P(x_{m,1:T}|pa_{m,1:T}, \theta_m, G) = \mathcal{N}(x_{m,2:T}; \mathbf{0}, \mathbf{C}_m) \tag{5}$$

ここに，$\mathcal{N}(x_{m,2:T}; \mathbf{0}, \mathbf{C}_m)$ は平均 $\mathbf{0}$，共分散行列 $\mathbf{C}_m$ のガウス分布を意味する．$\mathbf{C}_m$ は，時刻 1 から時刻 $T-1$ の全遺伝子の発現データによって後述のように構成される．ガウス過程の定式化で中心的役割を演じるのがこの共分散行列であり多くの方法が提案されている．この研究では，比較的柔軟性と近似能力が高いといわれている[17]infinite neural network を考える．[2] いま，$w_{inm,i}, w_{inbm}, w_{outm}, w_{outbm}$ を infinite neural network のパラメータ，ϵ_m を観測雑音とし，パラメータの事前分布と観測雑音の分布が次のように仮定できたとする：

$$P(w_{inm,i}) = \mathcal{N}(0, {\alpha_{inm,i}}^{-1}), \quad i = 1, 2, \cdots, n \tag{6}$$
$$P(w_{inbm}) = \mathcal{N}(0, {\alpha_{inbm}}^{-1}) \tag{7}$$
$$P(w_{outm}) = \mathcal{N}(0, {\alpha_{outm}}^{-1}) \tag{8}$$
$$P(w_{outbm}) = \mathcal{N}(0, {\alpha_{outbm}}^{-1}) \tag{9}$$
$$P(\epsilon_m) = \mathcal{N}(0, {\beta_m}^{-1}) \tag{10}$$

perceptron はパラメータ数が多い基底関数族のひとつである．特に中間素子数の増大とともにパラメータ数の爆発が起き，予測アルゴリズムが機能しなくなることもままある．ノンパラメトリック・ベイズによる定式化では，対象とするパラメータを自然共役事前分布で周辺化してしまうためパラメータが存在しない．

しかし，この研究で扱うガウス過程を含め，ノンパラメトリックとはいってもその上位階層のハイパーパラメータは存在する．上式では，$\alpha_{inm,1}, \cdots, \alpha_{inm,n}$，$\alpha_{inbm}$，$\alpha_{outm}$，$\alpha_{outbm}$，$\beta_m$ がそれに相当し，perceptron の入力層，入力層のバイアス，出力層，出力層のバイアス，遺伝子 m の発現データに伴う観測雑音の精度を表すハイパーパラメータである．後述するように，これらはデータから学習しなければならない．これらをまとめたものが式 (4) 及び式 (5) の θ_m である．これらをもとに，式 (5) に現れる共分散行列を次のように定義する：

$$\mathbf{C}_m = {\alpha_{outm}}^{-1}\mathbf{K}_m + {\alpha_{outbm}}^{-1}\mathbf{E} + {\beta_m}^{-1}\mathbf{I} \tag{11}$$

ここで，$\mathbf{K}_m$ は遺伝子 m に関する perceptron の共分散行列[1),2)]，$\mathbf{E}$ は perceptron のバイアスに由来しする要素が全て 1 の行列，$\mathbf{I}$ は単位行列を表すものとし，以下のように定義する．

$$[\mathbf{K}_m]_{t,t'} := \frac{2}{\pi}\arcsin\left(\frac{2\tilde{x}_{1:n,t}^\top \tilde{\Sigma}_m \tilde{x}_{1:n,t'}}{\sqrt{(1+2\tilde{x}_{1:n,t}^\top \tilde{\Sigma}_m \tilde{x}_{1:n,t})(1+2\tilde{x}_{1:n,t'}^\top \tilde{\Sigma}_m \tilde{x}_{1:n,t'})}}\right) \tag{12}$$
$$[\mathbf{E}]_{t,t'} = 1 \tag{13}$$
$$[\mathbf{I}]_{t,t'} = \delta_{t,t'} \tag{14}$$

ただし，$\tilde{x}_{1:n,t} = (1, x_{1,t}, \cdots, x_{n,t})^\top$ は，perceptron のバイアス成分と時刻 t における遺伝子 1 から遺伝子 n の発現データ，δ はクロネッカーのデルタである．

また，$\tilde{\Sigma}_m$ は，次を意味する：

$$\tilde{\Sigma}_m := \tilde{\Sigma}(G) = \mathrm{diag}({\alpha_{inbm}}^{-1}, {\alpha_{inm,1}}^{-1}, \cdots, {\alpha_{inm,n}}^{-1}) \quad (15)$$

もし遺伝子 m と遺伝子 i の間にリンクが存在しない場合は，対応するハイパーパラメータ $\alpha_{inm,i}$ は存在しないものと考える．すなわち，$\tilde{\Sigma}_m$ は遺伝子 m を子ノードとするグラフの部分構造を表現していると考えることができる．

2.3 ハイパーパラメータ

この研究では，グラフ構造の事前分布には一様分布と仮定する．それ以外も考慮することは可能である．[25)]

ハイパーパラメータの事前分布は，しばしば独立なガンマ分布を仮定することが多く，この研究でもそのように仮定する：

$$P(\theta) = \prod_{m=1}^{n} P(\theta_m) \quad (16)$$

$$P(\theta_m) = \prod_{i=1}^{n} P(\alpha_{inm,i}) P(\alpha_{inbm}) P(\alpha_{outm}) P(\alpha_{outbm}) P(\beta_m) \quad (17)$$

$$P(\alpha_{inm,i}) = \mathrm{Gamma}(\psi_{in}, \kappa_{in}), \quad i = 1, 2, \cdots, n \quad (18)$$

$$P(\alpha_{inbm}) = \mathrm{Gamma}(\psi_{inb}, \kappa_{inb}) \quad (19)$$

$$P(\alpha_{outm}) = \mathrm{Gamma}(\psi_{out}, \kappa_{out}) \quad (20)$$

$$P(\alpha_{outbm}) = \mathrm{Gamma}(\psi_{outb}, \kappa_{outb}) \quad (21)$$

$$P(\beta_m) = \mathrm{Gamma}(\psi_{noise}, \kappa_{noise}) \quad (22)$$

ただし，$\psi_.$ はガンマ分布の形状パラメータを，$\kappa_.$ は尺度パラメータを表す．

2.4 事 後 分 布

遺伝子発現データ $x = (x_{1:n,1}, x_{1:n,2}, \cdots, x_{1:n,T})$ 及びハイパーパラメータ θ が与えられたもとでのグラフ構造 G の事後確率はベイズ公式から以下のように表せる．

$$P(G|x,\theta) = \frac{P(x|G,\theta)P(G|\theta)}{\sum_{G' \in \mathcal{G}} P(x|G',\theta)P(G'|\theta)} \quad (23)$$

ガウス過程がノンパラメトリックと呼ばれる理由は，その第一階層のパラメータ，perceptron でいえば入力層—中間層，中間層—出力層のパラメータが周辺化されて存在しないことから来る．しかし，上述したようにその背後にある第二階層でのハイパーパラメータ θ は存在しており，この研究でもこれらを学習する．その事後分布は次のように与えられる：

$$P(\theta|x,G) = \frac{P(x|G,\theta)P(\theta|G)}{\int_{\Theta} P(x|\theta',\theta)P(\theta'|G)d\theta'} \quad (24)$$

ただし，Θ はハイパーパラメータの空間である．

グラフ構造の事後確率からのサンプリング及びハイパーパラメータの事後確率からのサンプリングはマルコフ連鎖モンテカルロ法 (メトロポリス法) により行う．

2.5 提案アルゴリズム

提案アルゴリズムを以下にまとめる．グラフ G のリンクは，0 または 1 をとる離散確率変数と考える．0 はリンクなない事を意味し，1 はリンクがあることを意味する．リンクの方向は時間の進む向きに決まる点に注意したい．

提案アルゴリズム

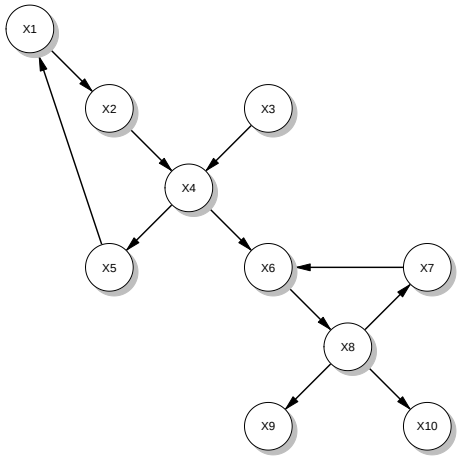
(1) ネットワークのグラフ構造の初期値として空のグラフ構造を，ハイパーパラメータの初期値はガンマ乱数から取得する．
(2) ハイパーパラメータ θ を所与として，グラフ構造 G をマルコフ連鎖モンテカルロ法によりサンプリングする．
(3) グラフ構造 G を所与として，ハイパーパラメータ θ をマルコフ連鎖モンテカルロ法によりサンプリングする．
(4) 2，3 を繰り返す．
(5) グラフ構造の事後分布の平均を計算し，それを推定結果 $\hat{G}$ とする．

3. 実　　験

3.1 人工データによる実験

提案手法を人工データに適用する．目標とするネットワーク（正解）は図 3 の左に示すネットワークであり，これは 19) を参考にした．この図は時間方向をつぶしてあり，図 2 に対応している．このネットワークでは第 8 番目と 10 番目に非線形項があり，各々平方根と 2 乗の非線形依存関係をもっている．図 4 に，このモデルから発生された典型的な時系列データを示す．それほど自明な問題ではないのではないかと考えられる．以下に示す実験条件で推定を 10 回行い，ROC 曲線を用いて考察を行う．[⋆1]

⋆1 数値実験，実データによる実験ともに ROC 曲線の算出には，http://oku.edu.mie-u.ac.jp/ okumura/stat/ROC.html で公開されているものを用いた．



- $x_{1,t} = x_{5,t-1} + \epsilon_{1,t}$
- $x_{2,t} = x_{1,t-1} + \epsilon_{2,t}$
- $x_{3,t} = -0.6x_{2,t-1} + x_{4,t-1} + \epsilon_{3,t}$
- $x_{4,t} = \epsilon_{4,t}$
- $x_{5,t} = -x_{3,t-1} + \epsilon_{5,t}$
- $x_{6,t} = -x_{3,t-1} + x_{8,t-1} + \epsilon_{6,t}$
- $x_{7,t} = x_{6,t-1} + \epsilon_{7,t}$
- $x_{8,t} = -4\sqrt{x_{7,t-1}} + 20 + \epsilon_{8,t}$
- $x_{9,t} = -2x_{7,t-1} + \epsilon_{9,t}$
- $x_{10,t} = 0.1{x_{7,t-1}}^2 - 50 + \epsilon_{10,t}$

図 3 目標とするグラフ構造（正解）.
Fig. 3 The target network.

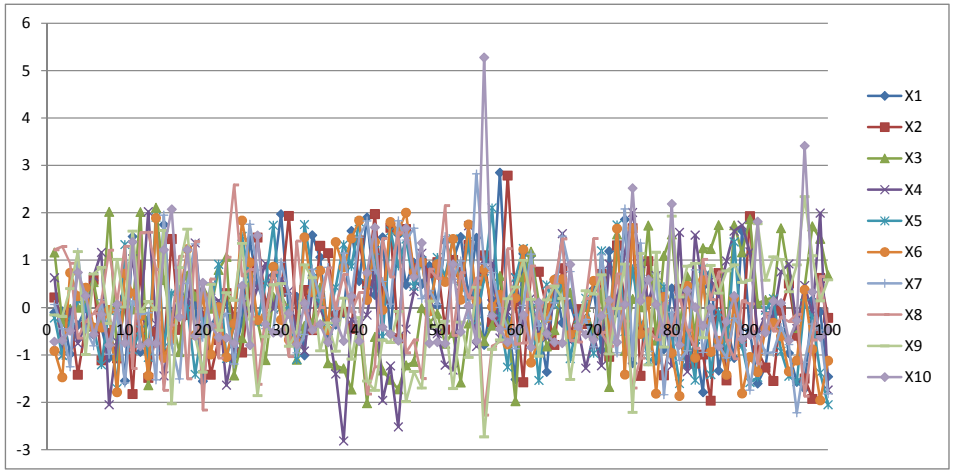


図 4 実験に用いる時系列人工データ.
Fig. 4 Synthetic time series data.

ただし，$\epsilon_{i,t}$，$(i = 1, \cdots, 10)$ は独立な正規雑音であり，

- $\epsilon_{1,t}, \epsilon_{2,t}, \epsilon_{3,t}, \epsilon_{6,t}, \epsilon_{9,t}, \epsilon_{10,t} \sim \mathcal{N}(0, 12^2)$
- $\epsilon_{4,t}, \epsilon_{5,t}, \epsilon_{7,t} \sim \mathcal{N}(0, 24^2)$
- $\epsilon_{8,t} \sim \mathcal{N}(0, 6^2)$

である.

実験条件

- 遺伝子数：10
- 時点数：100
- グラフ構造の事前分布：一様分布
- ハイパーパラメータの事前分布：ガンマ分布
- ハイパーハイパーパラメータ：ψ : 0.5，κ : 1.0
- サンプリング回数：50000
- 棄却数：45000

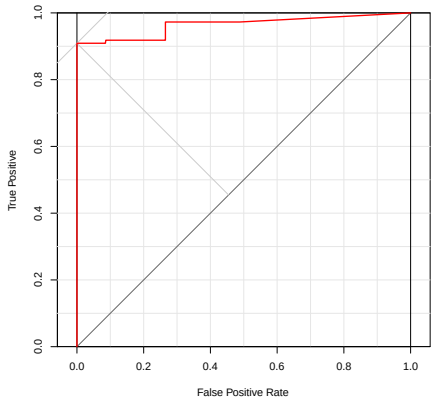


図 5 ROC 曲線.
Fig. 5 ROC curve.

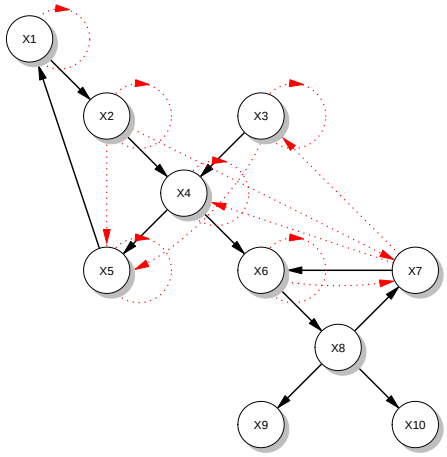


図 6 閾値を 0.5 として推定したグラフ．赤く示したリンクは誤検出 (false positive) のリンクである.
Fig. 6 An example of predicted networks(threshold:0.5). A red link represents a false positive link.

Vol.2011-BIO-25 No.6
2011/6/23

ROC 曲線を図 5 に示す．AUC 値は 0.964 であった．また，閾値を 0.5 として推定したグラフ構造を図 6 に示す．図中の黒いリンクは true positive であり，全て正しく推定されている．赤く示したリンクは誤検出 (false positive) のリンクである．この実験で false negative はなかった．

図 5 や AUC 値から，提案アルゴリズムはある程度有効に働いていると考えられる．しかしながら，リンクの誤検出も散見されるので，更なる精度の向上を目指して改良をしていく必要があると考えられる．

3.2 実データによる実験

次に，実際の遺伝子発現データを用いた実験を行う．ここでは，成体マウスの側脳室の脳室下帯 (subventricular zone; SVZ) 領域に存在する神経幹細胞・前駆細胞 (neural stem/progenitor cell; NSPC) から採取した時計遺伝子群の発現データを用いている．これは，本研究グループが定量 RT-PCR 法により採取したものであり，実験開始時とその 2 時間後に採取して以降は 4 時間おきに 70 時間採取し，計 19 点ある．同様の実験を 3 回繰り返し，データを 3 セット用意する．[24)]

時計遺伝子は概日リズムを遺伝子レベルで制御していると考えられており，様々な疾病との関連も指摘されている．そのため，時計遺伝子に関する研究は世界中でなされており，その相互作用が比較的わかっている．[5),11),14),26)] 概日リズムの形成に関わると考えられている遺伝子は，およそ 20 種類程存在するが，今回の実験ではそのうち特に中心的役割を果たしていると考えられている *Bmal1*，*Clock*，*Per1,2,3*，*Cry1,2* の 7 遺伝子に関するネットワーク推定を行うこととする．

幹細胞には，増殖と分化の 2 つの状態があるが，どちらの状態にあるかで発現レベルが変わるものがある．24) によると，今回の実験で扱う 7 遺伝子のうち，*Bmal1*，*Per2*，*Cry2* は分化時により強く発振する．その一方で，*Clock*，*Per1*，*Cry1* は増殖時と分化時で大きな差異は見られない．[⋆1] ここで行う実験では，分化時の遺伝子発現データを用いて推定を行う．

目標とするネットワーク（正解）を図 7 の左に示す．これは 5)，11)，14) を参考に作成したものである．以下の実験条件で予測実験を 10 回繰り返し，ROC 曲線を用いて考察を行う．

⋆1 *Per3* に関しては言及がないが，データから分化時により強く発振することを確認した．

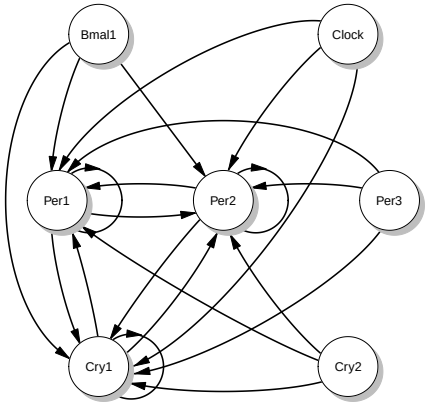


図 7 時計遺伝子の目標とするグラフ構造（正解）．5)，11)，14) から作成．
Fig. 7 The target network of clock genes reconstructed 5), 11), 14)

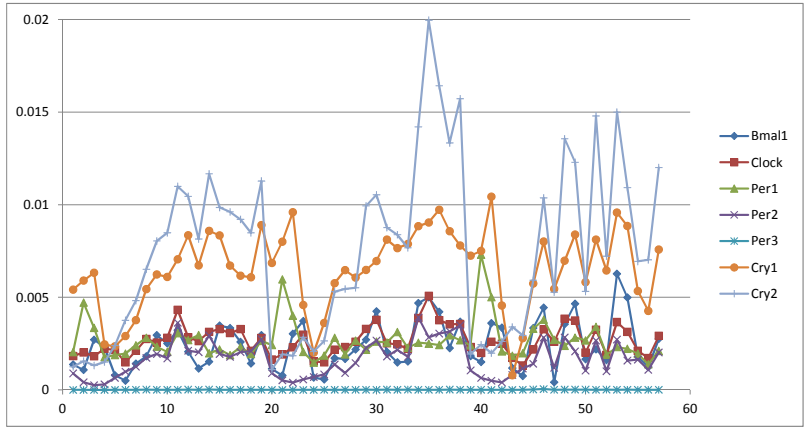


図 8 対象とする時計遺伝子の発現時系列データ．
Fig. 8 Expression time series data of the clock genes.

実験条件

- 遺伝子数：7
- 時点数：57
- グラフ構造の事前分布：一様分布
- ハイパーパラメータの事前分布：ガンマ分布
- ハイパーハイパーパラメータ：ψ : 0.5，κ : 1.0
- サンプリング回数：50000
- 棄却数：45000

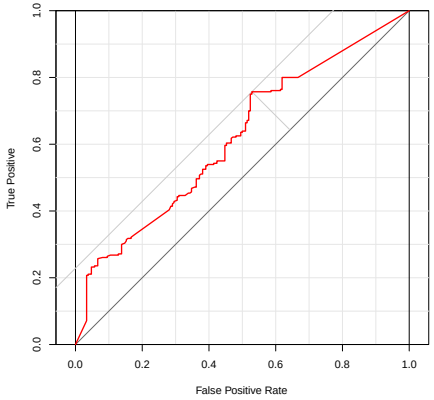


図 9 ROC 曲線.
Fig. 9 ROC curve.

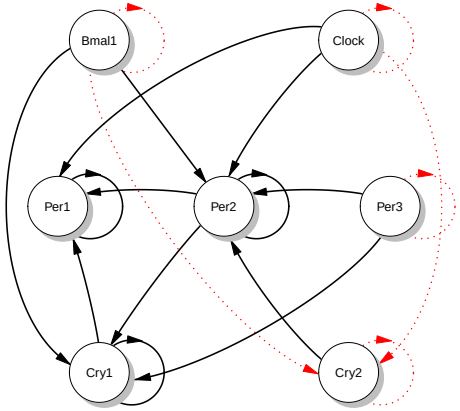


図 10 閾値を 0.5 として推定したグラフ．赤く示したリンクは誤検出 (false positive) のリンクである．
Fig. 10 An example of predicted networks associated with clock genes (threshold:0.5). A red link represents a false positive link.

ROC 曲線を図 9 に示す．AUC 値は 0.617 であった．また，閾値を 0.5 として推定したグラフ構造を図 10 に示す．図中に赤く示したリンクは誤検出 (false positive) のリンクである．

両図から提案手法により，ある程度の推定ができていると考えられる一方，改善の余地はいくつかあると考えられる．それらのいくつかを列挙する：

- 上述のように，人工データによる実験では 10 ノード，100 時点のデータセットを用意し，比較的良好な結果を得たが，実データによる実験では 7 遺伝子，57 時点のデータセットであり，データ数の少なさも原因の一つではないかと考えられる．
- 提案手法ではグラフ構造のサンプリング，ハイパーパラメータのサンプリングともマルコフ連鎖モンテカルロ法のメトロポリス法により行った．メトロポリス法はマルコフ連鎖モンテカルロ法でも最もベーシックなアルゴリズムであり，例えばこれをメトロポリス・ヘイスティングス法や拡張アンサンブル法などへの拡張を図ることで推定結果の改善が期待できるかもしれない．
- 時計遺伝子は朝，昼，夜で発現状態，すなわち制御関係が変わると考えられている[5),11),14),26)] が，今回の定式化ではそのようなグラフ構造の時間依存性は考慮されていない．制御関係の時間変化を捉える枠組みを推定アルゴリズムに組み込むことで，更に精度を向上させられる可能性がある．

4. 結　　論

遺伝子発現データの持つ非線形性，ダイナミクス，そして不確定性を捉える枠組みとして，ダイナミック・ガウス過程によるノンパラメトリック・ベイズ的なアプローチを提案した．マルコフ連鎖モンテカルロ法により実装し，人工データ及び本研究グループが採取した実際の時計遺伝子の発現データによる予測実験を行った．人工データに対しては比較的良好な推定結果を得た．実データに対してもある程度の推定できることを確認できた．しかし，改善点は多々あり，更に精度を上げる方策を検討していきたい．全体を通じて実験データが 19 点からなる 3 本の時系列のみであることはチャレンジングであり，相手にとって不足のない問題と考えている．

参　考　文　献

1) C. E. Rasmussen and C. K. I. Williams. ”Gaussian Process for Machine Learning”. The MIT Press, pp.1-31, 79-104, 2005.
2) C. K. I. Williams. ”Neural Computation”. The MIT Press, pp.1203-1216, 1998.
3) C. M. Bishop, et.al. ”Pattern Recognition and Machine Learning”. Springar New York, pp.291-323, 359-372, 523-546, 2006.

4) D. J. C. MacKay. "Information Theory, Inference and Learning Algorithms". Cambridge University Press, pp.534-548, 2003.
5) E. A Susaki, J. Stelling and H. R Ueda. "Challenges in synthetically designing mammalian circadian clocks". Current Opinion in Biotechnology 21:1?10. 2010
6) F. Harary and E. Palmer. "Graphical Ennumeration". Academic Press, 1973.
7) H. Miyachika, Y. Kitamura, T. Kimiwada, J. Maruyama, T. Kaburagi, T. Matsumoto and K. Wada. "Monte Carlo-Based Bayesian Prediction of Gene Regulatory Networks with Zipf Distribution: Mouse Nuclear Recepter Superfamily". 17th Annual International Conference on Intelligent Systems for Molecular Biology and 8th European Conference on Computional Biology, June 27-July 2, 2009, Stockholm, Sweden.
8) H. Miyachika, Y. Kitamura, T. Kimiwada, J. Maruyama, T. Kaburagi, T. Matsumoto and K. Wada. "A Gene Regulatory Networks Prediction Algorithm Using a Gaussian Bayesian Network Model with a Box-Cox Transformation". 18th Annual International Conference on Intelligent Systems for Molecular Biology, July 9-13, 2010, Boston, United States of America.
9) H. Miyachika, J. Maruyama, T. Kaburagi, Y. Nakada, T. Matsumoto, T. Kimiwada and K. Wada. "An MCMC Algorithm for Gene Regulatory Network Prediction with Bayesian Network". The 9th International Conference on Bioinformatics. September, 26-28. 2010, Tokyo, Japan.
10) H. Miyachika, J. Maruyama, T. Kaburagi, Y. Nakada, T. Matsumoto, T. Kimiwada and K. Wada. "A Gene Regulatory Network Prediction Algorithm with a Gaussian Bayesian Network". The 13th Slovenia-Japan seminar on nonlinear science and Waseda AICS sym- posium on nonlinear and nonequilibrium phenomena in complex systems. November 4-6, 2010, Tokyo, Japan.
11) H. Ukai and H. R Ueda. "System Biology of Mammalian Circadian Clocks". Annual Review of Physiology 72, pp.579-603, 2010.
12) J. M. Wang, D. J. Fleet and A. Hertzman. "Gaussian Process Dynamical Models". Proc. NIPS 2005. December, 2005. Canada Vancouver. pp.1441-1448.
13) J. M. Wang, D. J. Fleet and A. Hertzman. "Gaussian Process Dynamical Models for Human Motion". IEEE Transactions on Pattern Recognition and Machine Intelligence. February, 2008. pp.283-298.
14) M. Ukai-Tadenuma, T. Kasukawa and H. R. Ueda. "Proof-by-synthesis of the transcriptional logic of mammalian circadian clocks". Nature Cell Biology, 2008.
15) N. Friedman and M. Goldszmidt. "Learning Bayesian network with local structure". Proc. Twelfth Conference on Uncertainty in Artificial Intelligence, pp. 252?262, 1996.
16) N. Friedman, K. Murphy and S. Russell. "Learning the Structure of Dynamic Probabilistic Networks". Proc. Fourteenth Conference on Uncertainty in Artificial Intelligence (UAI 98), pp.139?147, 1998.
17) R. M. Neal. "Bayesian Learning for Neural Networks (Lecture Notes in Statistics, vol.118)". Springar New York, 1996.
18) S. Kikuchi, D. Tominaga, M. Arita, K. Takahashi and M. Tomita. "Dynamic modeling of genetic networks using genetic algorithm and S-system". Bioinformatics, vol. 19, No. 5, pp. 643-650, 2003.3.
19) S. Kim, S. Imoto and S. Miyano. "Dynamic Bayesian network and nonlinear regression for nonlinear modeling of gene networks from time series gene expression data". BioSystems 75 pp.57-65, 2004.
20) S. Kimura, K. Ide, A. Kashihara, M. Kano, M. Hatakeyama, R. Masui, N. Nakagawa, S. Yokoyama, S. Kuramitsu, and A. Konagaya. "Inference of S-system Models of Genetic Networks using a Cooperative Coevolutionary Algorithm". Bioinformatics 21(7), pp.1154-1163, 2005.
21) T. Akutsu, S. Kuhara, O. Maruyama, S. Miyano. "Identification of gene regulatory networks by strategic gene disruptions and gene over-expressions". Proc 9th ACM-SIAM SODA, pp.695-702, 1998.
22) T. Akutsu, S. Miyano and S. Kuhara. "Algorithms for identifying Boolean networks and related biological networks based on matrix multiplication and ?ngerprint function". Journal of Computational Biology. 7(3-4), pp.331-343, 2000.
23) K. Murphy and S. Mian. "Modelling Gene Expression Data using Dynamic Bayesian Networks". Technical report, Computer Science Division, University of California, Berkeley, 1999.
24) T. Kimiwada, M. Sakurai, H. Ohashi, S. Aoki, T. Tominaga and K. Wada. "Clock genes regulate neurogenic transcription factors, including NeuroD1, and the neuronal differentiation of adult neural stem/progenitor cells". Neurochem Int. 2009 May-Jun;54(5-6):277-85. Epub 2008 Dec 11.
25) Y. Kitamura, T. Kimiwada, J. Maruyama, T. Kaburagi, T. Matsumoto. "Monte Carlo based Mouse Nuclear Receptor Superfamily Gene Regulatory Network Prediction: Stochastic Dynamical System on Graph with Zipf Prior." IPSJ Transactions on Bioinformatics, 2009.
26) 岡村均, 深田吉孝編. 「時計遺伝子の分子生物学」. シュプリンガーフェアラーク東京, pp.41-101, 2004.