Vol.2010-BIO-21 No.17
2010/6/19

Zipf事前分布を用いたマウス核内受容体遺伝子発現制御ネットワーク予測アルゴリズムのモンテカルロ法による実装

北村悠輔[†1]　君和田友美[†2]　丸山隼[†1]
鏑木崇史[†3]　松本隆[†1]　和田圭司[†2]

本講演では、ベイジアンネットワークモデルを用いてマウス核内受容体の遺伝子発現ネットワーク構造を予測するアルゴリズムを説明する。これらの遺伝子は複数の疾患にかかわることが報告されているが、ネットワーク構造は未だに十分解明されていない。提案アルゴリズムではグラフを構成するリンクがZipf分布に従うものと考え、ベイズ的枠組みで定式化し、交換モンテカルロ法により実装した。数値実験で提案アルゴリズムの精度を検証したのち、生物学的観点からも考察を行う。

Monte Carlo-based Mouse Nuclear Receptor Superfamily Gene Regulatory Network Prediction: Stochastic Dynamical System on Graph with Zipf Prior

Yusuke Kitamura ,[†1] Tomomi Kimiwada ,[†2]
Jun Maruyama ,[†1] Takashi Kaburagi ,[†1]
Takashi Matsumoto [†1] and Keiji Wada[†2]

A Monte Carlo based algorithm is proposed to predict gene regulatory network structure of mouse nuclear receptor superfamily, about which little is known although those genes are believed to be related with several difficult diseases. The gene expression data is regarded as sample vector trajectories from a stochastic dynamical system on a graph. The problem is formulated within a Bayesian framework where the graph prior distribution is assumed to follow a Zipf distribution. Appropriateness of a graph is evaluated by the graph posterior mean. The algorithm is implemented with the Exchange Monte Carlo method. After validation against synthesized data, an attempt is made to use the algorithm for predicting network structure of the target, the mouse nuclear receptor superfamily. Several remarks are made on the feasibility of the predicted network from a biological viewpoint.

1. Introduction

1.1 Introduction

Genes in cells code for one or more proteins, many of which, in turn, regulate expression of genes through regulatory pathways. Deciphering such regulatory networks from experimental data is extremely important for understanding biological processes. Since the behaviors of genes are interrelated in complex ways rather than acting in isolation, the deciphering method needs to consider experimental data as a whole, which is far from trivial. This is currently one of the exciting challenges for machine learning. As such, there is much literature on gene regulatory network prediction[1)−18)].

This study regards gene expression time-series data as trajectories of stochastic dynamical systems on a graph G defined by genes. A gene is represented by a *node* in the graph. An arc between nodes is present if a gene influences another gene, and an arc is not present otherwise.

Let $x(t) := (x_1(t), \cdots, x_N(t))$ be the expression vector at time t, where N denotes the number of nodes (genes), and let $x := (x(0), \cdots, x(T))$. The Bayes rule gives

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

where $\mathcal{G}$ stands for the set of all possible graph configurations with a given number of nodes, and $P(G)$ is the prior distribution for the graph. The first factor of the numerator $P(x|G)$, the marginal likelihood for G, typically comes from

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

where θ is a parameter vector describing the likelihood function for the data, and $P(\theta)$ is the prior distribution for θ.

Three novel aspects of this study should be noted:

(i) Recent studies on topological structures of a variety of networks, including gene

†1 Waseda University
†2 National Center of Neurology and Psychiatry
†3 Gakushuin University

regulatory networks, revealed the presence of the Zipf law, that is, a power law.[16),17)] This can be thought of as a particular type of sparseness of the network topology. This study naturally incorporates such findings as prior information within a Bayesian framework and uses equation (1) for making predictions.

(ii) An important issue to be addressed in general prediction problems within a graphical setting, particularly gene regulatory network prediction, is the computational complexity of evaluating the performance criteria.[1)–15)] Most of the proposed algorithms, if not all, evaluate performance criteria to select optimum or good models for making predictions. Because of the nature of the problem, the performance criterion for gene regulatory network prediction is naturally a function of the graphical structure of the underlying model.

To be more specific, first note that both (1) and (2) are exact. Evaluation of (1) is performed in two steps. The first step is the marginalization (2). Generally, this is non-trivial; however, if one assumes a conjugate prior $P(\theta)$ for θ, then analytical marginalization in closed form is possible. Evaluation of (1) is harder than it looks. The computational cost of evaluating the denominator of (1) is often enormous. For instance, it is known[20)] that if the number of nodes of a graph is 40, the number of all possible directed non-cyclic graphs is in the order of 10^{276}. Our formulation presented in Section 2 does not exclude cyclic graphs. Therefore, exhaustive evaluation will be even more difficult. This study attempts to make predictions by computing the posterior mean instead of a single graph via a Monte Carlo method which avoids exhaustive evaluation while automatically searching for regions where probabilities are high.

More precisely, the proposed algorithm draws posterior samples of arcs from (1) via the Exchange Monte Carlo method without knowing the denominator and computes the Monte Carlo posterior mean for the graph structure prediction:

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

where $G^{(k)}$ stands for the k-th sample from the graph posterior (1).

After testing the prediction capabilities of the proposed algorithm, predictions are made for the target data.

A third novel aspect of this study is the target material:

(iii) Target experimental data used in this study is gene expression time-series data of nuclear receptors in proliferating neural progenitor cells (NPCs) derived from adult mouse brain, where little is known about the regulatory network structure. Those nuclear receptors are understood to be involved in several cancers, diabetes mellitus, hyperlipidemia, atherosclerosis, and immune system disorders, among others.

2. Algorithm

2.1 Formulation

This study regards gene expression time-series data as trajectories of stochastic dynamical systems on a graph G defined by genes. A gene is represented by a *node* in the graph. An arc with an arrow directed from node j to node i represents the fact that gene j influences gene i. No arc means no influence.

Recall that a graph G consists of a set of nodes $V := \{i\}_{i=1}^{N}$, and a set of arcs $B := \{b_{ij}\}_{i,j=1}^{N}$. We regard B, and hence G, as a set of random variables. An arc b_{ij} is regarded as a random variable with three discrete values: $b_{ij}\in\{-1, 0, +1\}$, where "0" means no influence, "+" means that gene j influences gene i , and "-" means that gene i influences gene j.

The expression value at time t of gene i is represented as the state variable $x_i(t)$. A gene i may influence other genes by influencing their expression via proteins, in addition to influencing gene i itself. Such an influencing mechanism is generally dynamic in the sense that the past expression values may influence the current expression values. Thus, $x_i(t)$ may depend on $x_j(t-\tau)$, for some j not equal to i, where $\tau > 0$, and also on the value $x_i(t-\tau)$, where $\tau > 0$, the past value of the expression of gene i itself. An *arc* between nodes is present if such an influencing mechanism exists, and an arc is not present otherwise. One of the important distinctions between the dynamic model assumed here and static models is that the underlying graph structure of the former can contain *loops*, whereas the latter does not allow loops. In particular, the dynamic model allows *self loops*. A dynamical system without self loops is a very restricted class of dynamics. Note that the exclusion of loops from static models is necessary for data consistency for statistical inference.

Vol.2010-BIO-21 No.17
2010/6/19

There could be at least two uncertainties associated with gene regulatory network prediction problems. One is the uncertainty incurred from measurement uncertainty in biological experiments. Another is the possible stochastic nature of the gene expression itself.[19] In order to take into account these uncertainties, transition of the states is assumed to be stochastic. This study assumes a first-order Markov process where the current expression value $x_i(t)$ depends on the values $x_j(t-1)$, as well as on $x_i(t-1)$, in a probabilistic manner, where time is discretized with an appropriate unit length, e.g. four hours. Generalizations are possible to multiple time delay cases.

The expression values, at least in this study, are discretized into finite discrete values so that nonlinearity is captured by state transition probabilities associated with stochastic dynamics. The prediction is formulated as a graph structure prediction problem within a Bayesian framework, where a score is computed using a graph posterior distribution.

Under such a graph structure, this study attempts to capture two other possible structures behind gene expression time-series data. First is the *dynamics*. We assume that $x_i(t)$ comes from a dynamical system, so that it could be influenced by the past expression values of other genes as well as its own. One of the important distinctions of this formulation from a static formulation, is that this formulation allows loops in the underlying graph, whereas a static formulation excludes loops. In particular, the current formulation considers self loops.

Second is *uncertainty* associated with the data. One of the uncertainties is measurement uncertainty of expression values, and another is the stochastic nature of the expression process itself[19].

In this study, we treat the target problem within a Bayesian framework, for which we need to define a likelihood function and prior distributions.

2.2 Likelihood: stochastic dynamical system on a graph

Recall that we regard each arc b_{ij} of a graph G as a ternary random variable $b_{ij} \in \{-1, 0, +1\}$, where "+" means that gene j influences gene i, "-" means that gene i influences gene j, and "0" means no influence. This study assumes the likelihood function of the form

$$P(x(0),\, x(1),\, ..., x(T)|G) := \prod_{t=1}^{T} P(x(t)|x(t-1),\, G) P(x(0)|G) \tag{4}$$

where

$$P(x(t)|x(t-1),\, G) \tag{5}$$

is the conditional probability of $x(t)$, the set of expression values at time t, given $x(t-1)$, those at the previous time t-1. Equation (4) amounts to considering the target gene expression time-series data as a realization of a first-order Markov process with transition probability (5) and initial state probability $P(x(0)|G)$. To be more precise, we need to consider the representation of $x_i(t)$. Although the gene expression values obtained from experiments are real valued, we will discretize them, at least in this paper, into K discrete values:

$x_i(t) \in \{1, 2, ..., K\}$.

If $x_i(t)$ is influenced by genes $x_j(t-1)$ for j belonging to some index set, the state variables associated with such indexes may be called the *parent* states of x_i. Define

$pa_i := \{\text{parent states of } x_i\}$,

let pa_i^j be the j-th configuration of pa_i, $j = 1, \cdots, q_i$ with q_i being the *number* of configurations[31] and set

$$P(x_i(t) = k|pa_i^j, \theta_i; G) := \theta_{i,j,k} \tag{6}$$

where

$$\theta_i := \{(\theta_{i,j,k})_{k=1}^{K}\}_{j=1}^{q_i}, \qquad \sum_{k=1}^{K} \theta_{i,j,k} = 1 \tag{7}$$

where G stands for the underlying graph structure.

Remarks

1. Note that this model is not a Hidden Markov Model because state is directly observable. Note also that, in contrast with static cases, cyclic graphs and self loops $\{b_{ii}\}_{i=1}^{N}$ are allowed. In fact, a dynamical system without self loops is highly restricted. Even without topological constraints, this class of dynamic models is sometimes called dynamic Bayesian networks[2],[3],[11].

2. The initial state distribution $P(x_i(0)|G)$ in this paper will be uniform over

Vol.2010-BIO-21 No.17
2010/6/19

$\{1, 2, ..., K\}$, where

$$P(x(0)|G) := \prod_{i=1}^{N} P(x_i(0)|G).$$

2.3 Prior Distributions

There are two unknown quantities of interest: $\theta := \{\theta_i\}_{i=1}^{N}$ and G. Due to the formulation of the problem, the prior distribution should be of the form

$$P(\theta, G) = P(\theta|G)P(G). \tag{8}$$

As a prior for a multinomial distribution with parameter $\theta_{i,j} := \{\theta_{i,j,k}\}_{k=l}^{K}$, a conjugate prior distribution, Dirichlet distribution

$$Dir(\theta_{i,j}; \alpha_{i,j}), \tag{9}$$

with hyperparameter $\alpha_{i,j} := \{\alpha_{i,j,k}\}_{k=l}^{K}$ is used. With this, one can analytically marginalize our likelihood (5). Because of the conjugacy, one sees that the analytical marginalization is possible:[31)]

$$P(x(t), \cdots, x(1)|G) = \prod_{i=1}^{N} \prod_{j=1}^{q_i} \frac{\Gamma(M_{ij})}{\Gamma(M_{ij} + s_{ij})} \prod_{k=1}^{K} \frac{\Gamma(\alpha_{ijk} + s_{ijk})}{\Gamma(\alpha_{ijk})} \tag{10}$$

where α_{ijk} is the hyperparameter associated with the Dirichlet distribution,

$$s_{ij} := \sum_{k=1}^{K} s_{ijk}, \qquad M_{ij} := \sum_{k=1}^{K} \alpha_{ijk},$$

and $\Gamma(\cdot)$ denotes a gamma function.

Note that one does not infer θ since it is integrated out in (10). Note also that (10) is computable via counting the number of cases that the state variables take.

This study assumes the Zipf prior distribution for G with hyperparameter γ:

$$P(G) = P(G|\gamma) = \prod_{i=1}^{N} P_{Zipf}(k_i^{(G)}; \gamma) = \prod_{i=1}^{N} \frac{(k_i^{(G)})^{-\gamma}}{\sum_{z=1}^{k_{\max}^{(G)}} z^{-\gamma}}, \tag{11}$$

where $k_i^{(G)}$ denotes the degree, or the number of arcs with node i, $i = 1, \cdots, N$, and $k_{\max}^{(G)}$ is the maximum number of arcs of a single node.

Generally, the Zipf law states that the number of elements with rank order k with respect to some ordering is proportional to $k^{-\gamma}$. This law has been reported in a variety of disciplines, including biological networks. See, for example,[21)]. Observe that the Zipf law is a form of sparseness of arcs in a graph. While many of the nodes have small degrees, there are nodes that have large degrees, although they are few. Fig. 2.3 (a) is a simple illustration of the Zipf law, and Fig. 2.3 (b) is a schematic diagram showing a "random" graph, where the degree distribution is centered around a particular value.

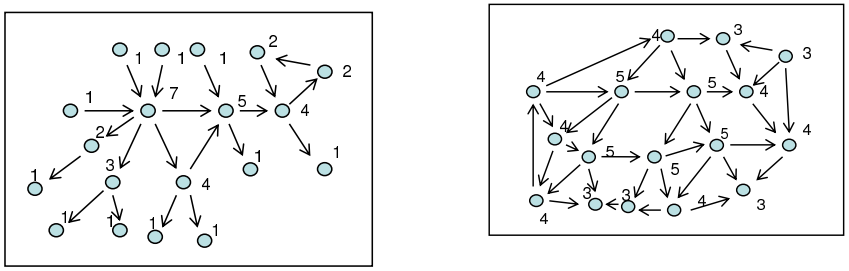

図 1 The schematic diagrams demonstrating network structures. The numerals show degrees at nodes. Self loops are omitted for simplicity. (a) is the Zipf law. While many of the nodes have small degrees, there are nodes that have large degrees, although they are few. (b) is for a random graph. The degree distribution is centered around a particular value.

2.4 Posterior Distributions

Given a time-series data set $(x(0), x(1), ..., x(T))$, $x(t) := (x_1(t), x_2(t), ..., x_N(t))$, the posterior distribution of G is given by

$$P(G|(x(0), x(1), .., x(T)), \gamma) = \frac{P((x(0), x(1), ..., x(T))|G)P(G|\gamma)}{\sum_{G' \in \mathcal{G}} P((x(0), x(1), ..., x(T))|G')P(G'|\gamma)} \tag{12}$$

where the first factor in the numerator is available from (10), and the second factor in the numerator is the Zipf prior alluded earlier.

2.5 Implementation

Our goal in this paper is to evaluate (12) for predicting plausible G. In order to achieve this goal, generally speaking, we need three different quantities: (a) the marginal like-

Vol.2010-BIO-21 No.17
2010/6/19

lihood for G, which is the denominator of (12); (b) the hyperparameter γ for the Zipf prior (11); and (c) the hyperparameters $\alpha_{i,j}$ for the Dirichlet prior (9). Each of them is generally nontrivial to set. Note that θ has been integrated out in (10). In this paper, we adopted a Markov Chain Monte Carlo (MCMC) method to evaluate (12).

2.6 MCMC Procedure for Graph Posterior Samples

Let

$$P(G|x,\gamma) \propto P(x|G)P(G|\gamma) \tag{13}$$

be the target distribution with $x := (x(0), ..., x(T))$, and consider the parameterized family of Q different distributions

$$P_q(G|x,\gamma) \propto (P(x|G)P(G|\gamma))^{1/T_q}, \quad q = 1, ..., Q, \quad 0 < T_1 < \cdots < T_Q = 1. \tag{14}$$

And set

$$P_q(x|G,\gamma) := (P(x|G,\gamma))^{1/T_q}, \quad q = 1, ..., Q. \tag{15}$$

Then, implementation follow that reported in[34].

Remarks

(1) Our prediction in this paper is performed using the Monte Carlo posterior mean of the graph:

$$\frac{1}{S}\sum_{n=\tau+1}^{\tau+S} G_n, \tag{16}$$

where G_n is the posterior sample obtained by the Monte Carlo method described above with a burn-in period τ.

(2) It should be noted that the graph posterior consists of

$$P(\{b_{ij}\}_{i,j=1}^{N}|(x(0),\, x(1),\, ..., x(T)), \gamma), \tag{17}$$

where $b_{ij} \in \{-1, 0, +1\}$ together with the set of nodes V.

(3) We do not need to know the normalization constants for $P(G|x,\gamma)$, $P_q(G|x,\gamma)$, $q = 1, ..., Q$, because β_q and w_q are the ratios.

(4) Hyperparameter for the Dirichlet distribution in principle, may be learned from the data. Because of the scarcity of the data for learning, we select not to perform learning, at least in the present paper.

3. Experiment 1: Synthesized Data

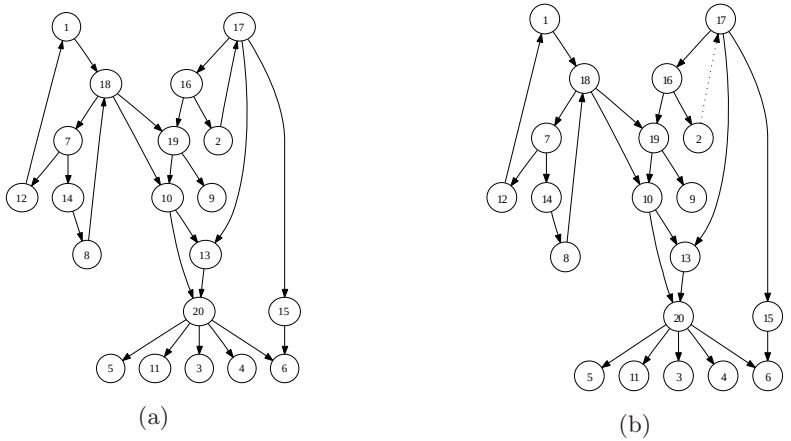

図 2 (a) Target graph for prediction experiment with synthesized data. Self loops are omitted for simplicity. (b) Predicted network with the proposed algorithm. There is one false negative indicated by a dotted arrow. The solid arrows indicate that the predictions are correct. Self loops are omitted for simplicity.

This section examines the prediction performance of the proposed algorithm against synthesized data. Fig. 2 (a) shows the target graph with the number of nodes N=20; many of the nodes have small degrees, and some have larger degrees. Time-series data is generated according to (4)-(6) and the number of discrete values of the state $x_i(t)$ is K=3. Since the purpose of the experiment with simulated data is to validate the performance of our proposed algorithm, the number of data points is set to emulate the real data experiment described in Section 4. Generally, the more the number of nodes the network has, the more difficult it is to estimate its topology. To have equal footing among the experiments, the ratio of the number of nodes N to the number of time points T is fixed, i.e. $\frac{T}{N} = const.$ The number of nodes N is restricted by the selection of genes one is interested in. There is more flexibility in the selection of time points T, as well as the number of experiments. In the biological experiment, $N = 35$, and *three*

Vol.2010-BIO-21 No.17
2010/6/19

sets of $T = 19$ data points are obtained. Therefore, for our experiment in the synthesized data, given the number of nodes $N = 20$, we set *three sets* of $T = \frac{19}{35} \times 20 \fallingdotseq 11$ as the number of data points. To explain the implementation of our prediction algorithm, first let $b_{ij}^{(k)}$ be an arc of a posterior sample graph defined by (17) and note that it has one of three possible values: 0, +1, and -1. We predict b_{ij} according to

$$b_{ij} = \begin{cases} 0, & \text{if } \frac{1}{S}\{\#b_{ij}^{(k)} | b_{ij}^{(k)} \neq 0, k = 1, ..., S\} < \xi \\ +1, & \text{if } \frac{1}{S}\{\#b_{ij}^{(k)} | b_{ij}^{(k)} \neq 0, k = 1, ..., S\} \geq \xi \\ & \text{and } \frac{1}{S}\{\#b_{ij}^{(k)} | b_{ij}^{(k)} = +1, k = 1, ..., S\} \\ & \qquad < \frac{1}{S}\{\#b_{ij}^{(k)} | b_{ij}^{(k)} = -1, k = 1, ..., S\} \\ -1, & \text{else.} \end{cases} \tag{18}$$

where S is the number of samples and ξ is a certain threshold.

Fig. 2 (b) shows our prediction with

- γ (the Zipf prior hyperparameter) = 2.2
- number of samples: 500000 (burn in 1000000)
- number of replicas: 5 with temperature 1/1.0, 1/0.9, 1/0.8, 1/0.7, 1/0.6

The predicted result has only one error: a false negative indicated by the dashed arrow. The remaining solid arrows indicate true positives.

4. Experiment 2: Real Data

Nuclear receptors represent a superfamily of ligand-dependent transcription factors that regulate essential biological processes, including development, reproduction, and metabolism. The classical endocrine receptors that mediate the actions, such as steroid hormones, thyroid hormones, vitamins A and D, and orphan receptors whose endogeneous ligands are unknown, are included in this superfamily.

In this study, nesural progenitor cells derived from adult mouse brains were applied as an experimental sample of particular interest and we aimed to investigate the possible network of nuclear receptors by using a bioinformatics approach. The number of members in the nuclear receptor superfamily has been shown to be 48 in the human genome database.[32] We selected the 48 genes reported in the human genome database out of 49 genes in the mouse genome database.[33]

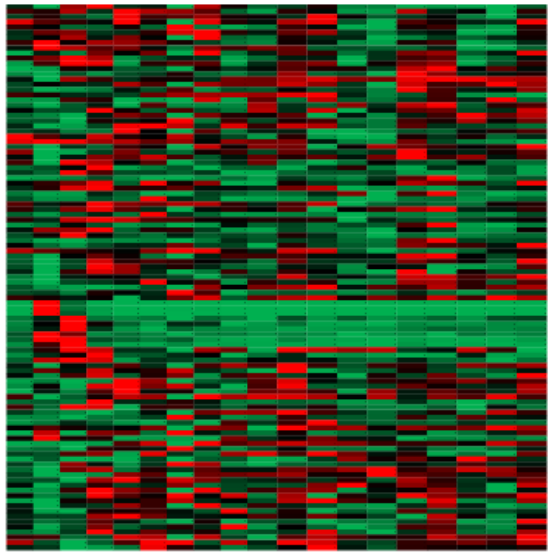

図 3 Target time series data for 72 hours with sampling period 4 hours consisting of 19 points. Of the 48 genes, 13 consistently exhibited very small amplitudes in their expression values so that they were discarded. The expression values are normalized between 0 and 1 at least in this study.

4.1 Prediction

There were 48 genes in our target superfamily. Three time-series data sets were measured, each covering a 72-hour period with a sampling frequency of 4 hours, so that each data set consisted of 19 points. Of the 48 genes, 13 consistently exhibited very small amplitudes in their expression values. Those 13 genes were discarded. Each expression time series was normalized between 0 and 1 and was then discretized into three equally spaced discrete values, at least in this study. The resulting time-series data is shown in Fig. 3. We will perform prediction experiments with other types of preprocessing in our future work.

The parameters were set as in our prediction described in Section 3.

In order to cope with the sparseness of the data, we drew posterior samples from 100 independent MCMC simulations, in which parameters were set as described in Section

Vol.2010-BIO-21 No.17
2010/6/19

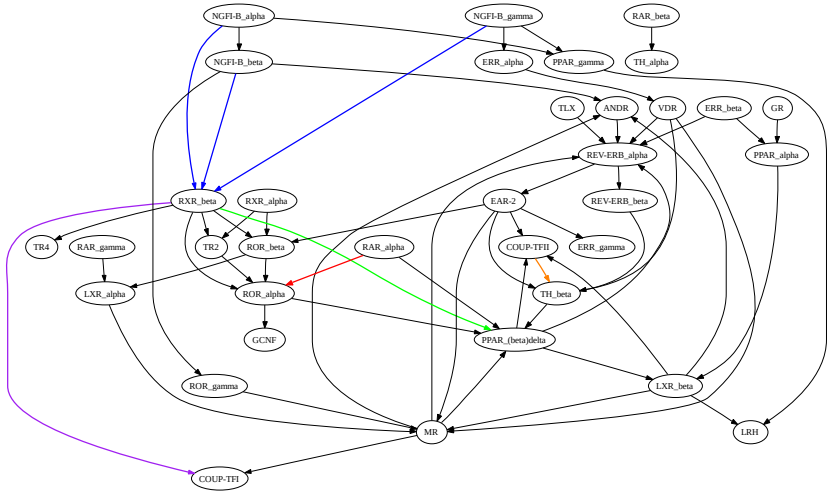

図 4 Prediction of the target network with the proposed algorithm.

3. We ranked ordered arcs b_{ij} by

$$\frac{1}{S^*}\{\#b_{ij}^{(k)}|b_{ij}^{(k)} \neq 0, k = 1, ..., S^*\}$$

where S^* denotes the total number of posterior samples from the 100 MCMC simulations, namely, $100S$. Those arcs ranked less than R were discarded and the remaining arcs were predicted in a manner similar to that defined in (18). Fig. 4 is our prediction with R=60.

4.2 Discussion

There is no ground truth for this network. There are, however, several well known biologically plausible scenarios in the target genes of this study. For instance, a previous biological study reported synergistic regulation of a cerebellum-specific gene by ROR_alpha and RAR[23]. The predicted network (Fig. 4) appears to indicate the functional relationship between ROR_alpha and RAR, as indicated by the red arrow. ROR_alpha has been shown to be a key regulator for the cerebellum development[24]. Another scenario is the RXR heterodimer arc between PPAR and LXR, as revealed in[25]. The green arrow in Fig. 4 appears to indicate this arc. It has also been verified that RXR with COUP-TF interactions modulates rethinoic acid signaling[26]. This relationship appears to be indicated by the purple arrow in Fig. 4. It is also known[27] that the orphan nuclear receptor NGFI-B (also called Nur77) can heterodimerize with RXR. This appears to be indicated by the blue arrow. Finally, COUP-TFII with TH beta heterodimer is also reported in[28]. The orange arrow of Fig. 4 appears to indicate this.

It seems reasonable to conclude that NGFI-B genes are located upstream of the gene regulation we predicted, since these genes are known to be immediate early genes. Besides NGFI-B genes, our prediction suggests that GR, TLX, and ERR_alpha genes are located upstream of some nuclear receptor genes, including PPAR_delta and LXR_beta. Both GR and TLX are suggested to play some role in neurogenesis[29],[30]. Our prediction may provide further clues to identify molecular cascades regulating the fate and function of neural progenitor cells.

For comparison, we tested the same data set using a greedy search method. The greedy method could not find any of the relationship biologically verified in the literature.

5. Conclusion

A Monte Carlo based algorithm was proposed for predicting a gene regulatory network from time-series expression data by formulating the problem as a graph prediction problem on which a stochastic dynamical system is defined. The prediction was performed by evaluating the graph posterior mean within a Bayesian framework, where the graph prior distribution is assumed to follow the Zipf prior, which is incorporated by taking into account several recent biological studies. The algorithm was first tested against synthesized data for which the ground truth is available, and the prediction was shown to be reasonable. Based on this, an attempt was made to predict the gene regulatory network structure of the mouse nuclear receptor superfamily. The prediction was discussed from a biological perspective.

The graphical representation of gene regulatory networks proposed in this study is from a bioinformatics view point. Even though further concreteness is not currently possible, the reported predictions can serve as guidelines for biological experiment de-

sign to verify the predicted interactions, which may lead to new biological findings.It is a challenging problem to infer γ, the hyperparameter for the Zipf prior, from the available data.

Acknowledgements

A part of this research is supported by Grant-in-Aid for Scientific Research(20810034).

参 考 文 献

1) N. Friedman and M. Goldszmidt. “Learning Bayesian networks with local structure,” Proc. Twelfth Conference on Uncertainty in Artificial Intelligence, pp. 252–262, 1996.
2) N. Friedman et al., “Learning the Structure of Dynamic Probabilistic Networks,” Proc. Fourteenth Conference on Uncertainty in Artificial Intelligence (UAI 98), pp.139–147, 1998.
3) K. Murphy and S. Mian, “Modelling Gene Expression Data using Dynamic Bayesian Networks,” Technical report, Computer Science Division, University of California, Berkeley, 1999.
4) N. Friedman et al., “Using Bayesian Networks to Analyze Expression Data”, J. Comput. Biol. 7, pp. 601–620, 2000.
5) P.T. Spellman et al., “Comprehensive Identification of Cell Cycle-regulated Genes of the Yeast Saccharomyces cerevisiae by Microarray Hybridization,” Molecular Biology of the Cell, vol. 9, pp. 3273-3297. 1998.
6) E.J. Moler et al., “Integrating naive Bayes models and external knowledge to examine copper and iron homeostasis in S. cerevisiae,” Physiol. Genomic vol. 4, pp. 127-135, 2000.
7) E.P. van Someren et al., “Linear Modeling of Genetic Networks from Experimental Data,” Proc. Eighth International Conference on Intelligent Systems for Molecular Biology, pp.355-366, 2000.
8) M. Eisen et al., “Cluster Analysis and Display of Genome-wide Expression Patterns,” Proc. Nat. Acad. Sci. USA, vol. 95(25), pp. 14863–14868, 1998.
9) X. Zhou et al., “A Bayesian Connectivity-based Approach to Constructing Probabilistic Gene Regulatory Networks,” Bioinformatics, vol. 20, no. 17, pp. 2918–2927, 2004.
10) M. Bittner et al., “Molecular Classification of Cutaneous Malignant Melanoma by Gene Expression Profiling,” Nature, vol. 406, pp. 536–540, 2000
11) S.Y. Kim et al., “Dynamic Bayesian Network and Nonparametric Regression for Nonlinear Modeling Gene Networks from Time Series Gene Expression Data,” Proc. 1st Computational Methods in Systems Biology, 104-113, 2003.
12) L. Mao and H. Resat, “Probabilistic Representation of Gene Regulatory Networks,” Bioinformatics, vol. 20, no. 14, pp. 2258-2269, 2004.
13) C.C. Guet et al., “Combinatorial Synthesis of Genetic Networks,” Science, vol. 296, pp. 1466-1470, 2002.
14) M. Middendorf et al., “Predicting Genetic Regulatory Response using Classification ,” Bioinformatics, vol. 20, no. 1, pp. 232–240, 2004.
15) M.J. Beal et al., “A Bayesian Approach to Reconstructing Genetic Regulatory Networks with Hidden Factors,” Bioinformatics, vol. 21, no. 3, pp. 349–356, 2005.
16) Jeong, H. et al., “The large-scale organization of metabolic networks”, Nature 407, 651–654, 2000.
17) K. Basso et al., “Reverse Engineering of Regulatory Networks in Human B cells,” Nature Genetics, vol. 37, pp.382 – 390, 2005.
18) J. Wildenhain and E.J. Crampin, “Reconstructing Gene Regulatory Networks: From Random to Scale-free Connectivity,” IEE Proc. Systems Biology, vol. 153, pp. 247-256, 2006.
19) K. Chen et al., “A stochastic differential equation model for quantifying transcriptional regulatory network in *Saccharomyces cerevisiae*” Bioinformatics, vol. 21(12), pp. 2883-2890, 2005.
20) http://www.research.att.com/~njas/sequences/table?a=3024&fmt=5
21) S.H Yook et al., “Functional and Topological Characterization of Protein Interaction Networks,” Proteomics, vol. 4, pp. 928–942, 2004.
22) A. Gelman et al., Bayesian Data Analysis, second edition, Chapman and Hall, 2004.
23) T. Matsui, “Transcriptional Regulation of a Purkinje Cell Specific Gene through a Functional Interaction between ROR Alpha and RAR,” Genes Cells, vol. 2, pp. 263–72, 1997.
24) A.M. Jetten et al., “The ROR nuclear orphan receptor subfamily: critical regulators of multiple biological processes,” Prog Nucleic Acid Res Mol Biol. vol. 69, pp. 205–47, 2001.
25) D.J. Mangelsdorf and R. M. Evans, “The RXR Heterodimers and Orphan Receptors”, Cell, vol. 83, pp. 841–50, 1995.
26) S.A. Kliewer et al., “Retinoid X receptor-COUP-TF Interactions Modulate Retinoic Acid Signaling,” Pro. Nat. Acad. Scie. USA, vol. 89, pp. 1448–1452, 1992.
27) Y. Katagiri et al., ”Modelling of Retinoid Signalling Through NGF-induced Nuclear Receptor Export of NGFI-B”, Nature Cell Biology, vol. 2. July 2000, pp. 435-440.
28) Berrodin TJ et al., ”Heterodimerization among Thyroid Hormone Receptor, Retinoic Acid Receptor, Retinoid X Receptor, Chicken Ovalbumin Upstream Promoter Transcription Factor, and an Endogenous Liver Protein”, Mol Endocrinol 6:1468–1478, 1992.
29) Y. Shi G et al., “Neural Stem Cell Self-renewal,” Crit. Rev. Oncol. Hematol., 2007 Jul 20, Epub ahead of print, 2007.
30) J.B. Kim et al., “Dexamethasone Inhibits Proliferation of Adult Hippocampal Neurogenesis in vivo and in vitro,” Brain Res. vol. 1027, pp. 1-10., 2004.
31) D. Heckerman, “Bayesian Networks for Data Mining ”, Data Mining and Knowledge Discovery, vol. 1, pp. 79-119, 1997.
32) Nuclear Receptors Nomenclature Committee, “A unified nomenclature system for the nuclear receptor superfamily”, Cell 97(2), pp. 161-3, 1999.
33) Zhang, Z. et al., “Genomic analysis of the nuclear receptor family: new insights into structure, regulation, and evolution from the rat genome”, Genome Res. 14(4), pp. 580-90, 2004.
34) Kitamura, Y. et al., “Monte Carlo based Mouse Nuclear Receptor Superfamily Gene Regulatory Network Prediction: Stochastic Dynamical System on Graph with Zipf Prior”, IPSJ Trans. BIO, Vol.3, pp.24-39, 2010.