Dynamic Bayesian Network and Nonparametric Regression for Nonlinear Modeling of Gene Networks from Time Series Gene Expression Data

SunYong Kim, Seiya Imoto, and Satoru Miyano

Human Genome Center, Institute of Medical Science, University of Tokyo
4-6-1 Shirokanedai, Minato-ku, Tokyo, 108-8639, Japan
{sunk, imoto, miyano}@ims.u-tokyo.ac.jp

Abstract. We propose a dynamic Bayesian network and nonparametric regression model for constructing a gene network from time series microarray gene expression data. The proposed method can overcome a shortcoming of the Bayesian network model in the sense of the construction of cyclic regulations. The proposed method can analyze the microarray data as continuous data and can capture even nonlinear relations among genes. It can be expected that this model will give a deeper insight into the complicated biological systems. We also derive a new criterion for evaluating an estimated network from Bayes approach. We demonstrate the effectiveness of our method by analyzing *Saccharomyces cerevisiae* gene expression data.

1 Introduction

The development of microarray technology provides us a huge amount of gene expression data and a new perspective of the analysis of whole genome mechanism. The estimation of a gene network from cDNA microarray gene expression data becomes one of the important topics in the field of bioinformatics and can be viewed as the first step of systems biology.

Using the Bayesian network model (Friedman *et al.* [13]; Imoto *et al.* [14, 15]; Pe'er *et al.* [19]) for estimating a gene network from cDNA microarray gene expression data has received considerable attention and many successful investigations have been reported. However, a shortcoming of the Bayesian network model is that this model cannot construct cyclic networks, while a real gene regulation mechanism has cyclic regulations. Recently, the dynamic Bayesian network model (Bilmes *et al.* [3]; Friedman *et al.* [12]; Murphy and Mian [18]; Someren *et al.* [21]) has been proposed for constructing a gene network with cyclic regulations. The dynamic Bayesian network is based on time series data, and usually the data has to be discretized into the several classes. Therefore, the resulting network of the dynamic Bayesian network model depends strongly on the thresholds that are chosen for the discretization. Unfortunately, the discretization leads to information loss. Rangel *et al.* [20] used the state space

model for constructing a gene network. Their method is based on linear models. However, there is no guarantee that the relationship between genes is linear. Imoto *et al.* [14, 15] proposed a network estimation method based on a Bayesian network and nonparametric regression to avoid discretization and for capturing nonlinear relations among genes. However, the Bayesian network and nonparametric regression model [14, 15] still has a remaining problem to be solved in the construction of cyclic regulations.

In this paper, we extend the Bayesian network and nonparametric regression model to the dynamic Bayesian network model, which can construct cyclic regulations when we have a time series gene expression data. We can include the time delay information into the proposed model naturally. The model can extract even nonlinear relations among genes automatically. For constructing a gene network with cyclic regulations based on time series gene expression data, an ordinal differential equation model (Chen *et al.* [5]; De Hoon *et al.* [8]) is an alternative method. However, this model is based on a linear system and probably unsuitable for capturing complex phenomena. We derive a new criterion for choosing an optimal network from the Bayesian statistical point of view [2]. The proposed criterion can optimize the network structure such that it gives the best representation of the gene interactions described by the data with noise. The efficiency of the proposed method is shown by analyzing *Saccharomyces cerevisiae* gene expression data.

2 Dynamic Bayesian Network and Nonparametric Regression

Suppose that we have an $n \times p$ microarray gene expression data matrix $\boldsymbol{X}$, where n and p are the numbers of microarrays and genes, respectively. Usually, the number of genes p is much larger than the number of microarrays, n. In the estimation of a gene network based on the Bayesian network, a gene is considered to be a random variable. When we model a gene network by using statistical models described by the density or probability function, the statistical model should include p random variables. However, we have only n samples and n is usually much smaller than p. In such case, the inference of the model is quite difficult or impossible, because the model has many parameters and the number of samples is not enough for estimating the parameters. The Bayesian network model has been advocated in such situations.

In the context of the dynamic Bayesian network, we consider time series data, with the ith row vector $\boldsymbol{x}_i$ of $\boldsymbol{X}$ corresponds to the states of p genes at time i. As for the time dependency, we consider the first order Markov relation described in Figure 1. Under this condition, the joint probability can be decomposed as

$$P(X_{11}, ..., X_{np}) = P(\boldsymbol{X}_1)P(\boldsymbol{X}_2|\boldsymbol{X}_1) \times \cdots \times P(\boldsymbol{X}_n|\boldsymbol{X}_{n-1}), \tag{1}$$

where $\boldsymbol{X}_i = (X_{i1}, ..., X_{ip})^T$ is a random variable vector of p genes at time i. The conditional probability $P(\boldsymbol{X}_i|\boldsymbol{X}_{i-1})$ can also be decomposed into the product

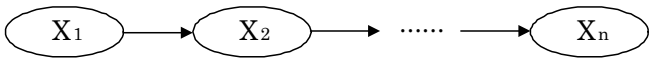


Fig. 1. Time dynamics. $\boldsymbol{X}_i$ is the states of the genes at time i for $i = 1, ..., n$.

of conditional probabilities of the form

$$P(\boldsymbol{X}_i|\boldsymbol{X}_{i-1}) = P(X_{i1}|\boldsymbol{P}_{i-1,1}) \times \cdots \times P(X_{ip}|\boldsymbol{P}_{i-1,p}), \tag{2}$$

where $\boldsymbol{P}_{i-1,j}$ is the state vector of the parent genes of jth gene at time $i-1$. The equations (1) and (2) hold when we use the density function instead of the probability measure. Hence, the dynamic Bayesian network can then be represented by using densities as follows:

$$\begin{aligned} f(x_{11}, ..., x_{np}) &= f_1(\boldsymbol{x}_1) f_2(\boldsymbol{x}_2|\boldsymbol{x}_1) \times \cdots \times f_n(\boldsymbol{x}_n|\boldsymbol{x}_{n-1}) \\ &= f_1(\boldsymbol{x}_1) \prod_{i=2}^{n} g_1(x_{i1}|\boldsymbol{p}_{i-1,1}) \times \cdots \times g_p(x_{ip}|\boldsymbol{p}_{i-1,p}) \\ &= f_1(\boldsymbol{x}_1) \prod_{j=1}^{p} \left\{ \prod_{i=2}^{n} g_j(x_{ij}|\boldsymbol{p}_{i-1,j}) \right\}, \end{aligned}$$

where $\boldsymbol{p}_{i-1,j} = (p^{(j)}_{i-1,1}, ..., p^{(j)}_{i-1,q_j})^T$ is a q_j-dimensional observation vector of parent genes. The decomposition is given by equation (2)

$$f_i(\boldsymbol{x}_i|\boldsymbol{x}_{i-1}) = g_1(x_{i1}|\boldsymbol{p}_{i-1,1}) \times \cdots \times g_p(x_{ip}|\boldsymbol{p}_{i-1,p}).$$

For modeling the relationship between x_{ij} and $\boldsymbol{p}_{i-1,j}$, we use the nonparametric additive regression model as follows:

$$x_{ij} = m_{j1}(p^{(j)}_{i-1,1}) + \cdots + m_{jq_j}(p^{(j)}_{i-1,q_j}) + \varepsilon_{ij},$$

where ε_{ij} depends independently and normally on mean 0 and variance σ_j^2. Here, $m_{jk}(\cdot)$ is a smooth function from R to R and can be expressed by using a linear combination of basis functions

$$m_{jk}(p^{(j)}_{i-1,k}) = \sum_{m=1}^{M_{jk}} \gamma^{(j)}_{mk} b^{(j)}_{mk}(p^{(j)}_{i-1,k}), \quad k = 1, \ldots, q_j,$$

where $\gamma^{(j)}_{1k}, ..., \gamma^{(j)}_{M_{jk}k}$ are unknown coefficient parameters and $\{b^{(j)}_{1k}(\cdot), ..., b^{(j)}_{M_{jk}k}(\cdot)\}$ is the prescribed set of basis functions. Then we define a dynamic Bayesian network and nonparametric regression model of the form

$$\begin{aligned} &f(x_{11}, ..., x_{np}; \boldsymbol{\theta}_G) \\ &\quad = f_1(\boldsymbol{x}_1) \prod_{j=1}^{p} \left[\prod_{i=2}^{n} \frac{1}{\sqrt{2\pi\sigma_j^2}} \exp\left\{ -\frac{(x_{ij} - \mu(\boldsymbol{p}_{i-1,j}))^2}{2\sigma_j^2} \right\} \right], \end{aligned}$$

where $\boldsymbol{\theta}_G$ is the parameter vector included in the Bayesian network model and $\mu(\boldsymbol{p}_{i-1,j}) = m_{j1}(p^{(j)}_{i-1,1}) + \cdots + m_{jq_j}(p^{(j)}_{i-1,q_j})$. When jth gene has no parent genes, $\mu(\boldsymbol{p}_{i-1,j})$ reduces to the constant μ_j.

We assume $f_1(\boldsymbol{x}_1) = g_1(x_{11}) \times \cdots \times g_1(x_{1p})$ and the joint density $f(x_{11}, ..., x_{np}; \boldsymbol{\theta}_G)$ can then be rewritten as

$$\begin{aligned} f(x_{11}, ..., x_{np}; \boldsymbol{\theta}_G) &= \prod_{j=1}^{p} \left[g_1(x_{1j}) \prod_{i=2}^{n} \frac{1}{\sqrt{2\pi\sigma_j^2}} \exp\left\{ -\frac{(x_{ij} - \mu(\boldsymbol{p}_{i-1,j}))^2}{2\sigma_j^2} \right\} \right] \\ &= \prod_{j=1}^{p} \prod_{i=1}^{n} g_j(x_{ij} | \boldsymbol{p}_{i-1,j}; \boldsymbol{\theta}_j), \end{aligned} \tag{3}$$

where $\boldsymbol{p}_{0j} = \emptyset$. Thus, $g_j(x_{ij} | \boldsymbol{p}_{i-1,j}; \boldsymbol{\theta}_j)$ represents the local structure of jth gene and its parent genes.

3 Derivation of a Criterion for Selecting Network

The dynamic Bayesian network and nonparametric regression model introduced in the previous section can be constructed when we fix the network structure. However, the gene network is generally unknown and we should estimate an optimal network based on the data. This problem can be viewed as a statistical model selection problem (see e.g., Akaike [1]; Burnham and Anderson [4]; Konishi [16]; Konishi and Kitagawa [17]). We solve this problem from the Bayesian statistical approach and derive a criterion for evaluating the goodness of the dynamic Bayesian network and nonparametric regression model.

Let $\pi(\boldsymbol{\theta}_G | \boldsymbol{\lambda})$ be a prior distribution on the parameter $\boldsymbol{\theta}_G$ in the dynamic Bayesian network and nonparametric regression model and let $\log \pi(\boldsymbol{\theta}_G | \boldsymbol{\lambda}) = O(n)$. The marginal likelihood can be represented as

$$\int f(x_{11}, ..., x_{np}; \boldsymbol{\theta}_G) \pi(\boldsymbol{\theta}_G | \boldsymbol{\lambda}) d\boldsymbol{\theta}_G.$$

Thus, when the data is given, the posterior probability of the network G is

$$\pi_{post}(G | \boldsymbol{X}) = \frac{\pi_{prior}(G) \int f(x_{11}, ..., x_{np}; \boldsymbol{\theta}_G) \pi(\boldsymbol{\theta}_G | \boldsymbol{\lambda}) d\boldsymbol{\theta}_G}{\sum_G \left\{ \pi_{prior}(G) \int f(x_{11}, ..., x_{np}; \boldsymbol{\theta}_G) \pi(\boldsymbol{\theta}_G | \boldsymbol{\lambda}) d\boldsymbol{\theta}_G \right\}}, \tag{4}$$

where $\pi_{prior}(G)$ is the prior probability of the network G. The denominator of (4) does not relate to model evaluation. Therefore, the evaluation of the network depends on the magnitude of numerator. Hence, we can choose an optimal network as the maximizer of

$$\pi_{prior}(G) \int f(x_{11}, ..., x_{np}; \boldsymbol{\theta}_G) \pi(\boldsymbol{\theta}_G | \boldsymbol{\lambda}) d\boldsymbol{\theta}_G.$$

It is clear that the essential point for constructing a network selection criterion is how to compute the high dimensional integral. Imoto *et al.* [14, 15] used the Laplace approximation for integrals (see also Tinerey and Kadane [23]; Davison [6]). This technique can be applied to the dynamic Bayesian network and nonparametric regression model directly. Hence, we have a criterion, named $\mathrm{BNRC}_{dynamic}$, of the form

$$\begin{aligned}
&\mathrm{BNRC}_{dynamic}(G) \\
&\quad = -2\log\left\{\pi_{prior}(G)\int f(x_{11},...,x_{np};\boldsymbol{\theta}_G)\pi(\boldsymbol{\theta}_G|\boldsymbol{\lambda})d\boldsymbol{\theta}_G\right\} \\
&\quad \approx -2\log\pi_{prior}(G) - r\log(2\pi/n) + \log|J_\lambda(\hat{\boldsymbol{\theta}}_G)| - 2nl_\lambda(\hat{\boldsymbol{\theta}}_G|\boldsymbol{X}), \qquad (5)
\end{aligned}$$

where r is the dimension of $\boldsymbol{\theta}_G$,

$$\begin{aligned}
l_\lambda(\boldsymbol{\theta}_G|\boldsymbol{X}) &= \log f(x_{11},...,x_{np};\boldsymbol{\theta}_G)/n + \log\pi(\boldsymbol{\theta}_G|\boldsymbol{\lambda})/n, \\
J_\lambda(\boldsymbol{\theta}_G) &= -\partial^2\{l_\lambda(\boldsymbol{\theta}_G|\boldsymbol{X})\}/\partial\boldsymbol{\theta}_G\partial\boldsymbol{\theta}_G^T
\end{aligned}$$

and $\hat{\boldsymbol{\theta}}_G$ is the mode of $l_\lambda(\boldsymbol{\theta}_G|\boldsymbol{X})$. The optimal graph is chosen such that the criterion $\mathrm{BNRC}_{dynamic}$ (5) is minimal.

4 Estimation of a Gene Network

In this section, we show the concrete strategy for estimating a gene network from cDNA microarray time series gene expression data.

4.1 Nonparametric Regression

We use the basis function approach for constructing the smooth function $m_{jk}(\cdot)$ described in Section 2. In this paper we use B-splines [7] as the basis functions. De Boor's algorithm (see, de Boor [7], Chapter 10, p.130 (3)) is a useful method for computing B-splines of any degree. We use 20 B-splines of degree 3 with equidistant knots (see also, Dierckx [10]; Eiler and Marx [11] for the details of B-spline).

4.2 Prior Distribution on the Parameter

For the prior distribution on the parameter $\boldsymbol{\theta}_G$, suppose that the parameter vectors $\boldsymbol{\theta}_j$ are independent of one another. The prior distribution can then be decomposed as $\pi(\boldsymbol{\theta}_G|\boldsymbol{\lambda}) = \prod_{j=1}^{p}\pi_j(\boldsymbol{\theta}_j|\boldsymbol{\lambda}_j)$. Suppose that the prior distribution $\pi_j(\boldsymbol{\theta}_j|\boldsymbol{\lambda}_j)$ is factorized as $\pi_j(\boldsymbol{\theta}_j|\boldsymbol{\lambda}_j) = \prod_{k=1}^{q_j}\pi_{jk}(\boldsymbol{\gamma}_{jk}|\lambda_{jk})$, where λ_{jk} are hyper parameters. We use a singular M_{jk} variate normal distribution as the prior distribution on $\boldsymbol{\gamma}_{jk}$,

$$\pi_{jk}(\boldsymbol{\gamma}_{jk}|\lambda_{jk}) = \left(\frac{2\pi}{n\lambda_{jk}}\right)^{-(M_{jk}-2)/2}|K_{jk}|_+^{1/2}\exp\left(-\frac{n\lambda_{jk}}{2}\boldsymbol{\gamma}_{jk}^T K_{jk}\boldsymbol{\gamma}_{jk}\right),$$

where K_{jk} is an $M_{jk} \times M_{jk}$ symmetric positive semidefinite matrix satisfying $\boldsymbol{\gamma}_{jk}^T K_{jk} \boldsymbol{\gamma}_{jk} = \sum_{\alpha=3}^{M_{jk}} (\gamma_{\alpha k}^{(j)} - 2\gamma_{\alpha-1,k}^{(j)} + \gamma_{\alpha-2,k}^{(j)})^2$. This setting of the prior distribution on $\boldsymbol{\theta}_G$ is the same as Imoto *et al.* [14, 15].

4.3 Proposed Criterion

By using the prior distributions in Section 4.2, the $\mathrm{BNRC}_{dynamic}$ can be decomposed as follows:

$$\mathrm{BNRC}_{dynamic} = \sum_{j=1}^{p} \mathrm{BNRC}_{dynamic}^{(j)}, \tag{6}$$

where $\mathrm{BNRC}_{dynamic}^{(j)}$ is a local criterion score of jth gene and is defined by

$$\begin{aligned}
&\mathrm{BNRC}_{dynamic}^{(j)} \\
&\quad = -2 \log \left\{ \int \pi_{prior}(L_j) \prod_{i=1}^{n} g_j(x_{ij} | \boldsymbol{p}_{i-1,j}; \boldsymbol{\theta}_j) \pi_j(\boldsymbol{\theta}_j | \boldsymbol{\lambda}_j) d\boldsymbol{\theta}_j \right\} \\
&\quad \approx -2 \log \pi_{prior}(L_j) - r_j \log(2\pi/n) + \log |J_{\lambda_j}^{(j)}(\hat{\boldsymbol{\theta}}_j)| - 2n l_{\lambda_j}^{(j)}(\hat{\boldsymbol{\theta}}_j | \boldsymbol{X}),
\end{aligned}$$

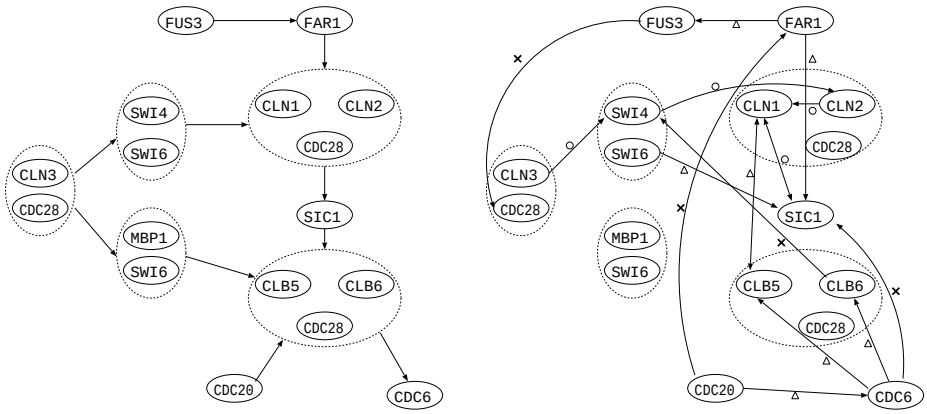
where r_j is the dimension of $\boldsymbol{\theta}_j$,

$$\begin{aligned}
l_{\lambda_j}^{(j)}(\hat{\boldsymbol{\theta}}_j | \boldsymbol{X}) &= \sum_{i=1}^{n} \log g_j(x_{ij} | \boldsymbol{p}_{i-1,j}; \boldsymbol{\theta}_j)/n + \log \pi(\boldsymbol{\theta}_j | \boldsymbol{\lambda}_j)/n, \\
J_{\lambda_j}^{(j)}(\hat{\boldsymbol{\theta}}_j) &= -\partial^2 \{ l_{\lambda_j}^{(j)}(\hat{\boldsymbol{\theta}}_j | \boldsymbol{X}) \} / \partial \boldsymbol{\theta}_j \partial \boldsymbol{\theta}_j^T
\end{aligned}$$

and $\hat{\boldsymbol{\theta}}_j$ is the mode of $l_{\lambda_j}^{(j)}(\boldsymbol{\theta}_j | \boldsymbol{X})$. Here $\pi_{prior}(L_j)$ are prior probabilities satisfying $\sum_{j=1}^{p} \log \pi_{prior}(L_j) = \log \pi_{prior}(G)$. We set the prior probability of local structure $\pi_{prior}(L_j)$ as $\pi_{prior}(L_j) = \exp\{-(\text{The number of parent genes ofthe } j \text{ th gene})\}$.

4.4 Algorithm for Learning Network

By using the dynamic Bayesian network and nonparametric regression model together with the proposed criterion, $\mathrm{BNRC}_{dynamic}$, we can formulate the network learning process as follows: it is clear from (3) and (6) that the optimization of network structure is equivalent to choosing the parent genes that regulate the target genes. However, it is a time-consuming task to consider all possible gene combinations as the parent genes. Therefore, we reduce the learning space by selecting candidate parent genes. After this step, a greedy hill-climbing algorithm is employed for finding better networks. Our algorithm can be expressed as follows:



(a)Target pathway (b)Result of the proposed method

Fig. 2. Cell cycle pathway compiled in KEGG.

Step1: *Preprocessing stage*

We make the $p \times p$ matrix whose (i, j)th element is the $\mathrm{BNRC}^{(j)}_{dynamic}$ score of the graph "$\mathrm{gene}_i \rightarrow \mathrm{gene}_j$" and we define the candidate set of parent genes of gene_j that gives small $\mathrm{BNRC}^{(j)}_{dynamic}$ scores.

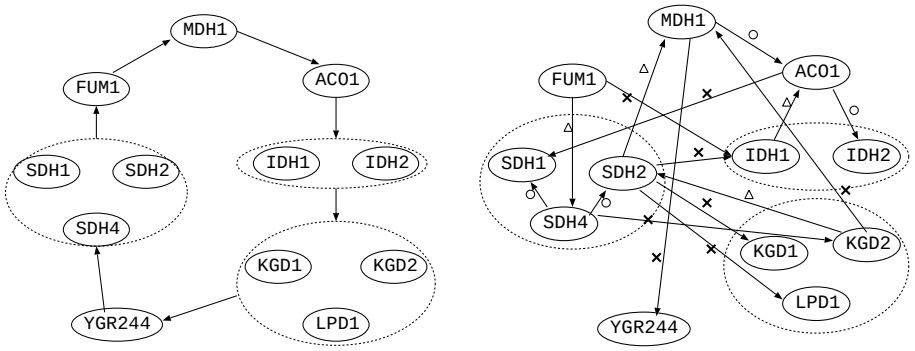
Step 2: *Learning stage*

For a greedy hill-climbing algorithm, we start form the empty network and repeat the following steps:

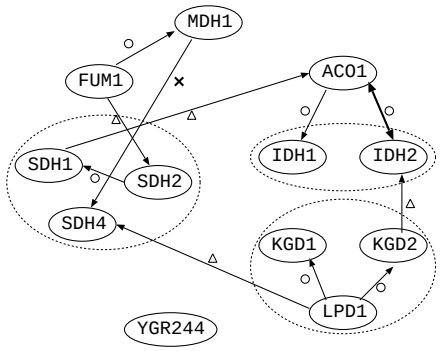
Step2-1: For gene_j, implement one from two procedures that *add* a parent gene or *delete* a parent gene, which gives smaller $\mathrm{BNRC}^{(j)}_{dynamic}$ score.
Step2-2: Repeat Step2-1 until we find the best set of parent genes of jth gene.
Step2-3: Repeat Step2-1 and 2-2 for all genes.
Step2-4: We choose the optimal network that gives the smallest $\mathrm{BNRC}_{dyanmic}$ score.

5 Computational Experiment

We demonstrate our proposed method through the analysis of the *Saccharomyces cerevisiae* cell cycle gene expression data collected by Spellman *et al.* [22]. This data contains two short time series (two time points; cln3, clb2) and four medium time series (18, 24, 17 and 14 time points; alpha, cdc15, cdc28 and elu). In the estimation of a gene network, we use four medium time series. For combining four time series, we ignore the first observation of the target gene and last one of the parent genes for each time series when we fit the nonparametric regression model. We set the number of the candidate parent genes to 10, since the resulting network did not change while using a larger set of candidate parents.



(a)Target pathway (b)Result of the Bayesian network



(c)Result of the proposed method

Fig. 3. Metabolic pathway reported by DeRisi *et al.* [9].

At first, we focus on the cell cycle pathway stored in the KEGG database [24]. The target network is around CDC28 (YBR160w; cyclin-dependent protein kinase). This network contains 45 genes. The partial pathway registered in KEGG is shown in Figure 2 (a) and the estimated network is in Figure 2 (b). The edges in the dotted circles can be considered as the correct edges. We can model some correct relations by using the proposed method. We denote the correct estimation by a circle next to the edge. A triangle represents the edge incorrectly directed or the edge which bypassed one gene, while a cross represents an estimated edge that is not present in the target graph.

Our second example is the metabolic pathway reported by DeRisi *et al.* [9]. This network contains 57 genes and the target pathway is partially shown in Figure 3 (a). We apply the Bayesian network and nonparametric regression model [14, 15] to these data. The resulting network is shown in Figure 3 (b). The network of Figure 3 (c) is obtained by the dynamic Bayesian network and

nonparametric regression model. It is difficult to estimate the metabolic pathway from cDNA microarray data. However, our model can detect some correct relations. Comparing with the Bayesian network and nonparametric regression, the number of false positives of the proposed method in Figure 3 (c) is much smaller than those in Figure 3 (b).

We observed that the Bayesian network and nonparametric regression can work well in many cases. However, when there is a cyclic gene regulation, the Bayesian network and nonparametric regression model tends to estimate many false positives in the cyclic regulation. In such case, the proposed method can reduce the number of false positives and estimate gene regulations effectively.

6 Discussion

In this paper, we proposed a new statistical gene network estimation method based on the dynamic Bayesian network and nonparametric regression model. Our proposed method has several advantages compared with other network estimation method such as the Bayesian and Boolean networks. First, our model can take time information into account naturally. Second, our model can analyze the microarray data as the continuous data without additional data pretreatments such as discretization. Last, even nonlinear relations can be detected and modeled by our proposed method.

Simulating a genetic system is one of the central topics in systems biology. Since the simulation is based on biological knowledge, our network estimation method can support the biological simulation by constructing the unknown regulations. In this paper, we only demonstrate the model based on the first-order Markov relation between time points described in Figure 1. However, the relationship between time points is arbitrary and we can choose the time dependency structure based on our proposed criterion. We would like to discuss this topic in our future work.

References

1. Akaike, H.: Information theory and an extension of the maximum likelihood principle. In: Petrov, B.N., Csaki, F. (eds.): 2nd International Symposium on Information Theory. Akademiai Kiado, Budapest (1973) 267–281
2. Berger, J.O.: Statistical Decision theory and Bayesian analysis. Springer-Verlag New York (1985)
3. Bilmes, J.A.: Dynamic Bayesian multinets. Proc. 16th Conf. on Uncertainty in Artificial Intelligence (2000) 38–45
4. Burnham, K.P., Anderson, D.R.: Model selection and inference, a practical information-theoretical approach. Springer-Verlag New York (1998)
5. Chen, T., He, H.L., Church, G.M.: Modeling gene expression with differential equations. Proc. Pacific Symposium on Biocomputing **4** (1999) 29–40
6. Davison A.C.: Approximate predictive likelihood. Biometrika **73** (1986) 323–332
7. De Boor, C.: A practical guide to splines. Springer-Verlag Berlin (1978)

8. De Hoon, M.J.L., Imoto, S., Kobayashi, K., Ogasawara, N., Miyano, S.: Inferring gene regulatory networks from time-ordered gene expression data of *Bacillussubtilis* using differential equations. Proc. Pacific Symposium on Biocomputing **8** (2003) in press
9. DeRisi, J., Lyer, V.R., Brown, P.O.: Exploring the metabolic and gene control of gene expression on a genomic scale. Science **278** (1997) 680–686
10. Dierckx, P.: Curve and surface fitting with splines. Oxford (1993)
11. Eiler, P.H.C., Marx, B.: Flexible smoothing with B-splines and penalties (with discussion). Statistical Science **11** (1996) 89–121
12. Friedman, N., Murphy, K., Russell, S.: Learning the structure of dynamic probabilistic networks. Proc. Conf. on Uncertainty in Artificial Intelligence (1998) 139–147
13. Friedman, N., Linial, M., Nachman, I., Pe'er, D.: Using Bayesian network to analyze expression data. J. Comp. Biol. **7** (2000) 601–620
14. Imoto, S., Goto, T., Miyano, S.: Estimation of genetic networks and functional structures between genes by using Bayesian network and nonparametric regression. Proc. Pacific Symposium on Biocomputing **7** (2002) 175–186
15. Imoto, S., Kim, S., Goto, T., Aburatani, S., Tashiro, K., Kuhara, S., Miyano, S.: Bayesian network and nonparametric heteroscedastic regression for nonlinear modeling of genetic network. Proc. IEEE Computer Society Bioinformatics Conference (2002) 219–227
16. Konishi, S.: Statistical model evaluation and information criteria. In: Ghosh, S. (ed.): Multivaliate Analysis, Design of Experiments and Survey Sampling. Marcel Dekker, New York (1999) 369–399
17. Konishi, S., Kitagawa, G.: Generalized information criteria in model selection. Biometrika **83** (1996) 875–890
18. Murphy, K., Mian, S.: Modelling Gene Expression Data using Dynamic Bayesian Networks. *Technical report, Computer Science Division, University of California, Berkeley, CA* (1999)
19. Pe'er, D., Regev, A., Elidan, G., Friedman, N.: Inferring subnetworks from perturbed expression profiles. Bioinformatics **17** (ISBM2001) 215–224
20. Rangel, C., Wild, D.L., Falciani, F.:Modelling biological responses using gene expression profiling and linear dynamical systems. Proc. 2nd International Conference on Systems Biology (2001)
21. Someren, E.V., Wessels, L., Reinders, M.: Linear modeling of genetic networks from experimental data. Bioinformatics **18** (ISMB2002) 355–366
22. Spellman, P.T., Sherlock, G., Zhang, M.Q., Iyer, V.R., Anders, K., Eisen, M.B., Brown, P.O., Botstein, D., Futcher, B.: Comprehensive identification of cell cycle-regulated genes of the yeast *Saccharomyces cerevisiae* by microarray hybridization. Molecular Biology of the Cell **9** (1998) 3273–3297.
23. Tinerey, L., Kadane, J.B.: Accurate approximations for posterior moments and marginal densities. J. Amer. Statist. Assoc. **81** (1986) 82–86
24. http://www.genome.ad.jp/kegg/