

IEICE Proceeding Series

The Parameter Optimization in the Inference of Gene Regulatory
Network by Neural Networks Adopting Majority Rule

Yasuki Hirai, Naoyuki Kizaki, Hiroshi Yoshino, Hiroaki Kurokawa

Vol. 1 pp. 648-651

Publication Date: 2014/03/17

Online ISSN: 2188-5079

Downloaded from www.proceeding.ieice.org



The Parameter Optimization in the Inference of Gene Regulatory Network by Neural Networks Adopting Majority Rule

Yasuki Hirai[†], Naoyuki Kizaki[†], Hiroshi Yoshino[†] and Hiroaki Kurokawa[†]

[†]School of Computer Science, Tokyo University of Technology
1404-1 Katakura, Hachioji, Tokyo, 192-0982 Japan
Email: hkuro@cs.teu.ac.jp

Abstract—In order to infer the Gene Regulatory Network (GRN) described by the differential equation, it is required that the function approximation of a lot of unknown functions using the time course data of gene expressions. Recently, the inferring method of the GRN using neural network had been proposed. We also had been proposed another method using neural networks that can show the various results depending on the parameter that defined in our method. Although our method can show the preferable results depending on the requirement, it cannot decide the parameter to show the reasonable result of the inference automatically. In this paper, we propose the method to decide the parameter for the GRN inference using our method. In simulations, the results show that the method can decide the parameter appropriately, and the reasonable result of the inference is obtained.

1. Introduction

The Gene Regulatory Network (GRN) shows the summary of the interaction among genes. In order to infer the unknown GRN, we use the experimental data from DNA microarray or other biological experiments. We can consider that the inferring the GRN from gene expression data is the reverse engineering of the biological systems. The inference of the GRN is important technique to discover the principle of the life[1]~[3].

Recently, various models of the GRN had been proposed. The differential equation model is one of the major model that can represent the time series dynamics of the gene expressions. In the differential equation models, the derivative of each gene expression in the network is defined by arbitrary function[4]~[8].

In general, these arbitrary function is unknown, so that the function approximation of these unknown functions are required for the inference of the GRN that shown by the differential equations. The experimental data of gene expression is used for the function approximation of these unknown functions. Recently, experimental technologies such as DNA microarrays are developed and the time course data of gene expressions can be easily obtained. Then it is assumed that the inference of GRN with differential equation is achieved by function approximation by the time course data of gene expressions.

The method using neural networks had been proposed to infer the GRN that described by the differential equations[6]. We also had been proposed another GRN inference method using the neural networks[9]. In our method, the valid regulations among genes are decided through a lot of trials of the inference using neural networks with the majority rule. Where we assumed that the frequently inferred regulations in the trials are valid regulations. Under this assumption, our method show the valid regulations by determining the threshold of the inferred frequency.

The merit of our method is that the characteristics of the result depends on the threshold of the inferred frequency. In order to obtain a preferable result, our method vary the threshold depending on the requirement. However, in our proposed method, the threshold of the frequency is considered as a parameter, and we cannot decide the threshold automatically. Therefore, the method does not suitable for the purpose to show the reasonable result of the inference automatically.

In this paper, we propose the method to decide the threshold automatically. This can be considered as the supplement of our proposed GRN inferring method[9]. In simulations, our method is applied to the inference of an artificial GRN model and SOS network of the Escherichia coli bacterium. From the results, it is shown that our method can show the reasonable result of the GRN inference.

2. Inference of the Gene Regulatory Networks using the Neural Networks

2.1. The inference method by the Neural Network

In the differential equation model of the GRN, the derivative of the gene expression X_i is formulated as following differential equations.

$$\frac{dX_i}{dt} = G_i(X_1, X_2, \dots, X_N). \quad (1)$$

Where $G_i(X_1, X_2, \dots, X_N)$ is an arbitrary unknown function and we assumed that the target GRN has N genes, *i.e.*, $i = 1, 2, \dots, N$. Therefore, in the differential equations model, determining N unknown functions G_1, G_2, \dots, G_N enables the calculation for the dynamical characteristics of gene expressions.

The GRN inference is achieved by the function approximation of G_1, G_2, \dots, G_N . In the method using neural networks[6], these function are approximated using the neural networks with expression data of each gene in the GRN.

To show the regulations among genes in the GRN, we have to derive them from a set of approximated functions G_i . To derive the regulation, the conventional method uses the sensitivity coefficient S_i that is defined as follows.

$$S_i(j) = \frac{\partial}{\partial X_j} \left(\frac{dX_i}{dt} \right) = \frac{\partial G_i(X_1, \dots, X_N)}{\partial X_j} \quad (2)$$

The sensitivity coefficient $S_i(j)$ represents an influence of the gene expression of j th gene to that of i th gene. Where the sensitivity coefficients is time varying. However, in general, the presentation of the static regulations among genes are required to the GRN. According to the method in the conventional study[6], the positive and negative sensitivity coefficient, $S_i^p(j)$ and $S_i^n(j)$, are calculated as follows. These are considered as an averaged value of the positive and negative sensitivity coefficients.

$$S_i^p(j) = \frac{1}{T} \sum_{k=1}^T p \left(\left. \frac{\partial G_i}{\partial X_j} \right|_{t_k} \right), \quad (3)$$

$$S_i^n(j) = \frac{1}{T} \sum_{k=1}^T n \left(\left. \frac{\partial G_i}{\partial X_j} \right|_{t_k} \right), \quad (4)$$

where,

$$p(x) = \begin{cases} x & \text{if } x > 0 \\ 0 & \text{otherwise} \end{cases}, \quad (5)$$

$$n(x) = \begin{cases} x & \text{if } x < 0 \\ 0 & \text{otherwise} \end{cases}. \quad (6)$$

Where, T is a number of time course data that is presented to the neural network for the function approximation, and t_k is the time that the k th data is sampled. Note that the sensitivity coefficient, $S_i(j) = \left. \frac{\partial G_i}{\partial X_j} \right|_{t_k}$, is calculated using the parameters in learned neural network that approximate the function G_i .

According to the $S_i^p(j)$ and $S_i^n(j)$, we can infer that the regulation between the j th gene to the i th gene is TRUE or FALSE. In the case that the regulation is TRUE, we can also infer that the regulation is either positive regulation(acceleration) or negative regulation(repression). These judgments for the inference is based on some user defined criteria. In this study, we use the criteria that had been shown in the conventional study[6].

2.2. Adopting the majority rule to the method using the neural networks

The result of the inference using the neural network includes the error depending on the accuracy of the function approximation. Also, it is well known that the expression

data from experiments include a lot of errors, and this will become a cause of the error in inference. Therefore, it can be considered that the perfect inference is impossible, so that the improvement of the accuracy of the inference is considered as the problem of GRN inference. Also, the accuracy of the inference cannot evaluate by simple criterion. Thus, the valuable result of the inference will vary depending on the various requirements.

Under these assumptions, we had proposed the GRN inference method using neural networks adopting a majority rule[9]. The merit of our method is that the method can show a preferable result for the requirement depending on the parameter. The procedure of our proposed algorithm is summarized as follows.

1. Approximate the function of the differential equation model given by eq.(1) using neural networks with the measured time course data.
2. Infer the gene regulatory network using sensitivity coefficients calculated from approximated function.
3. Repeat the procedure of 1) - 2) for sufficient number of trials.
4. Count the frequency of the regulation that inferred as TRUE for every possible regulation.
5. The regulation which is inferred more than the threshold is assumed to be a proper regulation.

2.3. Characteristics of the sensitivity and the specificity of inferred GRN

The inferred GRN is evaluated using the *sensitivity*, S_n , and the *specificity*, S_p , as criteria to show the plausibility of the inferred network. These criteria are defined as follows.

$$S_n = \frac{TP}{TP + FN}. \quad (7)$$

$$S_p = \frac{TN}{FP + TN}. \quad (8)$$

Where, TP, FN, TN and FP are the numbers of regulations that is true-positive, false-negative, true-negative and false positive, respectively. In other words, S_n shows a degree of correctly inferred regulations, and it doesn't depend on the incorrectly inferred regulations. On the other hand, S_p shows how few the incorrect regulations is in the inferred network, and it doesn't depend on the number of the regulations that is not inferred incorrectly.

For example, if the result of the inference shows that all the possible regulations are TRUE, S_n and S_p would become 1 and 0, respectively. On the contrary, in the case that no regulations are inferred, they would become 0 and 1, respectively. If the GRN is inferred perfectly, both S_n and S_p become 1. However, as we mentioned above, the perfect inference is almost impossible. In the case that the inference is not perfect, we often cannot determine the better inference from a set of S_p and S_n .

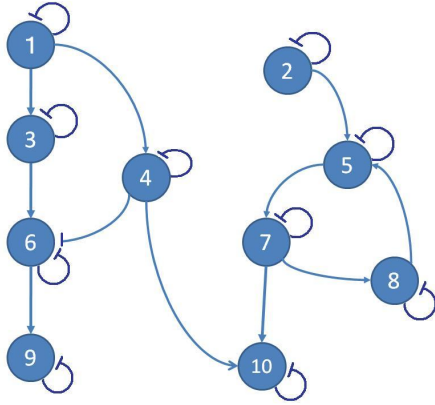


Figure 1: 10-gene artificial GRN model. The GRN is defined by the S-system

In the previous paper[9], we considered that the threshold of the frequency is a parameter of our method, and the characteristics of the threshold were showed, that is, the large threshold gives large specificity, and small threshold gives large sensitivity. To consider these characteristics, our method can show the preferable results of inferring depending on the requirement. However, our method cannot decide the threshold automatically.

In the next section, we show the results of our GRN inference method, and describe the method to decide the threshold that gives reasonable result automatically.

3. Examples and the decision of the parameter

3.1. The inference of the artificially defined GRN

In this section, we show the simulation results of the GRN inference by our method and describe the method to decide the threshold of the inferred frequency. In this simulation, we assume that the 10-gene artificial GRN model shown in the Fig.1 is the target network of the inference. This GRN is defined by S-system, which is given by,

$$\frac{dX_i}{dt} = \alpha_i \prod_{j=1}^N X_j^{g_{i,j}} - \beta_i \prod_{j=1}^N X_j^{h_{i,j}} \quad (i = 1, 2, \dots, N). \quad (9)$$

Where the α_i , $g_{i,j}$, β_i , and $h_{i,j}$ are the parameters, and they are defined corresponding to the regulations shown in the Fig.1, appropriately. In the simulation, the gene expression data generated by S-system is used for the GRN inference.

Figure 2 shows the histogram of the number of regulations for each inferred frequency, where the number of the trials of inference is 100. In Fig.2, the results of the fitting with quadratic polynomial curve is also shown. As shown in Fig.2, the inferred frequency shows clustered distribution. A lot of regulations are clustered in the range of low frequency and a few regulations are clustered in the range of high frequency. To decide the threshold automatically,

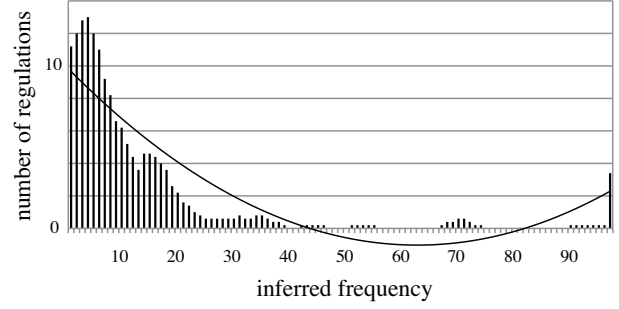


Figure 2: The histogram of the number of regulations for each inferred frequency and the fitting curve. The number of the trials of the inference is 100.

Table 1: The proper threshold and the threshold given by the proposed method

| trials | the proper threshold | by proposed method |
|--------|----------------------|--------------------|
| 50 | 30,31 | 31 |
| 100 | 55~70 | 63 |
| 250 | 166,167 | 140 |
| 500 | 338~355 | 348 |
| 1000 | 669~697 | 584 |

we assumed that the threshold is set to the minimum of the fitting curve.

Table 1 shows the proper threshold and the threshold obtained by proposed method. Where the proper threshold is empirically decided from the result of the simulation. Comparing to the proper threshold, the proposed method can give same or similar threshold. However, some of the threshold obtained by our method has small difference to the proper threshold.

Then we evaluate the sensitivity and the specificity of the inferred GRN by our proposed method. Table 2 shows the S_n and S_p of the inferred GRN using the threshold by our method and the proper threshold in Table 1. As shown in the Table 2, almost the same results of S_n and S_p are obtained even if the threshold that decided by our method has some differences to the proper threshold. Therefore, it is considered that the threshold decided by our method can give a reasonable result of the inference.

3.2. The inference of SOS network by DNA microarray data

In this section, we show the simulation results to apply our method to the practical example. Figure 3 shows the SOS network of the Escherichia coli bacterium, and we assumed that the GRN is the target network of the inference in this section. Note that each regulation in the SOS network is revealed by various results of experiment, and we referred to the conventional studies[7][8] to obtain the SOS network shown in Fig.3.

Table 2: The sensitivity and the specificity of inferred GRN

| trials | by proper threshold | | by proposed method | |
|--------|---------------------|---------|--------------------|---------|
| | S_n | S_p | S_n | S_p |
| 50 | 0.98324 | 0.80952 | 0.98324 | 0.80952 |
| 100 | 0.97765 | 0.80952 | 0.97765 | 0.80952 |
| 250 | 0.98324 | 0.80952 | 0.97765 | 0.80952 |
| 500 | 0.98883 | 0.80952 | 0.98883 | 0.80952 |
| 1000 | 0.98883 | 0.80952 | 0.97765 | 0.80952 |

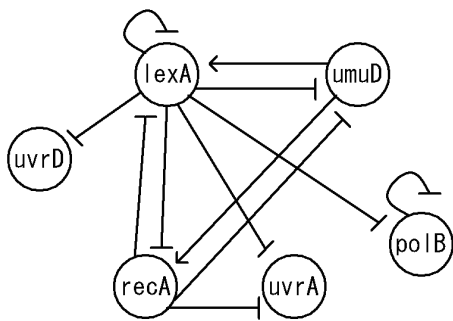


Figure 3: SOS network of the Escherichia coli bacterium

This target network is inferred by our method using the DNA microarray data[10]. Where, the number of the trials of the inference is 100. From the results of the inference by our method, we obtained the results that S_n and S_p are 0.750 and 0.667, respectively. Where, the reasonable result using the empirically decided threshold is $S_n=0.833$ and $S_p=0.667$. Therefore, we can know that the threshold of the frequency is decided almost appropriately.

To compare with the major conventional methods, we show S_n , S_p , TP, TN, FP, and FN of the inference results from the conventional methods in Table 3. Actually, the results of our method cannot show the superiority to the other method. However the results shown in the Table 3 are almost equal and we consider that our method has significance as alternative method of the GRN inference.

4. Conclusion

In this study, we showed the method to decide the parameter for the gene regulatory network inference by our method using neural networks. In the simulations, we showed that the results of the inference of artificially defined 10-genes GRN and SOS network of Escherichia coli bacterium. The simulation results showed that the method gives the reasonable result of the GRN inference.

References

[1] D'haeseleer, P. and Liang, S. and Somogyi, R., "Genetic network inference: from co-expression cluster-

Table 3: The results of GRN inference using the proposed method and the conventional methods.

| method | S_n | S_p | TP | TN | FP | FN |
|-----------|-------|-------|----|----|----|----|
| proposed | 0.750 | 0.667 | 9 | 40 | 20 | 3 |
| Kimura[6] | 0.583 | 0.767 | 7 | 46 | 14 | 5 |
| Mondal[7] | 0.667 | 0.883 | 8 | 53 | 7 | 4 |
| Kabir[8] | 0.917 | 0.983 | 11 | 59 | 6 | 1 |

ing to reverse engineering", *Bioinformatics* 16, No.8, pp.707-726, 2000

- [2] T. Akutsu and S. Miyano, "Identification of genetic networks from a small number of gene expression patterns under the boolean network model", *1999 Pacific Symposium on Biocomputing*, pp.17-28, 1999.
- [3] Dirk Husmeier, "Sensitivity and specificity of inferring genetic regulatory interactions from microarray experiments with dynamic Bayesian networks", *Bioinformatics*, Vol.19, no.17, pp. 2271-2282, 2003.
- [4] 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.
- [5] N. Noman and H. Iba, "Inferring Gene Regulatory Networks Using Differential Evolution with Local Search Heuristics", *IEEE/ACM Transactions on computational biology and bioinformatics*, Vol. 4, No. 4, pp.634-647, 2007
- [6] S. Kimura, K.Sonoda, S. Yamane, K. Matsumura and M. Hatakeyama, "Function approximation approach to the inference of neural network models of genetic networks" *IPSJ Transactions on Bioinformatics*, Vol.48, No.SIG5(TBIO2), pp.9-19, 2007.
- [7] B. S. Mondal, A.K. Sarkar, M.M. Hasan and N. Noman, "Reconstruction of Gene Regulatory Networks using Differential Evolution," *Proc. of ICCT2010*, pp.440-445, 2010.
- [8] M. Kabir, N. Noman and H. Iba, "Reverse engineering gene regulatory network from microarray data using linear time-variant model," *BMC bioinformatics*, 11(Suppl 1):S56, 2010.
- [9] Y. Hirai, M. Kikuchi, and H.kurokawa, "Inferring method of the Gene Regulatory Networks using Neural Networks Adopting a Majority Rule," *Proc. of IJCNN2011*, pp. 2936 - 2943, 2011.
- [10] "Data of SOS system reporter plasmids" <http://www.weizmann.ac.il/mcb/UriAlon/Papers/SOSData/>