Estimation of Gene Regulatory Network Using Stochastic Differential Equation Model

Shinya Nabatame\textsuperscript{1} \hspace{1cm} Hitoshi Iba\textsuperscript{2}
nabatame@iba.k.u-tokyo.ac.jp \hspace{1cm} iba@iba.k.u-tokyo.ac.jp

\textsuperscript{1} Department of Electronics, School of Engineering, University of Tokyo, Japan
\textsuperscript{2} Department of Frontier Informatics, School of Frontier Science, University of Tokyo, Japan

Keywords: gene regulatory networks, stochastic differential equation, Escherichia coli, S.O.S. DNA Repair network

1 Introduction

In order to express the dynamics of gene expressions, various models, such as S-system and additive model, are proposed so as to estimate a gene regulatory network. One of the proposed models is a Stochastic Differential Equation (SDE) based on the irregular Brownian motion. This SDE model is so widely used in finance, neuroscience and physics that it seems prominent for expressing dynamic diffusion process. Expressing the process of transcription with this SDE model, dynamics of gene expression is expressed by a simple linear model. The previous results \cite{1} show that time series data of a cell cycle was effectively captured for Saccharomyces cerevisiae. In this paper, we report how to estimate S.O.S. DNA Repair network with SDE model.

2 Method and Results

2.1 Stochastic Differential Equation model

The gene transcription process is modeled according to the theory of stochastic differential equation. A simple linear SDE model is given as follows:

\[
\Delta X_i = \left[ c_0 + \sum_{i=1}^{n} c_i f_i(X_{i-}) \right] \Delta t + \varepsilon_{i,\Delta t},
\]

where \(X_i\) and \(X_{i-}\) are expression level of target gene and \(i\)-th regulator gene, \(c_i\) is contribution of \(i\)-th regulator, \(f_i\) is sigmoid function depending on \(i\)-th regulator and \(\varepsilon_{i,\Delta t}\) is a random error generated from the normal distribution, i.e. \(N(0, \sigma^2 \Delta t)\).

2.1 Estimation method

First, a target gene is selected from a given network and regulator genes are chosen for the target gene. Then, the maximum likelihood estimation is utilized with its SDE model for the purpose of fitting the time series data of the target gene along with those of the putative regulator genes. Fitting quality is evaluated with Akaike Information Criterion (AIC). If there are \(N\) genes in the network, the number of combinations of regulator genes is \(2^N\). In this paper, we assume a small network with 8 genes so that we can enumerate all combinations for the AIC derivation. The one which gives a minimum AIC value is identified as the combination of regulators for the target gene. This identification process is repeated for all genes.

2.2 Heuristics

Since we have to consider a huge number of combinations for the regulator identification, we introduce the following heuristics so as to obtain the relatively small AIC values. First, regulators' contributions are calculated with the maximum likelihood estimation and the contribution vector, \(c = [c_1, c_2, \ldots, c_n]\), where \(c_i\) means the contribution of \(i\)-th regulator to the target gene. Then, we calculate the average of the vector elements, i.e. \(\overline{c} = 1/n \sum c_i\), and remove all \(c_i\) whose values are smaller than \(\overline{c} \times k\). \(k\) is a use-defined parameter ranged from 0 to 1. Remaining \(c_i\)'s are identified as regulators of the target gene.
2.3 Gene Expression Data and Data arrangement
We use the S.O.S. DNA Repair network of E.coli bacterium for the experiment. This gene regulatory network is well known for the responsibility for repairing the DNA after some damage. The entire system consists of about 30 genes regulated at the transcriptional level. The expressed data are the kinetics of main 8 genes of the above-mentioned network. They were measured after the irradiation of UV light at the initial time. Four experiments were conducted for various light intensities. In each experiment, 8 genes, i.e., uvrD, lexA, umuD, recA, uvrA, uvrY, ruvA, and polB, are monitored with 50 instants evenly spaced every 6 minutes.

Since four experiments were conducted, four AIC values are calculated for each combination of regulators. In this research, the sum of those four values is used for the evaluation, i.e., the smaller, the better.

We arrange the data using the moving average method before the final estimation. More precisely, the \( j \)-th value of time series data is averaged from \( j-1, j \) and \( j+1 \)-th values.

2.4 Results

From the estimation results, lexA's regulation was detected to all genes except ruvA (2nd column). recA's regulation to lexA (element(2,4)) was also detected. These relations are consistent with the experimental observation shown in Figure 1. However, a few wrong relations contradictory to the known observation were detected as well (e.g., the 3rd, 4th, 5th and 8th columns, regulation by umuD, recA, uvrA, polB).

3 Discussion and Conclusion
In this paper, we tried to estimate the S.O.S DNA Repair network of E.coli with our proposed model, i.e., SDE model. The empirical results showed that the edge detection ability seems to be superior to some previous research. There are some limitations to be improved. For example, the regulator activation or repression cannot be judged by checking the contribution value of each regulator. Another problem is that more regulations are liable to be detected than expected. In spite of these disadvantages, however, we believe that the proposed method seems promising for the further extension.

References
