

Development and Evaluation of a Prototype System for the Inference of Genetic Networks Using Evolutionary Computation

Kouji TANAKA^{1,2}
k.tanaka@waseda.jp

Hayato YAMANA^{1,2,3}
yamana@waseda.jp

- ¹ Consolidated Research Institute for Advanced Science and Medical Care, Waseda University, 513 Wasedaturumaki-cho, Shinjuku-ku, Tokyo 162-0041, Japan
² Research Institute of Information Technological Biology, Waseda University, AIST Annex 8F, 2-42, Aomi, Koto-ku, Tokyo 135-0064, Japan
³ Faculty of Science and Engineering, Waseda University, 3-4-1 Okubo, Shinjuku-ku, Tokyo 169-8555, Japan

Abstract

Estimating mutual interactions of genetic networks is mainly to infer the mutual control relationships from multiple genes from the gene expression data. Such correlations are typically expressible in the form of nonlinear simultaneous differential equations. However, most work to date has employed S-systems as an expression of such differential equations, allowing only rough approximations of Mass-action, and as such it was difficult to determine the actual correlations between genes. We proposed hybrid mathematical model which fused each advantage of Mass-action model and S-system model, and developed a prototype by implementing our method, and evaluated the effectiveness of our method.

Keywords:

genetic network, genetic programming, genetic algorithm, hybrid mathematical model

1 Introduction

Measurement technologies such as DNA microarrays have seen rapid development in recent years, allowing large quantities of gene expression data to become more available. As a result, many researchers are interested in inferring the genetic network from the observed time-course data and it has become one of the most important missions of post-genome project.

The inference of genetic networks is mainly to infer the mathematical model relationship from multiple genes from the gene expression data of the wild type of all target genes, or one of the mutant of some genes. Such mathematical model is typically expressible in the form of nonlinear simultaneous differential equations. Most work to date has employed Mass-action model or S-system model[1] as an expression of such differential equations[2, 3]. However, although Mass-action model can express detailed network structure, it is difficult to speed up the calculation time because it has a lot of optimization parameters. On the other hand, although S-system model can speed up the calculation time because it has few optimization parameters compared with Mass-action model, it can express only the outline of network structure because it is rough approximation of Mass-action model.

Our previous work, we formulate the mutual interactions as actual simultaneous differential equations (Mass-action), and automatically determine its structure and constants (coefficients) using genetic programming (GP) from a given data series[3]. This system has resulted in precise determination of the equations in some cases. However, this system had the problem which optimization efficiency and accuracy was not so good, and was not solved in the large-scale network. In this work, we propose a new method to improve the optimization efficiency and accuracy.

2 Method and Results

We propose hybrid mathematical model which fused each advantage of Mass-action model and S-system model. Concretely, the first stage of optimization infers the outline of network structure by using S-system model and genetic algorithm (GA) as a mathematical model and optimization method, respectively. The second stage of optimization uses a result of the first stage as a constraint condition, and infers detailed network structure by using Mass-action model and genetic programming (GP) as a mathematical model and optimization method, respectively. This proposal method decreases the optimization parameter by using the result of the first stage at the second stage, and achieves the improvement of the optimization efficiency and accuracy.

As the target equation, we used simultaneous differential equation shown in Eq.(1). In the experiment, 20 points data of each variable, which calculated target equation using the Runge-Kutta method on the sampling point 0.5 ($0 \leq t < 10$), was given as a data series. The one set of the data series was given to our system, and it experimented by 50 kinds of random seed.

$$\begin{cases} dX_0/dt = -1.2X_0 + 0.6X_1 + 0.2 \\ dX_1/dt = 0.5X_0X_1 - 1.6X_1 + 1.2 \\ dX_2/dt = 2.0X_0 + 1.7X_1 - X_2 \\ X_0(0) = 3, X_1(0) = 2, X_2(0) = 1 \end{cases} \quad (1)$$

As a result, the fitness reached a definition value in 46 kinds of random seed. In all success case, obtained equations are similar to an target equation. The obtained equation of best case (highest fitness) are as follows:

$$\begin{aligned} dX_0/dt &= -1.21X_0 + 0.59X_1 + 0.19 \\ dX_1/dt &= 0.55X_0X_1 - 1.64X_1 + 1.17 \\ dX_2/dt &= 1.96X_0 + 1.73X_1 - 1.02X_2 + 0.11 \end{aligned} \quad (2)$$

In case of our previous work (Mass-action + GP only), the number of optimization success were 9 (one data set) and 41 (four data sets). In case of proposal method, the number of optimization success was 46 (one data set). This result is almost the same as the result of our previous work in four data sets. Our proposal method had efficiency and accuracy that was more excellent than our previous work.

Acknowledgment

This study was partly supported by “Establishment of Consolidated Research Institute for Advanced Science and Medical Care”, Encouraging Development Strategic Research Centers Program, and Grant-in-Aid for Young Scientists (B) (No. 20700144) from the Ministry of Education, Culture, Sports, Science and Technology, Japan.

References

- [1] Savageau, M.A., *Biochemical Systems Analysis: A Study of Function and Design in Molecular Biology*, Addison-Wesley, 1976.
- [2] Ueda, T., Ono, I., Okamoto, M., Development of System Identification Technique Based on Real-Coded Genetic Algorithm, *Genome Informatics*, 13:386–387, 2002.
- [3] Tanaka, K. and Yamana, H., The Development and Evaluation of a Prototype System for the Inference of Genetic Networks Based on Genetic Programming, *The 11th World Multi-Conference on Systemics, Cybernetics and Informatics (WMSCI2007) Proceedings*, 4:13–18, 2007.