

Pathway Finding from Given Time-Courses Using Genetic Algorithm

Shinichi Kikuchi^{1,2} Daisuke Tominaga¹
 shinichi.kikuchi@aist.go.jp tominaga-daisuke@aist.go.jp
 Masanori Arita^{1,2} Masaru Tomita²
 m-arita@aist.go.jp mt@sfc.keio.ac.jp

- ¹ CBRC, National Institute of Advanced Industrial Science and Technology (AIST),
 2-41-6, Aomi 135-0064, Japan
² Institute for Advanced Biosciences, Keio University, 14-1, Baba 997-0035, Japan

Keywords: genetic networks, pathway finding, S-system, genetic algorithm

1 Introduction

We have developed an efficient algorithm based on the Genetic Algorithm (GA) for estimating parameters of S-system [4]. The S-system belongs to power-law formalism, and is rich enough in structure to capture all relevant dynamics. For example, it is known that S-system can express genetic networks or metabolic pathways.

In this work, we propose a method which can not only approximate time-courses but also find skeletal pathways. Our method is adopted by E-CELL project [5].

2 Method and Results

2.1 S-system

S-system [2] is a type of power-law formalism. It is based on a particular type of ordinary differential equation in which the component process are characterized by power-law functions;

$$\frac{dX_i}{dt} = \alpha_i \prod_{j=1}^n X_j^{g_{ij}} - \beta_i \prod_{j=1}^n X_j^{h_{ij}} \quad (1)$$

where n is the number of state variables or reactants X_i , i, j ($1 \leq i, j \leq n$) are suffixes of state variables. The terms g_{ij} and h_{ij} are interactive effectivity of X_j to X_i . In a biological engineering context, the non-negative parameters α_i and β_i are called rate constants, and real valued exponents g_{ij} and h_{ij} are referred to as kinetic orders.

It is known that S-system is rich enough in structure to capture all relevant dynamics. However, its formalism includes a large number of parameters that must be estimated (α_i , β_i , g_{ij} and h_{ij}). The number of estimated parameters in S-system formalism is $2n(n+1)$, where n is the number of state variables X_i .

2.2 Optimization using GA

We employ GA [1] for estimating parameters of S-system. For optimization problems, ease set of parameter values to be estimated is evaluated using the following procedure; Suppose that $X_{i,cal,t}$ is

numerically calculated time-course at time t of state variable X_i , and $X_{i,exp,t}$ represents the experimentally observed time-course at time t of X_i . Sum the relative error between $X_{i,cal,t}$ and $X_{i,exp,t}$ to get the error. Add the cost function of wasteful parameters to the error. Calculate the total error function E .

$$E = \sum_{i=1}^n \sum_{t=1}^T \left(\frac{X_{i,cal,t} - X_{i,exp,t}}{X_{i,exp,t}} \right)^2 + c \sum_{i=1}^n \sum_{j=1}^n (|g_{ij}| + |h_{ij}|) \quad (2)$$

where c is the weighted coefficient for the additional term. Using the second term of E , wasteful parameters become 0. That is to say, skeletal pathway can be discovered.

2.3 Experimental results

Our proposed method is applied to determine a set of parameters of the S-system which represents a typical model of a genetic network. As a case study, we create time-course data artificially, which is numerically calculated using the scheme [3]. Here, $n = 5$ and 38 out of 60 parameters are 0. As computer simulation results, wasteful parameters become 0 and skeletal network is obtained using 25 time-courses.

3 Discussions

Because cost function to obtain skeletal networks is employed, a priori information that genetic networks are sparse combination can be used for parameter estimation. Using the function, our method can not only obtain skeletal structures but also raise convergence success rate and learning time. It is thought that this technique will be effective in pathway prediction.

Acknowledgments

This work was supported in part by Japan Science and Technology Corporation. And this method is developed mainly for E-Neuron project, Keio University, which is a neuron simulation project using E-CELL system. We thank E-Neuron project members.

References

- [1] Davis, L., *Handbook of Genetic Algorithms*, Van Nostrand Reinhold, 1991.
- [2] Savageau, M.A., *Biochemical System Analysis : A Study of Function and Design in Molecular Biology*, Addison-Wesley, 1976.
- [3] Savageau, M.A., Rules for the evolution of gene circuitry, *Proc. of the Pacific Symposium on Biocomputing*, 54–65, 1998.
- [4] Tominaga, D., Okamoto, M., Maki, Y., Watanabe, S., and Eguchi, Y., Nonlinear numerical optimization technique based on genetic algorithm for inverse problem: towards the inference of genetic networks, *Proc. of the German Conference on Bioinformatics*, 127–140, 1999.
- [5] Tomita, M., Hashimoto, K., Takahashi, K., Shimizu, T.S., Matsuzaki, Y., Miyoshi, F., Saito, K., Tanida, S., Yugi, K., Venter, J.C., and Hutchison, C.A., E-CELL: software environment for whole-cell simulation, *Bioinformatics*, 15(1):72–84, 1999.