Inference of Nonlinear Biological Systems by Using Linear Programming

Tatsuya Akutsu[1] **Satoru Miyano**[1] **Satoru Kuhara**[2]

takutsu@ims.u-tokyo.ac.jp miyano@ims.u-tokyo.ac.jp kuhara@grt.kyushu-u.ac.jp

[1] Human Genome Center, Institute of Medical Science, University of Tokyo, 4-6-1 Shirokanedai, Minato-ku, Tokyo 108-8639, Japan

[2] Graduate School of Genetic Resources Technology, Kyushu University, 6-10-1 Hakozaki, Higashi-ku, Fukuoka 812-8581, Japan

Keywords: genetic network, DNA microarray, linear programming, S-system

1 Introduction

Recently, many studies have been done in order to develop computational methods for reconstructing underlying *genetic networks* from time series data of gene expression patterns. We have been studying inference algorithms for genetic networks based on the Boolean network model [1, 2, 3]. However, the Boolean network may not be adequate for modeling all kinds of gene regulation mechanisms. Therefore, we have been studying other models based on differential equations. In particular, we have been studying a model based on the S-system (*synergistic* and *saturable* system), where the S-system consists of nonlinear differential equations of a special form. S-systems have been successfully applied to the analysis of various biological systems [4]. Recently, Tominaga and Okamoto [6] applied GA (Genetic Algorithm) to inference of S-systems. However, their method was time consuming and was limited to inference of S-systems with a few parameters.

In this short article, we briefly present a novel method [2] for inferring S-systems from time series data. In this method, each instance of the inference problem is reduced to a linear program by means of focusing on the signs of differentials (or finite differences).

2 Inference of S-systems

First we briefly review the definition of the S-system [4, 6]. Let $\{X_1, \ldots, X_n\}$ be a set of genes and/or chemical substances in the underlying biological network. Let $X_i(t)$ be the value (expression level or concentration) of a gene or a chemical substance X_i at time t.

An S-system is a set of *nonlinear* differential equations of the form

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

where α_i and β_i are multiplicative parameters called *rate constants* and $g_{i,j}$ and $h_{i,j}$ are exponential parameters called *kinetic orders*.

The inference problem is, given time series data $X_i(t)$ that are assumed to be generated from an S-system $\mathcal{S}$, to estimate to parameters α_i, β_i, $g_{i,j}$ and $h_{i,j}$ of $\mathcal{S}$.

The proposed inference method is quite simple. Assume that $\frac{dX_i(t)}{dt} > 0$ at time t. By taking 'log' of each side of $\alpha_i \prod X_j(t)^{g_{i,j}} > \beta_i \prod X_j(t)^{h_{i,j}}$, we have

$$\log \alpha_i + \sum_{j=1}^{n} g_{i,j} \log X_j(t) > \log \beta_i + \sum_{j=1}^{n} h_{i,j} \log X_j(t) .$$

Since $X_j(t)$'s are known data, this inequality is linear if we treat $\log \alpha_i$'s and $\log \beta_i$'s as parameters. In the case of $\frac{dX_i(t)}{dt} < 0$, we can obtain a similar inequality. Therefore, solving these linear inequalities by LP (*linear programming*), we can determine parameters (we call this method the LP-based method).

However, parameters are not determined uniquely even if a lot of data are given, because the inequality can be re-written as $(\log \alpha_i - \log \beta_i) + \sum_{j=1}^{n} (g_{i,j} - h_{i,j}) \log X_j(t) > 0$. Therefore, only relative ratios of $\log \alpha_i - \log \beta_i$ and $g_{i,j} - h_{i,j}$'s are determined (for each i). But, this information is useful for qualitative understanding of S-systems. Since $\prod X_j(t)^{g_{i,j}}$ contributes to the net production of X_i, $\prod X_j(t)^{h_{i,j}}$ contributes to the net degradation of X_i and it is not usual that X_j contributes to both the net production and the net degradation, either $g_{i,j} = 0$ or $h_{i,j} = 0$ holds for each (i, j) in most cases. Thus, the fact that $|g_{i,j} - h_{i,j}|$ is large means that X_i is influenced by X_j.

3 Conclusion

The proposed LP-based method is much faster than the GA-based: it took less than one second in order to compute parameters for an S-system with two nodes, whereas the GA-based method took about two hours. Moreover, the LP-based method could infer most of qualitative relations in an S-system with 100 nodes (i.e., $n = 100$) within 6 hours on a SUN ULTRA-2 Workstation with 1 CPU (296MHz), where we employed SOPT [5] in order to solve LP. These results show that the LP-based method is efficient. For details of simulation results, see [2].

However, the LP-based method requires many time series data beginning from different sets of initial values, where different sets correspond to different environments or different conditions. Therefore, the LP-based method is superior to the GA-based method only when a large amount of time series data are available.

Another drawback of the LP-based method is that complex enzymatic reactions (for example, three-stage enzymatic reactions) can not be handled: these reactions can not be represented in the form of the S-system. Therefore, development of methods to infer complex enzymatic reactions is important future work.

References

[1] Akutsu, T., Miyano, S. and Kuhara, S., Identification of genetic networks from a small number of gene expression patterns under the boolean network model, *Proc. Pacific Symp. Biocomputing'99*, World Scientific, 17–28, 2000.

[2] Akutsu, T., Miyano, S. and Kuhara, S., Algorithms for inferring qualitative models of biological networks, *Proc. Pacific Symp. Biocomputing 2000*, World Scientific, 290–301, 2000.

[3] Akutsu, T., Miyano, S. and Kuhara, S., Algorithms for identifying Boolean networks and related biological networks based on matrix multiplication and fingerprint function, *Proc. RECOMB 2000*, ACM Press, 2000.

[4] Irvine, D.H. and Savageau, M.A., Efficient solution of nonlinear ordinary differential equations expressed in S-system canonical form, *SIAM J. Numer. Anal.*, 27:704–735, 1990.

[5] SAITECH Inc., *Smart Optimizer User's Guide*, 1998.

[6] Tominaga, D. and Okamoto, M., Design of canonical model describing complex nonlinear dynamics, *Proc. IFAC Int. Conf.*, CAB7:85–90, 1998.