

Interactive Evolution of a Gene Regulatory Network Model

Atsushi Mimura¹

amimura@miv.t.u-tokyo.ac.jp

Hitoshi Iba²

iba@miv.t.u-tokyo.ac.jp

¹ Graduate School of Engineering, The University of Tokyo, Hongo 7-3-1 Bunkyo-ku, Tokyo, Japan

² Graduate School of Frontier Sciences, The University of Tokyo, Hongo 7-3-1 Bunkyo-ku, Tokyo, Japan

Keywords: gene regulatory network, genetic algorithms, interactive evolutionary computing

1 Introduction

Inferring a gene regulatory network from gene expression data obtained by DNA microarray is considered as one of the most challenging problems in the field of bioinformatics [1, 2].

To infer a network efficiently while avoiding numerous local solutions, human intervention such as employing heuristics and adding sufficient search restriction or data is necessary. Hence, a computer can be used interactively for the inference – this seems to be an effective approach instead of conventional top-down approaches.

2 Method

When inferring a network, Genetic Algorithms (GAs) optimize model parameters. In this optimization, several population groups are generated and the parameters for optimization in each group are determined and the other parameters consistently fixed at zero.

The best individuals are picked up from each group after the specified generations. Individuals with better fitness are reserved as current solution candidates, replacing poor candidates. This process is executed until some terminal condition is satisfied.

Fitness value of each individual is defined as follows:

$$F = \sum_{c=0}^{\#c} \sum_{t=0}^{\#DP} (x_{c,i(exp)}(t) - x_{c,i(calc)}(t))^2, \quad (1)$$

where $x_{c,i}(t)$ is expression level at t under the condition of c (e.g., disruption of gene X_j).

Then, the score $S_{i,j} \in \mathbf{S}_i$, i.e., the reliability of the effect of gene X_j ($j = 1, 2, \dots, n$) on search node X_i , is calculated as follows:

$$S_{i,j} = \frac{Max(\mathbf{V}) - Avr(\mathbf{V})}{\sum V - Avr(\mathbf{V})}, \quad \text{and} \quad V_0 = \sum_{j=1}^{\#Sol} \left(\frac{1 - \{F_j | M_{i,j} = 0\}}{\sum_{l=1}^{\#Sol} F_l} \right), \quad (2)$$

where $\mathbf{V} = \{V_0, V_+, V_-\}$.

The system requests expression data or search restriction concerning the relationship between gene X_j and gene X_i with the lowest score.

3 Experimental Results

We used randomly generated networks composed of 10 nodes and 20 edges for the inference targets and used a model based on S-System [3] with the parameters restricted below:

$$\frac{dx_i}{dt} = a \prod_{j=1}^n x_j^{M_{ij}} - bx_i \quad (i = 1, 2, \dots, n), \quad (3)$$

where values of the influence matrix $\mathbf{M}[\mathbf{n}, \mathbf{n}]$ are real number in the range of $[-2.0, 2.0]$. a and b are set to be 1.0 for simplicity. It is assumed that a wild-type expression data are available for each node. Each time series is composed of 10 points with 0.5 time intervals and initial value is consistently set to be 0.10. Data points were interpolated by the spline method for accurate calculation.

For each node X_i ($i = 1, 2, \dots, 10$), we applied the following process:

Process1. After applying the GA algorithm, calculate $S_{i,j}$ using the parameters of solution candidates of which the fitness value is less than θ times of that of the best one (maximum 5 candidates).

Process2. If $\text{Min}(\mathbf{S}_i) = S_{i,j} \leq \sigma$ ($0 \leq \sigma \leq 1$), give the search restriction from true control relationship of X_j over X_i (positive, negative or none) as well as expression data of gene X_i when X_j is disrupted. Both of them are assumed to be available here. Then go back to Process1.

Process3. If $\sigma < \text{Min}(\mathbf{S}_i)$, the regulatory relationship as the final solution of X_j over X_i is determined by using the values of $\text{Max}(\mathbf{V})$ (positive, negative or none).

Table 1: GA parameters.

# of Population Groups	5
# of Generation of New Structures	5
# of Individuals in A Population	300
# of Generation	60
Crossover Rate	0.90
Mutation Rate / Elite Rate	0.02 / 0.01
Search Range	$[-2.0, 2.0]$
θ / σ Values	3 / 0.80

Table 2: Results using randomly generated network RR: Amount of required restriction/data, ST: Search time (Min.) by Pentium4 1.4GHz with 512MB Memory.

Approach	SN/SP	RR	ST
Single Population	0.600/0.769	-	74
Giving Random Data	0.825/0.863	19	123
Proposed Method	0.925/0.944	12	71

Table 2 shows the results. GA and other parameters are shown in Table.1. The result by the proposed method shows that adding the information in accordance with the requirements leads to inferring more accurate network structures with the small amount of data.

4 Conclusion

We proposed an interactive approach to infer a gene regulatory network and showed the possibility of the proposed method in achieving an efficient network inference with a very little amount of the expression data. The study will be continued with applying actual expression data with the consideration of other appropriate network models.

References

- [1] Ando, S. and Iba, H., Inference of gene regulatory model by genetic algorithms, *Proc. Congress on Evolutionary Computation*, 712–719, 2001.
- [2] Maki, Y. *et al.*, Development of a system for the inference of large scale genetic networks, *Proc. Pacific Symp. on Biocomputing '01*, World Scientific, 6:446–448, 2001.
- [3] Savageau, M.A., Rules for the evolution of gene circuitry, *Proc. Pacific Symp. on Biocomputing '98*, World Scientific, 54–65, 1998.