Proceedings of the 28th IEEE
EMBS Annual International Conference
New York City, USA, Aug 30-Sept 3, 2006

Inference of Genetic Network of Xenopus Frog Egg : Improved Genetic Algorithm

Shinq-Jen Wu [1], Chia-Hsien Chou*, Cheng-Tao Wu , and Tsu-Tian Lee*

Department of Electrical Engineering Da-Yeh University, Chang-Hwa, Taiwan, R.O.C
*Department of Electrical and Control Engineering National Chiao-Tung University Hsin-Chu, Tainwan, R.O.C.
[1] jen@cn.nctu.edu.tw, dau@cn.nctu.edu.tw, u9312623@cn.nctu.edu.tw, ttlee@cn.nctu.edu.tw

Abstract – *An improved genetic algorithm (IGA) is proposed to achieve S-system gene network modeling of Xenopus frog egg. Via the time-courses training datasets from Michaelis-Menten model, the optimal parameters are learned. The S-system can clearly describe activative and inhibitory interaction between genes as generating and consuming process. We concern the mitotic control in cell-cycle of Xenopus frog egg to realize cyclin-Cdc2 and Cdc25 for MPF activity. The proposed IGA can achieve global search with migration and keep the best chromosome with elitism operation. The generated gene regulatory networks can provide biological researchers for further experiments in Xenopus frog egg cell cycle control.*

Keywords*: IGA, Michaelis-Menten model, S-system, Xenopus frog egg cell cycle*

1. Introduction

As the rapid development in cDNA microarray technologies, the time-course gene data becomes available day by day. And hence, the construction of gene networks and signal transduction cascades for complicated biological systems has come of age. With the mathematical model, we can realize the genes-genes interaction, simulate the gene network without experiment and hence predict the gene behavior.

Numerous models are proposed to describe the gene network such as Boolean network, Bayesian network, Michaelis-Menten model, and S-system. Boolean network is to reconstruct gene regulatory network via Boolean function and express gene relationship in graphical way [1, 2], which distinguish gene states to be INPUT and OUTPUT. At any time points, the state values of chosen INPUT genes are set to be 1 and 0 for non-chosen genes; the states values are given in the similar way as OUTPUT. Further, Bayesian network can also the probabilistic relationships of genes [2, 3]; joint probability distributions among genes are calculated to construct the graphical model. Michaelis-Menten model is nonlinear differential equations to describe the metabolic concentration in the biological system. S-system is another nonlinear differential type expressed in power-law formalism [4, 5]. S-system describes gene regulation not only in mathematical description but also can further express into graphical form to show the activatory and inhibitory operation directly. Each equation is composed by synthesis and degradation flux; and the activation and inhibition relationship are shown in positive and negative kinetic order, respectively.

In these years, some researchers are devoted to infer gene regulatory network with various intelligent computation technologies such as hybrid differential evolution, genetic algorithm, genetic programming, ..., etc. Wang use Hybrid differential evolution and genetic algorithm to obtain the global optimal solution for highly nonlinear system and various biochemical system [6, 7, 8]. Kikuchi and coauthors use a genetic algorithm to transform parameters into individuals first and solve optimal parameters via evolution procedure [9]. Sakamoto and coauthors use genetic programming to develop the gene regulatory network in a tree form [10]. In this work, we shall adopt improved genetic algorithm (IGA) to infer the gene regulatory network of the Xenopus frog egg cell cycle in S-system. Improved evolutionary direction operator (IEDO), migration operation and elitism are combined into genetic program for global optimal, fast and best-optional searching. The input/output datasets, generated from Michaelis-Menten model of mitotic control in Xenopus frog eggs [11], is used to train the genetic network for searching the optimal parameters of the corresponding S-system model.

This paper is organized as follows. IGA algorithm is shown in Section 2. Section 3 is the construction of the cell cycle control model of the Xenopus frog egg. Section 4 is the conclusion.

2. IGA Learning Algorithm

Based on general genetic algorithm, IGA includes improved evolutionary direction operator to speed the searching; migration operator to escape from bogging down into local solution and elitism to keeping the best to be passed down always. The follows are the main procedures.

Step 1. Initialize a population.
Step 2. Individuals fitness
Step 3. IEDO
Step 4. Crossover operation
Step 5. Probability mutation and Elitism operations

Acknowledgement: This research is supported by the National Science Council of the R.O.C. under Grant NSC 94-2213-E-009 -124 -

[1] To whom all correspondence should be addressed.

Step 7. Migration operation
Step 8. Repeat to Step 3.

2.1 IEDO

Three preferred fitness value and its associated individuals are chosen to decide the evolutionary direction in the population. IEDO has the ability for both local and global search synchronously. We can use the IEDO operator to search quickly and converge toward the global optimal solution. Besides, the IEDO operator can avoid bogging down in the local optimal solution, which conventional GA is usually stuck into.

After completing all fitness values of individuals, we choose preferred fitness values denote F_b, F_s, F_t and its associated individuals denote I_b, I_s, I_t. And then, the new individual denote I_{ideo} is calculated as

$$I_{iedo} = I_b + r_1 * D_1 * (I_b - I_s) + r_2 * D_2 * (I_b - I_t), \quad (1)$$

$$I_{iedo} = max(min(I_{iedo}, I_{max}), I_{min}), \quad (2)$$

where r_1 and r_2 are two random numbers, r_1, $r_2 \in [0, 1]$; D_1 and D_2 are the magnitude of two evolutionary directions to be 1; $I_{\max}$ and $I_{\min}$ are the upper and lower bound, respectively. The new fitness value F_{new} of the I_{ideo} is calculated; if the F_{new} is better than one of the three preferred fitness values, it would be replaced that.

2.2 Reproduction

The probability of reproduction directly depends on the fitness of the individuals. The individual with better fitness has high probability of reproduction. In contrast, the individual with worse fitness has low probability of reproduction.

2.3 Crossover

Two-point crossover operation is adopted and operates according to crossover probability, which involves selection of two crossover cut-points randomly and then exchanges the chosen two cut-points genes of parent individuals to generate two child individuals. And further, randomly select one of the new child to replace father or mother individuals.

2.4 Probability Mutation and Elitism

Different from the conventional GA to choose only one gene in individual randomly for mutation operation, all genes in individuals are chosen and their mutation probability are assigned by the designer for exchanging their original values in IGA. This will bring excessive diversity in population and hence may fail to converge to temperately optimal solution. Therefore, we adopt elitism operator to decrease this effect. Elitism operator is to keep the best individual to survive for each generation and hence to ensure good characteristic to pass down always.

2.5 Migration

To wider the search space, a migration operator is done to get a new and diverse population. The degree of population diversity η is to check if the migration should be performed.

$$temp_{ij} = \begin{cases} 0, \text{ if } \left| \frac{x_{ij} - x_{bj}}{x_{bj}} \right| < \varepsilon_2 \\ 1, \text{ otherwise} \end{cases}, \quad (3)$$

$$\eta = \sum_{i=1}^{NP-1} \sum_{j=1}^{Dim\_I} temp_{ij} \Big/ Dim\_I \times (NP-1), \quad (4)$$

where $\varepsilon_2 \in [0,1]$ is the tolerance of the real-valued gene diversity; x_{ij} and x_{bj} are respectively the j-th chromosome in the i-th individual and the best individual; NP is the number of individual; Dim_I is the dimension of individual;ηis in the range between 0 and 1. $\varepsilon_1 \in [0,1]$ is a tolerance-threshold of population diversity for migration; if ηis small than ε_1, migration operate to generate a new chromosome as follows.

$$x_{ij} = \begin{cases} x_{bj} + r_2 \times \left(x_{j,\min} - x_{bj} \right), \text{ if } \frac{x_{bj} - x_{j,\min}}{x_{j,\max} - x_{j,\min}} > r_1 \\ x_{bj} + r_2 \times \left(x_{j,\max} - x_{bj} \right), \quad \text{otherwise} \end{cases}, \quad (5)$$

where $x_{j,\max}$ and $x_{j,\min}$ are the upper and lower bound of the j-th chromosome, respectively. The r_1 and r_2 are two random numbers, r_1, $r_2 \in [0, 1]$.

2.6 Fitness

Every individual is evaluated by its fitness value, which keeps the better individuals and eliminates the worse individuals. The fitness of a individual is defined as

$$fitness = \frac{\sum_{i=1}^{Dim\_I} \sum_{j=1}^{N-1} \left(X_{ei}(t_0 + j\Delta t) - X_i(t_0 + j\Delta t) \right)^2}{Dim\_I \times N}, \quad (6)$$

where N is the number of time-points; X_{ei} and X_i are the experiment value and the estimated value of i-th reactant, respectively.

3. Gene regulatory network modeling

Michaelis-Menten model [11] are concerned to describe the mitotic control in cell-cycle of Xenopus frog egg,

$$\dot{x}_1 = k_1 - k_2 x_1 - k_3 x_1, \quad (7)$$

$$\dot{x}_2 = k_{pp} x_5 - \left(k_{wee} + k_{cak} + k_2 \right) x_2 + k_{25} x_3 + k_3 x_1, \quad (8)$$

$$\dot{x}_3 = k_{wee} x_2 - \left(k_{25} + k_{cak} + k_2 \right) x_3 + k_{pp} x_4, \quad (9)$$

$$\dot{x}_4 = k_{wee} x_5 - \left(k_{pp} + k_{25} + k_2 \right) x_4 + k_{cak} x_3, \quad (10)$$

$$\dot{x}_5 = k_{cak} x_2 - \left(k_{pp} + k_{wee} + k_2 \right) x_5 + k_{25} x_4, \quad (11)$$

$$\dot{x}_6 = \frac{k_a x_5 (1 - x_6)}{1 + K_a - x_6} - \frac{k_b x_6}{K_b + x_6}, \quad (12)$$

$$\dot{x}_7 = \frac{k_e x_5 (1 - x_7)}{1 + K_e - x_7} - \frac{k_f x_7}{K_f + x_7}, \quad (13)$$

$$\dot{x}_8 = \frac{k_g x_5 (1 - x_8)}{1 + K_g - x_8} - \frac{k_h x_8}{K_h + x_8}, \quad (14)$$

$$\dot{x}_9 = \frac{k_c x_8 (1 - x_9)}{1 + K_c - x_9} - \frac{k_d x_9}{K_d + x_9}, \quad (15)$$

with

$$k_2 = V_2' + x_9\left(V_2'' - V_2'\right),$$
$$k_{wee} = V_{wee}'' + x_7\left(V_{wee}' - V_{wee}''\right),$$
$$k_{25} = V_{25}' + x_6\left(V_{25}'' - V_{25}'\right),$$

where x_i, i = 1, 2, ···, 9, are the concentrations or activities of cyclin, unphosphorylated cyclin-Cdc2, Tyr-15 phosphorylated cyclin-Cdc2, doubly phosphorylated cyclin-Cdc2, Thr-161 phosphorylated cyclin-Cdc2 activated by MPF, Cdc25 enzyme, Wee1 enzyme, IE enzyme and APC enzyme, respectively. Eqs. (8) ~ (11) describe four phosphorylation states of the cyclin-Cdc2 dimer.

We shall generate the training dataset from the above equations to generate the corresponding S-system model of the frog cell cycle to further realize the gene-gene inhibitory and activatory operation for gene and enzyme synthesis and decomposition. S-system use power-law flux to describe the synergism and saturation of the biological system,

$$\dot{x}_i = \alpha_i \prod_{j=1}^{n} x_j^{g_{ij}} - \beta_i \prod_{j=1}^{n} x_j^{h_{ij}}, \quad \text{for } i = 1, 2, ..., n, \tag{16}$$

where x_i is the state variable or reactant; n is the number of x_i. α_i is the production rate-constant and β_i is the degradation rate- constant; both can be positive or zero. g_{ij} and h_{ij}, are kinetic orders; their values can be positive to indicate activating influences or negative to denote inhibition. We now construct our S-system structure for Xenopus frog egg as

$$\dot{x}_1 = \alpha_1 x_1^{g_{11}} x_9^{g_{19}} - \beta_1 x_1^{h_{11}} x_9^{h_{19}},$$
$$\dot{x}_2 = \alpha_2 x_1^{g_{21}} x_2^{g_{22}} x_3^{g_{23}} x_5^{g_{25}} x_6^{g_{26}} x_7^{g_{27}} x_9^{g_{29}} - \beta_2 x_2^{h_{22}} x_3^{h_{23}} x_6^{h_{26}} x_7^{h_{27}} x_9^{h_{29}},$$
$$\dot{x}_3 = \alpha_3 x_2^{g_{32}} x_3^{g_{33}} x_4^{g_{34}} x_6^{g_{36}} x_7^{g_{37}} x_9^{g_{39}} - \beta_3 x_2^{h_{32}} x_3^{h_{33}} x_6^{h_{36}} x_7^{h_{37}} x_9^{h_{39}},$$
$$\dot{x}_4 = \alpha_4 x_3^{g_{43}} x_4^{g_{44}} x_5^{g_{45}} x_6^{g_{46}} x_7^{g_{47}} x_9^{g_{49}} - \beta_4 x_4^{h_{44}} x_5^{h_{45}} x_6^{h_{46}} x_7^{h_{47}} x_9^{h_{49}},$$
$$\dot{x}_5 = \alpha_5 x_2^{g_{52}} x_4^{g_{54}} x_5^{g_{55}} x_6^{g_{56}} x_7^{g_{57}} x_9^{g_{59}} - \beta_5 x_4^{h_{54}} x_5^{h_{55}} x_6^{h_{56}} x_7^{h_{57}} x_9^{h_{59}},$$
$$\dot{x}_6 = \alpha_6 x_5^{g_{65}} x_6^{g_{66}} - \beta_6 x_5^{h_{65}} x_6^{h_{66}},$$
$$\dot{x}_7 = \alpha_7 x_5^{g_{75}} x_7^{g_{77}} - \beta_7 x_5^{h_{75}} x_7^{h_{77}},$$
$$\dot{x}_8 = \alpha_8 x_5^{g_{85}} x_8^{g_{88}} - \beta_8 x_5^{h_{85}} x_8^{h_{88}},$$
$$\dot{x}_9 = \alpha_9 x_8^{g_{98}} x_9^{g_{99}} - \beta_9 x_8^{h_{98}} x_9^{h_{99}}, \tag{17}$$

Note that since concentrations of x_1, x_2, x_3, x_4 and x_5 are too small as compared to other variables. The scale-up operation is adopted to normalize all states variables to a computation reasonable range to improve the computation error. Another test data from Michaelis-Menten model is used to demonstrate the performance of the IGA program, Figure 1 is the simulation results with the estimated fitness 3.0766122E-08 for N=80,000, Dim_I=83. The low fitness value ensures the good fitting of the simulation results with the datasets and also guarantees the reliability of the generated S-system. From the constructed S model, we can realize the interaction between various genes in Xenopus frog egg. For instance, the concentration of x_2 increases rapidly as concentrations of x_1, x_3 and x_5 increase; the concentration of x_2 decreases rapidly as concentrations of x_7 and x_9 increase.

Conclusion

IGA technique is to construct S-system model for Xenopus frog egg cell cycle. The time-course training and test dataset are generated from Michaelis-Menten metabolic model of mitotic cell-cycle control of Xenopus frog egg. The proposed gene regulatory network reveals activatory and inhibitory operations for gene/enzyme synthesis and decomposition. Hence, the network can provide biological researchers for further experiments in Xenopus frog egg cell cycle control.

Acknowledgements

The authors thank to Dr. Feng-Sheng Wang for suggestion of genetic technology. Dr. Wang is a professor at the department of Chemical Engineering in National Chung Chen University. His researches are in the systems biology.

References

[1] Akutsu, T., Miyano, S., Kuhara, S., "Identification of genetic networks from a small number of gene expression patterns under the Boolean network model", In: Altman, R.B., Lauderdale, K., Dunker, A.K., Hunter, L., Klein, T.E. (Eds.), *Proceedings of the Pacific Symposium Biocomputation*, Vol. 4, World Scientific, Singapore, pp. 17–28, 1999.

[2] de Jong, H., "Modeling and simulation of genetic regulatory systems: a literature review", *J. Comput. Biol*, Vol. 9, pp. 67–103, 2002.

[3] Pe'er, D., Regev, A., Elidan, G., Fridman, N., Inferring subnetworks from expression profiles. *Bioinformatics*, Vol. 17, 215–224, 2001.

[4] Savageau, M.A., *Biochemical System Analysis : A Study of Function and Design in Molecular Biology*, Addison-Wesley, 1976.

[5] Voit, E.O., *Computational Analysis of Biochemical System*, Cambridge University Press, 2000

[6] Tsai, K.Y. and Wang, F.S., "Evolutionary Optimization with Data Collocation for Reverse Engineering of Biological Networks", *Bioinformatics*, Vol.21, p.p. 1180-1188, 2005.

[7] Chiang, C.L., Su, C.T. and Wang, F.S., "Augmented Lagrangian Method for Evolutionary Optimization of Mixed-Integer Nonlinear Constrained Problems", *Intern. Math. Journal*, vol.2, no.2, pp.119-154, 2002.

[8] Chiou, J. P. and Wang, F. S., "A hybrid method of differential evolutionary with application to optimal control problems of a bioprocess system", *Proceedings 1998 IEEE on Evolutionary Computation Conference*, pp.627-632.

[9] Kikuchi, S., Tominaga, D., Arita, M., Takahashi, K., Tomita, M., "Dynamic modeling of genetic networks

using genetic algorithm and S-system", *Bioinformatics*, Vol.19, p.p. 643-650, 2003.

[10] Sakamoto, E. and Iba, H., "Inferring a System of Differential Equations for a Gene Regulatory Network by using Genetic Programming", *Proceedings of the Evolutionary Computation Congress*, Vol.1, p.p. 720-726, 2001.

[11] Schmidt, H. and Jacobsen E.W., "Linear systems approach to analysis of complex dynamic behaviors in biochemical networks" *IEE Syst. Biol.*, vol.1, no.1, pp.149-158, June 2004.

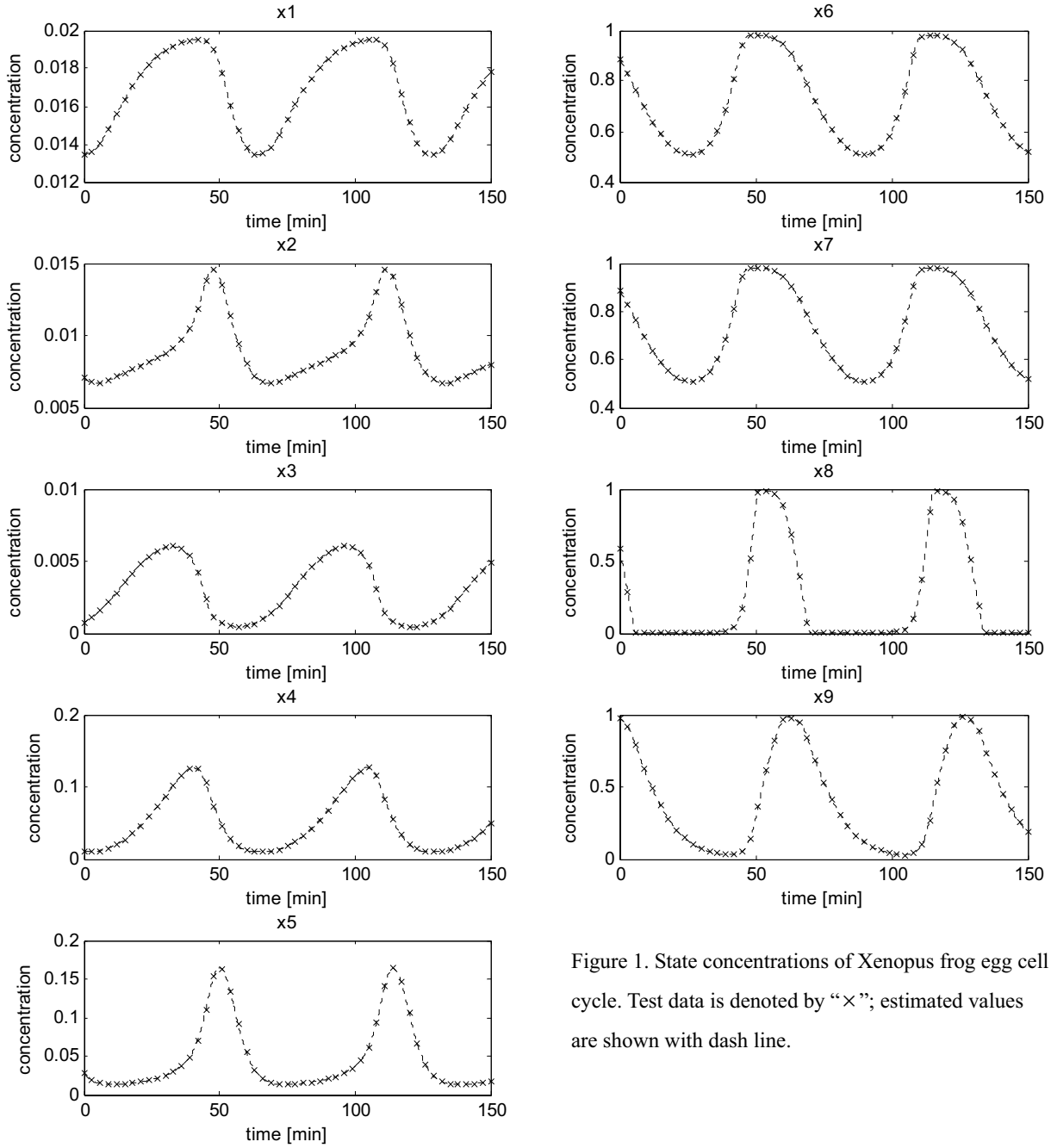

Figure 1. State concentrations of Xenopus frog egg cell cycle. Test data is denoted by "×"; estimated values are shown with dash line.