2008－BIO－14 (3)
2008／9／18

Two-phase search (TPS) 法: 動的生化学ネットワークモデルの偏りのない効率的なパラメータ探索

前田和勲[1]、倉田博之[2]
[1]九州工業大学大学院情報工学府
[2]九州工業大学大学院情報工学研究院

ダイナミックシミュレーションは環境ストレスや遺伝的変化に対してロバストネスを生み出す生化学ネットワークのメカニズムの解明に不可欠である。しかし、これまでに行われた多くのシミュレーションや解析は特定の速度パラメータ値に依存している。速度パラメータの正確な測定は困難であるので、特定のパラメータ値に依存するのではなく、パラメータ値の変化が与える影響を考慮した解析が必要である。これを可能にするためには細胞内の挙動を再現しうる全ての速度パラメータ値を広い探索空間から偏りなく探し出さなければならない。本稿では、ランダム探索と進化的探索を組み合わせた新しいパラメータ探索法、Two-phase search 法を提案する。そして、この手法が偏りなく効率的に速度パラメータを探索できることを示す。

Two-phase search (TPS) method: Nonbiased and efficient parameter search for dynamic models of biochemical networks

Kazuhiro Maeda and Hiroyuki Kurata
Department of Bioscience and Bioinformatics, Kyushu Institute of Technology

Dynamic simulations are essential for understanding the mechanism of how biochemical networks generate robust properties to environmental stresses or genetic changes. However, typical dynamic modeling and analysis yield only local properties regarding a particular choice of plausible values of kinetic parameters. Global and firm analyses are needed that consider how the changes in parameter values affect the results. A typical solution is to systematically analyze the dynamic behaviors in large parameter space by searching all plausible parameter values without any biases. In this paper, we propose the two-phase search method that consists of a random search and an evolutionary search to effectively explore all possible solution vectors of kinetic parameters satisfying the target dynamics. It enables a nonbiased and high-speed parameter search for dynamic models of biochemical networks.

1 Introduction

Computer simulations enable one to capture dynamic behavior of complex biochemical networks. In principle, both molecular network architecture and the values of kinetic parameters determine the dynamic behavior of systems. In biology, molecular network structures are being built, but it is still hard to measure the accurate values of kinetic parameters *in vivo* due to experimental complexity. In many studies, a particular set of local kinetic parameters has been determined for convenience so that dynamic models reproduce target data. Thus, the simulated results sometimes depend on the values of kinetic parameters, or reflect only local view of the system. There have been only a few simulation methods that extensively investigated how a systematic change in the parameter values alters the prediction of dynamic behaviors [1, 2]. These random or systematic searches are a great step for approaching to global analysis, but they restricted the search space of parameters or the size of models due to calculation complexity.

To overcome the problems, we developed an efficient search algorithm, the two-phase search (TPS) method that smoothly combines a random search with an evolutionary search to achieve both nonbiased and high-speed searches. To demonstrate the feasibility of this method, we apply it to benchmark problems and reveal the performance of it in terms of the efficiency and solution distributions. Finally, the effectiveness of the proposed method is verified through the parameter search of the *E. coli* heat shock response model.

2 Methods

2.1 Numerical optimization for dynamic models

Generally a dynamic model for biochemical networks is formulated by differential-algebraic equations (DAEs). Numerical optimization for a dynamic model is used to estimate the values of kinetic parameters so that the model reproduces the behaviors of the existing experimental data [3]. A certain fitness function is necessary to characterize the degree to which the model reproduces the target experimental behaviors.

2.2 Two-phase search (TPS) method

Since biological data contain different types of errors, it is meaningless to seek a global minimum for the fitness function defined for a given dynamic model. The objective in this study is not to find such a global minimum, but to explore all possible plausible solutions of kinetic parameter vectors that produce the target dynamics.

The TPS method is proposed that combines a random search with a search by genetic algorithms (GAs), as shown in **Figure 1**. First, the random search explores a large parameter space without any biases to find a coarse solution showing a good fitness value. In this phase, it is not necessary to find any solutions providing lowest fitness values. Note that a lower value of fitness is better in this search. The resultant coarse solution is employed to generate the initial populations for the subsequent search by GAs. Second, after the initial population is created around the coarse solution vectors, use of GAs intensively searches all plausible solution vectors that show a low fitness value or provide the target features. This two-phase search is iterated to obtain the sufficient number of the plausible solutions. The *i*-th resultant solution vector of kinetic parameters $\mathbf{P}(i)$ is given by:

$$\mathbf{P}(i) = (p(i,1), p(i,2), p(i,3), \ldots\ldots, p(i,N)), \quad (1)$$

where $p(i, j)$ is the value of the *j*-th parameter of the *i*-th solution vector and N is the number of search parameters.

The TPS method has two critical control parameters: The allowable error for the coarse solution (AEC) obtained by a random search in the first phase and the region of initial population for the search by GAs (RIG) in the second phase. The end condition of the first phase search for a coarse solution is provided by:

$$Fitness < AEC. \quad (2)$$

The end condition of the second phase search for a plausible or final solution is given by:

$$Fitness < AE, \quad (3)$$

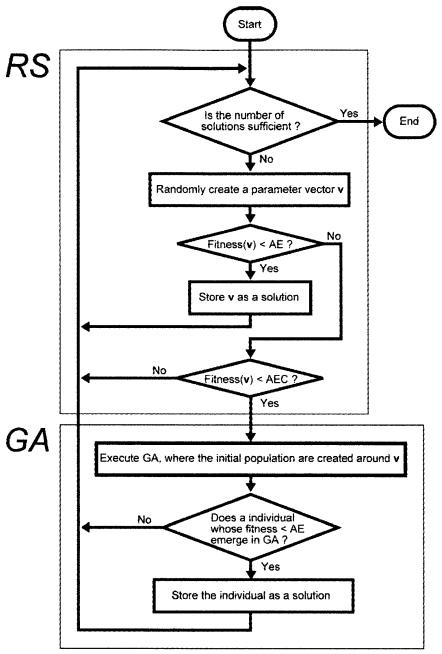

Figure 1: A flow chart for TPS that consists of a random search (the first phase) and a search by GAs (the second phase)

where AE is the allowable error of the plausible solution ($AE < AEC$). The initial population for the GA search is randomly generated within the hypercube whose length of the edges is RIG and whose centroid is the coarse solution in the first phase.

2.3 Characterization of the solutions

Three standards are defined to characterize the search results, the number of evaluation necessary for obtaining a given number of the final solutions (EVA), the centroid vector (CRV) and standard deviations vector (SDV) for the solutions. The CRV and SDV characterize the distribution profile for the parameter solution vector, is defined by:

$$\mathbf{CRV} = (c(1), c(2), \ldots, c(N)), \quad (4)$$

$$\mathbf{SDV} = (sd(1), sd(2), \ldots, sd(N)), \quad (5)$$

where,

$$c(j) = \frac{1}{M}\sum_{i=1}^{M} p(i,j), \quad (6)$$

$$sd(j) = \sqrt{\frac{1}{M}\sum_{i=1}^{M}(p(i,j) - c(j))^2}, \quad (7)$$

and M is the number of solutions. A small value of EVA indicates an efficient or high-speed search. A

Table 1: Benchmark functions

Objective function		Search region	Allowable error (AE)
Rosenbrock	$f(\mathbf{x})=\sum_{i=1}^{n-1}\left[100\left(x_{i+1}-x_i^2\right)^2+\left(x_i-1\right)^2\right]$	$-2.048<x_i<2.048$	0.676
ANFM	$f(\mathbf{x})=\left\lvert\frac{-x_2+\sqrt{x_2^{\,2}+4x_1x_2}}{2}-1\right\rvert$	$0.02<x_1<200$ $0.01<x_2<100$	0.0001

Table 2: Search performance by RS, SGA, and TPS
Presented data by TPS are best cases in our experiments.

		EVA	CRV		SDV	
			c(1)	c(2)	sd(1)	sd(2)
Rosenbrock	RS	9.57×10^{5}	8.71×10^{-1}	8.68×10^{-1}	3.29×10^{-1}	5.67×10^{-1}
	SGA	9.51×10^{5}	6.59×10^{-1}	5.37×10^{-1}	3.22×10^{-1}	4.82×10^{-1}
	TPS	8.96×10^{5}	8.61×10^{-1}	8.53×10^{-1}	3.33×10^{-1}	5.69×10^{-1}
ANFM	RS	3.09×10^{8}	1.51×10^{1}	7.45	2.26×10^{1}	1.76×10^{1}
	SGA	5.64×10^{6}	5.70	2.59	8.67	6.24
	TPS	2.37×10^{6}	1.50×10^{1}	7.70	2.25×10^{1}	1.78×10^{1}

search can be regarded nonbiased, when two standards of CRV and SDV are close to those in a random search.

2.4 Experiments

In order to demonstrate the feasibility of the TPS method, we designed a test problem. We applied it to five benchmark functions to search different solution vectors that give a smaller fitness value than a defined AE and investigated how the two control parameters, AEC and RIG, affect the search performance of EVA, CRV, and SDV. In **Table 1**, only two benchmark functions are shown. ANFM benchmark is our original function based on a typical biological model with an autogenous negative feedback model. The TPS method aims at both high-speed and nonbiased searches. Therefore, TPS is expected to achieve a smaller EVA than that by a random search and to provide the same CRV and SDV as those by a random search. Three types of the searches: a random search (RS), a search by GAs (SGA), and the TPS method, were iterated until the number of the solutions reaches to 10,000. RS and SGA were performed as controls. In SGA, one search was stopped when a solution was obtained or the search reached to the maximum generation. For the next search, the initial population was newly generated.

To verify the effectiveness of TPS, we applied it to parameter search of a dynamic model of the *E. coli* heat shock response [4]. Heat shock denatures or unfolds proteins, compromising cellular function. To counter heat shock, heat-shock proteins (hsps), chaperones and protease, are produced to refold the denatured proteins to their native state and to degrade them. The regulation of the synthesis, degradation, and activity of the σ^{32} factor plays a major role in heat shock response. We designed fitness function to capture behavior of σ^{32} concentration and refolding ability. In the heat shock response model 11 binding constants were searched. The searches are iterated until the number of solutions reaches to 1000.

3 Results and Discussion

3.1 TPS application to benchmark functions

The results for the Rosenbrock and ANFM benchmark functions are summarized in **Table 2**. In both Rosenbrock and ANFM, the CRV and SDV of SGA are different from those of RS, indicating that the solutions by SGA are biased. On the other hand, when AEC and RIG were well-designed, TPS provided the CRV and SDV close to those of RS. While the EVA value by the well-designed TPS was 94% of that of RS for Rosenbrock, it was 0.8% of that of RS for ANFM (**Table 2**). TPS showed a high efficiency for the ANFM function.

AEC is the critical control parameter that determines which search TPS becomes close to, RS or SGA. With the increase in AEC, the performance of TPS changes from RS to SGA. On the other hand, with an increase in RIG, the TPS method extends the region for the search in its second phase. The TPS method with a small AEC approaches to RS regard-

Table 3: EVA for the heat shock response model

	EVA
RS	2.45×10^5
SGA	1.75×10^5
TPS	9.22×10^4

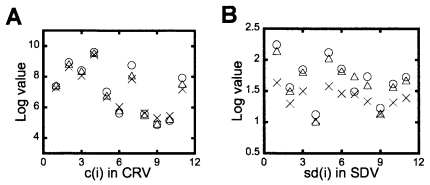

Figure 2: Solution distribution properties of three types of searches.
(A) CRV. (B) SDV. The searches were performed by RS (○), SGA (×), TPS (△).

less of the value of RIG. In this case, the value of EVA is large while nonbiased search is performed. When the value of AEC is not small, RIG affects the performance of TPS. When the values of AEC and RIG are adequately large, the performance of TPS is approximately the same as that of SGA. Thus, TPS achieves a small EVA, but the CRV and SDV are away from those of RS. The combination of a large AEC and a small RIG moves the TPS method far away from both RS and SGA. This search spends an enormous computational time, which is sometimes larger than that of RS and provides a different solution distribution from that of both RS and SGA.

3.2 Application to the heat shock response

The results are summarized in **Table 3** and **Figure 2**. The CRVs and SDVs were calculated in logarithmic space. The TPS method reduced the EVA value to 38% of that of RS and 53% of SGA (**Table 3**). While the CRVs were approximately the same for three searches (**Figure 2A**), the SDVs of SGA were smaller than those of RS (**Figure 2B**), indicating that SGA was biased. By contrast, the SDV distribution by the TPS method was approximately the same as that of RS, indicating that the solution distribution by TPS is non-biased. TPS carried out high-speed and nonbiased searches for the heat shock response.

The heat shock response model and the ANFM benchmark function are built based on molecular kinetics. The search results for both the biological models are much better in terms of calculation efficiency than those of the typical benchmark functions such as Rosenbrock. The landscape of the search space in both the biological models seems to differ from that of the typical benchmark functions. The TPS method is suggested to be suitable for search problems based on molecular kinetics.

4 Conclusion

We propose the TPS method that consists of a random search and an evolutionary search to effectively explore all possible solution vectors of kinetic parameters satisfying the target dynamics, which greatly enhances the search efficiency without any biases in biological problems. The proposed method enables one to approach to global and firm analyses that consider how the changes in parameter values affect the results. We investigated the effects of two critical control parameters, AEC and RIG, for the TPS method on search performance. When an appropriate value of AEC is selected, which depends on target functions, a small value of RIG enables the TPS method to achieve both high-speed and nonbiased searches. The TPS method does not show so high performance for typically-employed benchmark functions, but provides a great advantage in dynamic models of protein synthesis. The effectiveness of the TPS method for biochemical networks is verified thorough parameter search of the heat shock response model.

Acknowledgements

This work was supported by KAKENHI (Grant-in-Aid for Scientific Research) on Priority Areas "Systems Genomics" from the Ministry of Education, Culture, Sports, Science and Technology of Japan and partially by KAKENHI (Grant-in-Aid for Scientific Research (B), 2006, 18300098).

References

1. Alves, R. and M.A. Savageau, *Systemic properties of ensembles of metabolic networks: application of graphical and statistical methods to simple unbranched pathways.* Bioinformatics, 2000. **16**(6): p. 534-47.
2. Stelling, J., E.D. Gilles, and F.J. Doyle, 3rd, *Robustness properties of circadian clock architectures.* Proc Natl Acad Sci U S A, 2004. **101**(36): p. 13210-5.
3. Mendes, P. and D. Kell, *Non-linear optimization of biochemical pathways: applications to metabolic engineering and parameter estimation.* Bioinformatics, 1998. **14**(10): p. 869-83.
4. Kurata, H., et al., *Module-based analysis of robustness tradeoffs in the heat shock response system.* PLoS Comput Biol, 2006. **2**(7): p. e59.