

Inference of Genetic Networks using a Cooperative Coevolutionary Algorithm

協調的共進化アルゴリズムを用いた遺伝子ネットワークの同定

Shuhei Kimura^{1,2}, Mariko Hatakeyama^{1,2}, Akihiko Konagaya^{1,2}

木村周平^{1,2}, 畠山真里子^{1,2}, 小長谷明彦^{1,2}

(¹RIKEN GSC, ²RIKEN Harima Inst.)

(¹理研 ゲノム科学総合研究センター, ²理研 播磨研究所)

e-mail: {[skimura](mailto:skimura@gsc.riken.jp),[marikoh](mailto:marikoh@gsc.riken.jp),[konagaya](mailto:konagaya@gsc.riken.jp)}@gsc.riken.jp

DNA microarray technology allows us to measure gene expression patterns on a genomic scale, but to exploit this technology we must find ways to extract useful information from the massive amount of data. Among the possible solutions for extracting information, many researchers have taken an interest in the inference of genetic networks. The inference of genetic networks is a problem in which mutual interactions among genes are estimated using time-series data of gene expression patterns. The inferred model of the genetic network is conceived as an ideal tool to help biologists generate hypotheses and facilitate the design of their experiments. On another level, it may also shed light on the biological functions of genes.

In this study, we use an S-system model to describe the genetic network, as this model is rich enough in structure to capture various dynamics. Several network inference algorithms based on the S-system model have been proposed. These algorithms estimate the S-system parameters from observed time-series data of gene expression patterns. The number of S-system parameters is proportional to the square of the number of network components. Thus, the algorithms must estimate a large number of S-system parameters simultaneously if they are to be used to infer large-scale network systems containing many network components. This is why the inference algorithms based on the S-system model have been applied only to small networks that include only a few (less than five) genes.

To resolve the high-dimensionality of the genetic network inference problem in the S-system model, a problem decomposition strategy, that divides the original problem into several subproblems, has been proposed. This approach enables us to infer S-system models of larger-scale genetic networks. However, when the given time-series data contain the measurement noise, the inferred model cannot be used to computationally simulate a genetic network. This is one of disadvantages of the problem decomposition approach, since the computational simulation is needed to analyze and understand the genetic network.

In this study, we propose a new method that eliminates the disadvantage of the problem decomposition approach. The proposed method simultaneously solves the decomposed subproblems using a cooperative coevolutionary algorithm. In the proposed coevolutionary algorithm, all of the subproblems interact with each other through the gene expression curves which are updated when more reasonable curves are obtained. Because of this interaction, the proposed method has the ability to infer an S-system model that is ready for the computational simulation. We are now applying our method to DNA microarray data of *T. thermophilus* HB8 strains.

Reference

[1] S.Kimura, M.Hatakeyama, A.Konagaya, *CBI J.*, Vol. 4, No. 1, pp. 1-14 (2004)