A bio-inspired optimization for inferring interactive networks: Cockroach swarm evolution

Shinq-Jen Wu *, Cheng-Tao Wu

Department of Electrical Engineering, Da-Yeh University, Chang-Hwa, Taiwan, ROC

ABSTRACT

Most diseases are the result of interactions between gene products. Inferring the *quantitatively dynamic* interactive mechanisms of underlying systems has become a central part of the ongoing revolution in biology. Biological systems are sparely connected and always noise contaminated. The connection cannot be neglected even the strength is down to 0.001. Therefore, the used computational technologies are not only good in exploitation and exploration but also robust to noise. The development of effective top-down modeling technologies to largely increase the gap between redundant and true connections is a significant challenge. Cockroaches are found in nearly all habitats and survive in extreme environments and when there is food scarcity. We previously introduced cockroaches' competition behavior to improve GAs' exploitation for the parameter estimation of biological systems. However, cockroaches always work together for foraging. Competition occurs only during food is extremely scarcity. In this study, we further mimick cockroaches' cooperated-based swarm behavior, wherein their competitive behavior and migration are event-induced. The proposed cockroach-inspired swarm evolution (CSE) was tested through dry-lab experiments under noise environment. We successfully identified 80 connections (62 redundant and 18 true connections) with truncated interactive strengths smaller than 10^{-14}. Only one to three pruning steps for noise-free systems and five steps for noise-contaminated systems are needed even in a wide searching space. The estimated parameters are almost perfectly equal to the true ones.

© 2014 Elsevier Ltd. All rights reserved.

1. Introduction

A complete wiring architecture (the bowlike networks in Oda, Matsuoka, Funahashi, and Kitano (2005) is valuable but is insufficient to decipher how genes' contribution to a system. Mathematical modeling from time-series data to extract *quantitatively dynamic* information is more practical and valuable. Various models and their respective modeling techniques were proposed. Different models have unique advantages and inherent modeling issues. (Voit and collaborators reviewed some canonical structures such as generalized-mass-action models, S-system models and ling-log models (Chou & Voit, 2009; Voit, 2013). Some inherent issues and possible search directions were addressed. The recasting of ordinary differential equation (ODE)-based systems to S-systems is also discussed in Voit (2013).) How to choose a suitable model such that the regulatory strength of constitutes (genes, proteins and metabolites) is obtained is important. We discuss those popular or developing models and their respective modeling technologies in the next section. In this study, the S-system is chosen as a

prototype to infer the interactive network of underlying systems because S-systems take vantages of good generalization to large systems and **directly** show the **net** interactive information. However, S-system modeling is a multiobjective and multiconstraint optimization problem (as shown in Eq. (A2) of the Appendix A). The inferential methodologies and associated optimization algorithms for identifying S-systems' structures and parameters become central problems. As systems are contaminated with noise and searching spaces become wider (see Table 1), the identification of such highly *coupled* nonlinear systems becomes increasingly challenging. Various strategies to integrate swarm, evolutionary and local-search optimization were proposed (Chou & Voit, 2009; Sun, Garibaldi, & Hodgman, 2012; Voit, 2013). However, the relative values between the value of the smallest preserved term and that of the largest pruned terms are too small (see Table 2). Cockroaches' intelligence in self organization and labor division let them efficiently adapt to environment change. Instead of addressing how often or how deep the local-search methods were used, we focus on mimicking cockroaches' evolution and their swarm intelligence in cooperative-and-competitive foraging behavior, habit migration during harsh climatic conditions and marking territory. Swarm behaviors are hierarchical and some are

* Corresponding author. Fax: +886 4 8511245.
  E-mail address: jen@mail.dyu.edu.tw (S.-J. Wu).

event-induced to let the proposed swarm evolutionary algorithm have the ability to intensify and diversify the searching at the same time.

Sun et al. (2012) provided a review table that shows various meta-heuristic approaches in systems biology models. Some researchers recently identified S-systems' interactive structure and the respective parameters ***separately*** (*two-stage modeling technologies*). Hsiao and Lee (2012) introduced sensitivity-based incremental evolution to identify S-system's structure and then used particle swarm optimization (PSO), differential evolution (DE), or hybrid genetic algorithms (GAs) and PSOs to estimate their respective parameters. However, no inferred structure information is provided in Hsiao and Lee (2012). Zuniga (2011) used, respectively, discrete and continuous ant colony optimization (ACO) for structure and parameter identification of S-systems. However, the truncated criterion (the number of deleted edges) for the structure inference is unclear. (The number for a redundant edge is 21 but is 19 for the smallest preserved edge, as shown in Table 1 in Zuniga (2011)). Yang, Jiang, and Chen (2011) proposed a restricted additive tree (similar to a genetic programming structure) for S-system structure identification and used PSO for parameter identification.

One-stage modeling technologies identify the structure and parameters of biological systems simultaneously. Chou and Voit (2012) proposed a systematic approach of dynamic flux estimation to achieve the parameter and function-form identification. This type of S-system modeling becomes a multi-objective optimization problem and is increasing challenging as the number of the involved components increases. Various penalty terms were proposed to obtain the sparely interactive networks (Kikuchi, Tominaga, Arita, Takahashi, & Tomita, 2003; Kimura et al., 2005; Liu & Wang, 2008; Noman & Iba, 2005; Noman & Iba, 2007; Thieffry, Huerta, Perez-Rueda, & Collado-Vides, 1998; Wu, Wu, & Chang, 2012). It is rather difficult to use only a swarm or an evolutionary algorithm to achieve satisfactory results. Wang, Qian, and Dougherty (2010) considered two extreme cases and proposed a two-step method to determine the ranges and the mean values of the parameters. Marino and Voit (2006) proposed algorithms to gradually increase the model complexity. Wang and collaborators proposed hybrid differential evolution (Liu & Wang, 2009; Wang & Liu, 2010). Ho, Hsieh, Yu, and Huang (2007) introduced an intelligent two-stage evolutionary algorithm. Chen, Lee, Chuang, Wang, and Shieh (2010) hybridized genetic algorithms and simulated annealing. However, the gap between the true and redundant interaction was too small to guarantee that safety pruning actions were taken. We previously developed an exploration-and-exploitation genetic algorithm (EEGA) to ensure a safe pruning where a general search space is used (Wu et al., 2012). The approach of developing an algorithm that integrates various intelligent technologies is still in development.

Memetic algorithms (MAs) incorporate efficient local improvement operations within evolution- and swarm-based algorithms (GA, DE, PSO, ACO and artificial bee colony (ABC)) to achieve a subtle trade-off between exploration and exploitation. MAs have been applied to other fields for solving various optimal problems ([S1–S19] in the supplement file). Recently, Jadhav and Roy (2013) introduced the globally best solution to direct the ACO's search path towards global optima for power dispatch problems. Nedjah, Silva Junior, and de Macedo Mourelle (2013) proposed a congestion-aware colony algorithm to develop a network-on-chip platform. ACO's exploitation and exploration were improved through multiple colonies, pseudo-random propositional rules and a double pheromone update. To solve PSO's deficiency in premature convergence Wu et al. (2014) introduced a superior-solution set and individual-level-based mutation. Son (2014) introduced a passive-congregation term (the velocity variation of two randomly chosen

particles) and used chaotic parameters (generated through a Chirikov standard map). Cuevas, Cienfuegos, Zaldívar, and Pérez-Cisneros (2013) pointed out a serious problem in PSO and ABC: as the algorithms evolve, the entire population concentrates around the best particle or diverges without control. Cuevas et al. (2013) proposed a new social-spider optimization (SSO), wherein spiders were divided into male and female groups. Different cooperative operations were used in these two groups. In this study, we propose a diversity-ensured cooperative mechanism. The proposed algorithm initiates a competition mechanism periodically and performs migration for an emergency. Cockroaches are divided into leaders and members through the golden section method. All individuals cooperate for foraging, wherein individuals follow the strongest leader and take the variation between two leaders as a modified term. Leaders take an extra work (searching neighborhood). *Periodically*, leaders are responsible for exploitation and members for exploration. *In an emergency*, except the strongest leader all individuals explores for food.

In this study, we first addressed some canonical models and their related inherent modeling issues in Section 2. The cockroach-inspired swarm evolution (CSE) is proposed in Section 3. This algorithm possesses both swarm intelligence and an evolutionary vantage to improve the best-so-far individual and to maintain the population diversity at the same time. In Section 4, the performance of CSE was tested through dry-lab experiments, where both noise-free and noise-contaminated environments are considered. In-depth discussion of the performance and a comparison of the state-of-the-art technologies are given in this section. Section 5 summaries our research and the possible future work.

2. Data-driven modeling

Time-series data reveal the dynamic behavior that characterizes living systems. Therefore, the first motivation for modeling a biological network from a time-course response is to obtain a synthetic view of biological knowledge and to construct it in a way that highlights relevant properties that might remain hidden without a model or with a less relevant model. The development of system-level models from biological data is the cornerstone of computational systems biology. The choice of the model prototype is the first step for data-driven modeling. However, the respective top-down approaches (data-driven approaches or computational approaches) about network inferences require intensive research, and the techniques are challenging.

Some researchers used artificial neural networks (ANNs) or recurrent neural networks (RNNs) for modeling the dynamics of biological systems. Wang, Qian, and Dougherty (2007) choose a noise-free sigmoidal-function model and proposed a nested GA and Kalman-filter hybrid optimization. Kentzoglanakis and Poole (2012) used discrete ACO to infer the recurrent-neural-network (RNN)-based biological model and applied PSO to estimate the RNN's parameters. However, the accuracy of the structure is still an issue. Some researchers assumed a *fully-connected* RNN model and estimated the unknown parameters through artificial bee colony algorithm (Yeh & Hsieh, 2012), through the hybrid of HDE (DE with a migration operation) and a PSO algorithm (Xu, Ganesh, Venayagamoorthy, & Wunsch, 2007).

They addressed the capability that ANN or RNN-based models not only capture the dynamic behavior but also capture genes' regulatory interactions, where the connecting factor w_{ij} denotes the effect of the *j*th gene on the *i*th gene. NNs are inspired by the parallel-distributed biological neurons in the brain. The connection weights denote the synaptic strengths in the brain, the values of which change proportionally to the correlation between the firing of the pre- and postsynaptic neurons (Lin & Lee, 1996). However,