Estimating Reaction Constants in Stochastic Biological Systems with a Multi-swarm PSO Running on GPUs

Marco S. Nobile
Università degli Studi di Milano-Bicocca, Dip. Informatica, Sistemistica e Comunicazione, Milano, Italy
nobile@disco.unimib.it

Daniela Besozzi
Università degli Studi di Milano, Dip. Informatica e Comunicazione, Milano, Italy
besozzi@dico.unimi.it

Paolo Cazzaniga
Università degli Studi di Bergamo, Dip. Scienze della Persona, Bergamo, Italy
paolo.cazzaniga@unibg.it

Giancarlo Mauri
Università degli Studi di Milano-Bicocca, Dip. Informatica, Sistemistica e Comunicazione, Milano, Italy
mauri@disco.unimib.it

Dario Pescini
Università degli Studi di Milano-Bicocca, Dip. Statistica, Milano, Italy
dario.pescini@unimib.it

ABSTRACT

We present a parameter estimation method, based on particle swarm optimization (PSO) and embedding the tau-leaping algorithm, for the efficient estimation of reaction constants in stochastic models of biological systems, using as target a set of discrete-time measurements of molecular amounts sampled in different experimental conditions. To account for the multiplicity of data, we consider a multi-swarm formulation of PSO. The whole method is developed for GPGPU architecture to reduce the computational costs.

Categories and Subject Descriptors

I.6 [**Computing Methodologies**]: Simulation and Modeling; J.3 [**Computing Applications**]: Life and Medical Sciences; C.1.2 [**Processor Architecture**]: Multiprocessors—*SIMD*

Keywords

Particle Swarm Optimization, Parameter Estimation, Tau Leaping, GPU Computing, Systems Biology

1. INTRODUCTION

The development of computational methods for the analysis of biological systems is one of the foremost goals of Systems Biology. In order to gain insights into the functioning of these complex systems, we need to identify the system structure and a proper parameterization, which are indispensable to analyze the behavior of these systems in both physiological and perturbed conditions. Except for special cases where the experimental values of parameters are known, most of the times they are not available or inaccurate, since carrying out their measurement in vivo can be tangling or even impossible. Parameter estimation (PE) is a computational problem consisting in the automatic determination of the unknown parameters, which can be performed by means of optimization techniques [2]. The method for PE proposed hereby combines particle swarm optimization (PSO) [3] with stochastic simulation algorithms, and it is inspired by the common scenario of biological research, where multiple experiments on a system of interest are carried out under different conditions and each experiment is repeated a number of times, in order to account for possible measurement errors. The result of this process is a set of *discrete-time target series* (DTTS) – corresponding to sampled amounts of some molecular species – which represent the input of our PE method. In particular, to account for the multiplicity of target data, we designed a *multi-swarm* version of PSO, where each swarm is assigned to a different experimental condition and, in a synergistic way, they cooperate for the estimation of a common set of kinetic values – that can simultaneously fit all the measures in the analyzed conditions – by exchanging their best particles at regular intervals. To account for the intrinsic stochastic fluctuations of molecular reactions, the fitness function of each candidate solution has been defined as the average distance between the target data and a set of independent stochastic simulations of the system dynamics. In order to provide a reduction of the computational costs, our PE method has been entirely developed for the GPGPU architecture.

GECCO'12 Companion, July 7–11, 2012, Philadelphia, PA, USA.
ACM 978-1-4503-1178-6/12/07.

2. METHODS

We consider mathematical models of biological systems defined according to the stochastic formulation of chemical kinetics, i.e., we specify the set $\mathcal{S}=\{S_1,\ldots,S_N\}$ of molecular species occurring in the system and the set $\mathcal{R}=\{R_1,\ldots,R_M\}$ of biochemical reactions. Each reaction R_μ is characterized by a stochastic constant $c_\mu \in \mathbb{R}^+$ that encompasses the physical and chemical properties of the reaction. The vector of parameters $\boldsymbol{\gamma}_c=(c_1,\ldots,c_M)$, able to generate a dynamics comparable to the DTTS, is the goal of the PE. In our PE implementation we consider D different initial conditions for the biological system of interest, each one characterized by a set of DTTS needed for the optimization process. In our multi-swarm version of PSO, each initial condition is independently processed by a swarm consisting of n particles.

The D swarms can then cooperate by means of a migration process, used to share information of the best solutions found. The fitness function used to evaluate the quality of the estimated solutions $\boldsymbol{\gamma}=(\gamma_1,\ldots,\gamma_M)$ is defined as a relative point-to-point distance between the DTTS and the dynamics simulated by means of stochastic simulation algorithms. With respect to the PE method presented in [4], to compute the fitness value of each particle we perform here G parallel simulations with tau-leaping [1], in order to tame the error due to stochastic fluctuations. The total number of fitness values computed at each PSO iteration is $D \times G \times n$. Since each fitness evaluation can be performed independently from the others, we implemented our method on a parallel architecture: the GPGPU computing (specifically, Nvidia's CUDA) which exploits the computational power of modern multi-core GPUs. In our CUDA implementation, for each initial condition we create G *blocks* composed of n *threads* (corresponding to G parallel tau-leaping simulations for each particle), obtaining a $D \times G$ *grid*.

3. RESULTS

We tested our method with two reference models: the well known *Michaelis-Menten* (MM) enzymatic kinetics and a *prokaryotic auto-regulatory gene network* (PGN) [5]. For both systems, the DTTS we considered correspond to D=4 initial conditions, E=3 experiment repetitions and C=10 sampling time instants. To determine how much the estimated values of stochastic constants of a particle $\boldsymbol{\gamma}$ are close to the vector $\boldsymbol{\gamma}_c$ of real parameters, we consider the *mean error* $\overline{\varepsilon}_\gamma$ [5] and the *average mean error* $\langle\overline{\varepsilon}_\gamma\rangle$, calculated by averaging the values of $\overline{\varepsilon}_\gamma$ obtained at the last iteration of each PE run [4]. We performed several tests on the MM model to determine the best setting of our PE method. The first test was focused on the importance of *swarm size* n on the estimation performances. Our results show that a number of particles n<32 produces the worst estimates, while for n>64 the improvement is so slight that it does not justify the computational effort. In the second test we analyzed the influence of the number G of *parallel tau-leaping simulations*, used to evaluate the fitness value of a particle. By increasing G, the method yields a lower $\langle\overline{\varepsilon}_\gamma\rangle$, but for G>40 the quality does not increase anymore, whilst requiring a greater amount of CUDA resources. We have also tested different settings for the *migration interval* IT_{mig}, trying values ranging from 0 (no migration) up to 40, and observing that values between 5 and 25 yield the best results.

According to these analyses we determined the following best setting: population size n=64, inertia w linearly decrementing from 0.9 to 0.4, cognitive and social factor C_{soc}=C_{cog}=1.9, IT_{mig}=20, dynamic topology of migration, "damping" boundary conditions, maximum velocity of particles limited to 1/3 of the search space dimensions, G=10 parallel simulations. This setting was used to carry out the PE of the PGN model. Since in general the values of stochastic constants are not known, the error $\overline{\varepsilon}_\gamma$ cannot be computed to assess the quality of the estimated solutions. Therefore, for the PGN model we rely on the fitness: we choose the particles (i.e., the parameter vectors) characterized by the best fitness value and compare a simulation of their dynamics with the available DTTS. Considering the availability of DTTS for either all molecular species occurring in the system, or only a subset of them (the latter case being a common feature of laboratory experiments), our results clearly show that the best solution found by our PE method allows to generate a dynamics that accurately matches the DTTS in all D conditions (see an example in Figure 1).

Concerning the performances, our implementation outperforms a corresponding CPU execution: a single PSO iteration (equivalent to 2560 simulations and particle updates) takes 6 sec on a GPU Tesla C1060, while on a CPU Intel Core Duo 6700 (2.66 GHz) it takes 392 sec.

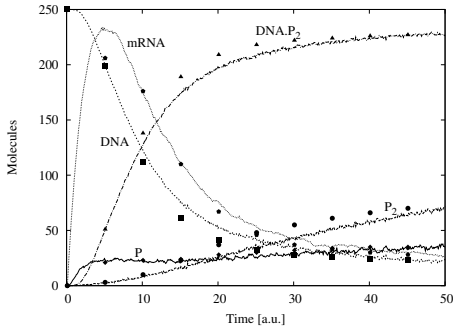

Figure 1: Dynamics of PGN species (lines) using the stochastic constants found by the best solution of multi-swarm PSO, compared to DTTS (dots).

4. CONCLUSION

We have proposed an efficient method for PE in stochastic biological systems, accelerated by means of tau-leaping algorithm and GPGPU computing. Our method can handle experimental discrete-time target series coming from multiple experiments executed under different initial conditions, and it does not require uniform sampling rates nor measurement of every molecular species. As a result, it estimates a common parameter vector, valid for all conditions, by exploiting a multi-swarm version of PSO. The fitness function averages the relative point-to-point distances between the experimental samples and a set of simulated dynamics, taking into account the effects of biological noise.

The method has been conceived around the GPGPU architecture to exploit the intrinsic parallelism of PSO, achieving a significant boost with respect to a strictly sequential implementation. This GPU-based method indeed represents a novel methodology in the context of PE, and a useful tool for the computational analysis of biological systems.

5. REFERENCES

[1] Y. Cao, D.T. Gillespie, and L.R. Petzold. Efficient step size selection for the tau-leaping simulation method. *J. Chem. Phys.*, 124:044109, 2006.

[2] I. Chou and E. Voit. Recent developments in parameter estimation and structure identification of biochemical and genomic systems. *Math. Biosci.*, 219:57–83, 2009.

[3] J. Kennedy and R. Eberhart. Particle swarm optimization. In *Proc. IEEE Int. Conference on Neural Networks*, volume IV, pages 1942–1948, Piscataway, NJ, 1995.

[4] M.S. Nobile, D. Besozzi, P. Cazzaniga, G. Mauri, and D. Pescini. A GPU-based multi-swarm PSO method for parameter estimation in stochastic biological systems exploiting discrete-time target series. In *Proc. EvoBIO2012, LNCS* 7246, 74–85, Springer, 2012.

[5] Y. Wang, S. Christley, E. Mjolsness, and X. Xie. Parameter inference for discretely observed stochastic kinetic models using stochastic gradient descent. *BMC Syst. Biol.*, 4(1):99, 2010.