BIOINFORMATICS APPLICATIONS NOTE Vol. 27 no. 7 2011, pages 1028–1029
doi:10.1093/bioinformatics/btr038

Systems biology

Advance Access publication February 8, 2011

SBML-PET-MPI: a parallel parameter estimation tool for Systems Biology Markup Language based models

Zhike Zi[1,2,3]

[1]BIOSS Centre for Biological Signalling Studies, [2]Center for Biological Systems Analysis (ZBSA) and [3]Freiburg Initiative in Systems Biology (FRISYS), University of Freiburg, Freiburg, Germany

Associate Editor: Olga Troyanskaya

ABSTRACT

Summary: Parameter estimation is crucial for the modeling and dynamic analysis of biological systems. However, implementing parameter estimation is time consuming and computationally demanding. Here, we introduced a parallel parameter estimation tool for Systems Biology Markup Language (SBML)-based models (SBML-PET-MPI). SBML-PET-MPI allows the user to perform parameter estimation and parameter uncertainty analysis by collectively fitting multiple experimental datasets. The tool is developed and parallelized using the message passing interface (MPI) protocol, which provides good scalability with the number of processors.

Availability: SBML-PET-MPI is freely available for non-commercial use at http://www.bioss.uni-freiburg.de/cms/sbml-pet-mpi.html or http://sites.google.com/site/sbmlpetmpi/.

Contact: zhike.zi@bioss.uni-freiburg.de

Received on September 20, 2010; revised on January 7, 2011; accepted on January 10, 2011

1 INTRODUCTION

Mathematical models, paired with experimental analysis, have been widely used for the studies on cellular signaling pathways, gene regulatory networks and metabolic networks. In order to perform simulation and kinetic analysis of the models, one must know the parameter values for the modeled biological system. However, some parameter values cannot be experimentally measured in practice and their values are unknown. Therefore, parameter estimation is a common issue and very important for the mathematical modeling of biological systems.

To address the parameter estimation problem, we previously developed a parameter estimation tool called Systems Biology Markup Language (SBML)-PET to automatically estimate the unknown parameters of SBML-based models (Hucka *et al.*, 2003) by fitting a variety of experimental datasets (Zi and Klipp, 2006). The availability of many parameter estimation tools has facilitated the development of models in systems biology community (Hirmajer *et al.*, 2009; Hoops *et al.*, 2006; Koh *et al.*, 2010; Maiwald and Timmer, 2008; Quach *et al.*, 2007; Zi and Klipp, 2006). However, implementing parameter estimation is still time consuming and computationally demanding. A parallel implementation of a global optimization algorithm was developed by Ji and Xu (2006), but it did not address the issue of parameter estimation for SBML models. A fast and convenient parameter estimation tool is yet to be developed. Here, we introduce a new parallel parameter estimation tool for SBML-based models, which is named as SBML-PET-MPI. This tool allows the user to perform parameter estimation and parameter uncertainty analysis by collectively fitting multiple experimental datasets. SBML-PET-MPI is parallelized using the message passing interface (MPI) protocol (Gropp *et al.*, 1996). It is fast and efficient as it provides a good speed up scalability for parameter estimation and analysis depending on the number of processors and the complexity of the model.

2 SOFTWARE FEATURES

Using the MPI protocol, SBML-PET-MPI implemented several algorithms for parameter estimation, and parameter uncertainty analysis, which include:

(1) Parameter estimation with stochastic ranking evolution strategy (SRES) global optimization algorithm (Runarsson and Yao, 2000). SBML-PET-MPI automatically estimates the parameter values that minimize the difference between model and data based on weighted least squares.

(2) Parameter uncertainty analysis with profile likelihood exploit algorithm. SBML-PET-MPI parallelizes the algorithm of profile likelihood exploit for the identifiability analysis and confidence intervals analysis of the estimated parameters (Raue *et al.*, 2009). Levenberg–Marquardt algorithm (Marquardt, 1963) is used for the optimization in profile likelihood exploit algorithm.

(3) Parameter uncertainty analysis with the bootstrap method. SBML-PET-MPI implements another algorithm to get the confidence limits of the estimated parameters based on a bootstrap approach, which generates new synthetic datasets D_s from the actual experimental datasets D_0 (Press *et al.*, 1992):

$$D_s(i) = D_0(i) + \delta_i \quad (1)$$

where δ_i is randomly taken from a normal distribution $N(0, \sigma_i)$, σ_i corresponds to the standard deviation in the i-th experimental data.

3 IMPLEMENTATION AND PERFORMANCE

SBML-PET-MPI was parallelized with the MPI protocol. To speed up the solving of the ordinary differential system, SBML-PET-MPI integrated ODEPACK library (Hindmarsh, 1983) and is coded in C programming language. It supports the models in SBML format by using libSBML package (Bornstein *et al.*, 2008).

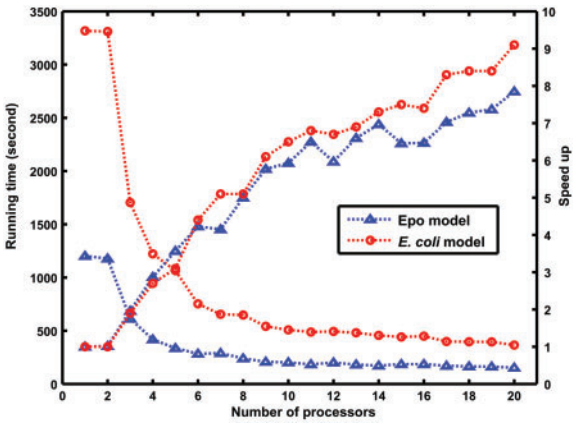

Fig. 1. The running time and speed up for the global optimization and parameter identifiability analysis of the example models using 1–20 CPUs (options: –with-gen = 1000 –with-ple). The running time might slightly vary at different execution of the program due to the variation of model simulations with different parameter sets.

To evaluate the performance of the parallelized parameter estimation analysis algorithms, we recorded the running time of SBML-PET-MPI for global optimization (parameter estimation) and profile likelihood exploit analysis (parameter uncertainty analysis) for the Epo model with multiple real experimental datasets (Becker *et al.*, 2010) and the *Escherichia coli* tryptophan operon model with datasets from different conditions. Details about the examples are described in SBML-PET-MPI user-guide document. The speed up (calculated as ratio of the running time with one processor to that with n multiple processors, T_1/T_n) is good with the increased number of processors in these examples (Fig. 1).

In practice, the speed up performance depends on the balance of the decreased time by task parallelization and the increased time for data exchange in different processors. One can assume that the total time for parameter estimation and uncertainty analysis is x seconds with one CPU, among which y seconds are required for the tasks that cannot be parallelized. In this tool, the first processor is used for algorithm coordination and data exchange. The rest tasks with $x-y$ seconds computation time are distributed into other $n-1$ processors. On the other hand, the addition of new processors requires more communication computation time, if one assume that it is linearly increased, there will be $z\times(n-1)$ more time. In this case, the speed up can be calculated with Equation (2). The speed up scalability is in general better for complex models with more estimated parameters and more experimental datasets.

$$s_n = \frac{T_1}{T_n} = \frac{x}{y + \frac{x-y}{n-1} + z\times(n-1)} \quad (n \geq 2) \qquad (2)$$

4 OUTLOOK

With the popularity of multi-core processors and clusters, there is an increasing demand for the parallelization of parameter estimation and simulation tools for the modeling and analysis of biological systems. A parallelized parameter estimation tool (e.g. SBML-PET-MPI) will decrease the time of parameter estimation for the model development and let the modelers focus on the model analysis and the investigated biological questions. The systems biology community will benefit from more parallelized tools for model simulation and analysis.

ACKNOWLEDGEMENTS

We would like to thank Andreas Raue for the helpful discussion of the profile likelihood exploit algorithm.

Funding: Excellence Initiative of the German Federal and State Governments (EXC 294); BMBF - Freiburg Initiative in Systems Biology (grant 0313921).

Conflict of Interest: none declared.

REFERENCES

Becker,V. *et al.* (2010) Covering a broad dynamic range: information processing at the erythropoietin receptor. *Science*, **328**, 1404–1408.

Bornstein,B.J. *et al.* (2008) LibSBML: an API library for SBML. *Bioinformatics*, **24**, 880–881.

Gropp,W. *et al.* (1996) A high-performance, portable implementation of the MPI message passing interface standard. *Parallel Comput.*, **22**, 789–828.

Hindmarsh,A.C. (1983) ODEPACK, a systematized collection of ODE solvers. In Stepleman,R.S. *et al.* (eds), *Scientific Computing*. North-Holland, Amsterdam, pp. 55–64.

Hirmajer,T. *et al.* (2009) DOTcvpSB, a software toolbox for dynamic optimization in systems biology. *BMC Bioinformatics*, **10**, 199.

Hoops,S. *et al.* (2006) COPASI–a COmplex PAthway Simulator. *Bioinformatics*, **22**, 3067–3074.

Hucka,M. *et al.* (2003) The systems biology markup language (SBML): a medium for representation and exchange of biochemical network models. *Bioinformatics*, **19**, 524–531.

Ji,X. and Xu,Y. (2006) libSRES: a C library for stochastic ranking evolution strategy for parameter estimation. *Bioinformatics*, **22**, 124–126.

Koh,C.H. *et al.* (2010) DA 1.0: parameter estimation of biological pathways using data assimilation approach. *Bioinformatics*, **26**, 1794–1796.

Maiwald,T. and Timmer,J. (2008) Dynamical modeling and multi-experiment fitting with PottersWheel. *Bioinformatics*, **24**, 2037–2043.

Marquardt,D.W. (1963) An algorithm for least-squares estimation of nonlinear parameters. *J. Soc. Ind. Appl. Math.*, **11**, 431–441.

Press,W.H. *et al.* (1992) *Numerical Recipes in C : The Art of Scientific Computing*. Cambridge University Press, Cambridge.

Quach,M. *et al.* (2007) Estimating parameters and hidden variables in non-linear state-space models based on ODEs for biological networks inference. *Bioinformatics*, **23**, 3209–3216.

Raue,A. *et al.* (2009) Structural and practical identifiability analysis of partially observed dynamical models by exploiting the profile likelihood. *Bioinformatics*, **25**, 1923–1929.

Runarsson,T.P. and Yao,X. (2000) Stochastic ranking for constrained evolutionary optimization. *IEEE Trans. Evol. Comput.*, **4**, 284–294.

Zi,Z. and Klipp,E. (2006) SBML-PET: a Systems Biology Markup Language-based parameter estimation tool. *Bioinformatics*, **22**, 2704–2705.