BioSystems 108 (2012) 73–77

Contents lists available at SciVerse ScienceDirect

BioSystems

journal homepage: www.elsevier.com/locate/biosystems

ConvAn: A convergence analyzing tool for optimization of biochemical networks

Andrejs Kostromins*, Ivars Mozga, Egils Stalidzans

Biosystems Group, Department of Computer Systems, Latvia University of Agriculture, Liela iela 2, LV3001 Jelgava, Latvia

ARTICLE INFO

Article history:
Received 22 August 2011
Received in revised form 2 December 2011
Accepted 19 December 2011

Keywords:
Software tool
Biochemical network
Optimization
Forecast
Convergence speed

ABSTRACT

Dynamic models of biochemical networks usually are described as a system of nonlinear differential equations. In case of optimization of models for purpose of parameter estimation or design of new properties mainly numerical methods are used. That causes problems of optimization predictability as most of numerical optimization methods have stochastic properties and the convergence of the objective function to the global optimum is hardly predictable.

Determination of suitable optimization method and necessary duration of optimization becomes critical in case of evaluation of high number of combinations of adjustable parameters or in case of large dynamic models. This task is complex due to variety of optimization methods, software tools and nonlinearity features of models in different parameter spaces.

A software tool ConvAn is developed to analyze statistical properties of convergence dynamics for optimization runs with particular optimization method, model, software tool, set of optimization method parameters and number of adjustable parameters of the model. The convergence curves can be normalized automatically to enable comparison of different methods and models in the same scale.

By the help of the biochemistry adapted graphical user interface of ConvAn it is possible to compare different optimization methods in terms of ability to find the global optima or values close to that as well as the necessary computational time to reach them. It is possible to estimate the optimization performance for different number of adjustable parameters.

The functionality of ConvAn enables statistical assessment of necessary optimization time depending on the necessary optimization accuracy. Optimization methods, which are not suitable for a particular optimization task, can be rejected if they have poor repeatability or convergence properties.

The software ConvAn is freely available on www.biosystems.lv/convan.

1. Introduction

Optimization of biochemical pathways is a developing field of systems biology and synthetic biology that enable better understanding of features of existing and designed systems. Optimization methods are widely used to solve the inverse problem of metabolic pathways (estimation of parameters from measured variables) and designing improved biochemical pathways (Mendes and Kell, 1998; Moles et al., 2003). Increasing number of optimization methods and tools (Hoops et al., 2006; Maiwald and Timmer, 2008; Mendes et al., 2009; Schmidt and Jirstrand, 2006) enable different optimization strategies.

In case of dynamic biochemical models optimization is a nonlinear programming (NLP) case where some of the constraints (including reaction kinetics) or objective functions are nonlinear causing potential existence of multiple local solutions (multimodality) (Banga, 2008). Therefore the global optimization methods can be used to avoid locally optimal solutions which may hide potentially better fitness of the model to the objective function: it would indicate a bad fit even for a model which could perfectly match a set of experimental data (Banga, 2008) if proper optimization method and duration would be chosen. On the other hand in some cases local optimization methods can be more efficient if the nonlinearity of the model is not critical. There are also hybrid methods which change methods during optimization (Balsa-Canto et al., 2008; Rodriguez-Fernandez et al., 2006). The mentioned range of opportunities regarding optimization methods make the choice complicated.

Actually the choice of a method include two decisions: (1) the optimization tool, method, it's settings and (2) the duration of optimization run for particular method and tool. Both decisions are depending on the peculiarities of particular model and perhaps even on the set of the adjustable parameters. Several investigations of convergence properties (Baker et al., 2010; Balsa-Canto et al., 2008; Moles et al., 2003; Rodriguez-Fernandez et al., 2006) indicate the complexity of choice in the case of parameter estimation (inverse problem). Similar problems to predict the model-specific convergence can be expected also in case of

---

* Corresponding author. Tel.: +371 28824464.
E-mail address: andrejs.kostromins@gmail.com (A. Kostromins).

0303-2647/$ – see front matter © 2011 Elsevier Ireland Ltd. All rights reserved.
doi:10.1016/j.biosystems.2011.12.004

optimizations with design task which are not studied that carefully so far. Inappropriate optimization method or tool can cause stagnation at local optima thus making wrong conclusions about properties or potential of the system and (or) lead to longer optimization runs compared to other optimization methods. Inappropriate duration of optimization run even in case of appropriate method and tool can lead to local optima in case of too short run or waste of time in case of too long run. Therefore the improvement of choice of the method, tool and duration of optimization run can increase the efficiency of optimization activities especially in case when potential of high number of combinations of adjustable parameters of the same model have to be explored (Mozga and Stalidzans, 2011).

There is a number of software tools (Hoops et al., 2006; Maiwald and Timmer, 2008; Mendes et al., 2009; Schmidt and Jirstrand, 2006) for interdisciplinary fields of systems biology and synthetic biology with graphical user interface (GUI) which is specialized for solving of modeling and optimization tasks and can be used by specialists with poor skills in mathematics and programming. On the other hand so far there is no software with biochemically friendly GUI for seeking of the best performing optimization method and it's settings for a particular model. The software tool ConvAn (Convergence Analysis) is developed to automate and facilitate comparison of convergence dynamics of different optimization methods and tools for particular models depending on the number of adjustable parameters. Several optimization runs with each method has to be performed to gain the average curves with descriptive statistics (Ennos, 2000) of convergence dynamics. The convergence dynamics data is needed in form of time series in a text (.txt) file. The curves are normalized to enable comparison of dynamics in case of different range of objective function values. Use of different methods minimizes the risk of stagnation in local optima and enables rejection of inappropriate methods for optimization of particular set of adjustable parameters of particular model.

A yeast glycolysis model and software tool COPASI is used as application example to compare two global stochastic optimization methods depending on the number of adjustable parameters.

ConvAn can be used in selection of appropriate optimization methods, tools and duration of optimization time for particular model. Generalization of conclusions about suitability of optimization method, tools, duration and other parameters can be done at own risk as the ConvAn cannot predict convergence based on the model features. Still statistics about experienced convergence behavior can be valuable to predict the behavior in similar cases and improve the optimization setup to save time and improve the reliability of optimization.

2. Software description

2.1. Software development and compatibility

ConvAn is developed for Windows, Linux and Mac operating system (OS) in C# programming language. ZedGraph dynamic-link library is used to draw result curves on the plots and is already implemented in the ConvAn installation package.

Microsoft SQL Server 2008 or PostgreSQL open source object-relational database system (available on http://www.postgresql.org/download/) must be installed on the machine in case of Windows OS (a free edition of SQL Server is available on http://www.microsoft.com/express/Downloads/). ConvAn works on machines running Windows XP or higher and Microsoft .NET Framework version 4.0 (available on http://www.microsoft.com/net). ConvAn works with point decimal separator that means that user must check under Regional Setting in Windows OS if decimal separator is point.

In case of Linux or Mac OS ConvAn works only with PostgreSQL and it is necessary to use an open source Mono Framework (available on http://www.mono-project.com/Main_Page).

2.2. Data importation into ConvAn and storage in database

The first steps before storage of data in the database are creation of optimization profile and method declaration. There is a form for method declaration and a form for profile creation. The user should create a separate profile to process each set of optimization runs where optimization conditions were equal. The presented software is universal for any optimization tool which supports output of time series in text file format where columns are separated by "tab" separator. User can freely choose which columns to import into ConvAn. When appropriate columns are chosen user can check data in graph and then save values into data base for further analysis.

2.3. Data processing

Several comparable optimization runs with identical settings (the same profile) should be applied to calculate the descriptive statistics: estimate of the mean, estimate of standard deviation, estimate of the standard error and 95% confidence interval (Ennos, 2000). The values of target function are compared in the same time moments which are calculated accordingly to the user defined discretization time step. Software automatically discretizes data starting from time value of 0.01 second and increasing time value by discretization time step. The meaning of error bars can be selected from estimate of standard deviation, estimate of standard error and 95% confidence interval.

2.4. Functionality of ConvAn

The main task of ConvAn is the statistical analysis of dynamic model optimization of a biochemical network with different optimization tools and methods to determine appropriate tool and optimization method with it's settings as well as duration of optimization runs.

To analyze dynamics of convergence of parameter estimation and design to optima by different optimization methods ConvAn has six menu items in the main screen. The main menu items "File" and "Data" operate with the previously stored and newly imported data files that are created by the optimization software as .txt documents.

Each data file (time series of dynamics of objective function value) has to be linked to a profile (main menu item "Profile"), which contain several runs of optimizations with identical initial conditions and settings. Depending on the optimization algorithm a number of independent algorithm parameters can be changed and affect the convergence speed as a consequence. The impact of algorithm parameters on the convergence dynamics can be conveniently assessed creating separate profiles for each set of parameters thus comparing them just the same way like different optimization methods. Each profile uses a particular optimization tool and method which have to be described in the main menu item "Method". Several profiles can use the same optimization methods to solve different optimization tasks of different models and sets of adjustable parameters. The contents of a profile can be seen and edited in the main menu item "Optimizations".

The main menu item "Analyze" is used to analyze both minimization and maximization dynamics in design tasks as well as minimization dynamics in case of parameter estimation.

The proposed tool is able to calculate average curves for each profile based on registered optimization data. It is possible to choose whether to show all data series including average curve