Engineering Applications of Artificial Intelligence 62 (2017) 164–180

Contents lists available at ScienceDirect

Engineering Applications of Artificial Intelligence

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

An enhanced scatter search with combined opposition-based learning for parameter estimation in large-scale kinetic models of biochemical systems

Muhammad Akmal Remli[a], Safaai Deris[b], Mohd Saberi Mohamad[a,*], Sigeru Omatu[c], Juan Manuel Corchado[d]

[a] Artificial Intelligence and Bioinformatics Research Group, Faculty of Computing, Universiti Teknologi Malaysia, 81310 Skudai, Johor, Malaysia
[b] Faculty of Creative Technology & Heritage, Universiti Malaysia Kelantan, Locked Bag 01, 16300 Bachok, Kota Bharu, Kelantan, Malaysia
[c] Department of Electronics, Information and Communication Engineering, Osaka Institute of Technology, Osaka 535-8585, Japan
[d] University of Salamanca, Biomedical Research Institute of Salamanca/BISITE Research Group, Salamanca, Spain

ARTICLE INFO

Keywords:
Bioinformatics
Artificial intelligence
Metabolic engineering
Evolutionary algorithm
Scatter search
Opposition-based learning

ABSTRACT

An enhanced scatter search (eSS) with combined opposition-based learning algorithm is proposed to solve large-scale parameter estimation in kinetic models of biochemical systems. The proposed algorithm is an extension of eSS with three important improvements in terms of: reference set (*RefSet*) formation, *RefSet* combination, and *RefSet* intensification. Due to the difficulty in estimating kinetic parameter values in the presence of noise and large number of parameters (high-dimension), the aforementioned eSS mechanisms have been improved using combination of quasi-opposition and quasi-reflection, which were under the family of opposition-based learning scheme. The proposed algorithm is tested using one set of benchmark function each from large-scale global optimization (LSGO) problem as well as parameter estimation problem. The LSGO problem consists of 11 functions with 1000 dimensions. For parameter estimation, around 116 kinetic parameters in Chinese hamster ovary (CHO) cells and central carbon metabolism of *E. coli* are estimated. The results revealed that the proposed algorithm is superior to eSS and other competitive algorithms in terms of its efficiency in minimizing objective function value and having faster convergence rate. The proposed algorithm also required lower computational resources, especially number of function evaluations performed and computation time. In addition, the estimated kinetic parameter values obtained from the proposed algorithm produced the best fit to a set of experimental data.

1. Introduction

Metabolic engineering is an important technique in analyzing metobolic pathway of microorganism to support the production and improvement of cellular properties (Keasling, 2012; Mendes and Kell, 1998). This technique which is commonly used in bioprocess engineering or/and genetic engineering is conducted through modeling, experimental and computational procedures (Cvijovic et al., 2011). The outcomes of metabolic engineering is sustainable bioproduct, specifically for industrial biotechnology application (Almquist et al., 2014). Many bioproducts are produced through capitalizing living cells as cell factories. Microorganisms are reported to be efficient cell factories that are able to convert sugar into chemical of interest (Liu et al., 2013a, 2013b). This method can be achieved either by using natural or genetically modified microorganisms. Genetically modified cells are proven to improve cells production, substrate utilization, product quality as well as process design (Almquist et al., 2014). Cell factories have multiple uses ranging from producing bacteria and yeast to developing therapeutic protein in mammalian cells. Nonlinear mathematical models are important tools in the development of this application as they represent the dynamic and mechanistic nature of the cellular processes. The models are used for understanding and analyzing, for example, the concentration changes in fermentation processes before they can be used for predicting and improving production in order to meet industrial demands (Smallbone et al., 2013). Mathematical models are important in metabolic engineering because they are used for the development of various bioproducts such as biofuels and other chemicals (Almquist et al., 2014). Among all mathematical models, kinetic model is considered to be the most efficient tool for in silico metabolic engineering (Cvijovic et al., 2011). This model has attracted a lot of attention from the research community and industrial biotechnology players. The model has several advantages over other models, namely it can describe a complex biological behavior and it can be used for rational design in cell factory.

* Corresponding author.
E-mail addresses: saberi@utm.my, mohd.saberi@gmail.com (M.S. Mohamad).

http://dx.doi.org/10.1016/j.engappai.2017.04.004
Received 2 November 2016; Received in revised form 13 February 2017; Accepted 4 April 2017

Building an efficient kinetic model that is beneficial in metabolic engineering is considered as an iterative task (Almquist et al., 2014) which involves these processes:

1) Determine the purpose of the model;
2) Design the model structure;
3) Estimate the parameter; and
4) Validate.

First, determining the model's purpose is a key step towards building a kinetic model. The modeler should identify various organisms and types of biological processes that can be used in metabolic engineering. Second, the general structure of the kinetic model is designed based on mathematical formulation using Ordinary Differential Equation (ODE). This design can be done by formulating enzymatic reaction and regulation using ODEs starting from small to large-scale reactions, specifically from small biological pathways to large microbial genomes. The ODEs contain time-dependence state variables (metabolites) and kinetic parameters that measure the rate of changes in metabolites concentrations. Third, values of kinetic parameters such as Michaelis-Menten constant K_m and rate of reaction V are subject to estimation; they empirically influence the model prediction or model output. The final stage of developing a kinetic model is model validation, which consists of various experiments and statistical analyses before the model can be routinely used in industrial biotechnology. Due to the highly nonlinear nature of biochemical reactions, building a kinetic model is a difficult and time consuming process. One of the most difficult tasks in this process is parameter estimation that is used to determine the best possible parameter values that are able to measure the goodness of the predictive model by reproducing the data that is as close to the experimental or real data. Also known as model calibration, system identification or inverse problem, this task is widely used in various application domains ranging from metabolic engineering (Copeland et al., 2012), signal processing (Perez-ramirez et al., 2016) and also control systems engineering (Alfi and Fateh, 2011a). It is important to use highly accurate nonlinear model together with optimal kinetic parameters value for the aforementioned domains to save both time and resources.

This work focuses on the task of parameter estimation of kinetic model in metabolic engineering field, assuming that the structures and experimental data for the kinetic models are provided. Due to the highly nonlinear nature of biological systems, parameter estimation is considered as a multimodal and non-convex optimization problem with the existence of several local minima. To estimate kinetic parameter values in ODEs, optimization methods are employed by minimizing the distance between prediction models (models with parameter estimates) and experimental data. The methods can be divided into two categories: local and global. Local optimization methods such as hill climbing and Newton methods can give unsatisfactory results because their local nature can cause the solutions to be easily stuck in local minima. In addition, their efficiency solely depends on the value of initial solutions that are commonly obtained by in vitro measurements or random guesses (Moles et al., 2003). Since most objective functions in real world problems have several local minima, initial solutions are crucial for local method in finding the global minima. If the initial solution is located far from global minima, the solution might be stuck in local minima although its convergence rate is high. This major drawback has spurred the development of global optimization methods in order to comprehensively find the global minima. Metaheuristic algorithm is one of the most efficient global optimization methods which can be divided into single-solution and population-based searches. Single-solution searches that include iterated local search (ILS), simulated annealing, and variable neighbourhood search are operated through improving single solution within the search space. On the other hand, population-based searches operate through maintaining and improving a set of candidate solutions. The set of solutions qualities are iteratively improved using a particular search mechanism to obtain a better solution. In global optimization, the search process can be divided into intensification (exploitation) and diversification (exploration) (Blum and Roli, 2003). Intensification in search process depends on information obtained from the problem to generate better solution from previous solutions using small changes. This is a typically local process which is suitable in local search method. One of the advantage of intensification process is it has very high convergence rate. However, it may be easily stuck in local minima. On the other hand, diversification process explores the broad search space more efficiently. Hence, it is capable of finding the global solution that is far from the initial point. However, the diversification process may cause slower convergence rate and sometimes leads to high computational cost. Thus, finding the balance between these two search processes is crucial in global optimization problem (Liu et al., 2013a, 2013b).

In parameter estimation and system identification problems, metaheuristic algorithms have been mainly applied in various areas including biochemical kinetic models, control systems engineering and aquatic ecosystems. Single-solution based search, namely differential simulated annealing (DSA) (Dai and Lai, 2014) is proposed to estimate biological network model and the proposed method seems robust and efficient compared to other metaheuristic methods. In order to investigate which methods perform well in this area, several comparative studies of state-of-the-art metaheuristic algorithms in parameter estimation is listed as below. A study in nonlinear dynamic model of an aquatic ecosystem has been presented by Tashkova et al. (2012). Several methods were tested and compared to obtain the most accurate model of ecosystems. Another comparative study of parameter estimation is obtained using crop growth model (Zúñiga et al., 2014). In this study, the authors compare state-of-the-art algorithms such as Differential Evolution (DE), Particle Swarm Optimization (PSO) and Artificial Bee Colony (ABC). Similar study has also been conducted by estimating reservoir parameter for predicting reservoir performance (Awotunde, 2015). Three global optimization methods are tested including Covariance Matrix Adaptation Evolution Strategy (CMAES), DE and PSO. The outcomes of this study indicated that DE and PSO are the most efficient algorithms to be used in parameter estimation problem. In system identification of control systems, Adaptive Particle Swarm Optimization (APSO) (Darabi et al., 2012) has been proposed to identify parameters of an exciter machine. Two modifications are made in order to avoid local convergence as well as to obtain excellent quality of final result. Another interesting work has been proposed based on a novel modified particle swarm optimization (MPSO) (Alfi and Fateh, 2011a) to identify nonlinear system for hydraulic suspension system applications. In their contributions, novel mutation mechanism is introduced in MPSO to enhance the global search ability and it is also capable to increase the convergence speed. Several other variants of PSO have also been applied in intelligent identification and control system using improved fuzzy particle swarm optimization (IFPSO) method (Alfi and Fateh, 2011b) and adaptive particle swarm optimization (APSO) (Alfi and Modares, 2011). Excellent results were obtained from these PSO variants compared to other state-of-the-art metaheuristic methods.

Another notable algorithm that has been proposed in the parameter estimation and bioprocess engineering field is scatter search (SS). SS is one of the most early evolutionary algorithms developed by Glover (1977) that is derived from surrogate constraints method. The main difference between this algorithm and modern metaheuristic algorithm is that search mechanism technique is applied to population members. Unlike other evolutionary algorithms, SS does not use crossover and mutation operator. Instead, it uses a solution combination method that operates among population members. New solutions are generated using systematic (partial random) combination rather than fully random solution. One of the main benefits of SS is it maintains a low number of population sizes, even for large problems. Since small population size is not preferred in many algorithms (because it may