UTM UNIVERSITI TEKNOLOGI MALAYSIA
RESEARCH UNIVERSITY

International Journal
of
Innovative Computing

Journal Homepage: http://se.cs.utm.my/ijic

An Improved Method of Newton Method, Genetic Algorithm and Cooperative Coevolutionary Algorithm for Optimization of Metabolic Pathway Production

Mohd Arfian Ismail, Safaai Deris
Artificial Intelligence and Bioinformatics Group (AIBIG)
Department of Software Engineering
Faculty of Computing
Universiti Teknologi Malaysia
81310 Skudai, Johor, Malaysia
marfian2@live.utm.my, safaai@utm.my

Mohd Saberi Mohamad, Afnizanfaizal Abdullah
Artificial Intelligence and Bioinformatics Group (AIBIG)
Department of Software Engineering
Faculty of Computing
Universiti Teknologi Malaysia
81310 Skudai, Johor, Malaysia
saberi@utm.my, afnizanfaizal@utm.my

***Abstract*— In this study, an improved method for optimization of metabolic pathway was presented. The proposed method combines Newton method, Genetic Algorithm (GA) and Cooperative Coevolutionary Algorithm (CCA). The aim of the proposed method was to improve the metabolic pathway production and at the same time reduce the total chemical reaction concentration involved. The proposed method started with Newton method that treated the metabolic pathway as a nonlinear equations system. Then, GA and CCA were used to represent the variables in nonlinear equations system as candidate solutions in the optimization process. GA was used to improve the production, while CCA minimized the total chemical reactions concentration involved. The proposed method was tested on *Escherichia coli* pathway, and several comparisons with previous works were made. The result showed that the proposed method perform well compared to previous works.**

Keywords — Newton method, Genetic Algorithm, Cooperative Coevolutionary Algorithm.

I. INTRODUCTION

Nowadays, many researchers have focused on microbial production in producing renewable biomass. One way to obtain renewable biomass from microbial production is by extracting biomass from metabolic pathway. Metabolic pathway can be defined as a series of chemical reactions that occur in microorganism cell. With the knowledge of biotechnology process, metabolic pathway can be modelled mathematically, thus experiment can be performed by computer simulation. There are many types of models for modelling metabolic pathway, and it has been found that ordinary differential equation (ODE) model is the most suitable approach to be applied [1].

Microbial production is beneficial in producing renewable biomass because it involves lower cost. However, there is a disadvantage due to low production [2]. Hence, microbial production needs to be improved, and this can be done through optimization process. The optimization of metabolic pathway can be defined as a process to increase the production and reduce the total chemical concentrations, which involves the fine-tuning process of chemical reaction [3].

Currently, many works have been published on the optimization of metabolic pathway, for example, in [4]–[6]. All of these works used geometric programming (GP) in improving the production of metabolic pathway. There is a limitation in applying GP as using GP requires expert knowledge in the definition process of decision variables,

where it could cause the convergence problem if the definition process is not performed correctly [7]. Because of that, this study proposed a stochastic method by applying Genetic Algorithm (GA) in fine-tuning the chemical reaction in metabolic pathway. Stochastic method has an advantage in the definition process of decision variables, where it uses a random method to determine the search direction in searching the best variables [8].

The optimization of metabolic pathway becomes difficult as it involves large metabolic pathway that have many components, which makes the fine-tuning process harder. Apart from that, the optimization of metabolic pathway needs to consider metabolic pathway constraints. The constraints of metabolic pathway need to be followed in order to maintain the survival of microorganism cell.

In order to handle this situation, this study proposed an improved method for optimization of metabolic pathway that comprise with Newton method, GA and Cooperative Coevolutionary Algorithm (CCA). The objective of the proposed method was to improve the production and simultaneously minimize the total chemical reaction concentration involved. The optimization process started with Newton method that views the metabolic pathway as a nonlinear equations system. Then, GA and CCA were used to fine-tune variables in nonlinear equations system in order to search the optimum result. The method of this study is presented in the following sections. This includes the modelling of metabolic pathway, problem formulation and the explanation of the proposed method. Then, the case study of the optimization of *tryptophan* (*trp*) biosynthesis in *Escherichia Coli* (*E.coli*) pathway was conducted, before the results and discussion, as well as the conclusion were made.

II. METHOD

This section describes the optimization method of metabolic pathway. The process started with modelling of metabolic pathway and followed by the problem formulation, as well as the optimization process using the proposed method.

A. Modelling of Metabolic Pathway

Metabolic pathway can be represented by ODE. In ODE, two types of models that are usually used are S-system and generalized mass action (GMA) models. This study used GMA model due to its capability to represent the nonlinearity of metabolic pathway [4]. The GMA model that represents the metabolic pathway has the form as follows:

$$\frac{dy}{dx} = Sv(\mathrm{x}) \tag{1}$$

where S is the stoichiometric and $v(x)$ is the vector of reaction rate. $v(x)$ is in a linear form as follows:

$$v_i = \gamma_i \prod_j x_j^{f_{ij}} \tag{2}$$

where coefficient γ_i is the rate constant and coefficient f_{ij} is the kinetic order. Coefficients γ_i and f_{ij} are derived from the Taylor series in the logarithmic space around the steady state [6]. These two coefficients are defined as follows:

$$\gamma_i = |v_i|_0 \tag{3}$$

$$f_{ij} = \left| \frac{\delta v_i}{\delta x_j} \frac{x_j}{v_j} \right| \tag{4}$$

B. Problem Formulation

The optimization of metabolic pathway requires fine-tuning process of metabolic pathway components. This fine-tuning process cannot be performed randomly as there are some constraints that must be followed. These constraints include the steady state constraint and the component of metabolic pathway constraint. The steady state constraint is a situation where all components in the metabolic pathway have a static value, thus making all GMA models (Equation 1) equal to 0. Hence, this leads to a nonlinear equations system [9] and can be formulated as follows:

$$\frac{dx_n}{dy} = [Sv(\mathrm{x})_1, \cdots Sv(\mathrm{x})_n] = 0 \tag{5}$$

In solving a nonlinear equations system, all of the equations are equal to 0 and can be formulated as follows:

$$f(x) = [f(x)_1, \cdots, f(x)_n] = 0 \tag{6}$$

This situation is similar with the steady state constraint in optimization of metabolic pathway. As a result, the optimization of metabolic pathway can be viewed as solving a nonlinear equations system. For the component of metabolic pathway constraint, it refers to the specific range that must be followed by the component concentration [3]. Therefore, the optimization of metabolic pathway can be formulated as follows:

$$max\, F_1(v) \tag{7}$$

$$min\, F_2\left(\sum_{j=1}^{n} x_j\right) \tag{8}$$

s.t. satisfying

$$sv(x)_i = 0, \qquad i = 1, 2, \ldots, n \tag{9}$$

$$x_j^L \le x_j \le x_j^U \qquad j = 1, 2, \ldots, m \tag{10}$$

where Equation 7 is the metabolic pathway production and Equation 8 is the total of chemical reaction involved. Equation 9 is the steady state constraint whereas Equation 10 is the constraint of metabolic pathway components. The superscripts

L and *U* in Equation 10 denote the specific range, where *L* is the lower range and *U* is the upper range.

C. The Combination of Newton method, Genetic Algorithm and Cooperative Coevolutionary Algorithm

The proposed method comprised of Newton method, GA and CCA. The details of the steps are the proposed method is given as follows:

Step 1: *N* initial sub-chromosome in *M* sub-population was generated randomly. The number of *N* and *M* depends on the number of variables in nonlinear equations system. The sub-chromosome was produced in a binary format.

Step 2: Sub-chromosome was evaluated. A representative from all sub-populations was selected to produce a complete chromosome. The selection process was based on their fitness value, where the lower fitness value was selected first. This is to ensure that all representatives from all sub-populations with the lowest fitness value are combined with each other in order to minimize the total chemical concentrations involved.

Step 3: A complete solution was formed. All of the representatives were combined with each other in order to form a complete solution.

Step 4: The complete solution was evaluated. At this step, the complete solution was tested by Newton method. The complete solution was encoded into variables in nonlinear equations system. Then, Newton method was used in solving nonlinear equations system. At this stage, the termination conditions occurred whether the maximum number of generations was achieved or not, and whether the constraint of metabolic pathway component was followed or not. If both conditions were fulfilled, then proceed to Step 7. Otherwise, go to the next Step.

Step 5: The complete solution was transformed back into sub-chromosomes. Then, the sub-chromosomes went back into their sub-population.

Step 6: A new generation was produced. The new generation was produced by three processes; selection, crossover and mutation. The purpose of this step was to produce a new generation with a better quality. The new generation then went back to Step 2.

Step 7: The final solution was produced. The best solution obtained during the optimization was given in this step.

III. CASE STUDY: OPTIMIZATION OF TRYPTOPHAN BIOSYNTHESIS IN ESCHERICHIA COLI PATHWAY

To demonstrate the capability of the proposed method, the method was tested in a case study, which was the optimization of *trp* biosynthesis in *E. coli* pathway. A library named JAMA version 1.3 was used and then integrated with Java program to test the proposed method.

The production that needs to be improved in this pathway was *trp*. A detailed description of this pathway can be found in [10]. This pathway has the following GMA models:

$$\begin{aligned}\frac{dX_1}{dt} &= V_{11} - V_{12}\\ \frac{dX_2}{dt} &= V_{21} - V_{22}\\ \frac{dX_3}{dt} &= V_{31} - V_{32} - V_{33} - V_{34}\end{aligned} \tag{11}$$

All the reaction rates (denoted by *V*) have the following values:

$$\begin{aligned}V_{11} &= 0.6403 X_3^{-5.87\times10^{-4}} X_5^{-0.8332}\\ V_{12} &= 1.0233 X_1 X_4^{0.0035} X_{11}^{0.9965}\\ V_{21} &= X_1\\ V_{22} &= 1.4854 X_2 X_4^{-0.1349} X_{12}^{0.8651}\\ V_{31} &= 0.5534 X_2 X_3^{-0.5573} X_6^{0.5573}\\ V_{32} &= X_3 X_4\\ V_{33} &= 0.9942 X_3^{7.0426\times10^{-4}} X_7\\ V_{34} &= 0.8925 X_3^{3.5\times10^{-6}} X_4^{0.9760} X_8 X_9^{-0.0240} X_{10}^{-3.5\times10^{-6}}\end{aligned} \tag{12}$$

The production is given by reaction V_{34}, thus, it becomes the fitness function for the complete solution. The problem formulation for this pathway is given as follows:

$$max\, F_1 = V_{34} \tag{13}$$

$$min\, F_2 = \sum_{j=1}^{13} X_j \tag{14}$$

s.t. satisfying

$$\begin{aligned}V_{11} - V_{12} &= 0\\ V_{21} - V_{22} &= 0\\ V_{31} - V_{32} - V_{33} - V_{34} &= 0\end{aligned} \tag{15}$$

$$\begin{aligned}&X_j^{0.8} \le X_j \le X_j^{1.2} \qquad j = 1,2,3\\ &0 \le X_4 \le 0.00624\\ &4 \le X_5 \le 10\\ &500 \le X_6 \le 5000\\ &X_7 = 0.0022 X_5\\ &0 \le X_8 \le 1000\\ &X_9 = 7.5\\ &X_{10} = 0.005\\ &X_{11} = 0.9\\ &X_{12} = 0.02\\ &X_{13} = 0\end{aligned} \tag{16}$$

where Equation 13 is the production, Equation 14 is the total of chemical reaction involved, Equation 15 is the steady state

constraint and Equation 16 is the constraint of metabolic pathway components.

IV. RESULTS AND DISCUSSIONS

Several experiments were conducted as many parameter settings were involved. Table 1 gives the parameter setting in obtaining the best result. For Newton method, fixed parameters were used, with the maximum number of iterations of 50 and tolerance of 10^{-6}.

In this pathway, the proposed method was able to improve the *trp* production to 3.979 times from its initial steady state. The proposed method was also able to reduce the total of chemical concentrations involved to 6,006.589. Table 2 gives the best result obtained by the proposed method and the comparison with other methods. It was found that all variables (X_1-X_6 and X_8) were in their optimal range. For the *trp* production (F_1), this work produced the highest result compared to other works. This might be due to the implementation of GA, as GA is known as a good optimization method [11]. For the total of chemical concentrations involved (F_2), the proposed method was able to produce the lowest result. This might be attributed by the CCA method where CCA method decomposed the chromosome into many sub-chromosomes and only allowed the sub-chromosome with lower fitness value to combine with each other to form a complete chromosome. In conclusion, the proposed method has demonstrated the effectiveness in improving the *trp* production and at the same time able to reduce the total of chemical concentrations involved.

V. CONCLUSIONS

In this study, an improved method that combined Newton method, GA and CCA was presented. The method was proposed to improve the metabolic pathway production and simultaneously reduce the total chemical concentrations involved. The proposed method works by viewing metabolic pathway as a nonlinear equations system. Newton method was used to deal with nonlinear equations system, GA for fine-tuning variables in nonlinear system, while CCA to reduce the total chemical concentrations involved. Several experiments were performed on *E.coli* pathway and the result showed that the proposed method performed well when compared to other methods.

The main idea of this work is by combining Newton method, GA and CCA. As it has been stated before, GA is a good optimization method. However, it performs poorly in producing the best result as it requires 500 generations and thus time consuming to perform experiments. It requires a strategy to be applied in order to reduce the number of generation and time, and applying Pareto approach seems to be a good idea [12].

TABLE 1. PARAMETERS SETTING IN OBTAINING THE BEST RESULT

Parameter	Value
Number of sub-population	7
Number of sub-chromosome in sub-population	150
Maximum generation	500
Crossover point	0.9
Mutation rate	0.2

ACKNOWLEDGMENT

The authors would like to express their appreciation for the support of the E-Science Project (Project No. Vot 4S102) entitled "*The Development of Computational Modelling Tool using Hybrid Optimization Methods for Bioproduct Improvement in Cell Factories*" from the Malaysian Ministry of Science, Technology and Innovation, managed by the Research Management Centre, Universiti Teknologi Malaysia, and also the Fellowship scheme under Universiti Malaysia Pahang.

REFERENCES

[1] Voit EO. 2013. Biochemical Systems Theory: A Review. *ISRN Biomathematics*, in press.

[2] Zheng YN, Li LZ, Xian M, Ma YJ, Yang JM, Xu X and He DZ. 2009. Problems with the microbial production of butanol. *Journal of Industrial Microbiology & Biotechnology*, 36(9), 1127–1138.

[3] Link H, Vera J, Weuster-Botz D, Darias N T, and Franco-Lara E. 2008. Multi-objective steady state optimization of biochemical reaction networks using a constrained genetic algorithm. *Computers and Chemical Engineering*, 32(8), 1707–1713.

[4] Marin-Sanguino A, Voit EO, Gonzalez-Alcon C, and Torres NV. 2007. Optimization of biotechnological systems through geometric programming. *Theoretical Biology and Medical Modelling*, 4, 38-54.

[5] Xu G. 2013. Steady-state optimization of biochemical systems through geometric programming. *European Journal of Operational Research*, 225(1), 12–20.

[6] Vera J, Gonzalez-Alcon C, Marin-Sanguino A and Torres N. 2010. Optimization of biochemical systems through mathematical programming: Methods and applications. *Computers & Operations Research*, 37(8), 1427–1438.

[7] Mariano AP, Costa CBB, de Angelis DF, Filho FM, Atala DIP, Maciel MRW, and Filho RM. 2009. Optimization strategies based on sequential quadratic programming applied for a fermentation process for butanol production. *Applied Biochemistry and Biotechnology*, 159(2), 366–381.

TABLE 2. THE BEST SOLUTION AND COMPARISON WITH PREVIOUS WORKS

Variable	Initial steady state value	This work	Marin-Sanguino *et al*. [4]	Vera *et al*. [6]	Xu 2013 [5]
X_1	0.1847	0.8462	1.1900	1.2000	1.200
X_2	7.9868	0.8756	1.1480	1.1500	1.1150
X_3	1,418.9319	0.8000	0.8000	0.80000	0.8000
X_4	0.0031	0.0054	0.0041	0.0040	0.0054
X_5	5	4.0613	4	4	4.0110
X_6	2283	5,000	5,000	5,000	5,000
X_8	430	1,000	1,000	1,000	1,000
F_1	1.3102	**3.979**	3.0620	3.0620	3.946
F_2	4,145.1065	**6,006.5885**	6,007.1421	6,007.1540	6,007.1314

[8] Balsa-Canto E, Banga JR, Egea JA, Fernandez-Villaverde A, and Hijas-Liste GM. 2012. Global Optimization in Systems Biology: Stochastic Methods and Their Applications. In: Goryanin II and Goryachev AB (Eds.) *Advances in Systems Biology*. Springer, New York, USA, pp409–424.

[9] Grosan C and Abraham A. 2008. A New Approach for Solving Nonlinear Equations Systems. *Systems, Man and Cybernetics, Part A: Systems and Humans, IEEE Transactions on*, 38(3), 698–714.

[10] Xiu ZL, Zeng AP, and Deckwer WD. 1997. Model analysis concerning the effects of growth rate and intracellular tryptophan level on the stability and dynamics of tryptophan biosynthesis in bacteria. *Journal of Biotechnology*, 58(2), 125–140.

[11] Elsayed SM, Sarker RA, and Essam DL. 2014. A new genetic algorithm for solving optimization problems. *Engineering Applications of Artificial Intelligence*, 27, 57–69.

[12] Zitzler E and Thiele L. 1999. Multiobjective evolutionary algorithms: a comparative case study and the strength Pareto approach. *IEEE Transactions on Evolutionary Computation*, 3(4), 257–271.