A Hybrid of Newton Method and Genetic Algorithm for Constrained Optimization method of the Production of Metabolic Pathway

Mohd Arfian Ismail, Safaai Deris*, Mohd Saberi Mohamad, Afnizanfaizal Abdullah

· Artificial Intelligence and Bioinformatics Group, Department of Software Engineering, Faculty of Computing, Universiti Teknologi Malaysia, Johor, Malaysia
marfian2@live.utm.my, safaai@utm.my, saberi@utm.my, afnizanfaizal@utm.my

Abstract: In this work, constrained optimization method of the production in metabolic pathway is presented. The optimization of the production in metabolic pathway is a difficult task due as there are many components in the metabolic pathway. In addition, the condition of the constraint in improving the production of metabolic pathway should also be considered. In order to overcome this situation, this study presents an improved method in constrained optimization of the production in metabolic pathway. The proposed method consists of the Newton method and Genetic Algorithm (GA). The proposed method works with the Newton method by treating metabolic pathway as a non-linear equation system. Then, GA was applied to represent the variables in the non-linear equations system as a chromosome and fine-tune the chromosome to improve the variables. The proposed method was applied on several metabolic pathways to assess its performance. Several comparisons were conducted to evaluate the performance of the proposed method, and it was shown that the proposed method works well compared to the other methods.

[Ismail, MA, Deris, S. **A Hybrid of Newton Method and Genetic Algorithm for Constrained Optimization method of the Production of Metabolic Pathway.** *Life Sci J* 2014;11(9s):409-414]. (ISSN:1097-8135). http://www.lifesciencesite.com. 85

Keywords: Newton method, Genetic algorithm, constrained optimization, metabolic pathway

1. Introduction

In order to reduce the dependence on limited biomass resources, many researchers focused on microbial production to produce renewable biomass. This is due to the lower cost and time involved in producing renewable biomass. A way to produce renewable biomass from microbial production is by extracting the biomass from metabolic pathway. In general, metabolic pathway can be defined as a series of chemical reactions that can be found in microorganism cell. However, there is a limitation in extracting renewable biomass from metabolic pathway, which is low production (Zheng et al., 2009). Therefore, many researchers studied the optimization of the production of metabolic pathway (Marin-Sanguino, Voit, Gonzalez-Alcon, & Torres, 2007; Rodriguez-Acosta, Regalado, & Torres, 1999; Vera, Gonzalez-Alcon, Marin-Sanguino, & Torres, 2010; Xu, 2013).

The optimization of the production of metabolic pathway can be performed as metabolic pathway can be represented by a mathematical model. The mathematical model that represents metabolic pathway is known as ordinary differential equation (ODE) (Voit, 2013). In ODE model, all the components in metabolic pathway are represented by numerical values. These values can be manipulated and changed. As a result, the optimization of metabolic pathway production involves a fine-tuning process of the metabolic pathway components with the aim of improving the production.

The optimization process becomes complicated as it involves a large of metabolic pathway that contain many components, therefore making the fine-tuning process becomes complicated. Besides that, steady-state constraints and metabolic pathway component constraints need to be considered (Link, Vera, Weuster-Botz, Darias, & Franco-Lara, 2008). In dealing with this situation, this study introduces an improved method that combines the Newton method and Genetic Algorithm (GA). The proposed method works with the Newton method views the metabolic pathway as a non-linear equation system. Then, GA is utilized to represent the variables in the metabolic pathway as a chromosome, and consequently fine-tune the variables in order to search for the best set of solution.

2. Modeling of metabolic pathway

Within ODE, there are two types of models in representing metabolic pathway into a mathematical model, which are S-system and generalized mass action (GMA) model. This work focuses on GMA because of its performance (Marin-Sanguino et al., 2007). The representation of metabolic pathway within GMA model is as follows:

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

In Equation 1, *S* is the stoichiometric matrix of the GMA model, while *v(x)* is the vector that contains reaction rate. The vector *v(x)* is in a linear form, which is shown as follows:

$$v_i = \gamma_i \prod_j x_j^{f_{ij}} \quad (2)$$

In Equation 2, there are two coefficients that are derived from the Taylor series in the logarithmic space around a steady state (Marin-Sanguino & Torres, 2003; Xu, 2013). These coefficients are denoted by γ_i and f_{ij}, and these coefficients represent the rate constant and kinetic order for constant v_i, respectively. These coefficients are in the following form:

$$\gamma_i = |v_i|_0 \quad (3)$$

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

In Equation 3, subscript 0 is assigned to the value at a steady state condition.

3. Problem statement

The constrained optimization method of the production of metabolic pathway involves the fine-tuning process of the metabolic pathway components in order to improve the production. The fine-tuning process cannot be done arbitrarily. This is because there are some constraints that must be followed, which are the steady state constraint and the constraint of the metabolic pathway components. The steady state constraint is a condition where all components in the metabolic pathway are in static value. This will make all ODE models (Equation 1) equal to 0, thus producing the following form of Equation 1:

$$\frac{dX_n}{dt} = \left[sv(x)_1, \ldots sv(x)_n \right] = 0 \quad (5)$$

As all ODE models are equal to 0, this leads to a non-linear equations system. In a non-linear equations system, all of the equations are equal to 0 (Baghmisheh, Mahmoudi, & Jahangirirad, 2013; Grosan & Abraham, 2008). This situation is similar to the constrained optimization of the metabolic pathway production. As a result, the constrained optimization of the production of metabolic pathway can be considered as solving a non-linear equations system. The system has the following form:

$$f(x) = \left[f(x)_1, \ f(x)_2, \ f(x)_3 \ \ldots f(x)_n \right] = 0 \quad (6)$$

where x= (x_1, x_2, x_3,... x_n) is n equations and n variables in the non-linear equations system. $f(x)_1$, $f(x)_2$, $f(x)_3$.... $f(x)_n$ are the nonlinear functions within the system.

For the constraint of the components in metabolic pathway, the components must remain in the optimal range to ensure the survival of the microorganism cell (Link et al., 2008). The constrained optimization method of the production of metabolic pathway can be formulated as follows:

$$\max f(v) \quad (7)$$

s.t. satisfying

$$sv(x)_i = 0, \qquad i = 1, 2, \ldots, n \quad (8)$$

$$x_j^L \le x_j \le x_j^U \qquad j = 1, 2, \ldots, m \quad (9)$$

where Equation 7 is the production of metabolic pathway (objective function). The steady state constraint is given by Equation 8. Meanwhile, Equation 9 is the constraint of the metabolic pathway component, where it denotes the specific range. L denotes the lower range while U represents the upper range.

4. The hybrid of Newton Method and Genetic Algorithm

The hybrid of the Newton method and GA is intended to improve the metabolic pathway production. The proposed method functions by treating metabolic pathway as a non-linear equations system. Then, GA is applied to fine-tune the variables in the non-linear equations system to search for the best set of solution. A flow chart of the proposed method is depicted in Figure 1. The steps of the proposed method are given as follows:

Step 1: Generating initial chromosome. This step generates N chromosome randomly in a population. The chromosome represents all variables in the non-linear equations system, and the chromosome is in binary format.

Step 2: Encoding chromosome into variables. In this step, the chromosome is transformed from the binary representation into the variables in a nonlinear equations system. The variables are in a real number format and known as candidate solution. Figure 2 shows the encoding process.

Step 3: Evaluating with the Newton method. In this step, all candidate solutions are tested by the Newton method. A termination condition occurred in this step, either the candidate solutions are optimized or not. The termination condition can occur in three situations, which are when the maximum number of generations is achieved, the steady state constraint is fulfilled, and the component constraints are fulfilled. If all of these situations are achieved, then the proceeds to the subsequent step.

Step 4: Decoding variables into chromosome. This step involves the transformation of the variables back into chromosome form. Figure 3 illustrates the decoding process of the variables into chromosome form.

Step 5: Selecting chromosome. This step is about choosing a pair of chromosome for reproduction. The selection process is based on the chromosome's fitness value, where the highest value will be chosen.
Step 6: Reproducing new generation. In this step, two genetic operators, crossover, and mutation are applied to produce a new generation. This step is intended to improve the quality of the candidate solution. Then, a new generation goes back to Step 2.
Step 7: Returning the final solution. This step gives the best set of solution that is achieved during the optimization process.

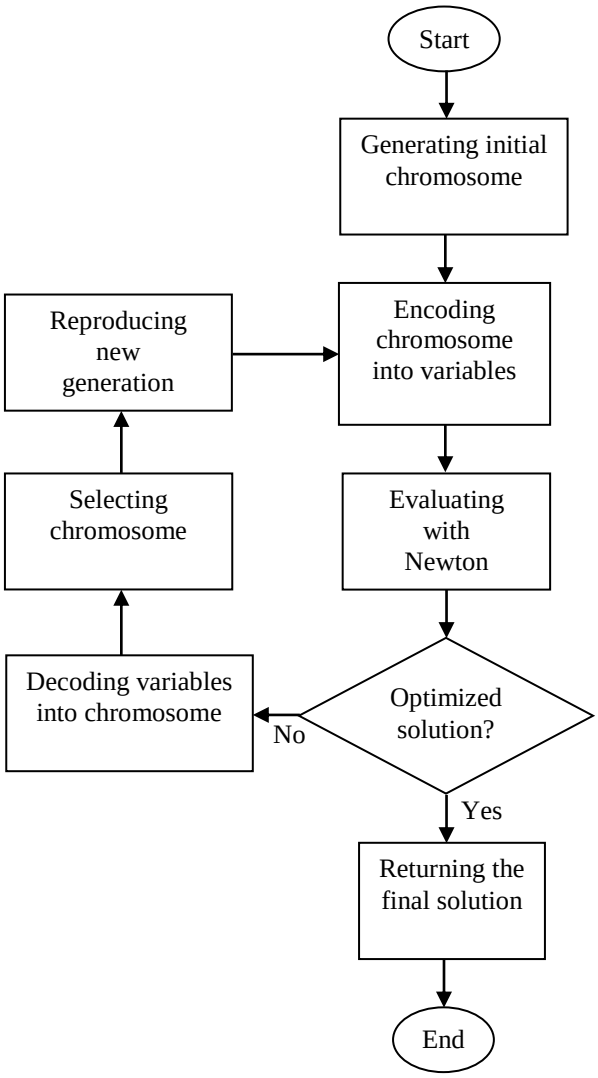

Figure 1 : The flow chart of the proposed method

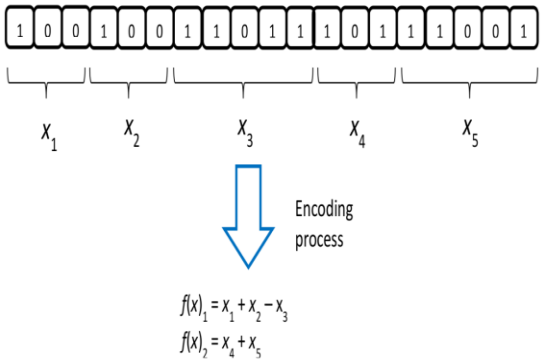

Figure 2 : The encoding process

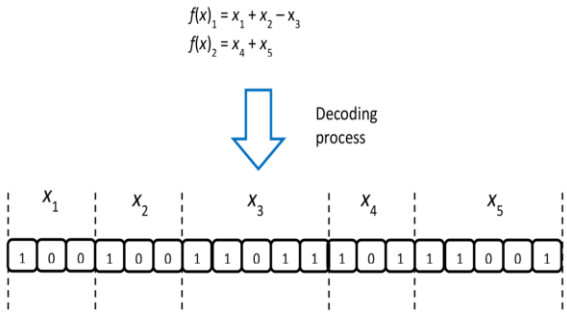

Figure 3 : The decoding process

5. Case studies

To illustrate the capability of the proposed method, the method was tested on two case studies. A Java program was developed and integrated with JAMA version 1.3 to test the proposed method.

5.1. Case study 1: Optimization of tryptophan biosynthesis in *Escherichia coli* pathway

In this case study, the end product of *Escherichia coli* (*E.coli*), *tryptophan* (*trp*) was optimized. A detailed description of this pathway can be found in the works done by Xiu and colleagues (Xiu, Zeng, & Deckwer, 1997). This pathway formulates the ODE models in the following form:

$$\frac{dX_1}{dt} = V_{11} - V_{12}$$
$$\frac{dX_2}{dt} = V_{21} - V_{22} \quad (10)$$
$$\frac{dX_3}{dt} = V_{31} - V_{32} - V_{33} - V_{34}$$

At the steady state condition, the rate of all reactions are as follows:

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

In case study 1, the optimization of *trp* was given by reaction V_{34} (Marin-Sanguino & Torres, 2000). Thus, it became the fitness function of the chromosome in GA. The problem statement for the optimization for this pathway can be formulated as follows:

$$max\, F = V_{34} \tag{12}$$

For the steady state constraint, Equation 8 of case study 1 is given as follows:

$$
\begin{aligned}
V_{11} - V_{12} &= 0 \\
V_{21} - V_{22} &= 0 \\
V_{31} - V_{32} - V_{33} - V_{34} &= 0
\end{aligned} \tag{13}
$$

In this pathway, not all reactions (components) were tuned. Thus, the constraints for all components were formulated as follows:

$$
\begin{aligned}
& X_j^{0.8} \le X_j \le X_j^{1.2} \qquad j = 1,2,3 \\
& 0 \le X_4 \le 0.00624 \\
& 4 \le X_5 \le 10 \\
& 500 \le X_6 \le 5000 \\
& X_7 = 0.0022 X_5 \\
& 0 \le X_8 \le 1000 \\
& X_9 = 7.5 \\
& X_{10} = 0.005 \\
& X_{11} = 0.9 \\
& X_{12} = 0.02 \\
& X_{13} = 0
\end{aligned} \tag{14}
$$

5.2. Case study 1: Optimization of ethanol production in *Saccharomyces cerevisiae* pathway

In case study 2, the proposed method was used to optimize ethanol production that was produced in *Saccharomyces cerevisiae* (*S. cerevisiae*) pathway. The detail of this pathway can be found in the work performed by Galazzo and Bailey (1990). The ODE models of this pathway have the following form:

$$
\begin{aligned}
\frac{dX_1}{dt} &= V_{in} - V_{HK} \\
\frac{dX_2}{dt} &= V_{HK} - V_{PFK} - V_{Carb} \\
\frac{dX_3}{dt} &= V_{PFK} - V_{GAPD} - 0.5 V_{Gro} \\
\frac{dX_4}{dt} &= 2V_{GAPD} - V_{PK} \\
\frac{dX_5}{dt} &= 2V_{GAPD} + V_{PK} - V_{HK} - V_{Carb} - V_{PFK} - V_{ATPase}
\end{aligned} \tag{15}
$$

At the steady state condition, all fluxes in Equation 15 have the following rate:

$$
\begin{aligned}
V_{in} &= 0.8122 X_2^{-0.2344} Y_1 \\
V_{HK} &= 2.8632 X_1^{0.7464} X_5^{0.0243} Y_2 \\
V_{PFK} &= 0.5232 X_2^{0.7318} X_5^{-0.3941} Y_3 \\
V_{Carb} &= 8.904\times10^{-4} X_2^{8.6107} Y_6 \\
V_{GAPD} &= 7.6092\times10^{-2} X_3^{0.6159} X_5^{0.1308} Y_4 \\
V_{Gro} &= 9.272\times10^{-2} X_3^{0.05} X_4^{0.533} X_5^{-0.0822} Y_7 \\
V_{PK} &= 9.471\times10^{-2} X_3^{0.05} X_4^{0.533} X_5^{-0.0822} Y_5 \\
V_{ATPase} &= X_5 X_8
\end{aligned} \tag{16}
$$

For case study 2, ethanol was given by the flux V_{PK}, thus it became the fitness function of chromosome in GA. The optimization for case study 2 was formulated as follows:

$$max\, F = V_{PK} \tag{17}$$

At steady state condition, Equation 1 became as follows:

$$
\begin{aligned}
V_{in} - V_{HK} &= 0 \\
V_{HK} - V_{PFK} - V_{Carb} &= 0 \\
V_{PFK} - V_{GAPD} - 0.5 V_{Gro} &= 0 \\
2V_{GAPD} - V_{PK} &= 0 \\
2V_{GAPD} + V_{PK} - V_{HK} - V_{Carb} - V_{PFK} - V_{ATPase} &= 0
\end{aligned} \tag{18}
$$

In this pathway, the components in metabolic pathway were divided into two categories; metabolite concentration and enzyme concentration. For the metabolite concentration constraint, the constraint was set to the range of 0.8 to 1.2 (Link et al., 2008; Rodriguez-Acosta et al., 1999) and given as follows:

$$X_j^{0.8} \le X_j \le X_j^{1.2} \qquad j = 1,2,3,4,5 \tag{19}$$

For the enzyme concentration constraint, the range was set to the range of 0 to 50 (Link et al., 2008; Rodriguez-Acosta et al., 1999). The enzyme concentration constraint is given as follows:

$$Y_j^0 \le Y_j \le Y_j^{50} \qquad j = 1,2,3,4,5,8 \tag{20}$$

6. Results and discussions

Several parameters were used in performing the experiments. Table 1 gives the GA parameters setting used in producing the best result. For the Newton method, fixed parameters were used, where the number of iterations was fixed to 100, and the tolerance was set to 10^{-6}.

Table 1: GA parameters used to produce the best result

Parameter	Case study 1	Case study 2
No of chromosomes	250	270
No of generations	300	300
Crossover probability	0.4	0.3
Mutation probability	0.1	0.1

For case study 1, the proposed method was able to increase the production to 3.96 times compared to its steady state value. Detailed results, including the average result and the comparison with other methods are given in Table 2. These results were produced by using the parameters in Table 1. From Table 2, it can be observed that the proposed method produced the result that satisfied all the components constraint. Besides that, the average result also satisfied all the component constraint. It can be concluded that the proposed method was able to produce reliable results. In accessing the performance of the proposed methods, the results were compared with those obtained in other previous works. It was shown that the proposed method outperformed other methods in improving the *trp* production.

Table 2: Detailed results of case study 1

Variable	Initial steady state	The best result of the proposed method	Average of the proposed method	(Marin-Sanguino et al., 2007)	(Vera et al., 2010)	(Xu, 2013)
X_1	0.184654	1.11	1.09	1.99	1.99	1.20
X_2	7.986756	1.114	1.133	1.148	1.148	1.115
X_3	1418.931944	0.8	0.8	0.8	0.8	0.8
Y_1	0.00312	0.00538	0.00538	0.00414	0.00414	0.00536
Y_2	5	4.754	4.543	4.000	4.000	4.011
Y_3	2283	5000	5000	5000	5000	5000
Y_5	430	1000	1000	1000	1000	1000
F	1.310202	3.957	3.955	3.062	3.062	3.946

In case study 2, the production obtained using the proposed method was 52.57 from its steady state value. Table 3 gives the best result obtained, the average result, and the comparison with other methods. These results were obtained by using parameters in Table 1. An observation that can be made from Table 3 is, all components of the best result fulfill their constraint. Moreover, the components of the average result also fulfill their constraint, thus confirms that the proposed method was able to produce reliable results. Based on the performance of the method, the proposed method was able to produce more ethanol compared to other methods.

Table 3: Detailed results of case study 2

Variable	Initial steady state	The best result of the proposed method	Average of the proposed method	(Rodriguez-Acosta et al., 1999)	(Xu, 2013))
X_1	0.0345	1.11	1.03	1.14	1.10
X_2	1.0110	1.03	0.95	1.05	1.05
X_3	9.1440	1.13	1.00	1.15	1.14
X_4	0.0095	1.18	1.17	1.17	1.17
X_5	1.1278	1.14	1.01	1.12	1.11
Y_1	19.70	49.99	49.98	49.97	50.00
Y_2	68.50	45.83	45.50	44.77	45.95
Y_3	31.70	49.92	49.92	49.89	50.00
Y_4	49.90	47.97	47.71	47.26	47.77
Y_5	3440.00	48.30	48.21	48.00	48.37
Y_8	25.10	49.79	49.88	49.75	50.00
F	30.11	52.57	52.47	52.31	52.38

7. Conclusion

In this work, a hybrid of the Newton method and GA was presented. The proposed method was employed in the optimization of the production of metabolic pathway. The optimization process becomes complicated when the steady state

constraint and the constraint of metabolic pathway component are involved. To overcome this situation, the Newton method was applied in dealing with metabolic pathway, and then GA was used to fine-tune the components in the metabolic pathway. The proposed method was applied on two case studies, and it was shown that the proposed method works better compared to other works. In addition, the proposed method was also able to produce reliable results.

Acknowledgements:

The authors would like to express their appreciation for the support of the sponsors of the E-Science Project (Project No. Vot 4S102), entitled "*The Development of Computational Modeling Tool using Hybrid Optimization Methods for Bioproduct Improvement in Cell Factories*" from the Malaysian Ministry of Science, Technology and Innovation and the Potential Academic Staff grant, entitled "*Synthetic Modelling of Hyperosmotic Stress in Yeast Glycerol Production*" from Universiti Teknologi Malaysia. This work was also supported by the Fellowship scheme under Universiti Malaysia Pahang and the Malaysian Ministry of Education. Thanks are expressed to the editor and anonymous reviewers who reviewed this paper.

Corresponding Author:
Safaai Deris
Artificial Intelligence and Bioinformatics Group (AIBIG),
Department of Software Engineering,
Faculty of Computing, Universiti Teknologi Malaysia,
Johor, Malaysia.
E-mail: safaai@utm.my

References

1. Baghmisheh, M., Mahmoudi, Y., & Jahangirirad, M. (2013). A new modification of Newton's method by Gauss integration formula. Life Science Journal, 10(5), 288–291.
2. Grosan, C., & 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.
3. Link, H., Vera, J., Weuster-Botz, D., Darias, N. T., & 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., & Torres, N. V. (2003). Optimization of biochemical systems by linear programming and general mass action model representations. Mathematical Biosciences, 184(2), 187–200.
5. Marin-Sanguino, A., Voit, E. O., Gonzalez-Alcon, C., & Torres, N. V. (2007). Optimization of biotechnological systems through geometric programming. Theoretical Biology and MedicalModelling, 4.
6. Rodriguez-Acosta, F., Regalado, C. M., & Torres, N. V. (1999). Non-linear optimization of biotechnological processes by stochastic algorithms: Application to the maximization of the production rate of ethanol, glycerol and carbohydrates by Saccharomyces cerevisiae. Journal of Biotechnology, 65(1), 15–28.
7. Vera, J., Gonzalez-Alcon, C., Marin-Sanguino, A., & Torres, N. (2010). Optimization of biochemical systems through mathematical programming: Methods and applications. Computers & Operations Research, 37(8), 1427–1438.
8. Voit, E. O. (2013). Biochemical Systems Theory: A Review. ISRN Biomathematics, 2013.
9. Xiu, Z.-L., Zeng, A.-P., & Deckwer, W.-D. (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.
10. Xu, G. (2013). Steady-state optimization of biochemical systems through geometric programming. European Journal of Operational Reseach, 225(1), 12–20.
11. Zheng, Y.-N., Li, L.-Z., XIan, M., Ma, Y.-J., Yang, J.-M., Xu, X., & He, D.-Z. (2009). Problems with the microbial production of butanol. Journal of Industrial Microbiology & Biotechnology, 36(9), 1127–1138.

7/5/2014