Optimization of the Culture Medium for Rhodospirillum Rubrum S1 with an Artificial Neural Network Model and GA

Yan Pan, Weijian Wang, haijuan Tian, Shihao Li, Zhu Zhu*

JiLin province key Laboratory of Grain and Oil processing, Jilin Business and Technology College
Changchun, 130062, P. R. China
E-mail: zzabc2008@126.com

Abstract: on the basis of the data from the previous box-behnken central composite design, an Artificial Neural Network (ANN) model was constructed for the prediction of outputs of carotenoids. GA (genetic algorithm) was used to search for the optimal culture medium for Rhodospirillum Rubrum S1:citric acid 3.678g/L, Beef extract 3.407 g/L, $MgSO_4$ 0.524g/L, $FeSO_4$0.023g/L. In the optimal culture medium, it was predicted that the outputs of the carotenoids were13.85 mg/ L.After three verification experiments, the outputs of the carotenoids were 13.72mg/L, the error between the expected value and the experimental value was 0.93%.

Keyword: ANN; GA; Rhodospirillum Rubrum S1

Ⅰ. INTRODUCTION

Rhodospirillum Rubrum S1 is one of the oldest and most primitive bacteria, abundant in photosynthetic pigments such as Cars and Bphls. Bphls are mainly responsible for capturing light, whereas Cars are a major source of nutrients. Cars are also an important kind of pigment, widespread in nature.[1]. Nowadays, Photosynthetic bacteria (PSB) are the main kind of microorganisms that produce Cars. They have the following advantages, nutrition requirements being low, raw materials being cheap and abundant, production cycles being short, non-toxic and not restricted by seasons [2,3] It is estimated that the output of carotenoids in nature is up to 100 million tons every year [4,5].Several methods have been investigated for optimization of the culture medium for Rhodospirillum Rubrum S1 [6,7].Satistical techniques like response surface methodology (RSM) are also being increasingly used in the Optimization of the Culture Medium [8].The development of accurate models for a biological system on a chemical and physical basis is still a critical challenge, mainly due to the complex and highly nonlinear nature of the extraction conditions. An artificial neural network, based on GA, is an algorithm which is good at simulating and modeling various aspects of biological information in food science and biochemical engineering [9]. The forecast accuracy of an ANN model is superior to that of RSM in the same experiment design [10].GA, the implementation of computer retrieval and nonlinear optimization, adopts the principles of natural genetics and natural selection [11]. It simulates the process of biological evolution known as "survival of the fittest", improving solutions from the previous generation before reaching an almost optimal solution [12]. In the paper, the culture medium for Rhodospirillum Rubrum S1 was optimized by following the steps below. (ⅰ) In developing the ANN model, the influential procedure variables are used for input data, and the extraction rate of the protein for output data. (ⅱ) The optimization of input space of the ANN model was conducted with GA in accordance with the maximization of the extraction rate. (ⅲ) Experiments were carried out to test the feasibility of the ANN model.

Ⅱ. MATERIALS AND METHODS

A. Materials

The strains of *Rhodospirillum Rubrum* S1 were provided by .College of Life Science, Jilin University

B. Methods

a)NN-Based Modeling of Protein Extraction

An Artificial Neural Network (ANN) is a nonlinear computing model based on biological neural networks. It simulates the learning process of human brains by the use of a mathematical model of an interconnected group of artificial neurons in a network. It is reported that the precision for prediction of ANN is higher that response surface methodology (RSM) if with the same experiment design [13] .In this experiment, the Artificial Neural Network consists of three layers, the input layer, the hidden layer and the output layer, which are exploited to search for the optimal data of Box-behnken. Sigmoid function is employed between the input layer and the hidden layer, and purelin function is for the hidden and the output layers. ANN derives from Levenberg-Marquardt algorithm. The data in Box-Behnken experiment design are classified into three groups, including a calibration set, a test set and a pretest set. The calibration set is used to establish an artificial neural network model; for the testing and pretest sets, they are aimed at examining the induction and forecast abilities of the model. As some hidden neurons can have a powerful influence on the abilities of the artificial neural network model, the following function (1) used to select appropriate neurons.

$$D_a = \frac{c}{\frac{n_c}{n}\times MSE_c + \frac{n_t}{n}\times MSE_t + \left|MSE_c - MSE_t\right|} \quad (1)$$

The function, MSEC and MSET refer to MSE of the calibration set and the test set and n_c and n_t are their numbers. N is the total number of the calibration set and the test set. C is a continuous number obtained after Da is adjusted. (in this study, the value of c is 1000.) it can be learned from the above that the bigger the value of Da becomes, the closer the artificial neural network model gets to the experimental data. In the optimal ANN model, initial

weights are extremely important. In this experiment, the ANN model has practiced 30 times with the initial weights which were given randomly so that the value of Da has selected the fittest ANN.When an ANN model is set up, its memory space will be taken full advantage by GA, which is a nonlinear optimization. GA has been regarded as the most effective technique in solving various problems for optimization in biological engineering [6, 7].

b) Data analysis and software

SAS 8.02 was used to design the experiment, establishing a linear regression model and a two regression model. Meanwhile, it was also used to count and analyze the data in the experiment. Matlab 7.6.0.324 was for setting up an ANN model and implement GA.

Ⅲ. RESULTS AND DISCUSSION

A. The results of RSM design experiments

The culture medium for *Rhodospirillum Rubrum* S1 had been optimized previously using RSM method.. The culture medium for *Rhodospirillum Rubrum* S1 was tested at three levels (-1, 0, 1) (table 1). The Experimental design and the outputs of the carotenoids were shown in Table 2, te maximum the outputs of the carotenoids by RSM model was13.8 mg/ L.

TABLE I. VARIABLES AND THEIR LEVELS USED IN RSM

Independent	Number	Variable levels		
		(−1)	(0)	(+1)
citric acid /g·L^{-1}	X_1	3.0	3.4	3.8
Beef extract /g·L^{-1}	*X2*	3.0	3.3	3.6
$MgSO_4$/g·L^{-1}	*X3*	0.4	0.5	0.6
$FeSO_4$/g·L^{-1}	*X4*	0.01	0.02	0.03

TABLE II. EXPERIMENTAL DESIGN AND RESULTS OF RSM

serial number	X_1	X_2	X_3	X_4	Y
1	-1	-1	0	0	9.54
2	-1	1	0	0	11.02
3	1	-1	0	0	12.46
4	1	1	0	0	11.54
5	0	0	-1	-1	10.12
6	0	0	-1	1	11.06
7	0	0	1	-1	11.84
8	0	0	1	1	10.42
9	-1	0	0	-1	9.84
10	-1	0	0	1	11.56
11	1	0	0	-1	12.34
12	1	0	0	1	11.68
13	0	-1	-1	0	10.58
14	0	-1	1	0	10.2
15	0	1	-1	0	11.06
16	0	1	1	0	11.86
17	-1	0	-1	0	11.12
18	-1	0	1	0	11.58
19	1	0	-1	0	11.98
20	1	0	1	0	12.22
21	0	-1	0	-1	11.68
22	0	-1	0	1	10.46
23	0	1	0	-1	11.28
24	0	1	0	1	12.6
25	0	0	0	0	13.76
26	0	0	0	0	13.52
27	0	0	0	0	13.66

*X_1, X_2, X_3 and X_{43} represent citric acid, Beef extract, $MgSO_4$, $FeSO_4$.

B. ANN-Based Modeling of Protein Extraction

A set of data was sampled as the forecast set from the box-behnken experiment, and another set was selected as the test set. All other data were used as calibration sets. Through the three layers, the data were returned to the ANN model. Then, proper hidden neurons were chosen in accordance with the value of Da. As shown from Figure 1, the value of Da reaches maximum at the twenty four neuron, so the number of the hidden neuron was set twenty-four.The coefficient of determination (R2) of the ANN model after optimization was 0.95766.the above results show that the ANN model is universally valid and has excellent forecasting abilities.

C. GA-ANN Model

Then, GA was used to search for the optimal values. The parameters for GA were as follows, the population type being double vector, the population size being 20. The initial population was provided randomly. As for the random selection function, exact algorithm was 2, crossing-over rate was 0.8, migration fraction was 0.2 and penalty factor was 100. The number of generations got by GA was 100. The number of the generation was shown in Fig. 2. As can be seen, when the number of generation was 38,the fitness reached the lowest levels; when the number of generation continued to increase, the fitness stayed at the lowest level.

The code values of the culture medium after optimization by means of GA included X1=-0.103 ; X2=0.178 , X3=0.235, X4=0.044. The corresponding experimental values consisted of X1=3.678, X2=3.407 , X3=0.524,X4=0.023. The expected outputs of carotenoids was 13.85 mg/ L in these optimized the culture medium. Then, the test experiment was conducted. The mean obtained after three parallel experiments was13.72g/ L,the error between the expected value and the experimental value was 0.93%, which shows the forecasting precision of the ANN model is better than that of the two regression model.

Ⅳ. CONCLUSIONS

It is shown in the paper that the modeling with the combination of RSM and ANN-GA is a good tool for Optimization of the culture medium for *Rhodospirillum Rubrum* S1. RSM is a statistics tool for identifying interactions among different factors, whereas an ANN model is a more precise simulation technique for its

nonlinearity. The results show that it was predicted that the outputs of the carotenoids were13.85 mg/ L.After three verification experiments, the outputs of the carotenoids were 13.72 mg/ L, the error between the expected value and the experimental value was 0.93%.Which shows the method can be applied to other system modeling and optimization in biological engineering.

REFERENCES

[1] Neli s H J, De Leenheer A P. Microbial Sources of carotenoids pigment s us ed in f oods and f eed s [J] . J Appl Bacteriol , 1991, 70: 181-191.

[2] Buzzini P. An optimization study of carotedoid production by Rhodotorula lutinis DBVPG 3853 from substrates containing concentrated rectified grape must as the sole carbohydrate source [J]. J Ind Microbial Biotechnol , 2000, (24) :41-45

[3] Frengova G, Simova E, Pavlova K, et al. Formation of carotenoids by Rhodotorula glutinis in whey ultrafiltrate [J] .Biotechnol Bioeng , 1994, 44(8) : 888-894.

[4] BARTLEY G E, SCOLNIK P A. Plant carotenoids: pigments for photoprotection,visual attraction and human health[J]. Plant Cell, 1995(7): 1027-1038.

[5] NIYOGI K K. Photo protection revisited J]. Annu Rev Plant Physlol Plant Mol Bio1, 1999, 50: 391-417.

[6] Xu,W.; Reddy, N.; Yang, Y, "An acidic method of zein extraction from DDGS." Agric. Food Chem, 2007, 55, pp. 6279–6284.

[7] Cookman, J; Glatz, C. E,"Extraction of protein from distiller's grain." Bioresour. Technol. 2009, 100, pp. 2012–2017.

[8] AlmeidaJ.S,"Predictive non-linear modeling of complex data by artificial neural networks."Curr. Opin. Biotechnol, 2002, 13, pp. 72–76.

[9] Venkatasubramanian, V., Sundaram, A, "Genetic algorithms: introduction and applications." Encyclopedia of Computational Chemistry, 1998, pp. 1115–1127.

[10] Anderson-Cook C. M., Borrir C. M. and Montgomery D. C. "Response surface design evaluation and comparison.Journal tatistical Planning and inference." 2009, 139, pp. 629-641.

[11] Adel Mellit, Soteris A. Kalogirou , Mahmoud Drif, "Application of neural networks and genetic algorithms for sizing of photovoltaic systems." Renewable Energy, 2010, 35, pp. 2881–2893.

[12] Stuart Russell and Peter Norvig, "Artificial Intelligence A Modem Approach- 2nd edition", Local Search Algorithms and Optimization Problems- Genetic algorithms, Prentice Hall, 2003. pp. 116-119.

[13] Desai K. M., Akolkar Y. P. and Badhe Y. P. (2006).Optimization of fermentation media for exopolysaccharide production from Lactobacillus plantarum using artificial intelligence-based techniques. Process Biochemistry 41: 1842-1848

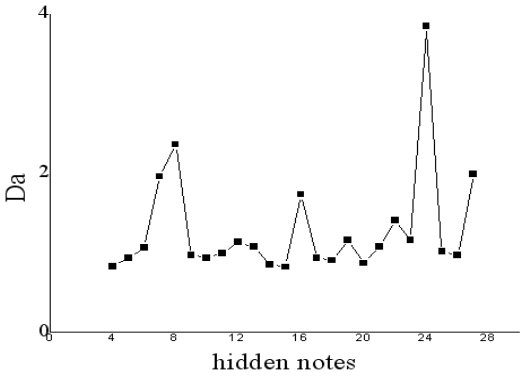

Figure 1 he influence by the number of hidden neurons

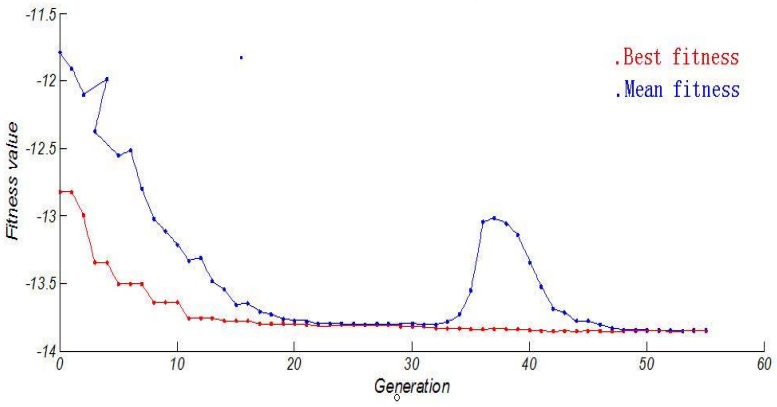

Figure 2 the influence of generations on simulation degree