Phellinus is a kind of fungus known as one of the elemental components in drugs to avoid cancers. With the purpose of finding optimized culture conditions for Phellinus production in the laboratory, plenty of experiments focusing on single factor were operated and large scale of experimental data were generated. In this work, we use the data collected from experiments for regression analysis and then a mathematical model of predicting Phellinus production is achieved. Subsequently, a gene-set based genetic algorithm is developed to optimize the values of parameters involved in culture conditions, including inoculum size, PH value, initial liquid volume, temperature, seed age, fermentation time, and rotation speed. These optimized values of the parameters have accordance with biological experimental results, which indicate that our method has a good predictability for culture conditions optimization.

1. Introduction

Phellinus is a kind of fungus having great medicinal value, since it is known as one of the elemental components in drugs with functions of avoiding cancers [1, 2]. Phellinus flavonoids are one of the most popular parasitifiers of Phellinus in nature [3], and the research on Phellinus focuses on polysaccharides, proteoglycans medicinal mechanism, composition, and so forth, which are mostly extracted from the fruiting bodies of Phellinus flavonoids [4]. Phellinus rarely exists in the wild environment [5], and it becomes a promising research branch to cultivate it in the laboratory. With mycelial growth by liquid fermentation, the fermentation broth flavonoids, polysaccharides, alkaloids, and other active substances can be produced, which have high level physical activity, short fermentation period, and mass productions, thus providing a possible way of producing Phellinus in the laboratory [6]. In recent years, updated machine learning approaches (see, e.g., [7, 8]) have been developed and applied in biological data processing.

From the understanding of the wild conditions of Phellinus, it is believed that PH value, temperature, and fermentation time have effect on the productions. Also, in general biochemical experiments, we need to consider the inoculum size, initial liquid volume, seed age, and rotation speed. In the laboratory, plenty of experiments have been designed and operated for maximizing the Phellinus production. The methods can be separated into two major groups.

(i) With biological technologies: it used optimum media on mycelial growth of Phellinus in [9] and liquid fermentation technology to cultivate Phellinus in [10]. Active ingredients in Phellinus and polysaccharide metabolism regulation are designed in [11].

(ii) With mathematical models: some researches focus on building mathematical models for the progress of producing Phellinus by differential equations [12], metabolic path and network [13], and complex network models [14].
Artificial algorithms and models have been used in the bioprocess, particularly for the optimization of culture conditions. In [15], artificial neural network (ANN) is used to optimize the extraction process of azalea flavonoids. Neural networks combined with evolutionary algorithms have been used to optimize the experimental environment, such that neural network and particle swarm optimization method were used for finding optimized culture conditions to maximize the production of Pleurotus mutillus in [16]. Recently, with the increment of biological data, regression analysis becomes a useful tool for the data analysis. In [17] the method of fitting models to biological data using linear and nonlinear regression is proposed, where some multivariate statistical analysis strategies from [18, 19] are formulated to be helpful and useful for biologists. These results give us hints of using regression analysis and artificial algorithms to optimize the culture conditions for Phellinus production. And, to the best of our knowledge, few work focuses on the optimization of culture conditions to maximize the production of Phellinus in the laboratory.

In this work, we start from operating 45 experiments for producing Phellinus from Phellinus flavonoids with different culture conditions, involving parameters PH value, temperature and fermentation time, inoculum size, initial liquid volume, seed age, and rotation speed. With the data collected during the experiments, we use regression analysis method to create a mathematical model, which can forecast the flavonoid yield and the most important element to the production of Phellinus. After that, a gene-set based genetic algorithm (GA) is proposed to optimize the culture condition, where the obtained mathematical model is used as fitness function for the evolution of individuals. Data experimental results show that predicted optimal values of the parameters have accordance with biological experimental results, which indicate that our method has a good predictability for culture conditions optimization.

2. Data Collected from Experiments

In this section, biological experiments are performed for finding optimal value of certain single factor.

In Table 1, experiments are operated for collecting data. In rows 1–14, it is associated with experiments with PH values ranging from 1 to 14, where the temperature is fixed to 28°C, initial volume is set to be 100 mL, the rotation speed is 140 r/min, and seed age is 8 days. Rows 15 to 20 are 6 experiments with initial volume ranging from 40 mL to 140 mL, where PH value is set to be 6, the best one obtained from experiments with PH values ranging from 1 to 14.

In Table 2, experiments with including inoculum ranging from 2% to 16% and temperature ranging from 25°C to 40°C are performed. In Table 3 the situations on experiments with fermentation time ranging from 1 to 12 hours are shown. From the in total 45 experiments, we collect data of culture conditions for production of Phellinus. Different culture conditions have a fundamental influence on the production of Phellinus, but the optimized culture conditions remain unknown.

3. Methods

We consider here using regression analysis and gene-set based genetic algorithm to find the optimized culture conditions for maximizing the production of Phellinus. In general, we convert the data collected in Section 2 to construct a mathematical model by regression analysis. And then, the obtained model can be used as fitness function for optimizing the culture condition with gene-set based genetic algorithm.

3.1. Regression Analysis. In statistical modeling, regression analysis is a statistical process for estimating the relationships among variables [20]. Regression analysis is one of the extremely versatile data analysis methods, which is appropriated to establish dependencies between variables based on observational data and widely used to analyze the data inherent law and to predict the result. Regression analysis can be divided into linear regression and nonlinear regression analysis [21], according to the type of relationship between the independent variables and dependent variables. In general, the relationship between variables is determined by the independent variables and dependent selected variables, by which regression models can be made. After that, it is used to solve the various parameters of the model based on the measured data and then evaluate whether the regression model can fit the observed data. If the model can fit the data well, then the model can be used to further predict based arguments [22]. The regression analysis is composed of the following steps [23, 24].

Regression analysis is widely used in data mining, particularly for biological data analysis in recent years, with the purpose of finding a feasible statistical law by the large amount of data of experiments. The general process is given as follows.

Step 1. Determine the variables.

Step 2. Establish the prediction model.

Step 3. Relate analysis.

Step 4. Calculate the prediction error.

Step 5. Determine the predicted value.

From the data collected in Section 2, it consists of seven independent variables and one dependent variable. The seven independent variables are inoculum size, PH values, initial liquid volume, temperature, seed age, fermentation time, and rotation speed. And, the dependent variable is flavonoid yield. From the observation of the experiments, it is found that some treatment conditions are not suitable for production of Phellinus. These data are taken as extreme data are removed from regression analysis. Extreme data refers to the data which were measured in extreme experimental environment. Also duplicate data were cancelled. Only the following data are selected in regression analysis.

(i) Inoculum size 0.5%~1.2%.

(ii) PH 5~7.
Table 1: Experiments with PH values ranging from 1 to 14 and initial volume ranges from 40 mL to 140 mL.

| Phellinus production | PH  | Temp. | Initial volume | Rotation speed | Including inoculum | Seed age | Fermentation time |
|----------------------|-----|-------|----------------|----------------|--------------------|----------|------------------|
| 45.929               | 1   | 28°C  | 100 mL         | 140            | 5%                 | 8        | 8                |
| 35.077               | 2   | 28°C  | 100 mL         | 140            | 5%                 | 8        | 8                |
| 45.654               | 3   | 28°C  | 100 mL         | 140            | 5%                 | 8        | 8                |
| 534.39               | 4   | 28°C  | 100 mL         | 140            | 5%                 | 8        | 8                |
| 702.81               | 5   | 28°C  | 100 mL         | 140            | 5%                 | 8        | 8                |
| 1467.7               | 6   | 28°C  | 100 mL         | 140            | 5%                 | 8        | 8                |
| 189.20               | 7   | 28°C  | 100 mL         | 140            | 5%                 | 8        | 8                |
| 91.049               | 8   | 28°C  | 100 mL         | 140            | 5%                 | 8        | 8                |
| 534.39               | 9   | 28°C  | 100 mL         | 140            | 5%                 | 8        | 8                |
| 702.81               | 10  | 28°C  | 100 mL         | 140            | 5%                 | 8        | 8                |
| 1467.7               | 11  | 28°C  | 100 mL         | 140            | 5%                 | 8        | 8                |
| 189.20               | 12  | 28°C  | 100 mL         | 140            | 5%                 | 8        | 8                |
| 91.049               | 13  | 28°C  | 100 mL         | 140            | 5%                 | 8        | 8                |
| 534.39               | 14  | 28°C  | 100 mL         | 140            | 5%                 | 8        | 8                |
| 508.495              | 6   | 28°C  | 100 mL         | 140            | 10%                | 8        | 8                |
| 900.662              | 6   | 28°C  | 100 mL         | 140            | 10%                | 8        | 8                |
| 1273.594             | 6   | 28°C  | 100 mL         | 140            | 10%                | 8        | 8                |
| 1153.937             | 6   | 28°C  | 100 mL         | 140            | 10%                | 8        | 8                |
| 1123.330             | 6   | 28°C  | 100 mL         | 140            | 10%                | 8        | 8                |
| 1088.064             | 6   | 28°C  | 100 mL         | 140            | 10%                | 8        | 8                |

Table 2: Experiments with including inoculum ranging from 2% to 16% and temperature ranging from 25°C to 40°C.

| Phellinus production | PH  | Temp. | Initial volume | Rotation speed | Including inoculum | Seed age | Fermentation time |
|----------------------|-----|-------|----------------|----------------|--------------------|----------|------------------|
| 546.609              | 6   | 28°C  | 100 mL         | 140            | 2%                 | 8        | 8                |
| 606.345              | 6   | 28°C  | 100 mL         | 140            | 4%                 | 8        | 8                |
| 1320.794             | 6   | 28°C  | 100 mL         | 140            | 6%                 | 8        | 8                |
| 1447.519             | 6   | 28°C  | 100 mL         | 140            | 8%                 | 8        | 8                |
| 1841.729             | 6   | 28°C  | 100 mL         | 140            | 10%                | 8        | 8                |
| 1631.990             | 6   | 28°C  | 100 mL         | 140            | 12%                | 8        | 8                |
| 481.1172             | 6   | 28°C  | 100 mL         | 140            | 14%                | 8        | 8                |
| 449.5187             | 6   | 28°C  | 100 mL         | 140            | 16%                | 8        | 8                |
| 1145.669             | 6   | 25°C  | 100 mL         | 140            | 10%                | 8        | 8                |
| 1506.055             | 6   | 30°C  | 100 mL         | 140            | 10%                | 8        | 8                |
| 1374.982             | 6   | 35°C  | 100 mL         | 140            | 10%                | 8        | 8                |
| 875.341              | 6   | 40°C  | 100 mL         | 140            | 10%                | 8        | 8                |

Table 3: Experiments with fermentation time ranging from 1 to 12 hours.

| Phellinus production | PH  | Temp. | Initial volume | Rotation speed | Including inoculum | Seed age | Fermentation time |
|----------------------|-----|-------|----------------|----------------|--------------------|----------|------------------|
| 56.606               | 6   | 28°C  | 100 mL         | 150            | 2%                 | 8        | 1                |
| 83.435               | 6   | 28°C  | 100 mL         | 150            | 4%                 | 8        | 2                |
| 303.984              | 6   | 28°C  | 100 mL         | 150            | 6%                 | 8        | 3                |
| 449.919              | 6   | 28°C  | 100 mL         | 150            | 8%                 | 8        | 4                |
| 777.331              | 6   | 28°C  | 100 mL         | 150            | 10%                | 8        | 5                |
| 1103.987             | 6   | 28°C  | 100 mL         | 150            | 12%                | 8        | 6                |
| 1619.554             | 6   | 28°C  | 100 mL         | 150            | 14%                | 8        | 7                |
| 1597.995             | 6   | 28°C  | 100 mL         | 150            | 16%                | 8        | 8                |
| 1546.336             | 6   | 28°C  | 100 mL         | 150            | 10%                | 8        | 9                |
| 1502.487             | 6   | 28°C  | 100 mL         | 150            | 10%                | 8        | 10               |
| 1489.364             | 6   | 28°C  | 100 mL         | 150            | 10%                | 8        | 11               |
| 1465.664             | 6   | 28°C  | 100 mL         | 150            | 10%                | 8        | 12               |
Table 4: Regression analysis results.

|                  | Sum of square | df | Mean square | F     | R     | R-squared | Standard error |
|------------------|---------------|----|-------------|-------|-------|-----------|----------------|
| Regression       | 3796787.42    | 14 | 249770.53   | 5.234 | 0.93  | 0.88      | 218.48         |
| Residuals        | 47719.89      | 10 | 47719.54    |       |       |           |                |
| Sum              | 3973983.26    | 24 |             |       |       |           |                |

(iii) Initial liquid volume 60–100 mL.

(iv) Temperature 25–30 °C.

(v) Seed age 4–9 days.

(vi) Fermentation time 6–12 days.

(vii) Rotation speed 140–200 r/min.

After data filtering, a statistical model is made to represent these data. It is known that there is a correlation between these data relationships, so we applied linear regression analysis to fit them. At this stage, a lot of models were tested one by one with IBM SPSS software and response surface methodology. The statistical model is:

\[ Y = A_1 \times X(1) + \cdots + A_7 \times X(7) + B_1 \times X(1) \times X(2) + B_2 \times X(1) \times X(3) + B_3 \times X(1) \times X(4) + B_4 \times X(1) \times X(5) + B_5 \times X(1) \times X(6) + B_7 \times X(1) \times X(7) + B_8 \times X(2) \times X(3) + B_9 \times X(2) \times X(4) + B_{10} \times X(2) \times X(5) + B_{11} \times X(2) \times X(6) + B_{12} \times X(2) \times X(7) + B_{13} \times X(3) \times X(4) + B_{14} \times X(3) \times X(5) + B_{15} \times X(3) \times X(6) + B_{16} \times X(3) \times X(7) + B_{17} \times X(4) \times X(5) + B_{18} \times X(4) \times X(6) + B_{19} \times X(4) \times X(7) + B_{20} \times X(5) \times X(6) + B_{21} \times X(5) \times X(7) + B_{22} \times X(6) \times X(7) + C, \]

where \( Y \) is a dependent variable of the flavonoid yield, \( X(1), X(2), \ldots, X(7) \) are the seven independent variables associated with inoculum size, PH value, initial liquid volume, temperature, seed age, fermentation time, and rotation speed, respectively, and \( A_j, B_k, \) and \( C \) are real numbers.

Although the relationship between the data may not be linear, we can put squared term for a type of data into these data. It is known that there is a correlation between these data relationships, so we applied linear regression analysis to fit them. At this stage, a lot of models were tested one by one with IBM SPSS software and response surface methodology. The statistical model is:

\[ Y = A_1 \times X(1)^2 + \cdots + A_7 \times X(7)^2 + B_1 \times X(1) \times X(2) + B_2 \times X(1) \times X(3) + B_3 \times X(1) \times X(4) + B_4 \times X(1) \times X(5) + B_5 \times X(1) \times X(6) + B_7 \times X(1) \times X(7) + B_8 \times X(2) \times X(3) + B_9 \times X(2) \times X(4) + B_{10} \times X(2) \times X(5) + B_{11} \times X(2) \times X(6) + B_{12} \times X(2) \times X(7) + B_{13} \times X(3) \times X(4) + B_{14} \times X(3) \times X(5) + B_{15} \times X(3) \times X(6) + B_{16} \times X(3) \times X(7) + B_{17} \times X(4) \times X(5) + B_{18} \times X(4) \times X(6) + B_{19} \times X(4) \times X(7) + B_{20} \times X(5) \times X(6) + B_{21} \times X(5) \times X(7) + B_{22} \times X(6) \times X(7) + C, \]

where \( Y \) is a dependent variable of the flavonoid yield, \( X(1), X(2), \ldots, X(7) \) are the seven independent variables associated with inoculum size, PH value, initial liquid volume, temperature, seed age, fermentation time, and rotation speed, respectively, and \( A_j, B_k, \) and \( C \) are real numbers.

3.2. Gene-Set Based Genetic Algorithm. Genetic algorithm (GA) was first proposed by J. Holland in 1975 [25, 26], whose general process is shown in Figure 1. In the mutation operation, if a short segment is selected in a mutation possibility and replaced by another segment, then the gene-set based GA is achieved [27].

In gene-set based GA, a chromosome is treated as a set of gene-sets, instead of a set of genes as in classical GAs. It starts with gene-sets of the largest size equal to half the chromosome length. It is most appropriate to genetics model because each gene-set represents a set of adjacent parameters of certain factor of the culture conditions.

It is noted that, in the selection, only the winning individuals from the population can be selected. Select operators are also known as reclaimed operator (reproduction operator), whose purpose is to optimize the selection of individuals (or solutions) to the next generation. Population can be updated by fitness ratio method and random sampling method to traverse, local selection. Cross operator refers to the part of the structure of the two parent individuals to generate new recombinant replacing individual operation. Variation is to make GA have local random search capability. When the GA crossover neighborhood is close to the optimal solution, the use of such a mutation operator of local random search capability can accelerate the convergence to the optimal solution.

The statistical model obtained by regression analysis is used as the fitness function here, and gene-set based GA is used to optimize the culture condition for maximizing the production of Phellinus. The data simulation is achieved by gatool in MATLAB. In the data experiments, we use a binary string composed of 7 segments to represent an individual in GA population, where each segment is associated with the value of one of the 7 parameters for the culture condition. Initial population size is 50, and cross rate is set to 0.8. Mutation rate is set to be 0.01, and selection method is roulette wheel selection. If the time is long enough then the GA process will halt by meeting the stopping conditions, such as generations limit or fitness limit.

After 156 iterations the gene-set based GA process returns the best individual and shuts down the process in Figure 2.

After the regression analysis and GA process, an optimized culture condition is obtained, shown in Table 5.

The results obtained by our method have accordance with experimental experience in literature of Phellinus growth environmental studies. Specifically, the suitable environment is neutral acidic environment, about PH value 6. The appropriate temperature range is from 22°C to 28°C [10]. Seed age and fermentation time of species vary due to the strain [3, 28, 29]. These optimized values of the parameters have accordance with biological experimental results, which indicate that our method has a good predictability for culture conditions optimization.
4. Conclusion

In this work, 45 experiments are firstly operated for collecting data related to the production of *Phellinus* from *Phellinus* flavonoids. We use regression analysis method to create a mathematical model with the collected data, and then a gene-set based GA is proposed to optimize the culture condition, where the obtained mathematic model is used as fitness function for the evolution of individuals. In the comparison results, it is believed that PH value is credible and
The temperature is also within the appropriate temperature range. Taking into account environmental factors in the laboratory, the temperature value we predicted is also reliable. The seed age and fermentation time predicted are close to the original data 8. Data experimental results show that predicted optimal values of the parameters have accordance with biological experimental results, which indicate that our method has a good predictability for culture conditions optimization.

Neural-like computing models, such as artificial neural networks [30], spiking neural networks [31], and spiking neural P systems [32–34], have been successfully used in pattern recognition and engineering practice. It is of interest to use these neural-like computing models for optimizing culture conditions for _Phellinus_ production. Our work would also guide for the “Precision Medicine” with personal SNP culture conditions for using these neural-like computing models for optimizing pattern recognition and engineering practice. It is of interest that neural P systems [32–34], have been successfully used in networks [30], spiking neural networks [31], and spiking neural systems.
[17] L. Harvey and C. Arthur, *Fitting Models to Biological Data Using Linear and Nonlinear Regression: A Practical Guide to Curve Fitting*, Oxford University Press, 2004.

[18] S. Hilary, *Multivariate Statistical Analysis for Biologists*, John Wiley & Sons, 1964.

[19] S. Robert and F. Rohlf, "The principles and practice of statistics in biological research," in *Multivariate Statistical Analysis for Biologists*, Methuen, London, UK, 1969.

[20] H. L. Seal, *Multivariate Statistical Analysis for Biologists*, Methuen, London, UK, 1964.

[21] B. Liu, F. Liu, X. Wang, J. Chen, L. Fang, and K. Chou, "Pse-in-One: a web server for generating various modes of pseudo components of DNA, RNA, and protein sequences," *Nucleic Acids Research*, vol. 43, no. 1, pp. W65–W71, 2015.

[22] R. Wang, Y. Xu, and B. Liu, "Recombination spot identification based on gapped k-mers," *Scientific Reports*, vol. 6, Article ID 23934, 2016.

[23] F. John, *Applied Regression Analysis, Linear Models, and Related Methods*, Sage, 1997.

[24] G. A. F. Seber and A. J. Lee, *Linear Regression Analysis*, vol. 936, John Wiley & Sons, 2012.

[25] D. Lawrence, *Handbook of Genetic Algorithms*, Van Nostrand Reinhold, 1991.

[26] D. Beasley, R. R. Martin, and D. R. Bull, "An overview of genetic algorithms: part 1. Fundamentals," *University Computing*, vol. 15, no. 2, pp. 58–69, 1993.

[27] T.-P. Hong, M.-T. Wu, Y.-F. Tung, and S.-L. Wang, "Using escape operations in gene-set genetic algorithms," in *Proceedings of the IEEE International Conference on Systems, Man, and Cybernetics (SMC ’07)*, pp. 3907–3911, Montreal, Canada, October 2007.

[28] J. Luo, J. Liu, C. Ke et al., "Optimization of medium composition for the production of exopolysaccharides from *Phellinus baumii* Pilát in submerged culture and the immuno-stimulating activity of exopolysaccharides," *Carbohydrate Polymers*, vol. 78, no. 3, pp. 409–415, 2009.

[29] D. B. Harper and J. T. Kennedy, "Effect of growth conditions on halomethane production by *Phellinus species*: biological and environmental implications," *Journal of General Microbiology*, vol. 132, no. 5, pp. 1231–1246, 1986.

[30] R. P. Lippmann, "An introduction to computing with neural nets," *IEEE ASSP Magazine*, vol. 4, no. 2, pp. 4–22, 1987.

[31] W. Maass, "Networks of spiking neurons: the third generation of neural network models," *Neural Networks*, vol. 10, no. 9, pp. 1659–1671, 1997.

[32] T. Song, J. Xu, and L. Pan, "On the universality and non-universality of spiking neural P systems with rules on synapses," *IEEE Transactions on NanoBioscience*, vol. 14, no. 8, pp. 960–966, 2015.

[33] T. Song, Q. Zou, X. Liu, and X. Zeng, "Asynchronous spiking neural P systems with rules on synapses," *Neurocomputing*, vol. 151, no. 3, pp. 1439–1445, 2015.

[34] X. Wang, T. Song, F. Gong, and P. Zheng, "On the computational power of spiking neural P systems with self-organization," *Scientific Reports*, vol. 6, Article ID 27624, 2016.

[35] P. Li, M. Guo, C. Wang, X. Liu, and Q. Zou, "An overview of SNP interactions in genome-wide association studies," *Briefings in Functional Genomics*, vol. 14, no. 2, pp. 143–155, 2015.