Journal of Materials, Processing and Design (2021) 5: 63-66
Clausius Scientific Press, Canada

DOI: 10.23977/jmpd.2021.050113
ISSN 2516-0923

Research on multiple regression -PSO algorithm for C4 olefin preparation

Huifang Chen, Qian Zhang

Jiangnan University, Wuxi, Jiangsu, 214122

Keywords: PSO algorithm, Spearman correlation coefficient, multiple regression

Abstract: In this paper, based on the experimental data of C4 olefin yield under different catalyst combinations and temperatures, the optimal experimental conditions for the preparation of olefin by ethanol catalytic coupling[1] were explored by establishing multiple regression model[2] and solving by particle swarm optimization algorithm. Firstly, Spearman correlation coefficients were calculated for each catalyst combination based on standardized data processing. It was found that ethanol conversion was positively correlated with C4 olefins selectivity and temperature under most preparation conditions. Then, multiple linear regression model was used to set experimental parameters and temperature range with C4 olefin yield as the objective function. Finally, particle swarm optimization algorithm was introduced to search for optimal solution.

1. Background

In modern chemical industry, C4 olefins are produced by ethanol coupling catalysis.During the reaction, the combination of catalysts (Co loading, Co/SiO2--HAP charging ratio, ethanol concentration) and the experimental temperature had an effect on the conversion of ethanol and the selectivity of C4 olefins. Therefore, it is of great significance to explore the optimum technological conditions for preparing C4 olefin through the study of catalyst combination design[4].

2. Modeling and solving of problem 1

2.1 Data preprocessing

First of all, centralized data processing can make the mean value of samples become 0.Then compress the data so that the variance of each variable is 1.The given data is transformed as follows:

$$\begin{cases} S_j = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(x_{ij}-\bar{x})^2} \\ x_{ij} = \frac{x_{ij}-\bar{x}}{S_j} \end{cases}, i = 1,2,\cdots,5, n = 4 \tag{1}$$

2.2 Spearman correlation coefficient analysis

Spearman's correlation coefficient is a nonparametric index to measure the dependence of two variables. Monotone equation was used to evaluate the correlation of two statistical variables[3]. If there are no duplicate values in the data, and when the two variables are completely monotonically correlated, the Spearman correlation coefficient is +1 or -1.

First, the data were arranged in order of size and the rank was given. When the values of two data were equal, the average value of the numerical grade was calculated as the rank number, and the grade difference between the two groups of data was calculated for Spearman correlation coefficient. The specific calculation method was as follows:

$$r_s = 1 - \frac{6\sum_{i=1}^{n} d_i{}^2}{n(n^2-1)}, n = 4, d_i is\ rank\ difference \tag{2}$$

The correlation coefficients of 20 types of catalyst combinations can be obtained by MATLAB calculation, and the final results are shown in the following table:

Table 1: The correlation coefficient of each catalyst combination

group	A1	A2	A3	A4	A5	A6	A7
conversion	1	1	1	1	0.9429	0.9	1
selectivity	0.9	0.9	0.9643	0.9429	1	1	1
group	A8	A9	A10	A11	A12	A13	A14
conversion	1	1	1	/	1	1	1
selectivity	1	1	0.7	/	1	1	1
group	B1	B2	B3	B4	B5	B6	B7
conversion	1	1	1	1	1	1	1
selectivity	1	1	1	0.8117	1	1	1

When Spearman correlation coefficient of two variables is 1, it can be directly considered that there is a monotone relationship between them.According to the data, it can be found that the conversion rate of ethanol is monotonically increasing with temperature.The selectivity of olefin increases monotonically with increasing temperature.

3. Modeling and solving of problem 2

3.1 Model Establishment

All catalyst combinations and temperatures are divided into five parts: Co loading capacity, Co/SiO2 amount, HAP amount, ethanol concentration and temperature, which are represented by x_1, x_2, x_3, x_4, x_5 respectively in this paper. y_1 is ethanol conversion, y_2 is C4 olefin selectivity,and y is C4 olefin yield.So that:

$$y = y_1 * y_2 \tag{3}$$

$$\begin{cases} y = \beta_0 + \beta_1 x_1 + \cdots \ldots + \beta_m x_m + \varepsilon \\ \varepsilon \sim N(0, \sigma^2) \end{cases} \tag{4}$$

$$\begin{cases} y_i = \beta_0 + \beta_1 x_{i1} + \cdots \ldots + \beta_m x_{im} + \varepsilon_i \\ \varepsilon_i \sim N(0, \sigma^2), i = 1, \ldots, n \end{cases} \tag{5}$$

$$X = \begin{bmatrix} 1 & x_{11} & \cdots & x_{1m} \\ \vdots & \vdots & \cdots & \vdots \\ 1 & x_{n1} & \cdots & x_{mm} \end{bmatrix}, Y = \begin{bmatrix} y_1 \\ \vdots \\ y_n \end{bmatrix} \tag{6}$$

$$\varepsilon = [\varepsilon_1 \quad \cdots \quad \varepsilon_n]^T, \beta = [\beta_0 \quad \beta_1 \quad \cdots \quad \beta_M]^T \tag{7}$$

$$\begin{cases} Y = X\beta + \varepsilon \\ \varepsilon \sim N(0, \sigma^2 E_n) \end{cases} \tag{8}$$

$$X^T Y = (X^T X)\hat{\beta} \tag{9}$$

$$\hat{\beta} = (X^T X)^{-1} X^T Y \tag{10}$$

Based on the above analysis, the following five-element linear regression model is established:

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \beta_4 x_4 + \beta_5 x_5 + \varepsilon \tag{11}$$

MATLAB was used to solve the multiple linear regression analysis model and carry out the least square analysis.The residual analysis figure is shown below:

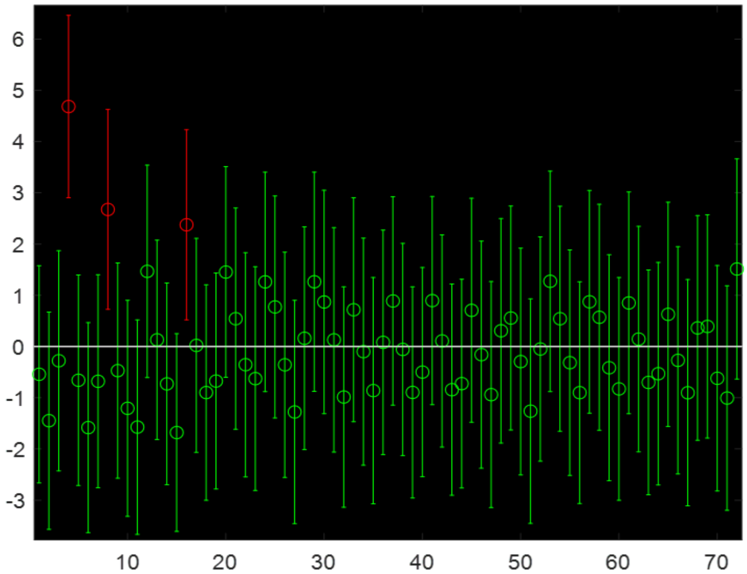

Figure 1: Residual analysis diagram of regression model

According to the case order diagram of residual error, it can be seen that the 4th, 8th and 15th points are outliers, so they are deleted and the coefficients of the improved regression model are obtained by regression again.Finally, the multiple regression equation of the ball was as follows:

$$y = 0.7260 - 0.0395x_1 - 0.0564x_2 + 0.2497x_3 - 0.0496x_4 + 0.4895x_5 \tag{12}$$

3.2 Model solution based on particle swarm optimization

Particle swarm optimization algorithm (PSO) is an optimization calculation method based on swarm intelligence.

Now suppose the search is carried out in an N-dimensional space, and the information of particle I can be represented by two N-dimensional vectors:

The position of the ith particle is $x_i = (x_{i1}, x_{i2}, \ldots, x_{iN})^T$, the velocity is $v_i = (v_{i1}, v_{i2}, \ldots, v_{iN})^T$, update its position and speed according to the following iterative formula:

$$v_{id}{}^{k+1} = v_{id}{}^{k} + c_1 * rand_1{}^{k} * (Pbest_{id}{}^{k} - x_{id}{}^{k}) + c_2 * rand_2{}^{k} * (Gbest_d{}^{k} - x_{id}{}^{k}) \tag{13}$$

Where, v_{id}^k is the velocity of the d-dimension of particle I in the Kth iteration; x_{id}^k is the d-dimensional position of particle i in the Kth iteration; $i = 1,2,3,\ldots,M$ is the population size.

That is, when Co loading is 2wt%,Co/SiO2: HAP= 1:1,and Co/SiO2 and HAP feeding are 200mg, ethanol concentration is 0.3 mL /min, and temperature is 400 degrees, the yield of C4 olefin is the highest,reaching 45.97%.

When the temperature is lower than 350 degrees, the temperature constraint condition $250 \leq x_5 < 350$ is added to the particle algorithm, and MATLAB code is run again.When the temperature is lower than 350 degrees, the Co loading capacity is 1wt% and the Co/SiO2 to HAP charging ratio is 1:1. The maximum yield of C4 reached 33.26% when the charging volume was 200mg and the concentration of ethanol was 0.48 mL /min.

4. Evaluation of Model

4.1 Advantages

(1) The pre-processing of the data makes the data more real and effective with high reliability.

(2) The control variables are accurate and appropriate, and the multi-factor problem is simplified

(3) The model makes a quantitative analysis of the influencing factors, and then realizes visualization of the data, which is convincing.

4.2 Disadvantages

(1) There are few influencing factors and some errors in solving the problem.

(2) The lack of data leads to the limitations of regression analysis.

References

[1] LU SHAopei.Synthesis of butanol and C_4 olefin by ethanol coupling [D]. Dalian University of Technology, 2018.

[2] Chen Xiaojun, Ji Fuxing. Customer relationship management, internal control and Comprehensive evaluation of m&a Performance Measurement based on Multiple Linear Regression Model [J]. Management Review, 2021.

[3] Chen Xi. Using control variable method to solve complex mass spectrometer problems [J]. Mathematical and Physicochemical Problem Solving Research, 2021.

[4] Wei Wei, YAN Xinghua. Research on investment efficiency of Chinese tourism listed companies based on multiple regression Analysis [J]. Journal of Chongqing Normal University (Natural Science Edition), 2013