Research and analysis on the preparation of C4 olefin by ethanol coupling method based on regression algorithm with computer technology

Bofan Lu, Zhixuan Yu, Lu Chen, Rongshui Qin*

Anhui Jianzhu University, School of Civil Engineering, Anhui, China

* Corresponding Author Email: 646974818@qq.com

Abstract. Computer technology has been used to study the different combination of catalyst and temperature on the ethanol conversion (one-way conversion rate of 100% per unit time ethanol (alcohol intake - ethanol residue)/ethanol intake) and C4 olefin selectivity (the proportion of all products of C4 olefin), it is necessary to decompose the catalyst composition into several indicators. The effects of several indexes on ethanol conversion and C4 olefins selectivity were analyzed by computer technology. Canonical correlation analysis was used to study them, and the coefficient before each component represented the degree of importance. The results showed that the higher the catalyst combination temperature, the higher the weight of hydroxyapatite, the higher the ethanol conversion and C4 olefins selectivity, and vice versa. When the C4 olefin yield (its value is the selectivity of C4 olefin to ethanol conversion) is as high as possible, the multiple linear stepwise regression function of catalyst parameters and temperature on C4 olefin yield is obtained. MATLAB software was used to calculate the function formula. The absolute value of coefficients determines the importance of each index to C4 olefin yield. Combined with chemical reaction, when the mass ratio is 200:200 and the temperature is 400℃, the yield of C4 olefins can be maximized, and the yield of C4 olefins is 44.728%. Through the establishment of decision tree regression model by computer technology, it can be intuitively concluded that when the temperature is lower than 350℃, the best experimental condition is the C4 olefins yield of 17.264%.

Keywords: Canonical correlation analysis, multiple linear stepwise regression, MATLAB.

1. Introduction

Using a Sio2-HAP catalyst with both acid and base activities, and covering its surface with Co metal with dehydrogenation activity, the surface pH of the catalyst was changed by changing the Co load (weight ratio of Co to SiO2), the mass ratio of Co/SiO2 to HAP, and the concentration of ethanol. The reaction process of preparing C4 olefin by ethanol was explored according to the combination of catalyst and temperature, and the relevant factors of the reaction were analyzed in detail by using the selectivity and yield of C4 olefin, which is of great significance and value to the preparation of C4 olefin by ethanol [2].

2. The modeling method

In order to ensure the rationality of the research results, the indexes in each catalyst group are not affected by each other in the research process, and the process is regarded as a complete reaction with no residual intermediates. Considering the effects of different catalyst combinations and temperature on ethanol conversion and C4 olefins selectivity, the catalyst components should be divided into four parts: Co/SiO2 mass, HAP mass, Co load and ethanol concentration. Therefore, the effects of different catalyst combinations and temperature on ethanol conversion and C4 olefins selectivity can be transformed into the effects of Co/SiO2 quality, HAP quality, Co loading, ethanol concentration and temperature on ethanol conversion and C4 olefins selectivity. In order to study the relationship between the two groups of variables, canonical correlation analysis[3] was adopted to solve it. First, the two groups of variables were assumed as follows: $x^{(1)}=(x_I^{(1)},x_2^{(1)},\cdots x_n^{(1)}),x^{(2)}=(x_1^{(2)},x_2^{(2)},\cdots x_n^{(2)})$,From the above, ethanol conversion and C4 olefins selectivity can be regarded as $x^{(2)}$,All other variables are taken as $x^{(1)}$ .Several representative comprehensive variables were selected from the two groups of variables respectively U_i, V_i .So that each synthetic variable is a linear combination of the original variables:

$$\begin{aligned} U_i &= a_1^{(i)} X_1^{(1)} + a_2^{(i)} X_2^{(1)} + \cdots + a_p^{(i)} X_p^{(1)} = a^{(i)'} X^{(1)} \\ V_i &= b_1^{(i)} X_1^{(2)} + b_2^{(i)} X_2^{(2)} + \cdots + b_p^{(i)} X_p^{(2)} = b^{(i)'} X^{(2)} \end{aligned} \tag{1}$$

To make the information found in the second set of data more effective, it is necessary to ensure that the second set of data is not related to the first:

$$\operatorname{cov}(U_1, U_2) = \operatorname{cov}(V_1, V_2) = 0 \tag{2}$$

The typicality related variables *r* were extracted,

$$\mathrm{r} \le \min(p, q) \tag{3}$$

r variables were gotten.

According to relevant knowledge, the C4 olefin yield is equal to the conversion rate of ethanol times the selectivity of C4 olefin. Therefore, the conversion rate of ethanol is multiplied by the selectivity of C4 olefin to obtain a new data, namely, the C4 olefin yield. The problem becomes the correlation analysis of dependent variable Y (olefin yield) and m independent variables $x_1, x_2, \cdots x_m$ .In order to make the yield of C4 olefin as high as possible, it is necessary to take each index of catalyst, temperature as independent variable and the yield of C4 olefin as dependent variable [4]. Its model is as follows:

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

In this problem, m independent observation data $(y_i, x_{i1}, \cdots, x_{im}), i = 1,2,\cdots,n, n > m$ are obtained, and it can be concluded from：

$$\begin{cases} y_i = \beta_0 + \beta_1 x_{i1} + \cdots + \beta_m x_{im} + \varepsilon_i \\ \varepsilon_i \sim N(0, \sigma^2), \operatorname{var}(\varepsilon_i) = \sigma^2, Cov(\varepsilon_a, \varepsilon_b) = 0(\text{当} a \ne b) \end{cases} \tag{5}$$

Where, $x_1, \cdots x_m$ is a non-random variable; *y* is a random dependent variable; $\beta_0, \cdots, \beta_m$ is the regression coefficient; ε is the error term. Because the number of independent variables, and the influence degree of each independent variable on the dependent variable Y is different, therefore, adopt the method of multiple stepwise regression analysis, one by one to introduce significant effect on Y of the independent variables, and the equation of the old variables tested individually, the not significant variable from the equation, retain a significant influence on Y variable, resulting in significant effect on Y of both variables, It doesn't include variables that don't have a significant effect on Y.In order to make the obtained results more intuitive, the decision tree regression model[5] is established again, and the attribute selection measure is gini index, namely CART decision tree. Decision tree can be used to more intuitively see the relationship between the significant influence of variables.

3. Overview of the modeling process

According to the model established above, SPSS software was used to conduct canonical correlation analysis on the pre-treated data. Set Co/SiO2 quality, HAP quality, Co load, ethanol concentration and temperature as set 1, ethanol conversion rate and C4 olefins selectivity as set 2. The typical correlations obtained are shown in the following table 1:

Table 1. Cnonical CORRELATION

	1	2
Relevance	.912	.504
Characteristic vaule	4.959	.340
Wilke statistics	.125	.746
F	39.080	9.189
Molecular freedom	10.000	4.000
Denominator freedom	214.000	108.000
Significance	.000	.000

By observing the significance of the two pairs of typical variables in the table above, it can be seen that the significance of the two pairs of variables is less than 0.01, which indicates that the correlation between the two pairs of typical variables is significant. It can be seen that the correlation coefficient of the first group is 0.912, which is significantly greater than that of the second pair of typical variables and the significance P value. Therefore, the first canonical correlation variable is taken to explain.

By observing the significance of the two pairs of typical variables in the table above, it can be seen that the significance of the two pairs of variables is less than 0.01, which indicates that the correlation between the two pairs of typical variables is significant. It can be seen that the correlation coefficient of the first group is 0.912, which is significantly greater than that of the second pair of typical variables and the significance P value. Therefore, the first canonical correlation variable is taken to explain.

Table 2. Standardized typicality correlation coefficient

Variable	**1**	**2**
HAP mass	-.549	.048
Co/Sio2 mass	.071	.469
Co load	.131	-.774
Ethanol concentration	.050	.882
Temperature	-.868	-.036

Table 3. Standardized typicality correlation coefficient

Variable	**1**	**2**
Ethanonl conversion	-.650	-1.325
C4 olefin selectivity	-.420	1.414

After the above analysis, correlation coefficients of a pair of typical variables are needed here, which can be expressed as:

$$\begin{aligned} U_1 &= -0.549X_1^{(1)} + 0.071X_2^{(1)} + 0.131X_3^{(1)} + 0.05X_4^{(1)} - 0.868X_5^{(1)} \\ V_1 &= -0.65X_1^{(2)} - 0.42X_2^{(2)} \end{aligned} \tag{6}$$

When looking at typical variables, HAP mass and temperature were positively correlated with ethanol conversion and C4 olefins selectivity, while Co/SiO2 mass, Co load, and ethanol concentration were negatively correlated with ethanol conversion and C4 olefins selectivity. The coefficient in front of each component represents the degree of importance, so it can be seen that temperature and HAP quality have a great influence on ethanol conversion. Therefore, it can be concluded that the higher the HAP mass in the catalyst combination and the higher the temperature during the experiment, the higher the ethanol conversion and C4 olefins selectivity, and vice versa. That is, when the catalyst combination "200mg 1wt%Co/ Sio2-200mg HAp-ethanol concentration 0.9 mL /min" at 450 c^0 , the ethanol conversion and C4 olefin selectivity are 86.4% and 49.9%, respectively. When the catalyst combination was "100mg 1wt%Co/ Sio2-100mg HAp-ethanol concentration 0.9 mL /min" and the temperature was 250 c^0 , the ethanol conversion rate and C4 olefin selectivity minimum 0.2% and 0.1%, respectively. The null hypothesis can be rejected. Therefore, at least one of λ_1, λ_2 is not 0, indicating that the first pair of typical correlation variables U_1, V_1 are correlated. It can be concluded that this pair of typical variables has extracted most information of correlation analysis of two groups of variables $X^{(1)}$, $X^{(2)}$,which is valuable. Therefore, the first pair of canonical correlation variables used in question 2 is reasonable.

According to the established multivariate stepwise regression model, there are six independently observed data, and the following functions can be established;

$$y = \beta_0 + \beta_1 x_1 + \cdots + \beta_5 x_5 \tag{7}$$

It can be obtained that the final retained variables are x_2 and x_5 , and their corresponding *y*. Therefore, the final functional relationship obtained is:

$$y = -39.528 + 0.04689 x_2 + 0.1297 x_5 \tag{8}$$

The goodness of fit of the regression function is observed, that is, $R^2 = 0.65$, and significance $p = 1.28 \times 10^{-26} \approx 0$, so the model is more accurate. According to the above formula, the retained indexes are HAP mass and temperature. It shows that these two indexes have strong significance on the yield of C4 olefin. Both of them are positively correlated with the yield of C4 olefin. Therefore, in order to make the C4 olefin yield as high as possible, the catalyst combination with high HAP quality should be selected, that is, the combination with the mass ratio of Co/SiO2 and *HAP* 200:200. At the same time, it can be seen from the above equation; With the increase of temperature, the yield of C4 olefins also increases. According to relevant data, when the temperature is higher than 400 degrees, the selectivity of ethylene is enhanced and the yield of C4 olefin decreases. Therefore, when the A3 catalyst combination is selected and the temperature reaches 400 degrees, the yield of C4 olefin can be increased as much as possible. At this time, the yield of C4 olefin is 44.728%. As can be seen from the decision tree, when the temperature is greater than 350 degrees,Co/SiO2 mass is greater than 100, C_0 loading capacity is not equal to 0.5, 2, and 5, the yield of C4 olefin is the highest. As can be seen from the table, the yield of C4 olefin is also 44.728%. Similarly, when the temperature is limited to 350 degrees and the mass ratio at 200:200, the use of a combination of A2 catalysts at 325 degrees can maximize the yield of C4 olefin. At this time, the yield of C4 olefins is 17.264%. In the decision tree, when the temperature is kept between 300-350 degrees, the Co/SiO2 mass is greater than 100, the concentration of ethanol is not equal to 0.3, and the Co load is not equal to 0.5, 1,5, the yield of C4 olefin reaches the maximum. Check the table for 17.264%. The results are the same as those obtained by stepwise regression.

4. Conclusion

When studying the effects of different catalyst combinations and temperature on ethanol conversion and C4 olefin selectivity, different catalyst combinations were divided into multiple targets to analyze the effects of these indicators on ethanol conversion and C4 olefin selectivity. Canonical correlation analysis was used to study it. Using SPSS software to obtain the corresponding standardized typical variables, it can be seen that HAP mass and temperature are positively correlated with ethanol conversion and C4 olefins selectivity, while Co/SiO2 mass, Co load, ethanol concentration are negatively correlated with ethanol conversion and C4 olefins selectivity. And the coefficient in front of each component represents the degree of importance. Therefore, it can be concluded that the higher the HAP mass in the catalyst combination and the higher the temperature during the experiment, the higher the conversion rate of ethanol and the C4 olefins selectivity, and vice versa.To maximize the yield of C4 olefin, multiple linear stepwise regression functions of various catalyst parameters and temperature on the yield of C4 olefin were obtained. Matlab software was used to calculate the coefficient corresponding to each indicator, and the functional relationship was determined as follows: $y = -39.528 + 0.0689x_2 + 0.1297x_5$; The HAP mass and temperature are positively correlated with C4 olefin yield. When the mass ratio is 200:200 and the temperature is 400 degrees, the yield of C4 olefin can be increased as much as possible, and the yield of C4 olefin is 44.728%.

References

[1] Chinese Society of Industrial and Applied Mathematics. National College Students Mathematical Contest in Modeling Problem B, 2021.

[2] Lv Shaopei. Synthesis of butanol and C_4 olefin by ethanol coupling [D]. Dalian University of Technology,2018.

[3] Duan_. Mathematical Modeling -- Canonical Correlation Analysis and relevant SPSS operation [EB/OL]. [2021.9.12].

[4] https://blog.csdn.net/NoBuggie/article/details/102830981?utm_source=app&app_version=4.15.0&code =app_1562916241&uLinkId=usr1mkqgl919blen.

[5] Jiang Qiyuan, Xie Jinxing, Ye Jun. Mathematical Model [M]. 4th edition. Higher Education Press, 2011.

[6] Fan Ming, Meng Xiaofeng. Data Mining concepts and Techniques [M]. First edition. China Machine Press, 2012.