International Core Journal of Engineering Volume 8 Issue 7, 2022
ISSN: 2414-1895 DOI: 10.6919/ICJE.202207_8(7).0025

Study on Catalytic Process of Olefin Preparation by XGBoost Ethanol Coupling based on PSO Particle Swarm Optimization

Wenrui Cui

School of Remote Sensing & Geomatics Engineering, Nanjing University of Information Science and Technology, Nanjing, JiangSu, 210044, China

Abstract

In this paper, aiming at the coupling preparation of C4 olefins, on the basis of the catalytic effect fitting model of XGBoost (parallel tree lifting), PSO particle swarm optimization algorithm is introduced to construct the combinatorial optimization model of XGBoost-PSO catalyst. the objective function is the fitting model of XGBoost training, and the fitness value is optimal after 21 iterations.

Keywords

C4 Olefin; Parallel Tree; XGBoost Method.

1. Introduction

In recent years, with the continuous development of chemical industry and automobile industry, the demand for coal and fossil fuels is increasing, which has caused a series of problems such as environmental pollution and resource shortage. Therefore, looking for an alternative renewable energy is one of the current research hotspots[1]. As a kind of clean energy, ethanol can be obtained from biomass fermentation such as straw and corn, and the source of raw materials is very wide. With the increase of ethanol production and the decrease of cost year by year, the application prospect of ethanol is more and more broad. C4 olefins are widely used in pharmaceutical products and chemical production[2].

Ethanol is the raw material for the production of C4 olefins, which can be prepared by biomass fermentation. It has a wide range of sources, green and clean, and has great environmental and economic benefits[3]. In the preparation process, temperature and catalyst combination (including the combination of CO loading, ethanol concentration, Co / SiO2 and HAP loading ratio) will affect the selectivity and yield of C4 olefins. Therefore, it is of great significance and value to design the catalyst combination and explore the process conditions of ethanol catalytic coupling to prepare C4 olefins.

2. Data Preprocessing

First of all, the digital features of the catalyst combination should be extracted. At the same time, temperature is also an important factor affecting the effect of catalysis, and the difference of temperature must be taken into account when studying the effect of catalyst. From the above two points of view, the data of Annex I is cleaned as follows (in part, see the supporting material for all the data):

ISSN: 2414-1895 DOI: 10.6919/ICJE.202207_8(7).0025

Table 1. Data cleaning results

Temperature(℃)	Ethanol conversion%	C4 olefin selectivity (%)	CoSiO2(mg)	Co load (wt%)	HAP(mg)	ethanol (ml/min)
250	2.06716945	34.05	200	1	200	1.68
275	5.851720987	37.43	200	1	200	1.68
300	14.96889149	46.94	200	1	200	1.68
325	19.6813591	49.7	200	1	200	1.68
350	36.80101697	47.21	200	1	200	1.68
250	4.602997235	18.07	200	2	200	1.68
275	17.19553891	17.28	200	2	200	1.68
300	38.9222453	19.6	200	2	200	1.68
325	56.38246023	30.62	200	2	200	1.68
350	67.87929575	39.1	200	2	200	1.68

3. Fitting Model of Catalytic Effect based on XGBoost

The essence of XGBoost[4-6] is to gradually upgrade and evolve a group of weak learners (mainly tree models) to a strong learner to improve the accuracy of prediction. By inputting the cleaned data into the trainer, the tree model ft(x) can be trained through t iterations:

$$\hat{y}_j^{(t)} = \sum_{k=1}^{t} f_k\left(x_i\right) = \hat{y}_j^{(t-1)} + f_t\left(x_i\right) \tag{1}$$

$\hat{y}_j^{(t)}$ is the prediction result of sample i after the t iteration, $\hat{y}_j^{(t-1)}$ is the prediction result of the former tmur1 tree.

The prediction accuracy of the model is determined by the deviation and variance of the model. All objective functions include two aspects: loss function and regular term. The loss function represents the deviation of the model, and if you want the variance to be small, you need to add a regular term to the objective function to prevent over fitting.

The loss function can be expressed by the predicted value and the real value:

$$L = \sum_{i=1}^{n} l\left(y_i, \hat{y}_i\right) \tag{2}$$

Where n is the number of samples.

Each time a tree model is established, a new decision tree must be defined:

$$f_t(x) = w_{q(x)}, \quad \mathrm{w} \in R^t, \quad \mathrm{q}: R^d \to \{1, 2, \ldots, \mathrm{T}\} \tag{3}$$

Where w is a one-dimensional vector with length T, representing the weight of each leaf node of tree q, and q represents the structure of a tree (the mapping relationship between samples and leaf nodes).

ISSN: 2414-1895 DOI: 10.6919/ICJE.202207_8(7).0025

The complexity Ω of the decision tree can be composed of the number of leaves T. The fewer leaf nodes, the simpler the model is. In addition, the leaf nodes should not contain too high weight. Therefore, the regular term of the objective function is determined by the vector L2 normal form composed of the number of leaf nodes and the weight of all nodes of the generated decision tree.

$$\Omega\left(f_t\right)=\gamma T+\frac{1}{2}\lambda\sum_{j=1}^{T}w_j^2 \tag{4}$$

Therefore, the objective function is:

$$Obj=\sum_{i=1}^{n}l\left(y_i,\hat{y}_i\right)+\gamma T+\frac{1}{2}\lambda\sum_{j=1}^{T}w_j^2 \tag{5}$$

Where, Gj is the cumulative sum of the first derivative gi of the loss function of leaf node j, which is a constant; Hj is the sum of the second derivative hi of leaf node j, which is also a constant.

For each leaf node j, it can be disassembled from the objective function:

$$\mathbf{G}_{\mathrm{j}}\mathbf{w}_{\mathrm{j}}+\frac{1}{2}\left(\mathrm{H}_{\mathrm{j}}+\lambda\right)\mathbf{w}_{\mathrm{j}}^2 \tag{6}$$

The solution of the objective function is:

$$Obj=-\frac{1}{2}\sum_{j=1}^{T}\frac{G_J^2}{H_j+\lambda}+\gamma T \tag{7}$$

4. Degree of Fit and Model Solving

In order to evaluate the influence of parameters in the model, this paper introduces the importance to evaluate it quantitatively. First, calculate the baseline when using all parameters to build the model:

$$\text{baseline}=\text{accuracy}/R^2 \tag{8}$$

Then, remove only one parameter at a time, retrain the model and calculate the accuracy of the new model accuracy/R2, Finally, calculate the difference between the feature line and baseline of each parameter. This value is the importance of the parameter. When the importance of the parameter is greater than 0, it indicates that the parameter has a positive impact on the model results, and the greater the importance, the greater the degree of impact. When the importance of the parameter is less than 0, it indicates that the parameter has a negative impact on the model results, and the greater the importance, the greater the degree of impact; The closer the importance of the parameter is to 0, the less its influence is.

Input the ethanol conversion, temperature and four catalyst combination parameters into the XGBoost model, and the fitting results are shown in the figure below:

ISSN: 2414-1895 DOI: 10.6919/ICJE.202207_8(7).0025

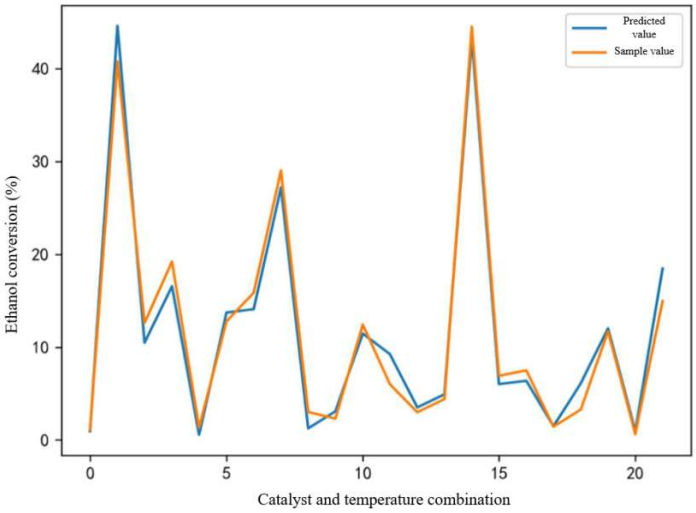

Fig.1 Ethanol conversion fitting

The goodness of fit R2 of the model is 98%. 98% of the samples can be explained by the trained model, and it can be seen from the image that its fitting degree is very high. Therefore, the model can be used to analyze the influence of parameters.

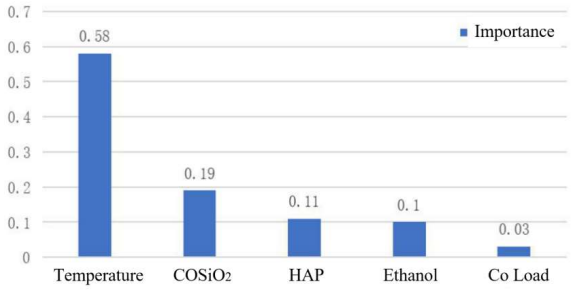

Fig.2 Effect of catalyst parameters on ethanol conversion

The increase of temperature has a very obvious effect on the improvement of ethanol conversion, while the change of CO loading has a close effect on the improvement of ethanol conversion, and the increase of CoSiO2, HAP and ethanol concentration has little positive effect on the improvement of ethanol conversion. Input C4 selectivity, temperature and four catalyst combination parameters into XGBoost model, and the fitting results are shown in the figure below:

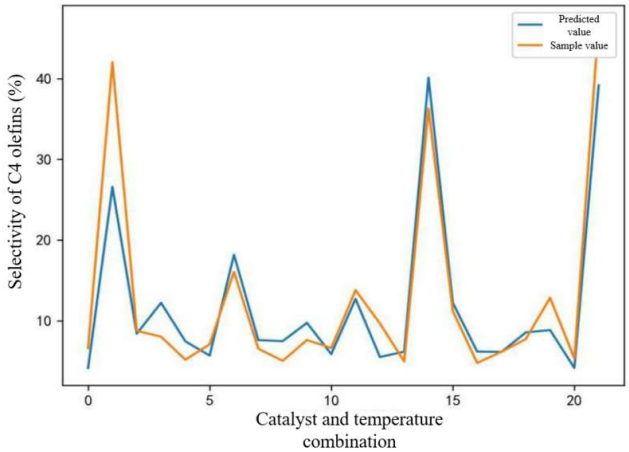

Fig.3 C4 selective fitting

ISSN: 2414-1895 DOI: 10.6919/ICJE.202207_8(7).0025

The goodness of fit R2 of the model is 0 92, that is, 92% of the samples can be explained by the trained model, and it can also be seen from the image that its fitting degree is high. Therefore, the model can be used to analyze the influence of parameters.

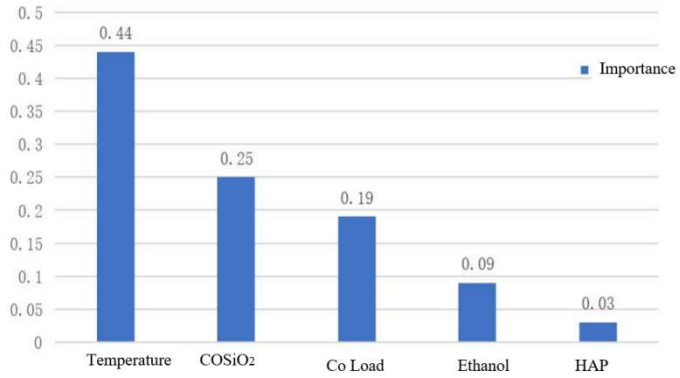

Fig.4 Effect of catalyst parameters on ethanol conversion

The increase of temperature also has an obvious effect on the improvement of C4 selectivity. The change of CO loading has a close impact on the ethanol conversion rate, and the increase of CoSiO2, HAP and ethanol concentration has little positive impact on the improvement of ethanol conversion rate

5. XGBoost-PSO Catalyst Combination Optimization Model

5.1 Steps

Input the sample set (C4 olefin yield, temperature and four catalyst parameters) into the XGBoost model, train the fitting model of C4 olefin yield, and then set the fitting model as the objective function in PSO particle swarm to evaluate the fitness of particles. Therefore, seeking the catalyst combination with the highest conversion rate of C4 olefins is transformed into finding the best position of fitness in PSO particle swarm.

Step 1: initialize the particle swarm, including the population size N, the position xi and velocity vi of each particle;

Step 2: Calculate the fitness value of each particle Fa[i].

Step 3: For each particle, compare its fitness value Fa[i] with the individual extreme value Pbest (i). If Fa[i]>Pbest(i), replace Pbest (i) with Fa[i].

Step4: for each particle, compare its fitness value Fa[i] with the global extremum gbest, and replace gbest with Fa[i] if Fa[i]>gbest.

$$v_i = v_i + c_1 * rand() * \left(P_{\text{best}}(i) - \text{x}_\text{i}\right) + c_2 * rand() * \left(g_{\text{best}} - \text{x}_\text{i}\right) \tag{9}$$

$$\text{X}_\text{i} = \text{x}_\text{i} + v_i \tag{10}$$

Step5: If the end condition (error is good enough or the maximum number of cycles is reached) is met, Step2 is returned

5.2 Solving the Maximum Yield of C4 Olefins by Direct Segmentation

The data of C4 olefin yield, temperature and catalyst combination were input into XGBoost model to train the fitting model of C4 olefin yield.

ISSN: 2414-1895 DOI: 10.6919/ICJE.202207_8(7).0025

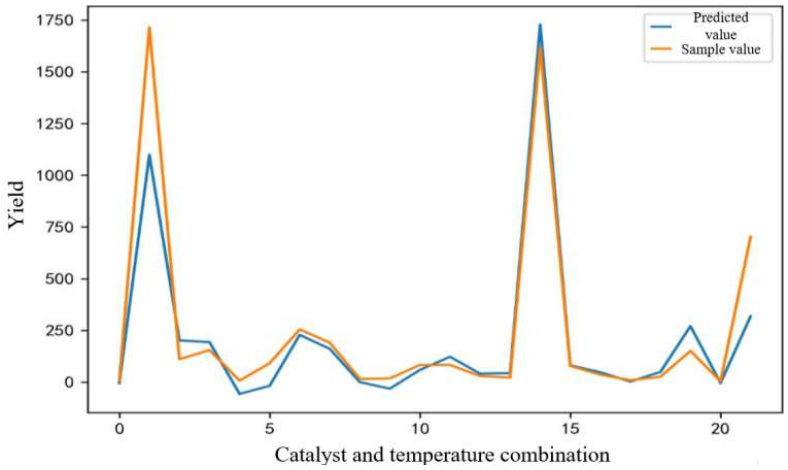

Fig.5 C4 olefin yield fitting

Without considering the optimization algorithm, the temperature and four catalyst parameters (co loading, Co / SiO2 and HAP loading ratio, ethanol concentration) are sliced in a certain step. Among them, the maximum value of temperature is 400 ℃, the minimum value is 250 ℃, and the step length is 5; The maximum value of CO load is 5wt%, and the minimum value is 0.5wt% in steps of 0.5wt%; The maximum mass of CoSiO2 is 200mg, the minimum mass is 50mg, and the step size is 10mg; The maximum value of HAP mass is 200mg, the minimum value is 50mg, and the step size is 10mg; The maximum ethanol concentration is 2.1ml / min, the minimum value is 0.3ml / min in steps of 0.2ml/min.

Table 2. Slice arrangement and combination data (part)

Temperature	CoSiO2	Co Load	HAP	Ethanol concentration
250	10	0.5	50	0.3
250	10	0.5	50	0.5
250	10	0.5	50	0.7
250	10	0.5	50	0.9
250	10	0.5	50	1.1
250	10	0.5	50	1.3
250	10	0.5	50	1.5
250	10	0.5	50	1.7
250	10	0.5	50	1.9
250	10	0.5	50	2.1
250	10	0.5	60	0.3
250	10	0.5	60	0.7

A series of temperature and catalyst combinations can be obtained by inputting the data of segmentation and combination into the model, so as to optimize the yield of C4 olefins. It shows that the optimal C4 olefin yield requires the highest temperature range, that is, temperature is the most key factor to achieve the optimal C4 olefin yield.

ISSN: 2414-1895 DOI: 10.6919/ICJE.202207_8(7).0025

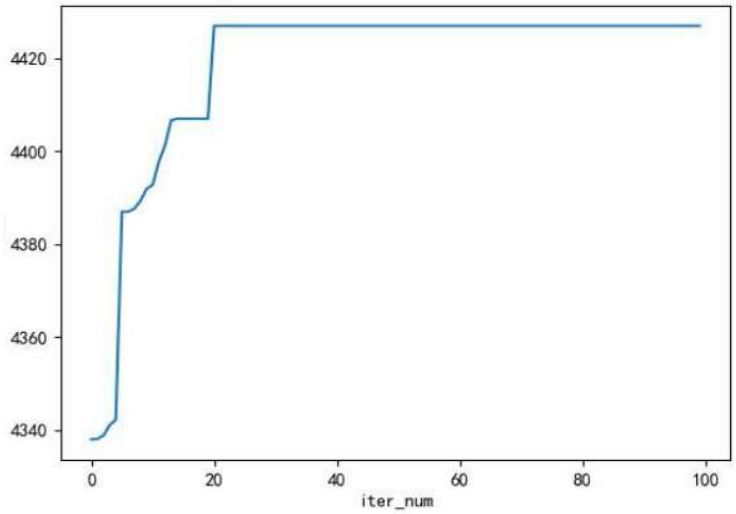

Fig.6 XGBoost PSO model optimization process

The maximum yield of C4 olefin is 4427 21, the corresponding catalyst combination is: 376 ℃ 196mg 2wt% Co / sio2300mg HAP ethanol concentration 1 38ml/min.

6. Conclusion

C4 olefins are widely used in pharmaceutical products and chemical production. Ethanol is the raw material for the production of C4 olefins, which can be prepared by biomass fermentation. It has a wide range of sources, green and clean, and has huge environmental and economic benefits. Therefore, in this paper, the combinatorial optimization model of XGBoost-PSO catalyst is constructed, and the objective function is the fitting model of XGBoost training. After 21 iterations, the fitness value is the best. The maximum yield of C4 olefin is 4427.21, and the reaction temperature is 376C. The catalyst combination is: 196mg 2wt% Cocontrol Sio _ 2-300mg HAP- ethanol concentration 1.38ml/min. When the temperature is limited below 350C, the adaptive value reaches the optimal value after 37 iterations, and the maximum yield of C4 olefin is 2521.39, the reaction temperature is 342C, and the catalyst combination is: 196mg 2.2wt% Coram SiO2-320mg HAP- ethanol concentration 1.89ml/min.

References

[1] Li Jinyou, Li Yuan, Wang Haixin, Li Chaoran. Health performance evaluation and prediction of wind turbine based on XGBoost-Bin automatic power limit calculation [J/OL]. Computer Integrated Manufacturing system: 1-21.

[2] Liao Bin, Wang Zhining, Li Min, Sun Ruina. The Football player's value Forecast and characteristic Analysis method based on XGBoost and SHAP Model [J/OL]. Computer Science: 1-13.

[3] Hu Zhihui, Jin Yongxing, Zhou Tianrui, Hu Qinyou. Real-time prediction of ship energy consumption based on XGBoost [J]. Journal of Shanghai Maritime University, 2022 743 (01): 23-29+37.

[4] Shi Yongsheng, Li Jin, Ren Jiarui, Zhang Kai. Prediction of remaining service life of lithium-ion battery based on WOA-XGBoost [J/OL]. Energy storage science and technology: 1-12.

[5] Chen Zhenyu, Liu Jinbo, Li Chen, Ji Xiaohui, Li Dapeng, Huang Yunhao, Di Fangchun, Gao Xingyu, Xu Lizhong. Ultra-short-term load forecasting based on LSTM and XGBoost combined model [J]. Grid Technology, 2020,544 (02): 614-620.

[6] Li Zhanshan, Liu Zhaoyi. Feature selection algorithm based on XGBoost [J]. Journal of Communications, 2019. 40 (10): 101-108.