ISSN: 2664-9640

Optimal catalyst combination and temperature study based on support vector machine for C4 olefin uptake rate

Linfeng Du[1], Jinguo Luo[2], Huili Chen[3]

[1]College of Computer Science, Sichuan University, Sichuan, Chengdu, 610207, China

[2]School of Mechanical Engineering, Sichuan University, Sichuan, Chengdu, 610207, China

[3]College of Electronics and Information Engineering, Sichuan University, Sichuan, Chengdu, 610207, China

Abstract

The production of C4 olefins from ethanol molecules as platform molecules has great application prospects and economic benefits. In this paper, we use mathematical methods and machine learning to establish mathematical models for catalyst combinations and temperatures to analyze the process conditions for the preparation of C4 olefins by ethanol coupling. In this paper, we search for the optimal catalyst combination and temperature to achieve the highest C4 olefin uptake rate; in addition to that, we search for the catalyst combination that can achieve the highest C4 olefin uptake rate when under the constraint of temperature conditions below 350 degrees. A support vector machine (SVM) model with oversampling and Bayesian tuning was used to obtain a model with a goodness-of-fit of 92%. The highest C4 olefin yield of 50.7% was predicted for all catalyst combinations and temperatures when the Co loading was 2 wt%, the ethanol concentration was 1.68 ml/min, the Co/SiO2 and HAP loading ratio was 1, the mass sum of Co/SiO2 and HAP was 400 mg, the loading method was I, and the temperature was 400 degrees. As well as the conclusions of increasing the Co loading, decreasing the ethanol concentration and reducing the catalyst dosage when the temperature was lower than 300 degrees Celsius.

Keywords

Oversampling; SVM regression prediction; Bayesian conditioning; Control variable method.

1. Introduction

In recent years, with the emergence of a series of problems such as resource shortage and environmental pollution, the search for an alternative renewable energy source is a current research hotspot. Ethanol, as a clean energy source, has a wide range of raw material sources. As the production of ethanol increases year by year and the production cost decreases, its conversion into other high-value-added products as a platform molecule has a broad application prospect. In this paper, we will analyze the effect of catalyst combination and temperature on the uptake rate of C4 olefins, and establish a mathematical model with Co loading, Co/SiO2 and HAP loading ratio, temperature, ethanol concentration and other important factors as variables to be able to make an accurate prediction of C4 olefins uptake rate values. After building the model, the model was used to find the combination of catalysts and temperature conditions for maximum C4 olefin uptake. Based on this, the catalyst combinations with maximum C4 olefin yield below 350 degrees were analyzed, and all the data obtained from the prediction were used for a comprehensive analysis of the data, and the highest C4 olefin yield was obtained at temperatures of 250 and 300 degrees[1].

ISSN: 2664-9640

2. Assumptions and notations

2.1. Assumptions

We use the following assumptions.

Assume that the given experimental data is good and the error is small

The explanatory variables are assumed to be non-random or fixed and the variables are not correlated with each other (no multicollinearity)

Random error terms have zero mean, homoscedasticity and no serial correlation

Explanatory variables are not correlated with the random term

2.2. Notations

The primary notations used in this paper are listed in Table 1.

Table 1 Notations

Symbols	Description
u_i	Perturbation term in the regression equation
β_0	The intercept term in the regression equation
β_i	Regression coefficients in the regression equation
R^2	Correlation coefficient
P	The probability corresponding to F
F	F
ω	Hyperplane decision factor
$C\sum_{i=1}^{m}\left(\xi_i+\xi_i^*\right)$	SVR loss value

3. Model construction and solving

3.1. Model Selection

In some catalyst combinations, there are still a few combinations with missing temperature conditions such as 400 degrees, based on which it is required to solve for the catalyst combination and temperature conditions when C4 olefin uptake is maximum, it is necessary to predict the unknown catalyst combination and temperature conditions. The regression model in machine learning is used to make the prediction.

Support vector machines (SVM), decision trees, and integration algorithms in machine learning are used in this paper for materials with little experimental data, large feature dimensions, and unbalanced samples[2].

3.2. Data processing (random oversampling)

Observation of the overall sample data shows that the proportional distribution of the number of variables in the sample is very uneven, with each sample containing seven variables. A schematic representation of the sample equilibrium distribution is shown in Table 2.

ISSN: 2664-9640

Table 2 Schematic representation of the sample equilibrium distribution

Variable types	Variable nature	Variable value	Proportion	Variable partition
ethanol concentration	independent	1.68	60.70%	majority
		0.9	20.50%	minority
		0.3	9.80%	minority
		2.1	8.90%	minority
CO capacity	independent	1	80.30%	majority
		2	5.30%	minority
		0.5	9.70%	minority
		5	4.40%	minority
Co/SiO2 to HAP charging ratio	independent	1	91.10%	majority
		2.03	4.40%	minority
		0.49	4.40%	minority
T	independent		Balance variable	
Co/SiO2 to HAP mass ratio	independent	100	40.70%	majority
		400	29.20%	minority
		20,50,140,200,150	30%	minority
loading pattern	independent		Balance variable	
C4 yield	dependent	>10	20%	minority
		<10	80%	majority

Random oversampling is used to reduce the problem of data imbalance in the sample. Random oversampling cannot simply be a subjective repetitive sample operation of the original data; fewer classes must be randomly sampled with a greater probability of sample repetition, and the data sets made superimposed. A random oversampling example diagram is shown in Figure 1.

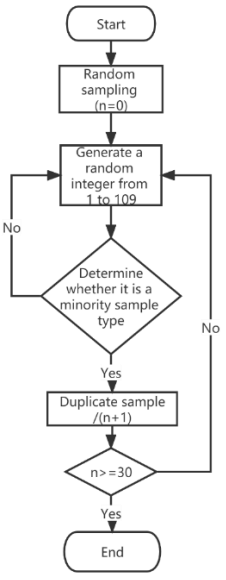

Figure 1 Random oversampling example diagram

The condition to determine whether the sample is a minority species sample: there are more species in the sample that are minority variables than majority species variables.

ISSN: 2664-9640

3.3. Model building (SVR)

SVR is support vector machine regression, using support vector machine algorithm to do regression, which belongs to supervised learning inside machine learning. The regression here is different from the traditional linear regression - the machine learning regression, with its own defined loss function, objective function and the optimization algorithm is also different from the traditional regression[3]. The supervised learning flow chart is shown in Figure 2.

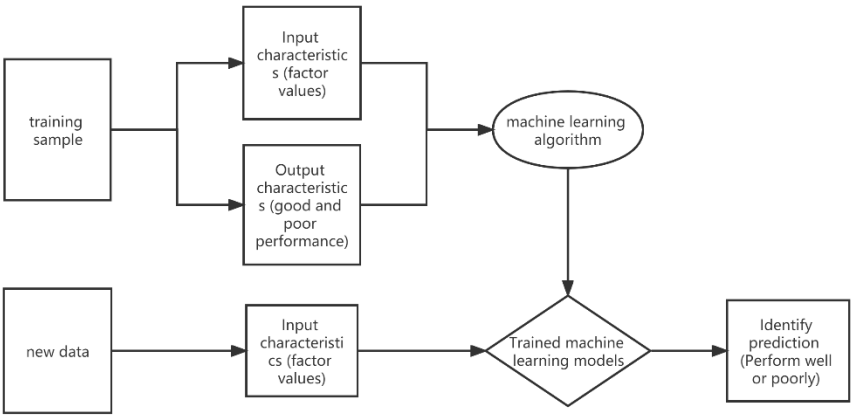

Figure 2 Supervised learning flow chart

The main task of SVR is to find an objective function, generate a hyperplane around the objective function, and make the sample points fall inside the hyperplane as much as possible. The points that fall outside the hyperplane are to be recorded as losses, and the losses are used to continuously optimize our model.

Mathematical description of the main task of the SVR model.

$$\begin{aligned} &\min_{\omega,b,\xi,\xi^*} \tfrac{1}{2}\|\omega\|^2 + C\sum_{i=1}^{m}\left(\xi_i + \xi_i^*\right) \\ &s.t. \qquad f(x_i) - y_i \le \varepsilon + \xi_i; y_i - f\left(x_i\right) \le \varepsilon + \xi_i^*; \xi_i \ge 0; \xi_i^* \ge 0, i = 1,2,\ldots,m \end{aligned} \tag{1}$$

3.4. Model performance testing

The SVR was implemented using Matlab's own tools and the data was imported for training. During the training process, we used the cross-validation method to check for us the level of the generalization ability of the model[4].

Through the predicted scatter plots for mass sum, Co loading, ethanol concentration, charging method, feeding ratio and temperature, we can observe that the predictions for temperature, charging ratio and charging method are more accurate; the predictions for Co loading, Co/SiO2 and HAP mass sum are average.

The cross-validation process diagram is shown in Figure 3.

ISSN: 2664-9640

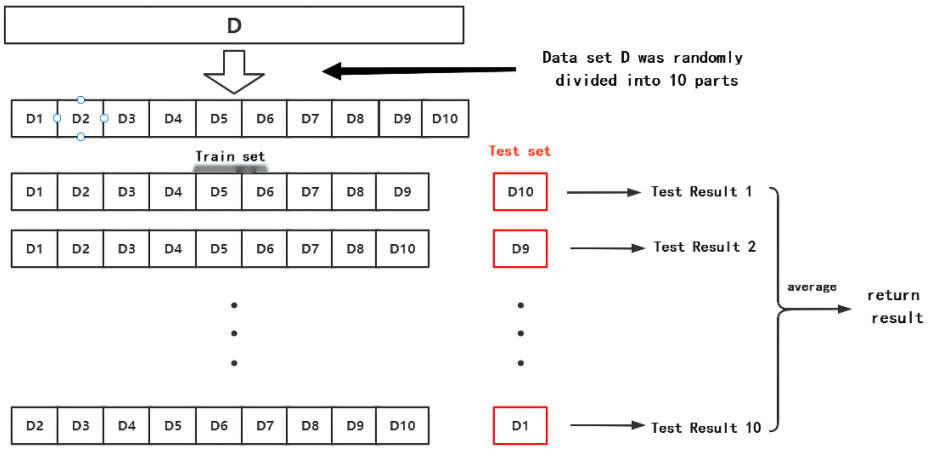

Figure 3 Cross-validation process diagram

The training results of the model are shown in Figure 4.

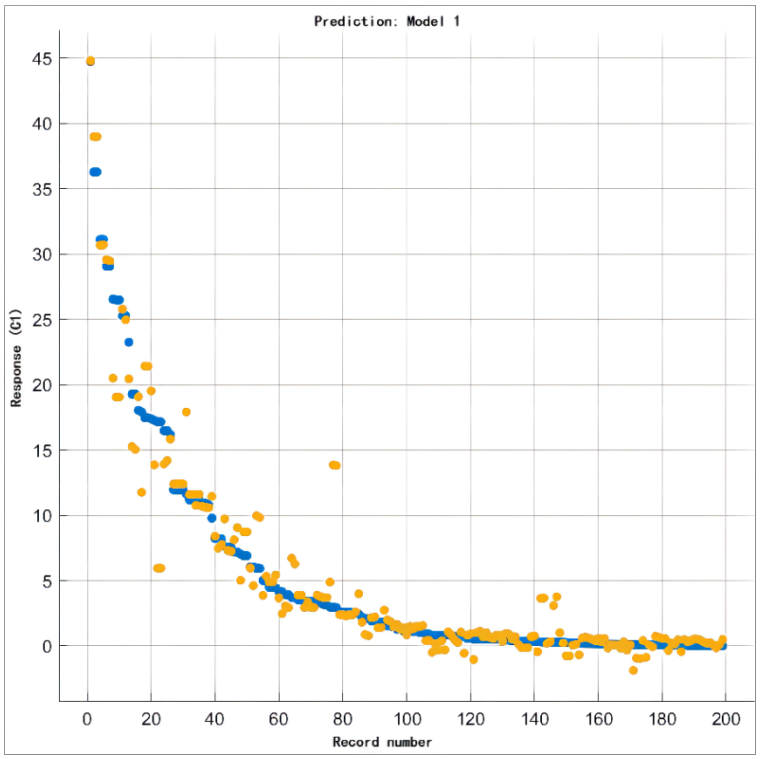

Figure 4 Model Results

3.5. Optimization of SVR models (Bayesian tuning)

According to the information returned by Matlab, the best results were obtained using a cubic SVM model, mainly for the "box constraint", "Epsilon", and "kernel scale" of the model. "Epsilon" and "kernel scale", the three parameters are Bayesian tuning[5].

Bayesian conditioning is a heuristic algorithm that uses a Gaussian process to continuously update the prior by considering previous parameter information. Compared with the traditional grid search, Bayesian tuning is more rapid and does not suffer from local optimal solutions during the tuning process for non-concave functions.

The model after the tuning reference is shown in Table 3.

Table 3 The model after tuning

R^2	0.93
MSE	5.8984

ISSN: 2664-9640

MAE	1.3246
RMSE	2.135
Kernel functions	Three times

3.6. Model Solution and Analysis

3.6.1 Data Prediction

The catalyst combinations were rearranged to obtain 4,032 catalyst and temperature combinations, which were brought into the model for prediction, and the following results were obtained in Table 4.

Table 4 Top 10 catalyst combinations for C4 olefin yields

CO capacity	Ethanol concentration	rate of charge	Q	loading pattern	T	C4 yield
2	1.68	1	400	1	400	50.69575038
2	2.1	1	400	1	400	44.79287148
2	0.9	1	400	1	400	44.40779861
1	0.9	1	400	1	400	44.32064372
0.5	0.9	1	400	1	400	42.98452236
2	1.68	0.5	400	1	400	42.82018186
1	1.68	1	400	1	400	41.40578185
2	2.1	0.5	400	1	400	37.91771643
2	0.9	0.5	400	1	400	37.82296789
1	0.9	0.5	400	1	400	37.58152624

3.6.2 Draw a conclusion

According to the results predicted by the SVR model, the highest yield of C4 olefins was predicted to be 50.7% when the Co loading was 2 wt%, the ethanol concentration was 1.68 ml/min, the Co/SiO2 and HAP loading ratio was 1, the mass sum of Co/SiO2 and HAP was 400 mg, the loading method was Ⅰ, and the temperature was 400 degrees. This result is in accordance with the analysis of the results of the previous model, the ethanol conversion and C4 olefin uptake are higher at high temperatures, the higher the Co, the larger the mass sum of Co/SiO2 and HAP, the closer the contact between catalyst and reactants, which is to make the reaction more adequate. when the loading ratio of Co/SiO2 and HAP is 1, the catalyst is more uniformly dispersed, which is more beneficial to increasing the activity.

ISSN: 2664-9640

3.7. Model solution under temperature condition constraints

All predictions were screened and analyzed when the temperature was less than 350 degrees. Due to the limitation of space, only the top ten catalyst combinations for C4 olefin yield at 325°C are shown in Table 5.

Table 5 Top 10 catalyst combinations for C4 olefin yield at 325°C

CO capacity	Ethanol concentration	rate of charge	Q	loading pattern	T	C4 yield
2	1.68	1	400	1	325	15.86806012
2	2.1	1	400	1	325	15.38891603
2	1.68	0.5	400	1	325	13.67100143
2	2.1	0.5	400	1	325	13.34736319
2	1.68	1	200	1	325	11.15481335
1	1.68	1	400	1	325	10.78510491
2	1.68	2	400	1	325	10.73607814
2	2.1	2	400	1	325	10.6492519
0.5	0.3	1	200	1	325	10.6376453
1	0.9	1	400	1	325	10.32492789

It was concluded that different catalyst combinations were selected for different temperature conditions as shown in Table 6.

Table 6 Catalyst combination selection for conditions below 350 degrees

Temperature (℃)	CO Load	Ethanol concentration	Feeding ratio	Sum of quality	Charging method
325	2wt%	1.68ml/min	1	400	1
300	2wt%	2.1ml/min	1	400	1
275	5wt%	0.3ml/min	2	40	1
250	5wt%	0.3ml/min	2	40	0

4. Conclusion

A model with a goodness-of-fit of 92% was obtained using oversampling and Bayesian conditioning. The highest yield of C4 olefins was predicted for all catalyst combinations and temperatures, with a prediction of 50.7% when the Co loading was 2 wt%, the ethanol concentration was 1.68 ml/min, the Co/SiO2 and HAP loading ratio was 1, the mass sum of Co/SiO2 and HAP was 400 mg, the loading method was Ⅰ, and the temperature was 400 degrees. As well as the conclusions of increasing the Co loading, decreasing the ethanol concentration and reducing the catalyst dosage when the temperature was lower than 300 degrees Celsius.

The SVR model was optimized by Bayesian tuning of the model, and the goodness of fit reached 0.93, which is an obvious fitting effect and makes the model prediction more scientific. The model generalization is insufficient, and the prediction of data that deviate too far from the range of values of the original experimental variables yields too large an error.

References

[1] Lv Shaopei. Preparation of butanol and C4 olefins by ethanol coupling [D]. Dalian: Dalian University of Technology Online Journal, 2018.06

ISSN: 2664-9640

[2] Meng, T. H., Lu, Y. H., Ding, S. J., HD, L. & Huang, R.. (2020). SVM-based regression model for temperature prediction. Modern Computer (20),3-6+13.
[3] Baoan Li,Yuqi Zhang & Pei-Luan Li. (2022). Bayesian optimization based SVM-Xgboost mobile payment risk warning model. Journal of Henan University of Science and Technology (Natural Science Edition) (05),50-56+7.
[4] Li Shaowei, Wang Yujie, Xiong Lang & Huang Shengqi. (2021). An optimization model for the preparation of C_4 olefins by ethanol coupling. Journal of Taizhou College (06), 26-32+77.
[5] Bong, K. and. (2011). Comparison of SVM classification kernel functions and parameter selection. Computer Engineering and Applications (03), 123-124+128.