

# 基于支持向量回归的乙醇催化反应分析与设计

## 摘要

$C_4$  烯烃是一种重要的基础化工原料，广泛应用于各行各业的生产，其需求量也逐年递增。乙醇俗称酒精，由于其化学键容易断裂，使其可用于各种化工产品合成。作为工业合成  $C_4$  烯烃的重要原料之一，如何调节反应温度以及催化剂如何搭配至关重要。因此本文在深入分析各因素对其化学反应影响的基础之上，通过优化算法探索乙醇催化偶合制备  $C_4$  烯烃的最佳工艺条件。

针对问题一，需要我们分析附件 1 中每种组合下乙醇转化率和  $C_4$  烯烃选择性与温度之间的关系，并对附件 2 中的测试结果展开分析。为此我们以温度为自变量，分别使用指数函数和二次多项式函数来对附件 1 中各组数据的乙醇转化率和  $C_4$  烯烃的选择性进行拟合，并对各组函数的特征进行分析，以此解决第 1 小问。对于第 2 小问，我们得到  $C_{4-12}$  脂肪醇的选择性与乙醇转化率之间存在显著正相关、乙醛选择性跟反应时间之间存在线性关系以及时间不是影响乙烯、 $C_4$  烯烃以及甲基苯甲醛和甲基苯甲醇选择性的关键因素这三条结论。

针对问题二，我们首先对催化剂组合进行分解得到三组新的自变量 ( $Co/SiO_2$  和 HAP 装料比、 $Co$  负载量和乙醇浓度)，随后剔除 A11、A13 和 A14 三组实验组合，最后分别对 A 组、B 组数据进行多元线性回归。经检验得，所有回归模型的存在均具有显著意义。通过对回归模型解释分析，以此得到不同温度与催化剂组合对  $C_4$  烯烃选择性和乙醇转化率之间的关系。

针对问题三，我们分离所给数据中的变量，并进行一定的处理。使用支持向量机回归拟合现有的实验数据以构建预测模型。通过构建的约束随机优化方法以及粒子群算法求得最优  $C_4$  烯烃收率为 51.13%。并通过对变量进行约束以满足反应温度低于  $350^{\circ}C$  的题目要求，在此约束条件下，得到最优  $C_4$  烯烃收率为 32.54%。

针对问题四，我们分析现有实验数据，根据加料方式及反应温度范围进行划分，同时考虑实际生产中可能遇到的极端情况。通过对部分条件进行约束，使用问题三构建的模型生成数组反应条件。并依此设计实验，以验证模型的准确性。

本文通过对现有实验数据进行分析，找出对实验结果影响较大的变量并探究其变化规律。在此基础上构建预测模型，对未来的实验设计提供重要参考。通过优化方法给出一定条件下的最佳实验方案，以供指导实验实施并验证模型。

**关键字：** 非参数检验   多元线性回归   支持向量机   粒子群算法

## 一、问题重述

### 1.1 问题背景

乙醇制备  $C_4$  烯烃反应需要经历脱氢、脱水、缩合的过程，在这个过程中不可避免地会生成其他产物，如：酮类、醛类，使得目标产物的选择性降低，适用性减弱。研究结果显示同时具有酸碱双功能性的催化剂能使提升目标产物的选择性，使其达到较高的收率。而  $Co/SiO_2$ 、 $HAP$  混合催化剂就具有这种功能， $Co$  具有脱氢活性， $SiO_2 - HAP$  同时具有酸碱活性位。使  $Co$  负载在  $SiO_2 - HAP$  表面，调节混合物的酸碱度，可以提高乙醇转化率以及  $C_4$  烯烃的选择性，以提高  $C_4$  烯烃的收率。[2]

在制备过程中，催化剂组合（ $Co$  负载量、 $Co/SiO_2$  和  $HAP$  装料比、乙醇浓度）与温度对乙醇转化率和  $C_4$  烯烃选择性产生较大影响。因此通过调整催化剂组合，探索乙醇催化偶合制备  $C_4$  烯烃的最佳工艺条件具有现实意义。

### 1.2 问题数据

附件 1 给出了性能数据表，给出了在不同催化剂组合、不同装料方式下、不同温度下，乙醇的转化率以及主要生成物和其他生成物的选择性；

附件 2 给出了  $350^\circ C$  时给定的某种催化剂组合在反应了不同时间后的测试数据。

### 1.3 问题要求

某化工实验室针对不同催化剂在不同温度下做了一系列实验，数据已经在附件 1 和附件 2 中给出，基于以上背景和数据，要求我们建立数学模型来解决以下的问题：

1. 对于附件 1 中给出的不同催化剂组合条件下得出的实验结果，研究乙醇转化率与温度、 $C_4$  烯烃选择性与温度之间的关系，并研究附件 2 中给出的在  $350^\circ C$  条件下某种催化剂组合在反应不同时间后的测试数据，对数据进行分析；
2. 研究不同催化剂组合及温度作用下，乙醇转化率、 $C_4$  烯烃选择性的变化规律；
3. 研究在相同实验环境条件下，应该如何选择催化剂组合与温度，使  $C_4$  烯烃的收率尽可能地高；如果将温度限制在  $350^\circ C$  以下，又应该选择哪种催化剂组合、在何种温度条件下，使  $C_4$  烯烃的收率尽可能地高；
4. 在得出以上结果后再次设计五组实验，并给出选择设计方案的理由。

## 二、问题分析

### 2.1 问题一的分析

针对问题一的第1小问，需要对附件1中每种催化剂组合下乙醇的转化率、 $C_4$  烯烃的选择性与温度的关系展开研究。我们决定选择温度作为自变量，分别以乙醇转化率和  $C_4$  烯烃的选择性为因变量，挑选合适的函数进行拟合。

查阅文献可知：将  $Co/SiO_2 - HAP$  与  $HAP$  催化剂以 1:1 比例机械混合应用于乙醇催化转化反应，乙醇浓度为  $WHSV = 0.48 \cdot g_{C_2H_5OH} \cdot g_{cat}$ ，在此条件下  $C_4$  烯烃的收率随反应温度的增加呈现出先增大后减小的趋势、乙醇转化率呈现出递增趋势。经检验，附件1中的数据除实验A5的乙醇转化率随温度的增加呈现先减小后增加的趋势以外，其余各组实验的观测数据均符合文献结论。

综上所述，我们选择使用具有随自变量的增加呈现先增加后减小趋势的二项式函数来作为拟合  $C_4$  烯烃的选择性与反应温度之间关系的函数表达式，使用指数函数作为拟合乙醇浓度与反应温度之间关系的函数表达式。分析两组函数表达式的数学特征，由此得出每种催化剂组合下乙醇的转化率、 $C_4$  烯烃的选择性与温度的关系。

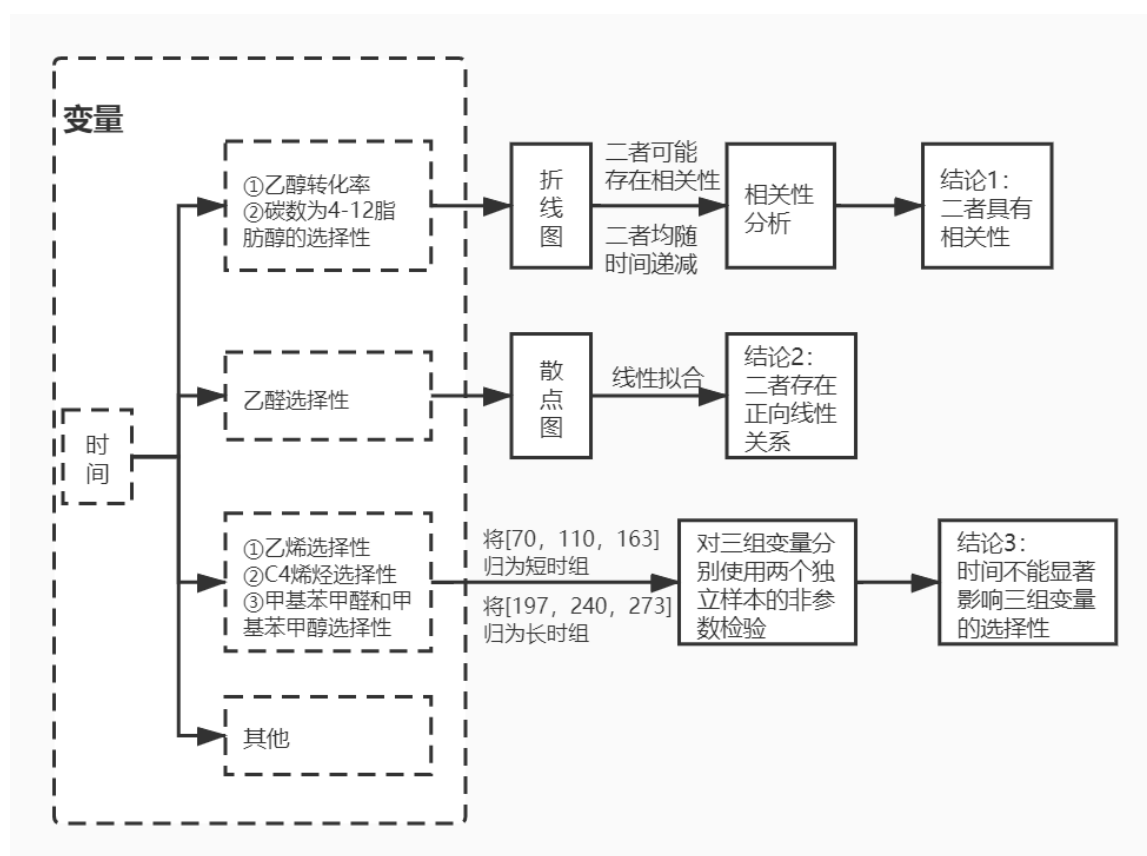


图1 问题一第2问模型

针对问题一的第2小问，需要对附件2中催化剂组合在不同时间的测试结果进行分析。经过对所给数据的观察，我们决定对  $C_{4-12}$  脂肪醇的选择性与乙醇转化率之间的相

关性、乙醛选择性跟反应时间之间的关系、时间是否是影响乙烯、 $C_4$  烯烃以及甲基苯甲醛和甲基苯甲醇选择性的关键因素这一系列问题展开研究。

## 2.2 问题二的分析

针对问题二，我们决定使用多元回归模型分别对乙醇转化率和  $C_4$  烯烃选择性进行回归分析，分析回归系数以此得出不同催化剂组合中各变量对乙醇转化率以及  $C_4$  烯烃选择性大小的影响程度。

## 2.3 问题三的分析

在已知不同催化剂组合及温度对试验结果的影响后，使用非线性机器学习模型构建催化剂与反应温度对  $C_4$  烯烃收率的特征模型。通过优化方法得到该特征模型在限定条件下的极大值即尽可能高的  $C_4$  烯烃收率，以及对应的催化剂组合与温度。再在加入题目要求温度限定条件的情况下再次使用优化方法处理模型。

## 2.4 问题四的分析

增加的 5 组实验可用于验证模型的正确性和预测结果的准确性。使用问题三中构建的模型，通过优化方法求出不同温度范围与装料方法对应的反应条件。如对温度和装料方式的限定以及特征模型输出的极端值，以这些反应条件为基础设计 5 组实验。

# 三、模型假设

**假设 1** 混合物中的石英砂不参与反应。

**假设 2** 忽略实验材料中的杂质对试验结果的影响。

**假设 3** 实验中催化剂的效果仅取决于各组分的质量而不取决于其形态。

**假设 4** 实验数据均为真实情况。

**假设 5** 试验结果仅受所给变量影响。

## 四、符号定义

符号	单位	定义	符号	定义
$lm$	$A, B$	装料方式	$X_{Ethanol}$	乙醇转化率
$m_{Co/SiO_2}$	$mg$	催化剂中 $Co/SiO_2$ 的质量	$S_{Ethanel}$	乙烯选择性
$w_{t\%}$	$wt\%$	$Co$ 负载量	$S_{alkene}$	$C_4$ 烯烃选择性
$m_{Co}$	$kg$	催化剂中 $Co$ 的质量	$S_{Ethanal}$	乙醛选择性
$m_{SiO_2}$	$kg$	催化剂中 $SiO_2$ 的质量	$S_{FattyAlcohols}$	$C_{4-12}$ 脂肪醇选择性
$m_{HAP}$	$kg$	催化剂中 $HAP$ 的质量	$S_{xylene}$	甲基苯甲醛/醇选择性
$\rho_{Ethanol}$	$m^3/s$	乙醇浓度	$S_{others}$	其他生成物的选择性
$T$	$K$	温度	$Y_{alkene}$	$C_4$ 烯烃收率

## 五、模型建立与求解

### 5.1 问题一

#### 5.1.1 模型的建立

通过文献资料以及观测每种催化剂组合下乙醇的转化率、 $C_4$  烯烃的选择性与温度的关系散点图可知应当分别使用递增函数、先增加后减小的函数来拟合附件 1 中的数据。

考虑到每种催化剂组合所给样本点数量较少，因此我们要尽可能选用形式简单的函数表达式来对样本点进行拟合，以减小出现过拟合现象的可能性。

①乙醇的转化率：通过观察乙醇转化率与温度的关系散点图，不难看出，他们之间并非是线性递增关系，而是指数增长关系。因此我们选择采用指数函数

$$y = a \cdot e^{bx} \quad (1)$$

作为拟合乙醇转化率与反应温度之间关系的函数表达式。

②  $C_4$  烯烃的选择性：通过观察  $C_4$  烯烃的选择性与温度的关系散点图可以看出，随温度的增加，有些催化剂组合<sup>1</sup>  $C_4$  烯烃的选择性呈现出先增加后减小的趋势，但大多数

<sup>1</sup> A3

的催化剂组合呈现出非线性递增趋势。结合文献资料不难得出出现这种情况的原因是实验次数过少，反应温度还未达到使 C4 烯烃选择性达到最大的阈值点。考虑到二项式函数具有先增加后减小的趋势，且具有非线性递增趋势，同时形式也较为简单，因此我们选择使用二项式函数

$$f(x) = c_2x^2 + c_1x + c_0 \quad (2)$$

来作为拟合乙醇浓度与反应温度之间关系的函数表达式。

### 5.1.2 模型的求解

①由于 (1) 并不是线性函数（对参数线性），因此我们不能采用决定系数  $R^2$  来对模型的好坏进行评价，为此我们选择  $SSE$  来对模型进行评判<sup>2</sup>。拟合结果如表 1、表 2 所示：

表 1 SSE 最小的五组催化剂组合

催化剂 组合编号	a	b	SSE
B5	0.01255	0.02046	0.2212
B3	0.000712	0.02574	0.3494
A11	0.000383	0.02838	0.4248
A9	0.006919	0.0217	0.4692
B4	0.002003	0.02431	0.8555

表 2 SSE 最大的五组催化剂组合

催化剂 组合编号	a	b	SSE
A7	3.723	0.007627	72.43
A6	0.7285	0.01193	122
A2	0.2377	0.01636	243.5
A4	0.7263	0.01213	345.4
A3	3.091	0.007643	602.9

虽然后五组所对应的  $SSE$  较大，但考虑到自变量与因变量数值较大，因此相对而言，即便是后五组的  $SSE$  仍在可接受范围内。综上，采用指数函数作为拟合乙醇转化率与反应温度之间关系的函数表达式较为合理。

经查阅资料可知：随着反应温度的增加，乙醇的转化率并不会一直增加，而是存在一个玻璃顶限制其转化上限，但由于所给数据数量有限，我们并不能根据已有数据预测何时出现玻璃顶，也许反应温度刚刚大于已测温度的最大值，乙醇的转化率就达到最大值。因此我们只能根据该函数解释反应温度在  $(T_{min}, T_{max})$  内的反应，乙醇转化率随温度呈现出指数增长的态势，并可对处于该温度区间内的其他各点反应转化率进行粗略预测，并不能用来预测  $T > T_{max}$  温度下所对应的反应转化率。

<sup>2</sup>误差平方和

②拟合结果如表 3、表 4所示：

表 3  $R^2$  最大的五组组合

催化剂 组合 编号	$c_2$	$c_1$	$c_0$	$R^2$
A12	9.02e-4	-0.3829	45.32	0.9993
A11	1.83e-4	-0.06745	5.586	0.9994
A8	7.47e-4	-0.2446	19.82	0.9995
A14	9.62e-4	-0.494	64.86	0.9995
A7	1.154e-3	-0.5651	74.78	0.9997

表 4  $R^2$  最小的五组组合

催化剂 组合 编号	$c_2$	$c_1$	$c_0$	$R^2$
A1	-2.11e-3	1.422	-190.8	0.916
A6	2.204e-3	-1.234	176.6	0.9454
A3	-9.5e-4	0.9244	-171.1	0.9551
B6	3.26e-4	-0.02167	-12.13	0.9709
B4	1.002e-3	-0.5487	81.07	0.9737

相比于上面所使用的指数模型，(2) 模型存在极大值点（文献指出随反应温度的增加， $C_4$  烯烃的选择性呈现先增加后减小的趋势），我们可以试图通过求导将其求出。但需注意的是由于所给数据量较少，且某些实验<sup>3</sup> 存在一两个与文献结论相悖的数据点，导致预测函数出现极小值点，在此情况下我们就不能用极值点去估计反应何时出现极值。但即便如此，由于拟合优度普遍较高，因此我们仍可使用拟合函数去预测  $T \in (T_{min}, T_{max})$  所对应的  $C_4$  烯烃的选择性。

虽然大多数的函数出现极小值点，但仍有一些函数与我们的预期相符。我们可以对其求导，预测  $C_4$  烯烃的选择性达到最大值时所对应的反应温度  $T_0$ 。符合预期的组合如表 5所示：

表 5 符合预期的组合的拟合值与极大值点处的  $C_4$  烯烃选择性 (%)

催化剂组合编号	$c_2$	$c_1$	$c_0$	$R^2$	$T_0(^{\circ}C)$	$f(T_0)$
A1	-0.00211	1.422	-190.8	0.916	337	48.78
A3	-0.00095	0.9244	-171.1	0.9551	487	53.77
A13	-0.00028	0.2873	-32.84	0.9817	513	40.86

<sup>3</sup>A10

同时，由图 2 可以看出，指数函数所对应的  $SSE$  具有很强的组别差异，A 组  $SSE$  普遍较大，B 组  $SSE$  普遍较小。说明 B 组实验数据的拟合程度较高，即 B 组实验乙醇转化率的增长更符合指数增长模型。而指数增长模型的增长态势呈现爆炸型，且常常出现在一个事物的发展初期，因此我们可以认为，B 组实验乙醇转化率相较于 A 组具有更长久的增长趋势。

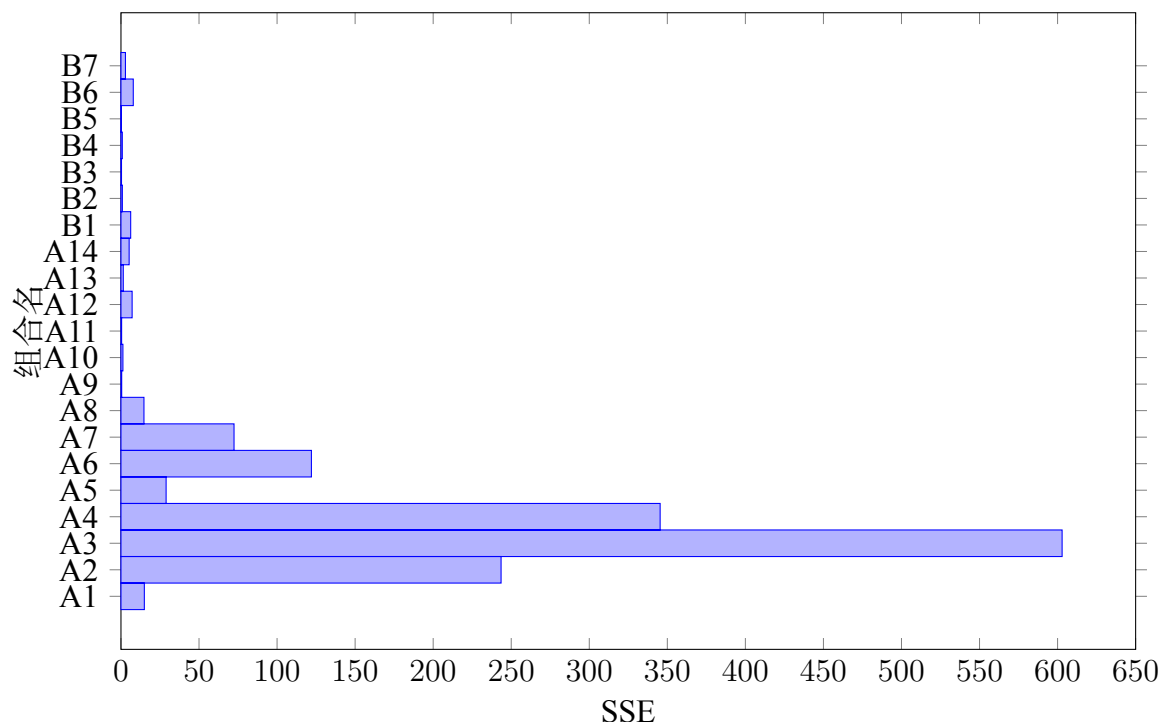


图 2 各组指数模型的  $SSE$

### 5.1.3 350°C 下给定催化剂组合单次实验对时间的测试结果分析

#### ① $C_{4-12}$ 脂肪醇的选择性与乙醇转化率之间存在显著正相关

通过观察以乙醇转化率和碳数为  $C_{4-12}$  脂肪醇的选择性为因变量，以反应时间为自变量的图 3 可知，随着反应的进行，乙醇转化率与碳数为  $C_{4-12}$  脂肪醇的选择性均呈现下降趋势，且二者曲线具有较强的相关性。为检验二者是否显著相关，我们对其进行相关性分析。

由于只有七组样本数据，且数据不服从正态分布，因此我们选择采用 Spearman 相关分析。分析结果如表 6 所示：

由表 6 可知： $C_{4-12}$  脂肪醇的选择性与乙醇转化率之间存在显著的相关关系， $P = 0.052 < 0.1$ （在 90% 的置信水平下）。更进一步， $C_{4-12}$  脂肪醇的选择性与乙醇转化率之间的相关系数为 0.750，为显著的正相关。正相关意味着同升同降，即高  $C_{4-12}$  脂肪醇的选择性意味着高乙醇转化率，低  $C_{4-12}$  脂肪醇的选择性意味着低乙醇转化率。

#### ② 乙醛选择性跟反应时间之间存在线性关系



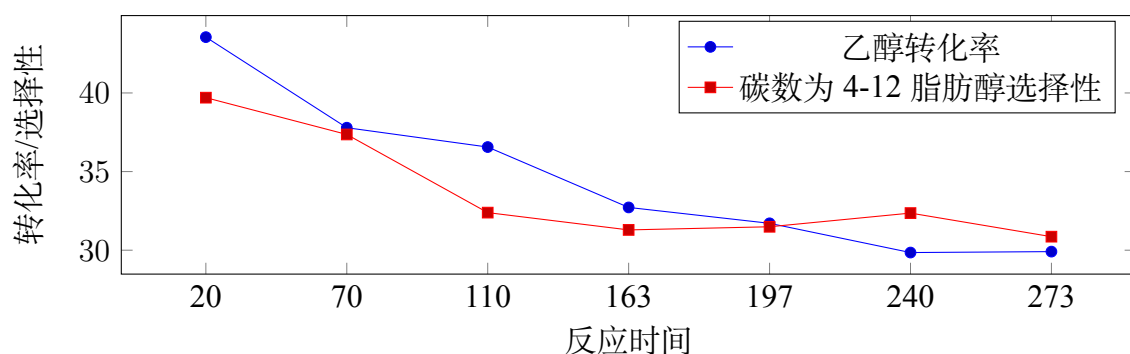


图3 转化率、选择性对反应时间

表6 相关性

		脂肪醇	乙醇转化率
脂肪醇	相关系数	1.000	.750
	显著性 (双尾)	.	.052
	个案数	7	7
乙醇转化率	相关系数	.750	1.000
	显著性 (双尾)	.052	.
	个案数	7	7

由乙醛选择性跟反应时间的散点图可知，二者之间存在一定的线性关系。  
利用 MATLAB 中的拟合工具箱对其进行拟合，得到方程

$$Y = 0.01557 \cdot T + 4.808$$

$R^2 = 0.9588$ ，拟合图像如图 4所示：

由图 4可知，乙醛选择性跟反应时间之间存在正向线性关系，即随着反应时间的进行，乙醇的选择性也在逐渐增加。

③时间不是影响乙烯、 $C_4$  烯烃以及甲基苯甲醛和甲基苯甲醇选择性的关键因素

#### 数据预处理

考虑到一共有七组样本数据，且第 70min 时，甲基苯甲醛和甲基苯甲醇的选择性明显高于第 20min，因此我们选择将第 20min 所对应的样本数据剔除。将 [70min, 110min, 163min] 归为短时组，将 [197min, 240min, 273min] 归为长时组。

分别对三组产物进行两个独立样本的非参数检验，以此判断短时组与长时组的中位数是否具有显著差异，倘若没有显著差异，即可认为反应时间不是影响某种产物选择性

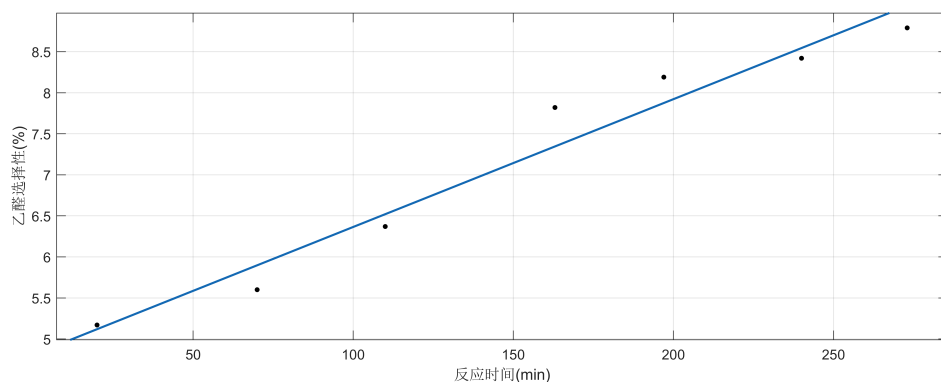


图 4 乙醛选择性

的关键因素。

表 7 乙烯选择性

组别	中位数	Z	显著性
短时组	4.46	-1.528	0.127
长时组	4.68		

表 8 C<sub>4</sub> 烯烃选择性

组别	中位数	Z	显著性
短时组	38.55	-1.091	0.275
长时组	39.04		

- 乙烯选择性

由表 7 可知：显著性  $P = 0.127 > 0.05$ ，因此我们选择接受原假设，认为时间不是影响乙烯选择性的关键因素。

- C<sub>4</sub> 烯烃选择性

由表 8 可知：显著性  $P = 0.275 > 0.05$ ，因此我们选择接受原假设，认为时间不是影响 C<sub>4</sub> 烯烃选择性的关键因素。

- 甲基苯甲醛和甲基苯甲醇的选择性

由表 9 可知：显著性  $P = 0.127 > 0.05$ ，因此我们选择接受原假设，认为时间不是影响甲基苯甲醛和甲基苯甲醇选择性的关键因素。

#### ④ 最佳反应时间

本次实验想要获得的产物为 C<sub>4</sub> 烯烃，由图 5 可知，C<sub>4</sub> 烯烃收率随反应的进行而逐步递减，因此不考虑其他因素，单从 C<sub>4</sub> 烯烃收率方面考虑，最佳的反应时间为 20min。

## 5.2 问题二

### 5.2.1 数据预处理

1. 剔除 Co/SiO<sub>2</sub> 和 HAP 质量不相等的组合 (A11、A13、A14)，原因如下：

表 9 甲基苯甲醛和甲基苯甲醇的选择性

组别	中位数	Z	显著性
短时组	4.63	-1.528	0.127
长时组	4.26		

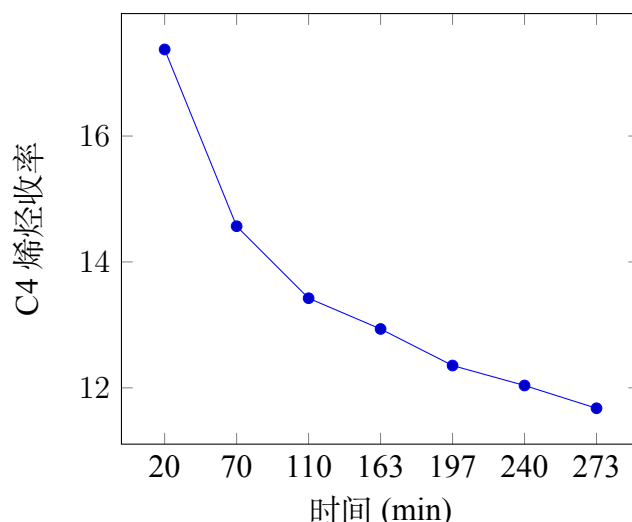


图 5 最佳反应时间

- 文献指出： $HAP$  呈现碱性，当  $Co/SiO_2$  和  $HAP$  装料比减小，即当  $HAP$  含量增加后，催化剂中碱性位含量逐渐增加，能够促进乙醛偶合转化为烯醛。而当催化剂比例为 1:1 时，催化剂表面具有能够促进烯醛转化为  $C_4$  烯烃合适的酸，所以此时  $C_4$  烯烃的选择性最高。而当进一步增加  $HAP$  含量时，催化剂表面酸性位降低，烯醛此时会更倾向于选择性加氢以生成  $C_4-OH$ ，从而进一步偶合生成高碳醇。同时这三组样本所对应的  $C_4$  烯烃收率较差，引入这三组数据可能会导致模型出现较大偏差。
  - 附件 1 中所给的 21 组组合，只有 3 组组合的  $Co/SiO_2$  和  $HAP$  质量不相等。剔除这三组变量以后，我们可以将  $Co/SiO_2$  和  $HAP$  装料比这一整体看作是一个变量，这样就能减少一个变量，这对于问题三这一优化问题来说无疑是巨大利好，能够大大提升问题求解的速度。
2. 将催化剂组合分解为三组自变量，即  $Co/SiO_2$  和  $HAP$  装料比、 $Co$  负载量和乙醇浓度。
  3. 由于 B 组所有数据的  $Co$  负载量均为 1wt%，无法进行回归分析，因此我们将其剔除 B 组的回归模型。

### 5.2.2 模型建立

且根据问题 1 中的结论，B 组实验乙醇转化率相较于 A 组具有更长久的增长趋势，我们选择分别对 A 组、B 组进行多元线性回归。

$$Y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \beta_4 x_4 + \mu$$

$\beta_i (i = 1, 2, 3, 4)$  称为回归系数。

将  $Co/SiO_2$  和  $HAP$  装料比、 $Co$  负载量、乙醇浓度和温度这四组变量作为自变量，分别以乙醇转化率和  $C_4$  烯烃选择性为因变量构建回归模型。

### 5.2.3 模型求解

将各组数据代入模型，回归结果如下所示：

#### A 组

##### 1、乙醇转化率：

由  $R^2 = 0.799$  可以认为拟合状况较好，表示  $Co/SiO_2$  和  $HAP$  装料比、 $Co$  负载量、乙醇浓度和温度这四组变量可以比较好解释乙醇转化率。同时， $F = 53.56$  方差分析结果  $P < 0.05$ ，意味着回归模型的存在具有显著意义。即四组自变量中至少有一个自变量可以显著影响因变量。

表 10 回归系数表 A1

	未标准化系数	标准化系数	t	P	VIF
常量	-92.288		-8.711	<0.001	
温度	0.384	0.807	13.198	<0.001	1.004
乙醇浓度	-8.250	-0.205	-3.179	0.002	1.119
Co 负载量	-0.386	-0.023	-0.361	0.719	1.102
装料比	0.087	0.258	4.189	<0.001	1.014

由表 10 可知，四组自变量的  $VIF$  均小于 5，意味着所有自变量之间不存在多重共线性，即回归模型的运算结果稳定可靠。且除  $Co$  负载量这一自变量的  $P = 0.719 > 0.05$ ，其余自变量的  $P$  值均小于 0.05，因此其余三组自变量均能显著影响因变量。

其中温度与装料比的回归系数大于 0，意味着这两组自变量可以显著正向影响乙醇转化率；乙醇浓度的回归系数小于 0，意味着乙醇浓度可以显著负向影响乙醇转化率（由于乙醇转化率的定义是单位时间内乙醇的单程转化率，倘若乙醇浓度不断增加，会使得反应速度不及进气速度，继而反向影响乙醇转化率）。

分析标准化回归系数可知，温度对于乙醇转化率的影响程度最大，其次是装料比，影响最小的是乙醇浓度。

但单纯考虑乙醇浓度或装料比对于乙醇转化率的影响毫无意义，必须综合考量。即装料比越大乙醇浓度越低，所对应的乙醇转化率越高。

##### 2、 $C_4$ 烯烃选择性：

由  $R^2 = 0.718$  可以认为拟合状况较好, 表示  $Co/SiO_2$  和  $HAP$  装料比、 $Co$  负载量、乙醇浓度和温度这四组变量可以比较好解释  $C_4$  烯烃选择性。同时,  $F = 34.375$  方差分析结果  $P < 0.05$ , 意味着回归模型的存在具有显著意义。即四组自变量中至少有一个自变量可以显著影响因变量。

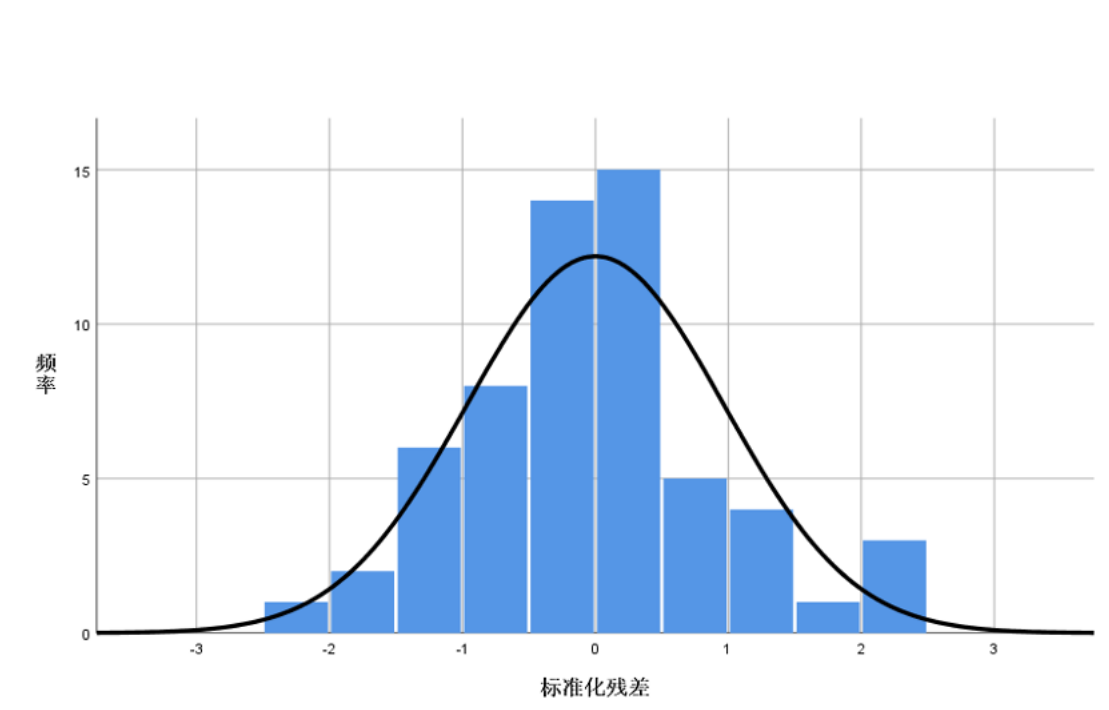


图 6 标准化残差

由图 6 可知: 回归模型的残差基本服从正态分布, 即回归模型的运算结果稳定可靠。

表 11 回归系数表 A2

	未标准化系数	标准化系数	t	P	VIF
常量	-50.907		-6.763	<0.001	
温度	0.199	0.698	9.634	<0.001	1.004
乙醇浓度	5.131	0.213	2.782	0.007	1.119
Co 负载量	-4.055	-0.405	-5.339	<0.001	1.102
装料比	0.063	0.312	4.292	<0.001	1.014

由表 11 可知, 四组自变量的  $VIF$  均小于 5, 意味着所有自变量之间不存在多重共线性, 即回归模型的运算结果稳定可靠。且所有自变量的  $P$  值均小于 0.05, 因此所有自变量均能显著影响因变量。

其中温度与装料比以及乙醇浓度的回归系数大于 0，意味着这三组自变量可以显著正向影响  $C_4$  烯烃选择性； $Co$  负载量的回归系数小于 0，意味着  $Co$  负载量可以显著负向影响  $C_4$  烯烃选择性。

分析标准化回归系数可知，温度对于  $C_4$  烯烃选择性的影响程度最大，其次是  $Co$  负载量，随后是装料比，影响最小的是乙醇浓度。

但单纯考虑乙醇浓度、装料比或  $Co$  负载量对于  $C_4$  烯烃选择性的影响毫无意义，必须综合考量。即装料比越大、乙醇浓度越高、 $Co$  负载量越低，所对应的  $C_4$  烯烃选择性越高。

## B 组

### 1、乙醇转化率：

$R^2 = 0.786$ ,  $F = 44.026$ ,  $P < 0.001$ ，该模型拟合状况较好，模型的存在具有显著意义。

表 12 回归系数表 B1

	未标准化系数	标准化系数	t	P	VIF
常量	-79.179		-6.459	<0.001	
温度	0.281	0.830	10.766	<0.001	1.000
乙醇浓度	-2.336	-0.046	-0.522	0.605	1.322
装料比	0.152	0.288	3.247	0.003	1.322

由表 12 可知，所有自变量之间不存在多重共线性，即回归模型的运算结果稳定可靠。且除乙醇浓度这一自变量不能显著影响因变量外，其余两组自变量均能显著影响因变量。

温度与装料比的回归系数大于 0，意味着这两组自变量可以显著正向影响乙醇转化率。

分析标准化回归系数可知，温度对于乙醇转化率的影响程度最大，其次是装料比。

### 2、 $C_4$ 烯烃选择性：

$R^2 = 0.848$ ,  $F = 67.136$ ,  $P < 0.001$ ，拟合状况较好，回归模型的存在具有显著意义。

由表 13 可知，所有自变量之间不存在多重共线性，即回归模型的运算结果稳定可靠。且除乙醇浓度这一自变量不能显著影响因变量外，其余两组自变量均能显著影响因变量。

其中温度与装料比的回归系数大于 0，意味着这两组自变量可以显著正向影响  $C_4$  烯烃选择性。

表 13 回归系数表 B2

	未标准化系数	标准化系数	t	P	VIF
常量		-44.982	-7.212	<0.001	
温度	0.182	0.888	13.682	<0.001	1.000
乙醇浓度	-1.623	-0.053	-0.713	0.481	1.322
装料比	0.069	0.217	2.912	0.006	1.322

分析标准化回归系数可知，温度对于  $C_4$  烯烃选择性的影响程度最大，其次是装料比。

A11组相比于其他组没有  $HAP$ ，但包含石英砂。由假设 1 可知，石英砂不参与反应，因此A11与A12唯一的区别就是A11不含  $HAP$ ，而A12含有 50mg  $HAP$ 。

由图 7可知，A11与A12两组的乙醇转化率与  $C_4$  烯烃选择性均存在很强的组间差异，且A12的乙醇转化率与  $C_4$  烯烃选择性显著高于A11。这表明  $HAP$  的存在能够显著提升乙醇转化率与  $C_4$  烯烃选择性。这也从侧面印证我们在构建多元回归模型时剔除A11组变量的做法具有一定合理性；同时，这也告诉我们在训练问题三模型时，也要剔除A11组数据。

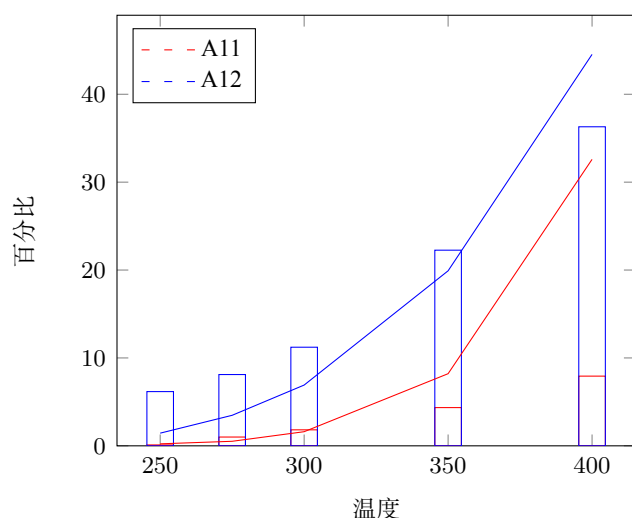


图 7 温度对转化率/选择性的影响

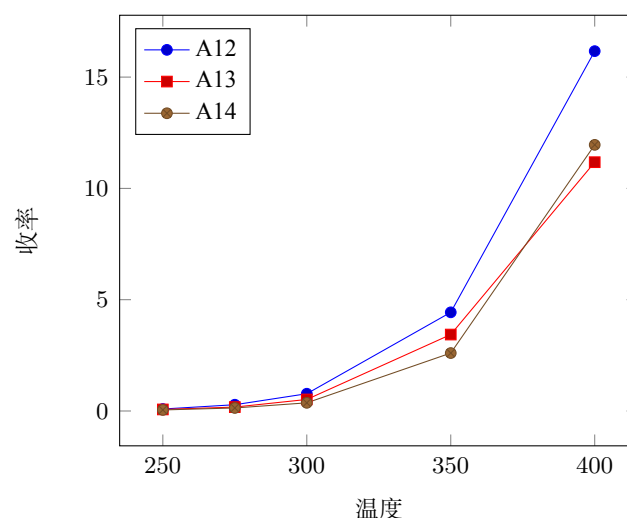


图 8 温度对收率的影响

A12、A13和A14三组，由图 8可见：虽然  $Co/SiO_2$  和  $HAP$  装料比不同，但他们的  $Co/SiO_2$  和  $HAP$  质量之和均为 100mg。虽然在同等温度下，A14的乙醇转化率要高于A12组 ( $Co/SiO_2$  和  $HAP$  质量之比为 1:1)，但A12组  $C_4$  烯烃收率要显著高于其他两组，因此对于问题三预测  $C_4$  烯烃收率的最大值，也应添加  $Co/SiO_2$  和  $HAP$  质量之比

为 1:1 这一条件以减小盲目搜索，加快求解速度。

### 5.3 问题三

#### 5.3.1 数据预处理

- 将“催化剂组合”一列量化分拆为  $Co$  质量、 $SiO_2$  质量、 $HAP$  质量、乙醇浓度四列，具体分拆方法下述。
- 将“催化剂组合编号”中体现装料方式的A、B标识提取为新的一列装料方式，并分别将A、B替换为0、1填入。
- 将所有百分数归一化至  $[0, 1]$  之间的浮点数
- 将所有物理量统一为以国际单位制计量的数值
- 剔除 A11 组实验数据

**例 1** 以 200mg 1wt%Co/SiO<sub>2</sub>- 200mg HAP-乙醇浓度 1.68ml/min 为例

首先我们可以得到 4 个数值变量，分别是  $Co/SiO_2$ 、 $HAP$  催化剂的质量， $Co$  负载量以及乙醇浓度。

根据假设 3，我们认为  $Co/SiO_2$  中的两种组分  $Co$  与  $SiO_2$  分别对反应起到催化作用。为使各变量具有可比较的价值，对变量进行如下操作：

$$m_{Co} = m_{Co/SiO_2} \cdot \frac{w_{t\%}}{(w_{t\%} + 100)} \cdot 10^{-6}$$

$$m_{SiO_2} = m_{Co/SiO_2} \cdot \frac{100}{(w_{t\%} + 100)} \cdot 10^{-6}$$

$$m_{HAP}(kg) = m_{HAP}(mg) \cdot 10^{-6}$$

$$\rho_{Ethanol}(m^3/s) = \rho_{Ethanol}(ml/min) \cdot 60^{-1} \cdot 10^{-6}$$

在问题二分析过的数据的基础上构建一个回归模型，用于通过反应条件预测  $C_4$  烯烃收率。再使用优化算法求得  $C_4$  烯烃收率的最佳情况。根据所给数据，我们知道生成物中含有乙烯、 $C_4$  烯烃、乙醛、碳数为 4-12 脂肪醇、甲基苯甲醛和甲基苯甲醇以及其他生成物。根据相关资料，我们知道乙醇在反应初期转化为乙醛等，乙醛参与了碳数为 4-12 脂肪醇、乙烯、 $C_4$  烯烃等的形成过程。在一系列过程中可能生成乙醚、丙酮、乙酸乙酯等其他物质。对于反应中生成的甲基苯甲醛与甲基苯甲酸，我们通常认为是由脂肪醇成环反应得来。故我们可以得到乙醇的预测反应线路图 9。（图中省略了其他生成物的反应路径）

同时通过对实验方法的分析，我们认为装料方式无法作为一个量化指标输入模型。故针对两种不同的装料方式，我们分别提出两组模型。对乙醇转化率而言，其与产物的选择率均为百分率数据，且根据相关资料，产物选择率不受反应物的转化率影响。对每一组模型，我们按照反应过程选择相应的自变量与因变量。



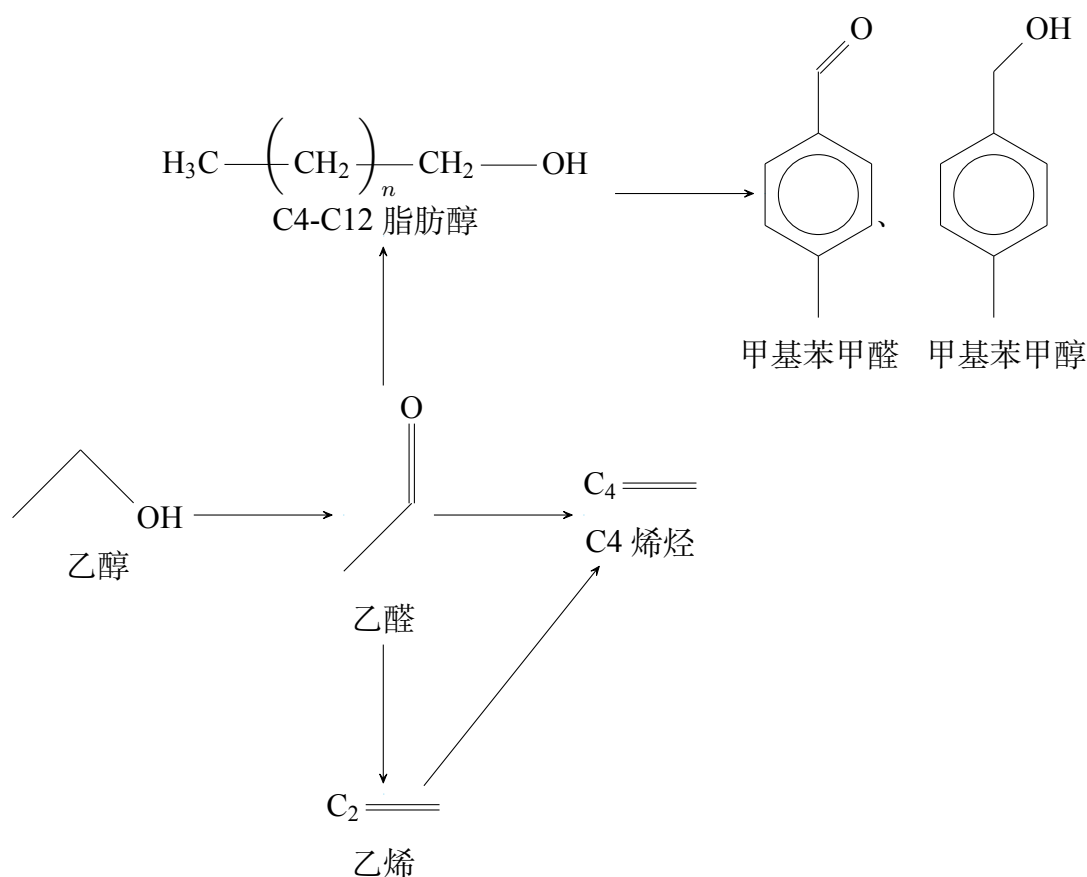


图9 反应路线

### 5.3.2 模型建立

问题二在剔除三组催化剂实验数据，并将 Co /SiO<sub>2</sub> 和 HAP 装料比看成一个变量的基础上构建出一个线性回归模型。但问题三我们为提升预测模型准确度，我们选择将 A13 与 A14 这两组表现较好的数据也代入模型进行预测。但此时自变量不再是三列，而是 Co 质量、SiO<sub>2</sub> 质量、HAP 质量、乙醇浓度四列变量。此时，由于自变量与因变量之间的关系不再是简单的线性，且所给数据不能认为连续数据。故我们使用 SVM（支持向量机）作为基准模型。

SVM 在高维或无限维空间中构造超平面或超平面集合。尽管原始问题可能是在有限维空间中陈述的，但用于区分的集合在该空间中往往线性不可分。为此，将原有限维空间映射到维数高得多的空间中，在该空间中进行分离会更加容易。[4] 就本任务而言，我们使用 SVR<sup>4</sup>，其基本定义如下：

给定  $x_i \in \mathbb{R}^p$   $i=1, \dots, n$  和一个向量  $y \in \mathbb{R}^n$   $\varepsilon - SVR$  以解决以下的主要问题：

$$\min_{w, b, \zeta, \zeta^*} \frac{1}{2} w^T w + C \sum_{i=1}^n (\zeta_i + \zeta_i^*)$$

<sup>4</sup>Support Vectors Regressor

表 14 F(·) 模型变量

反应过程	子模型
<i>GeneralConditions</i>	$T, m_{Co}, m_{SiO_2}, m_{HAP}, \rho_{Ethanol}$
1	$f_1(GeneralConditions) = X_{Ethanol}$
2	$f_2(GeneralConditions, X_{Ethanol}) = S_{Ethanol}$
3	$f_3(GeneralConditions, X_{Ethanol}, S_{Ethanol}) = S_{Ethanol}$
4	$f_4(GeneralConditions, X_{Ethanol}, S_{Ethanol}) = S_{FattyAlcohols}$
5	$f_5(GeneralConditions, X_{Ethanol}, S_{Ethanol}, S_{Ethanol}) = S_{alkene}$
6	$f_6(GeneralConditions, X_{Ethanol}, S_{Ethanol}, S_{FattyAlcohols}) = S_{xylene}$
7	$f_7(GeneralConditions, X_{Ethanol}, S_{Ethanol}, S_{alkene}, S_{Ethanol}, S_{FattyAlcohols}, S_{xylene}) = S_{others}$

$$\begin{aligned}
 &\text{subject to } y_i - w^T \phi(x_i) - b \leq \varepsilon + \zeta_i, \\
 &\quad w^T \phi(x_i) + b - y_i \leq \varepsilon + \zeta_i^*, \\
 &\quad \zeta_i, \zeta_i^* \geq 0, i = 1, \dots, n
 \end{aligned}$$

它的对偶是：

$$\begin{aligned}
 &\min_{\alpha, \alpha^*} \frac{1}{2} (\alpha - \alpha^*)^T Q (\alpha - \alpha^*) + \varepsilon e^T (\alpha + \alpha^*) - y^T (\alpha - \alpha^*) \\
 &\text{subject to } e^T (\alpha - \alpha^*) = 0 \\
 &\quad 0 \leq \alpha_i, \alpha_i^* \leq C, i = 1, \dots, n
 \end{aligned}$$

其中  $e$  是所有的向量,  $Q$  是一个大小为  $n \times n$  的半正定矩阵, 而  $Q_{ij} \equiv K(x_i, x_j) = \phi(x_i)^T \phi(x_j)$  是内核。所以训练向量是通过函数  $\phi$  , 间接反映到一个更高维度的 (无穷的) 空间。[1], 其决策函数是：

$$\sum_{i \in SV} (\alpha_i - \alpha_i^*) K(x_i, x) + b$$

通过训练, 我们得到两组模型。可以认为该模型一定程度上体现了实验条件与结果之间的部分关系。根据上述定义的模型, 我们可以得到  $C_4$  烯烃收率

$$\begin{aligned}
 Y_{alkene} &= X_{Ethanol} \cdot S_{alkene} \\
 &= F(T, m_{Co}, m_{SiO_2}, m_{HAP}, \rho_{Ethanol})
 \end{aligned}$$

根据题目要求，我们需要在两种条件下优化现有模型：

- 1、当温度不限时，求得使  $C_4$  烯烃收率最高的反应条件。
- 2、当温度小于  $350^{\circ}C$  时，求得使  $C_4$  烯烃收率最高的反应条件。

通过对现有实验数据的分析，得到  $C_4$  烯烃收率最高一组的反应条件。根据该条件中各变量的数量级分别设置学习率，设置迭代次数。以该条件为基准，开始迭代：每次迭代中随机调整反应条件中的一个变量的值；调整前判断变量的取值是否在该变量所属的上下界内，若出界则调整至临近的边界值；随机调整的幅度取决于学习率，调整的方向取决于随机的正负值。调整后输入模型得到对该组反应条件可达到的  $C_4$  烯烃收率的预测值。若预测值大于目前  $C_4$  烯烃收率，则将该组反应条件设为基准，该  $C_4$  烯烃收率作为参考值继续迭代。同时，为了避免随机过程出现较大误差，将每次迭代更新的  $C_4$  烯烃收率与参考值继续比较。当二者差值在连续迭代一定次数中均小于设定值时，中止优化过程。

该过程的伪代码如下

```
algorithm 随机优化 is
    input: 最大迭代次数 epoch,
           初始反应条件 X,
           学习率 lr,
           反应条件范围 limit(),
           预测模型 F()
    output: 最优化反应条件 X_best, 最优化C4烯烃收率 y_best

    (初始反应条件为已知数据中C4烯烃收率较高的反应条件)

    X_best ← X
    y_best ← F(X)

    for each in epoch do
        X ← limit(X) # 确保反应条件始终在合理区间
        X ← X + randint(-1, 1) * lr # 随机调整反应条件
        if F(X) > y_best
            x_best ← X
            y_best ← F(X) # 更新最优值
        early_stop(y, y_best) # 当y变化持续小于一定程度后提前停止优化
    return X_best, y_best
```

但是，我们所处理的实验数据的变量间关系存在诸多不确定性。经过实验，简单的随机优化算法并不能得到较好的结果，而粒子群优化算法相对跟容易搜索到有效解。粒子群算法的理论基础是于算法中赋予单一粒子记忆性，并能够透过与粒子群体中的其他粒子之间的互动而寻求到最适解。

任一个粒子皆可用有自身移动过程中所产生的记忆与经验，当个体移动的同时，能

依造自身的经验与记忆来学习调整自身的移动方向，由于在粒子群算法中，多个粒子是同时移动的，且同时以自身经验与其他粒子所提供的经验进行比对找寻最适当的解，并使自己处于最适解中，该粒子群算法的特性使得粒子不单单受自身演化的影响，同时会对群体间的演化拥有学习性、记忆性，并使粒子本身达到最佳调整。[3]

PSO 算法是基于群体的，根据对环境的适应度将群体中的个体移动到好的区域。然而它不对个体使用演化算子，而是将每个个体看作是  $D$  维搜索空间中的一个没有体积的微粒，在搜索空间中以一定的速度飞行，这个速度根据它本身的飞行经验和同伴的飞行经验来动态调整。第  $i$  个微粒表示为  $X_i = (x_{i1}, x_{i2}, \dots, x_{iD})$ ，它经历过的最好位置（有最好的适应值）记为  $P_i = (p_{i1}, p_{i2}, \dots, p_{iD})$ ，也称为  $p_{best}$ 。在群体所有微粒经历过的最好位置的索引号用符号  $g$  表示，即  $P_g$ ，也称为  $g_{best}$ 。微粒  $i$  的速度用  $V_i = (v_{i1}, v_{i2}, \dots, v_{iD})$  表示。对每一代，它的第  $d + 1$  维 ( $1 \leq d + 1 \leq D$ ) 根据如下方程进行变化：

$$v_{id+1} = w \cdot v_{id} + c_1 \cdot rand() \cdot (p_{id} - x_{id}) + c_2 \cdot Rand() \cdot (p_{gd} - x_{id}) \quad (3)$$

$$x_{id+1} = x_{id} + v_{id+1} \quad (4)$$

其中  $w$  为惯性权重， $c_1$  和  $c_2$  为加速常数， $rand()$  和  $Rand()$  为两个在  $[0, 1]$  范围里变化的随机值。此外，微粒的速度  $V_i$  被一个最大速度  $V_{max}$  所限制。如果当前对微粒的加速导致它的在某维的速度  $v_{id}$  超过该维的最大速度  $v_{max,d}$ ，则该维的速度被限制为该维最大速度  $v_{max,d}$ 。

标准 PSO 的算法流程如下：

1. 初始化一群微粒（群体规模为  $m$ ），包括随机的位置和速度；
2. 评价每个微粒的适应度；
3. 对每个微粒，将它的适应值和它经历过的最好位置  $p_{best}$  的作比较，如果较好，则将其作为当前的最好位置  $p_{best}$ ；
4. 对每个微粒，将它的适应值和全局所经历最好位置  $g_{best}$  的作比较，如果较好，则重新设置  $g_{best}$  的索引号；
5. 根据方程 (3) 变化微粒的速度和位置；
6. 如未达到结束条件（通常为足够好的适应值或一个预设最大代数  $G_{max}$ ），回到 2。

### 5.3.3 模型求解

对一个优化方法而言，较多的变量会导致较差的性能。根据前文对数据的分析，我们发现当  $m_{Co} + m_{SiO_2}$  等于  $m_{HAP}$  时， $C_4$  烯烃收率会远高于不等的情况。故我们加入一个限制条件，即  $m_{HAP} = m_{Co} + m_{SiO_2}$  这样将减少一个变量。

通过对数据的分析，我们筛选出现有实验数据中  $C_4$  烯烃收率最高的两组组合，并设定最大迭代次数为 1000000，设定学习率如下：

表 15 学习率

$lr_T$	$lr_{m_{Co}}$	$lr_{m_{SiO_2}}$	$lr_{m_{HAP}}$	$lr_{\rho_{Ethanol}}$
1	1e-7	1e-5	1e-5	1e-9

对题目条件 1，我们分别选用 A3 组 400 度、B7 组 400 度时的反应条件作为参考设定参数。对题目条件 2，我们选用 A2 组 325 度、B7 组 325 度时的反应条件作为参考设定参数。原始参数如表 18

表 16 实验设计

催化剂 组合编号	温度	催化剂组合	C4 烯烃收率 参考值
A3	400	200mg 1wt%Co/SiO <sub>2</sub> - 200mg HAP-乙醇浓度 0.9ml/min	0.4472805978
B7	400	100mg 1wt%Co/SiO <sub>2</sub> - 100mg HAP-乙醇浓度 0.9ml/min	0.2648998
A2	325	200mg 2wt%Co/SiO <sub>2</sub> - 200mg HAP-乙醇浓度 1.68ml/min	0.1726430932
B7	325	100mg 1wt%Co/SiO <sub>2</sub> - 100mg HAP-乙醇浓度 0.9ml/min	0.032841

对变量进行处理输入程序，待程序停止后，采集数据并转换后如表 17

表 17 实验结果

装料 方式	迭代 次数	温度	催化剂组合	C4 烯烃 收率 (%)
A	5998	387	411mg 0.0001wt%Co/SiO <sub>2</sub> - 411mg HAP-乙醇浓度 4.74ml/min	51.13
B	4193	456	376mg 0.0007wt%Co/SiO <sub>2</sub> - 376mg HAP-乙醇浓度 1.92ml/min	50.87
A	1814	312	430mg 0.0004wt%Co/SiO <sub>2</sub> - 430mg HAP-乙醇浓度 5.76ml/min	32.54
B	12743	338	371mg 0.0001wt%Co/SiO <sub>2</sub> - 371mg HAP-乙醇浓度 4.02ml/min	20.92

## 5.4 问题四

在问题三的基础上，我们通过控制输入回归与优化模型的变量构造了如下五组实验，实验的原始数据参见表 21，用于验证预测模型的准确性：

表 18 实验设计

序号	装料方式	温度	催化剂组合
1	A	387	411.1mg 0.3wt%Co/SiO <sub>2</sub> - 411mg HAP-乙醇浓度 4.74ml/min
2	A	312	429.7mg 0.8wt%Co/SiO <sub>2</sub> - 429.7mg HAP-乙醇浓度 5.76ml/min
3	A	346	5.700mg 0.0wt%Co/SiO <sub>2</sub> - 5.700mg HAP-乙醇浓度 3.24ml/min
4	B	313	652.4mg 0.5wt%Co/SiO <sub>2</sub> - 652.4mg HAP-乙醇浓度 4.98ml/min
5	B	359	345.6mg 1.6wt%Co/SiO <sub>2</sub> - 345.6mg HAP-乙醇浓度 4.26ml/min

上述实验分别体现了不同反应温度和装料方式，以及催化剂成分差异的影响，能够有效的验证模型在同一实验的不同条件下的稳定性和准确性。

实验 1、2、4、5 均为问题三中优化模型在不同限制条件下得出的催化剂组合最优解。

对实验 1，验证装料方式 A 与大于 350 度反应条件下模型预测的准确性。

对实验 2，验证装料方式 A 与小于 350 度反应条件下模型预测的准确性。

对实验 4，验证装料方式 B 与小于 350 度反应条件下模型预测的准确性。

对实验 5，验证装料方式 B 与大于 350 度反应条件下模型预测的准确性。

实验 3 的数据来源于装料方式 A 与小于 350 度反应条件下模型的预测结果，可以验证模型在极端条件下，如催化剂中某种成分含量为 0 的情况下模型的稳定性。

## 六、模型评价

### 6.1 优点

本文在构造线性回归模型的过程中，通过查阅资料，以及分析各组催化剂实验中的数据，创造性的将  $Co/SiO_2$  和  $HAP$  装料比这一整体看作为一个自变量放入模型，使模型得到简化。且经检验得，该模型能够较好的反应乙醇转化率和  $C_4$  烯烃选择性与各自变量之间的关系。

使用启发式算法来预测最佳实验方案，能够显著减小计算量，得到较为满意的解。

## 6.2 不足

由于每组催化剂实验中的数据数量较少，即便使用形式简单的函数模型，仍然会出现过拟合情况。

## 6.3 改进

可使用自适应惯性权重等方法来对粒子群算法进行改进。

## 参考文献

- [1] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay. Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, 12:2825–2830, 2011.
- [2] 吕绍沛. 乙醇偶合制备丁醇及 c<sub>4</sub> 烯烃, 2018.
- [3] 维基百科. 粒子群演算法 — 维基百科, 自由的百科全书, 2016.
- [4] 维基百科. 支持向量机 — 维基百科, 自由的百科全书, 2021.

## 附录 A 支撑材料目录

- 参考文献
- 经处理数据 (pickle 文件)
- Python 源代码

## 附录 B 问题一——典型拟合图像

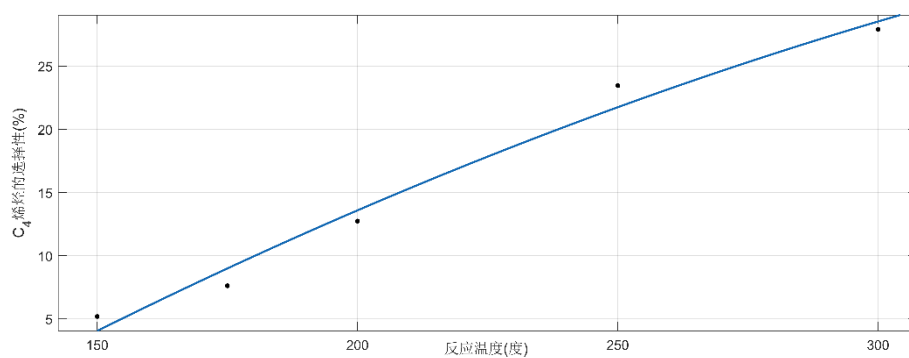


图 10 A13 组所对应的拟合图像

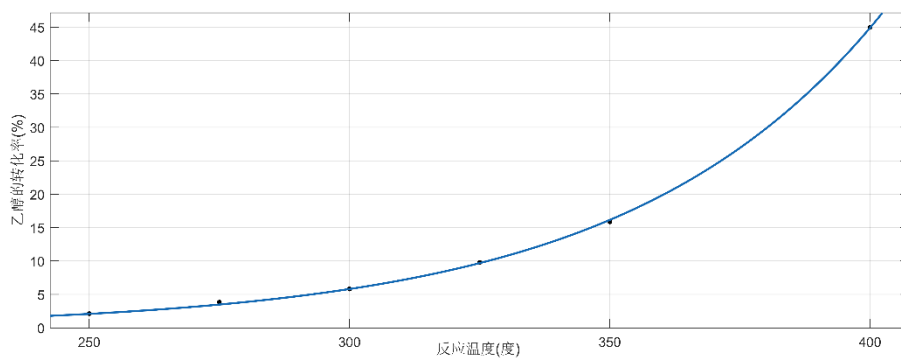


图 11 B5 组所对应的拟合图像



## 附录 C 问题三——原始实验数据

表 19 原始参数

自变量上下界	自变量初始值
{1000, 0.01, 0.99, 1, 1})	598.15, 3.9e-06, 1.96e-4, 0.0002, 2.8e-08
[{273.15, 0.01, 0.99, 1, 1},	673.15, 1.98e-06, 1.98e-4, 0.0002, 1.5e-08
{623.15, 0.01, 0.99, 1, 1})	598.15, 9.9e-07, 9.9e-05, 0.0001, 1.5e-08
[{273.15, 0.01, 0.99, 1, 1},	673.15, 9.9e-07, 9.9e-05, 0.0001, 1.5e-08

表 20 原始结果

装料方式	迭代次数	最优反应条件	最优 C4 烯烃收率
A	370	[768.15, 3.5e-6, 6.0802e-4, 6.1152e-4, 1.6e-8]	0.50150204
A	5998	[660.15, 1.1e-6, 4.1e-4, 4.111e-4, 7.9e-8]	0.51132841
A	111	[712.15, 3.5e-6, 6.1802e-4, 6.2152e-4, 4.4e-8]	0.4763758806422796
B	306	[720.15 3.79e-06 9.49e-04 9.528e-04 4.8e-08]	0.41733941186727525
B	4193	[729.15 6.5e-06 3.69e-04 3.755e-04 3.2e-08]	0.5087239294724344
B	9668	[695.15 8.39e-06 5.7e-04 5.7839e-04 8e-09]	0.5049493165910468
A	1814	[585.15, 3.622e-6, 4.26078e-4, 4.297e-4, 9.6e-8]	0.32540278
A	270	[619.15, 5.7e-06, 0, 5.7e-06, 5.4e-08]	0.223248
A	777	[607.15, 1.8e-6, 2.2e-4, 2.218e-4, 1.5e-8]	0.2336647
B	12743	[611.15 1.5e-06 3.69e-04 3.705e-04 6.7e-08]	0.20918571250988757
B	1638	[535.15 5.5e-06 6.1e-04 6.155e-04 2.9e-08]	0.2006762382286899
B	155	[556.15 1.289e-05 3e-05 4.289e-05 4e-08]	0.10275220798943015

## 附录 D 问题四——原始实验设计

表 21 原始实验设计

装料方式	$T$	$m_{Co}$	$m_{SiO_2}$	$m_{HAP}$	$\rho_{Ethanol}$
A	660.15	1.1e-6	4.1e-4	4.111e-4	7.9e-8
A	585.15	3.622e-6	4.26078e-4	4.297e-4	9.6e-8
A	619.15	5.7e-06	0	5.7e-06	5.4e-08
B	586.15	3.4e-6	6.490099e-4	6.524099e-4	8.3e-8
B	632.15	5.6e-6	3.4e-4	3.456e-4	7.1e-8

## 附录 E 数据处理（分离变量）——Python 源代码

```

附件1 = pd.read_excel('赛题/B/附件1.xlsx', '性能数据表')

附件1['催化剂组合编号'] = 附件1['催化剂组合编号'].ffill()
附件1['催化剂组合'] = 附件1['催化剂组合'].ffill()

old_columns = 附件1.columns
columns = ['乙醇转化率', '乙烯选择性', 'C4烯烃选择性', '乙醛选择性',
           '碳数为4-12脂肪醇选择性', '甲基苯甲醛和甲基苯甲醇选择性', '其他生成物的选择性']

附件1.rename(columns={old_columns[i+3]: columns[i] for i in range(len(columns))},
              inplace=True)

rc = ['装料方式', 'Co', 'SiO2', 'HAP', '乙醇浓度']

for i, c in enumerate(rc):
    附件1.insert(i + 3, c, 0)

def wt(desc):
    coc = float(re.search(r'(?<= )\d+(\.\d+)?(?=wt)', desc).group()) / 100
    cosio2 = float(re.search(r'\d+(\.\d+)?(?=mg )', desc).group()) / 10**6
    co = cosio2 * coc / (coc + 1)
    sio2 = cosio2 * 1 / (coc + 1)
    return co, sio2

for i in 附件1.iterrows():

```

```

附件1.loc[i[0], '温度'] = i[1]['温度'] + 273.15
附件1.loc[i[0], '装料方式'] = 0 if i[1]['催化剂组合编号'][0] == 'A' else 1
附件1.loc[i[0], 'Co'],附件1.loc[i[0], 'SiO2'] = wt(i[1]['催化剂组合'])
if re.search(r'无HAP', i[1]['催化剂组合']):
    w = float(re.search(r'(?<= )\d+(\.\d+)?(?=mg石英砂)',
        i[1]['催化剂组合']).group()) / 10**6
    附件1.loc[i[0], 'SiO2'] += w
else:
    附件1.loc[i[0], 'HAP'] = float(re.search(r'(?<= )\d+(\.\d+)?(?=mg HAP)',
        i[1]['催化剂组合']).group()) / 10**6
附件1.loc[i[0], '乙醇浓度'] = float(re.search(r'(?<=乙醇浓度)\d+(\.\d+)?(?=ml)',
    i[1]['催化剂组合']).group()) / 6 / 10**7
for r in columns:
    附件1.loc[i[0], r] = i[1][r] / 100
附件1.drop(columns=['催化剂组合', '催化剂组合编号'], inplace=True)
附件1.to_pickle('附件1.pickle')

```

## 附录 F 回归模型–Python 源代码

```

import pandas as pd

附件1 = pd.read_pickle('附件1.pickle')

rc = ['温度', 'Co', 'SiO2', 'HAP', '乙醇浓度']
rs = ['乙醇转化率', '乙烯选择性', 'C4烯烃选择性', '乙醛选择性',
    '碳数为4-12脂肪醇选择性', '甲基苯甲醛和甲基苯甲醇选择性', '其他生成物的选择性']

from sklearn.svm import NuSVR
from sklearn.pipeline import make_pipeline
from sklearn.preprocessing import StandardScaler

exp = {
    '乙醇转化率': rc,
    '乙醛选择性': rc + ['乙醇转化率'],
    '乙烯选择性': rc + ['乙醇转化率', '乙醛选择性'],
    '碳数为4-12脂肪醇选择性': rc + ['乙醇转化率', '乙醛选择性'],
    'C4烯烃选择性': rc + ['乙醇转化率', '乙醛选择性', '乙烯选择性'],
    '甲基苯甲醛和甲基苯甲醇选择性': rc + ['乙醇转化率', '乙醛选择性',
        '碳数为4-12脂肪醇选择性'],
    '其他生成物的选择性': rc + ['乙醇转化率', '乙烯选择性', 'C4烯烃选择性',
        '乙醛选择性', '碳数为4-12脂肪醇选择性', '甲基苯甲醛和甲基苯甲醇选择性']
}

model = {m: {i: make_pipeline(StandardScaler(), NuSVR()) for i in exp.keys()} for m
    in [0, 1]}

```

```

for m in [0, 1]:
    for i in exp.items():
        df = 附件1.loc[附件1['装料方式'] == m]
        X = df[i[1]]
        y = df[i[0]]
        model[m][i[0]].fit(X, y)
        print(model[m][i[0]].score(X, y))

```

## 附录 G 优化方法–Python 源代码

```

from tqdm.notebook import trange
import numpy as np
import random

C350 = False
LM = 0

def E(x, lm=0):
    x = np.expand_dims(x, axis=0)
    y_ethanol = model[lm]['乙醇转化率'].predict(x)
    x = np.expand_dims(np.append(x, y_ethanol), axis=0)
    y_ethanal = model[lm]['乙醛选择性'].predict(x)
    x = np.expand_dims(np.append(x, y_ethanal), axis=0)
    y_ethylene = model[lm]['乙烯选择性'].predict(x)
    x = np.expand_dims(np.append(x, y_ethylene), axis=0)
    y_alkene = np.expand_dims(model[lm]['C4烯烃选择性'].predict(x), axis=0)
    t = y_alkene * y_ethanol
    return t

min_x = [273.15, 0, 0, 0, 0]
max_x = [1000, 0.01, 0.99, 1, 1]

if not LM and C350:
    best_x = [598.15, 0.000003922, 0.000196078, 0.0002, 0.000000028]
elif not LM and not C350:
    best_x = [673.15, 0.000001980, 0.000198020, 0.0002, 0.000000015]
elif LM and C350:
    best_x = [598.15, 0.000000990, 0.0000990099, 0.0001, 0.000000015]
elif LM and not C350:
    best_x = [673.15, 0.000000990, 0.0000990099, 0.0001, 0.000000015]

if C350:
    max_x[0] = 623.15

```

```

best_y = E(best_x)
pre_y = best_y

x = best_x[:]

lr = [1, 1e-7, 1e-5, 1e-5, 1e-9]
epoch = 10000000
cnt = 0

path_x = []
path_y = []

with trange(epoch) as tr:
    for i in tr:
        r = random.randint(0, 3)
        if r == 3:
            r += 1
        x[r] = max(x[r], min_x[r])
        x[r] = min(x[r], max_x[r])
        x[r] += (random.randint(-1, 1)) * lr[r]
        x[3] = x[1] + x[2]
        path_x.append(x)
        y = E(x)
        path_y.append(y[0][0])
        if y - pre_y < 1e-8:
            cnt += 1
        if cnt > 10000:
            break
        if y > best_y:
            best_y = y
            best_x = x
            tr.set_postfix(best_y=best_y)
        pre_y = y

path_y = np.array(path_y)
path_x = np.array(path_x)

e = path_y.argmax()

t = path_x[e]

print(e)
print(t)
print(best_y[0][0])

```

## 附录 H PSO 算法–Python 源代码

```
from sko.PSO import PSO

C350 = False
LM = 0

min_x = [273.15, 0, 0, 0, 0]
max_x = [1000, 0.01, 0.99, 1, 1]

def E(x, lm=0):
    x = np.expand_dims(x, axis=0)
    y_ethanol = model[lm]['乙醇转化率'].predict(x)
    x = np.expand_dims(np.append(x, y_ethanol), axis=0)
    y_ethanal = model[lm]['乙醛选择性'].predict(x)
    x = np.expand_dims(np.append(x, y_ethanal), axis=0)
    y_ethylene = model[lm]['乙烯选择性'].predict(x)
    x = np.expand_dims(np.append(x, y_ethylene), axis=0)
    y_alkene = np.expand_dims(model[lm]['C4烯烃选择性'].predict(x), axis=0)
    t = y_alkene * y_ethanol
    return t

if not LM and C350:
    best_x = [598.15, 0.000003922, 0.000196078, 0.0002, 0.000000028]
elif not LM and not C350:
    best_x = [673.15, 0.000001980, 0.000198020, 0.0002, 0.000000015]
elif LM and C350:
    best_x = [598.15, 0.000000990, 0.0000990099, 0.0001, 0.000000015]
elif LM and not C350:
    best_x = [673.15, 0.000000990, 0.0000990099, 0.0001, 0.000000015]

if C350:
    max_x[0] = 623.15

pso = PSO(func=E, n_dim=5, pop=50, max_iter=200, lb=[273.15, 0, 0, 0, 1e-14],
          ub=[1000, 1e-5, 1e-3, 1e-3, 1e-7], w=0.08, c1=5, c2=0.05)
pso.run()
print('best_x is ', pso.gbest_x, 'best_y is', pso.gbest_y)
```