

乙醇偶合制备 C4 烯烃

摘 要

C4 烯烃广泛应用于化工产品以及医药的生产,在制备过程中,催化剂组合与温度起到关键作用。本文通过建立合适的数学模型,探究不同催化剂组合及温度对 C4 烯烃指标的影响,得出最优方案,并对实验室已有的设计方案进行补充与改善。

针对问题一:首先,对不同方案下的温度与乙醇转化率、C4 烯烃的选择性分别进行皮尔逊相关性分析,结果表明均具有强相关性。接着,根据实验已有数据,对不同方案下的温度与乙醇转化率、C4 烯烃的选择性绘制散点图。根据曲线的特征趋势,基于最小二乘法的思想,对曲线进行拟合,结果表明:乙醇转化率与 C4 烯烃的选择性都更符合三次函数的变化趋势,并确定具体的表达式。最后,对拟合结果进行误差分析。

针对 350 度时给定的催化剂组合,具体探究不同生成物与时间的变化关系。首先,利用聚类分析对生成物进行降维,可得到以乙醇转化率与 C4 烯烃选择性为代表的主要成分,和以乙醛选择性和乙烯选择性为代表的次要成分。其次,分别对乙醇转化率与乙醛选择性关于时间进行曲线拟合,结果表明:随着时间的推移,乙醇转化率慢慢减小至稳定值,乙醛选择性逐渐增大至平稳。最后,对曲线拟合的结果进行误差分析。

针对问题二:首先,结合催化剂的组成成分与已有成分,确定 1 个自变量:Co 负载量、HAP 质量、Co/SiO₂ 和 HAP 装料比、乙醇浓度与温度,以及 1 个因变量:乙醇转化率与 C4 烯烃选择性,并进行相关性分析,构造相关系数矩阵。结果表明自变量与因变量均具有一定的相关性。其次,考虑到多元回归的多重共线性,则利用主成分分析对自变量进行降维,筛选出 2 个主成分进行多元回归与回归检验,结果表明拟合优度高且降低多重共线性。接着,采取逐步多元回归,并引入二次交互项以提高拟合优度,通过 5 次变量的遴选可得到关于乙醇转换率的表达式,以及 4 次变量的遴选得到关于 C4 烯烃选择性的表达式。最后,通过对比分析,发现含有交互项的逐步多元回归拟合优度最高且共线性最弱,更能衡量不同催化剂组合及温度对乙醇转化率以及 C4 烯烃选择性大小的影响。

针对问题三:首先,根据 C4 烯烃收率定义与实际变化趋势,对其进行数据处理与 logistic 变换。接着,基于变换后的 C4 烯烃收率数据,与催化剂组合与温度进行含有二次交互项的逐步回归,得到相应的表达式。然后以实验室条件、C4 烯烃收率关于催化剂组合与温度的表达式作为约束条件,以最大 C4 烯烃收率为目标函数,建立单目标优化模型。其次,分别利用遍历法与二分法对优化模型进行求解,并对各自方案进行对比分析,结果表明 200mg 2.147wt%Co/SiO₂-10mg HAP-乙醇浓度 0.3ml/min 的催化剂与 450℃温度的组合最优,C4 烯烃收率最大可达到 49.97%。最后,对实验室的温度上限调整为 350℃,对单目标优化模型的约束条件进行修改,并仍采用遍历法与二分法进行求解,结果表明 200mg 2.1wt%Co/SiO₂-10mg HAP-乙醇浓度 0.3ml/min 的催化剂与 350℃温度的组合最优,C4 烯烃收率最大可达到 49.81%。

针对问题四:在已有实验方案的基础上,综合考虑正交试验与简单对比实验,对实验设计进行补充与改善。首先,根据已有实验数据,提取 5 个因子:Co 负载量、HAP 质量、Co/SiO₂ 和 HAP 装料比、乙醇浓度、温度与 4 个水平。其次,针对不同因子的 4 个水平,进行均匀分散求具有代表性的取值。接着,利用正交表 L16.4.5 进行分析,得出理想的正交实验方案。最后,综合考虑已有实验方案,与实验的可操作性,筛选出添加的 5 组实验方案(如表 23 所示)。

关键词: 多元逐步回归 C4 烯烃收率 循环遍历法 主成分分析 正交试验

一、问题重述

1.1 问题背景

C4 烯烃是重要的基础化工原料,以乙醇为原料生产 C4 烯烃具有良好的应用前景及经济效益。在乙醇偶合制备 C4 烯烃的过程中,不同的催化剂组合和温度将对 C4 烯烃的选择性和乙醇转化率产生重大影响,进一步影响 C4 烯烃收率。因此,利用数学模型探究乙醇催化偶合制备 C4 烯烃的工艺条件具有非常大的意义,同时为化工行业提供新的思路。

1.2 待解决的问题

某化工实验室针对不同催化剂在不同温度下进行一系列实验,结果在附件中给出。我们需要建立数学模型解决以下问题:

问题一:其需要解决 2 个子问题。

- (1) 附件 1 为实验性能数据表,对其每种催化剂组合,分别研究乙醇转化率、C4 烯烃的选择性与温度的关系。
- (2) 附件 2 为 350 度时给定的某种催化剂组合的测试数据,分析 350 度给定时的催化剂组合在一次实验不同时间的测试结果。

问题二:探究不同催化剂组合及温度对乙醇转化率及 C4 烯烃选择性大小的影响。

问题三:其需要解决 2 个递进问题。

- (1) 如何选择催化剂组合与温度,使得在相同实验条件下 C4 烯烃收率尽可能高。
- (2) 在温度低于 350 度的情况下,如何选择催化剂组合与温度,使得 C4 烯烃收率尽可能高。

问题四:在允许的情况下,设计增加 5 次实验并给出理由。

二、问题分析

2.1 问题一的分析

首先,分析题目中的要求,问题一中包含两个子问题,因此此文中需要分布建立两个子模型。对于子问题一,为了探究乙醇转化率、C4 烯烃的选择性与温度的关系,我们首先分别对其进行相关性分析,若相关性较强,我们则使用非线性拟合的最小二乘法拟合曲线,求出对应的方程。针对子问题二,由于化学指标较多,为了更加清晰的分析其结果,我们使用聚类分析对各化学指标进行降维,其次根据最小二乘拟合分析结果。

2.2 问题二的分析

为了探究不同催化剂组合及温度对乙醇转化率及 C4 烯烃选择性大小的影响,首先我们需要分析不同催化剂组合的成分区别,因此我们选取指标衡量催化剂组合,同时计算相关性系数矩阵判断各指标间的相关性,为了建立多元线性回归方程,我们需要判断其共线性。为了消除指标间共线性的影响,首先建立基于主成分分析的多元回归模型,其次建立逐步多元回归模型,并对其结果进行对比,最后提出增加交互项多元回归的模型优化。

2.3 问题三的分析

问题三由 2 个递进的子问题构成，其二者均为优化问题，首先我们建立 C4 烯烃收率与催化剂组合和温度二者的函数表达式，其次，建立 C4 烯烃收率最大的单目标优化模型。第二小问则在此模型的基础上，修改约束条件，建立模型。

2.4 问题四的分析

为了增加合理且有效的实验，我们利用正交试验并结合模型三所得到的最优方案，在筛选并分析结果后，选定最终 5 组试验。

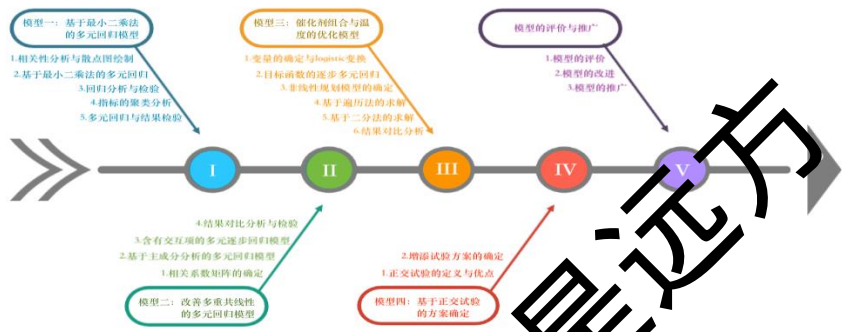


图 1：模型总体框架

三、模型假设

1. 假设不同的装料方式对产物无明显影响。
2. 假设产物仅由催化剂组合与温度影响，不受外界条件干扰。
3. 假设实验使用的材料是有限的，会有化学反应结束或无明显变化的时间节点。
4. 假设已完成的实验组别是实验室能提供的最高/最低条件。
5. 假设 C4 烯烃收率完全等于乙醇转化率以及 C4 烯烃选择性，没有实验误差。

四、符号说明

符号	说明
T_i	第 i 组催化剂温度
E_i	第 i 组催化剂乙醇转化率
O_i	第 i 组催化剂乙烯选择性
C_i	第 i 组催化剂 C4 烯烃选择性
Q_i	第 i 组催化剂乙醛选择性
Z_i	第 i 组催化剂碳数为 4-12 脂肪醇选择性
J_i	第 i 组催化剂甲基苯和甲基苯甲醇选择性
D_i	第 i 组催化剂其他生成物选择性
F	CO 负载量
α	CO/SiO ₂ 装料比

β	<i>HAP</i> 装料比
γ	乙醇浓度

注：其他局部符号在使用时给出各自具体说明

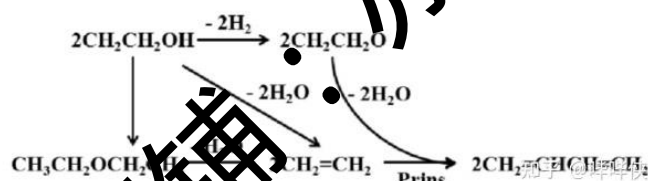
五、模型的建立与求解

5.1 模型的准备

5.1.1 化学反应机理分析

对乙醇生成 C4 的共性认识是反应需要经历脱氢、脱水和缩合过程，同时需要酸性位点催化乙醇脱水，碱活性位点促进醇类脱氢及中间体的偶合的协同催化。乙醇制备 C4 烯烃反应的催化剂主要有沸石/改性沸石和金属氧化物，不同催化剂上反应机理不同。在沸石上，乙醇脱水生成乙烯，乙烯随后进行二聚、异构化和复分解反应生成 C3-C4 烯烃。而在混合金属氧化物上，具有合适酸、碱对的金属氧化物催化乙醇生成乙醛和丙酮中间体，乙醛和丙酮通过进一步偶合脱水生成异丁烯和丁二烯等 C4 烯烃产物。尽管不同催化剂上反应机理不同，但从乙醇出发制备 C4 烯烃均需要经历脱水、偶合等过程。对于乙醇制备烯烃的过程，不同催化剂上反应机理不同，反应机理较为复杂。目前对反应机理的研究较多的是乙醇制备丁二烯的反应。对于该反应，有两种反应机理被提出，一种是普林斯机理，另一种是醇醛缩合机理。

普林斯机理



该反应机理认为反应物乙醇在碱性位的催化作用下脱氢生成乙醛，在酸性位的催化作用下直接脱水生成乙烯或经由乙醚生成乙烯，生成的乙醛和乙烯发生亲核加成反应生成丁二烯^[1]。在此化学反应中，由于反应物有限，在有限的实验条件内，各生成物的成分比例都会逐渐趋于平稳。

5.1.2 指标的选取

根据题目所述，A14 的催化剂组合“33mg 1wt%Co/SiO₂—67mg HAP—乙醇浓度 1.68ml/min”代表 Co/SiO₂ 和 HAP 装料比为 33mg : 67mg 且乙醇按每分钟 1.68 毫升加入。

因此，我们选取 Co 负载量、HAP 含量、Co/SiO₂ 和 HAP 装料比、乙醇浓度作为催化剂组合的衡量指标。问题二可转化为探究 Co 负载量、HAP 含量、Co/SiO₂ 和 HAP 装料比、乙醇浓度、温度此 5 个化学指标对乙醇转化率及 C4 烯烃选择性的影响。

表 1 变量的选取

自变量	Co 负载量、HAP 含量、Co/SiO ₂ 和 HAP 装料比、乙醇浓度、温度
因变量	乙醇转化率、C4 烯烃选择性

指标选取结果绘制为下图：

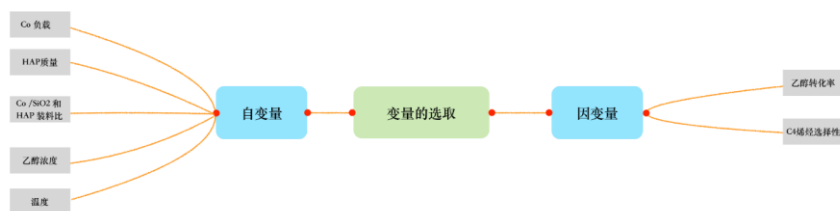


图 2：指标的选取

下面，我们将分别探究自变量 Co 负载量、 HAP 含量、 Co/SiO_2 和 HAP 装料比、乙醇浓度、温度与因变量乙醇转化率和 $C4$ 烯烃选择性的关系。

5.2 模型一的建立与求解

首先，分析题目中的要求，问题一中包含两个子问题，因此本文中需要分布建立两个子模型。针对子问题一，为了探究乙醇转化率、 $C4$ 烯烃的选择性与温度的关系，我们首先分别对其进行相关性分析，若相关性较强，我们则使用非线性拟合的最小二乘法拟合曲线，求出对应的方程。针对子问题二，由于化学指标较多，为了更加清晰的分析其结果，我们使用聚类分析对各化学指标进行降维，其次根据最小二乘拟合分析结果。

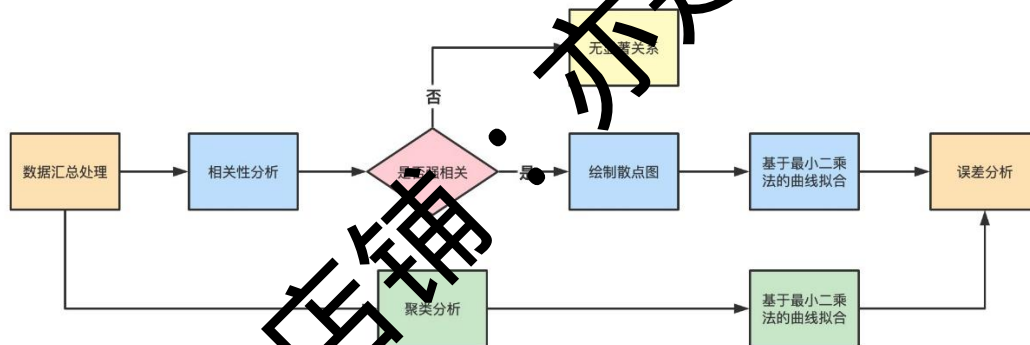


图 3：问题一流程图

5.2.1 子问题一：基于最小二乘法的曲线拟合

根据附件 1 中的数据，有 14 种 A 类催化剂组合，有 7 种 B 类催化剂组合。对于每一种催化剂组合，在 250 度、275 度、300 度、325 度和 350 度温度的试验下，分别对应乙醇转化率等各化学指标的具体数据。本题只需讨论各类催化剂组合温度与乙醇转化率和 $C4$ 烯烃的选择性的关系，下面我们通过求解皮尔逊系数判断其相关性强弱。

● 基于皮尔逊系数的相关性分析

我们分别计算 14 种 A 类催化剂组合和 7 种 B 类催化剂组合对应的温度—乙醇转化率的相关性系数以及温度— $C4$ 烯烃的选择性对应的相关性系数。皮尔逊相关系数的计算表达式为

$$\rho_{X,Y} = \frac{Cov(X,Y)}{\sigma_X \cdot \sigma_Y} = \frac{\sum_{i=1}^n (x_i - E(X))(y_i - E(Y))}{\sqrt{\sum_{i=1}^n (x_i - E(X))^2 \sum_{i=1}^n (y_i - E(Y))^2}}$$

其中， X 代表温度， Y 代表乙醇转化率或 C4 烯烃转化率， n 为数据量个数， n 在本题中取 5， x_i, y_i 具体数据见附件 1。

利用 SPSS 软件，对分别对每种催化剂组合分别计算温度与乙醇转化率和 C4 烯烃转化率的皮尔逊系数，由于篇幅限制，仅展示部分催化剂组合结果，具体数据表格如下：

表 2：相关性结果数据

A 类编号	乙醇转化率	C4 烯烃转化率	B 类编号	乙醇转化率	C4 烯烃转化率
1	0.965	0.887	1	0.962	0.986
2	0.995	0.914	2	0.929	0.985
3	0.982	0.955	3	0.890	0.971
...			4	0.895	0.895
12	0.903	0.989	5	0.912	0.978
13	0.936	0.988	6	0.940	0.982
14	0.964	0.959	7	0.936	0.994

当皮尔逊相关性系数大于 0.8 时即可认为具有强相关性，观察上述表格可以发现，相关性系数均大于 0.9，大多数数据大于 0.95，因此我们认为对于每组催化剂温度关于乙醇转化率和温度关于 C4 烯烃转化率均有强相关性。

● 散点图分析

根据皮尔逊系数的计算，我们发现每组催化剂温度关于乙醇转化率和温度关于 C4 烯烃转化率均有强相关性。因此，我们针对每一组催化剂组合分别画出温度关于乙醇转化率和 C4 烯烃转化率的散点图，为曲线拟合做准备。

由于催化剂组合数较多，我们选择 A4 组和 B3 组分析温度—乙醇转化率的散点图，具体图表如下：

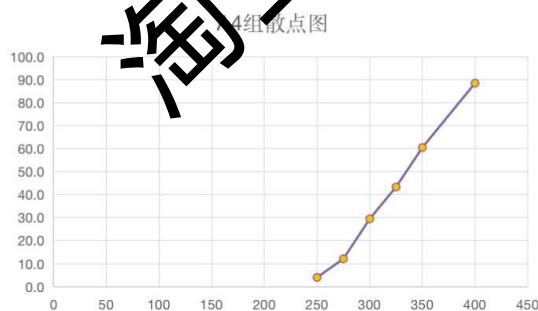


图 4：A4 组的温度—乙醇转化率

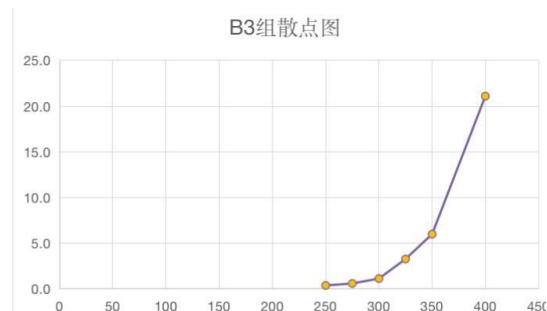


图 5：B3 组的温度—C4 烯烃转化率

由上述两图所示，温度与乙醇转化率均呈正相关趋势，温度逐渐升高时，乙醇转化率逐渐增大，具体趋势变化相同。我们可以采用幂函数和指数函数分别进行拟合。

同理，我们选择 A3 和 B6 两组催化剂组合分析温度—C4 烯烃转化率的散点图，具体图表如下：

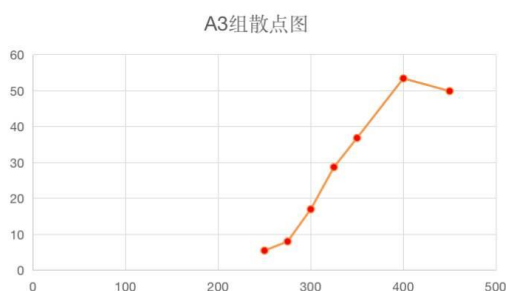


图 6: A3 组的温度—C4 烯烃转化率

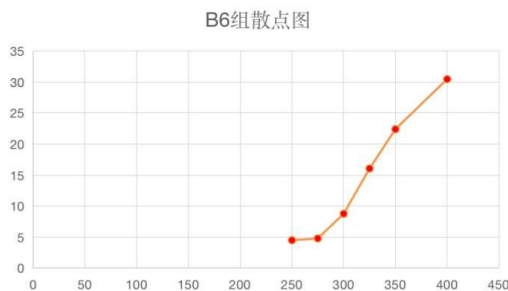


图 7: B6 组的温度—C4 烯烃转化率

观察上述两图我们发现其曲线变化趋势略有不同, A3 催化剂组合当温度逐渐升高时, C4 烯烃转化率先逐渐升高, 达到峰值后再下降, 而 B6 催化剂组合在温度逐渐升高时, C4 烯烃转化率逐渐升高。考虑到存在转折点, 我们考虑使用三次函数进行拟合。

● 基于 SPSS 的非线性回归曲线拟合

我们使用 SPSS 软件, 选取 4 组催化剂组合分别使用三次函数和指数函数对温度—乙醇转化率散点图进行拟合, 对比拟合精度。其次, 由于温度—C4 烯烃转化率存在拐点, 我们使用三次函数对每组催化剂组合分别进行拟合。

① 温度—乙醇曲线拟合

根据散点图我们分析使用三次函数和指数函数进行拟合, 由于篇幅限制, 以 A1 催化剂组合为例, 利用 SPSS 软件, 指数函数拟合与三次函数拟合结果对比图如下:

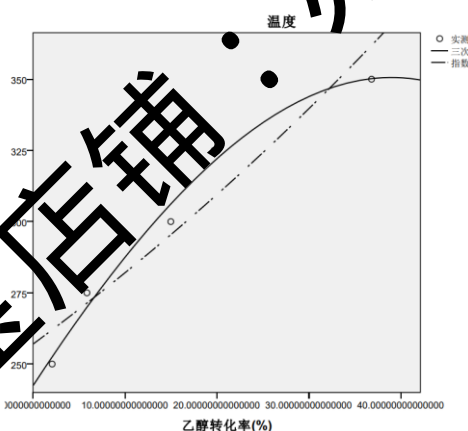


图 8: 拟合曲线对比

对于催化剂组合 A1 而言, 根据拟合图示可以发现, 三次函数拟合效果比指数函数拟合效果好, 为了更严谨, 更精确的分析, 我们对比拟合数据 R^2 , 选取部分表格数据如下:

表 3: 温度—乙醇转化率拟合数据

编号	R^2		F	
	三次函数	指数函数	三次函数	指数函数
A1	0.981	0.957	52.359	67.522
A2	0.967	0.898	58.618	44.014
A6	0.967	0.898	58.618	44.014
B6	0.967	0.898	58.618	44.014

根据表格数据可以发现，催化剂组合 A1, A2, A6, A8 的三次函数的 R^2 值均大于指数函数的 R^2 ，因此我们认为对于各催化剂组合温度—乙醇转化率三次函数的拟合精度更高，因此我们选取三次函数拟合。由于篇幅限制，我们选取催化剂组合 A1 和 B6 的拟合方程为

$$A1: y = -8.83 \times 10^{-7} x^3 + 0.55x - 126.06$$

$$B6: y = 4.72 \times 10^{-6} x^3 - 0.002x^2 + 39.87$$

②温度—C4 烯烃转化率

分析散点图可以发现，存在催化剂组合有拐点，于是我们利用 SPSS 使用三次函数对其进行拟合，选取 A3 和 B6 的温度—C4 烯烃转化率拟合曲线如下：

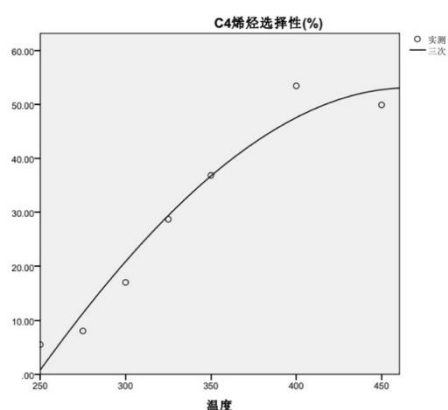


图 9：A3 组的温度—C4 烯烃转化率

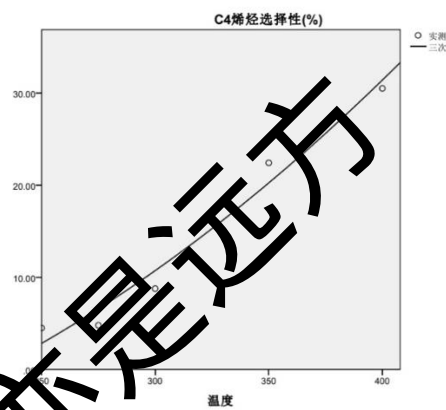


图 10：B6 组的温度—C4 烯烃转化率

有图像可知，三次函数拟合结果较好，我们分析其 R^2 数据得到更精确的结论，A3 组和 B6 组对应的拟合结果数据如下表所示：

表 4：拟合检验

编号	R^2
A3	0.967
B6	0.996

根据上述数据可知， R^2 的值均大于 0.95，因此可以认为用三次函数的拟合结果较为精确，由此得出 A3 组和 B6 组对应的拟合方程如下：

$$A3: y = 3.43 \times 10^{-6} x^3 + 0.71x + 147.13$$

$$B6: y = 3.26 \times 10^{-4} x^2 - 0.022x - 12.13$$

5.2.2 子问题二：基于聚类分析的曲线拟合

在本问题中，由于化学指标数量较多，因此我们使用聚类分析将乙醇等化学指标进行降维聚类。其次，根据分类分析其结果。

● 化学指标的聚类分析

附件 2 给出的数据有乙醇的选择性、乙烯的选择性等 7 个化学指标，由于指标数量较多分析各关系比较困难，因此我们使用聚类分析对化学指标进行降维处理，将其划分为不同类别，在此基础上进行下一步分析。

我们使用 SPSS 软件对其进行聚类分析，聚类结果及聚类结果谱系图如下所示：

表 5：聚类结果

反应物次要成分分类	反应物主要成分分类
乙烯选择性	乙醇转化率
甲基苯甲醇和甲基苯甲酸	碳数为 4-12 脂肪酸
乙醛选择性	C4 烯烃选择性
其他	

我们将反应物分为反应物主成分分类和反应物次要成分分类 2 种类别，反应物主成分分类包括乙醇、碳数为 4-12 脂肪酸以及 C4 烯烃选择性。反应物次要成分包括乙烯选择性、甲基苯甲醇和甲基苯甲酸、乙醛选择性和其他。结合实际实验结果，由于本实验为探究乙醇偶合制备 C4 烯烃，因此乙醇和 C4 烯烃为主要研究对象，因此化作主要成分分类是十分合理的。聚类结果谱系图如下：

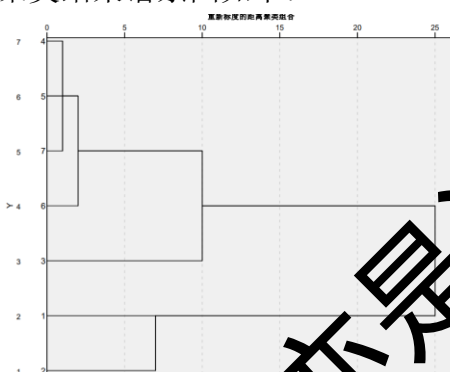


图 11：聚类结果谱系图

聚类结果谱系图与聚类结果一致。通过聚类分析，将化学指标分为主要成分分类和次要成分分类两类，实现了指标的降维，在此基础上，我们进行以下分析。

● 基于最小二乘法的曲线拟合

聚成一类的物质往往具有相同的性质。根据聚类分析的结果，我们分别从两类中各选取一个化学指标作为代表。我们选取乙醇转化率作为反应物主成分分类的代表，选取乙醇选择性作为反应物次要成分分类的代表，分别分析这两个指标与时间的关系，进而使用此关系作为时间与主成分分类和次要成分分类的关系。

首先，我们画出时间—乙醇转化率和时间—乙醇选择性的散点图，根据最小二乘法进行曲线拟合。

①时间—乙醇转化率

基于时间—乙醇转化率的散点图分析，我们发现其随时间的增加，转化率逐渐降低最终趋于一稳定值。结合实际实验情况分析，在乙醇偶合指标 C4 烯烃实验开始初期，乙醇量占比较高，随着不断反应，乙醇逐渐转化为 C4 烯烃，在这一时期，乙醇的转化率逐渐降低，由于乙醇的量是有限的，因此，乙醇不能无限向 C4 烯烃转化，当所有的乙醇全部转化完时，实验停止，此时乙醇转化率不再改变。此趋势与指数函数趋势相近，因此我们选择指数函数进行拟合。

我们假设拟合函数为 $f(x)$ ，其中 $f(x) = ae^{-bx} + c$ ，为了拟合曲线得到各系数的值，我们使用最小二乘法建立以下目标：使拟合曲线误差的平方和最小，此时可认为拟合出曲线满足散点的变化规律且误差最小，即

$$\min \varepsilon = \sum_{i=1}^n [f(x) - y_i^2]$$

其中， ε 为误差平方和。

②时间—乙醛选择性

我们分析乙醛选择性的散点图，随着时间的增加，乙醛的选择性不断增加，因此我们选用多项式进行拟合。

我们假设拟合函数为 $g(x)$ ，其中 $g(x) = a_3x^3 + a_2x^2 + a_1x + a_0$ ，为了拟合曲线得到各系数的值，我们使用最小二乘法建立以下目标：使拟合曲线误差的平方和最小，此时可认为拟合出曲线满足散点的变化规律且误差最小，即

$$\min \varepsilon' = \sum_{i=1}^n [g(x) - y_i^2]$$

其中， ε' 为误差平方和。

结合极值条件得出最小误差系数：

$$\frac{\partial \varepsilon}{\partial a_j} = 2 \sum_{i=0}^n \left(\sum_{k=0}^m a_k x_i^k - y_i \right) x_i^j = 0 \quad j = 0, 1, 2, 3$$

通过求解偏微分方程的解，得到各项系数。

● 最小二乘法拟合的求解

我们使用循环遍历算法进行求解，具体步骤如下：

➤ 温度—乙醇的转化率

Step 1: 对于系数 a 区间 $[20, 40]$ 进行步长为 0.01 的第一次搜索

Step 2: 对于系数 b 在区间 $[0, 30]$ 进行步长为 0.01 的第二次搜索

Step 3: 对于系数 c 区间 $[0, 1]$ 进行步长为 0.01 的第三次搜索

Step 4: 求出拟合方程为

$$f(x) = 18.15e^{-0.01x} + 29.04$$

➤ 温度—乙醛的选择性

Step 1: 对于系数 a_3 在区间 $[-7 \times 10^{-8}, -9 \times 10^{-8}]$ 进行步长为 10^{-10} 的第一次搜索

Step 2: 对于系数 a_2 在区间 $[2 \times 10^{-5}, 4 \times 10^{-5}]$ 进行步长为 10^{-7} 的第二次搜索

Step 3: 对于系数 a_1 在区间 $[-2 \times 10^{-4}, -4 \times 10^{-4}]$ 进行步长为 10^{-6} 的第三次搜索

Step 4: 对于常数 a_0 在区间 $[0, 10]$ 进行步长为 0.01 的第四次搜索

Step 5: 求出拟合方程为

$$g(x) = -7.95 \times 10^{-8} x^3 + 2.91 \times 10^{-5} x^2 - 2.76 \times 10^{-4} x + 4.22$$

5.2.3 模型一的检验

我们首先对最小二乘法的算法精度进行检验，其次，根据拟合做催化剂组合 A1，A6，A11，B2 和 B6 的拟合误差分析。

①算法精度检验

对于定步长的循环遍历算法，误差的来源主要为变量的步长，因此，我们可以通过减小步长，根据最优解变化幅度来判断步长是否合理。

取变量 h 的步长为原步长的 $\frac{1}{10}$ ，则算法精度提高 10 倍，定义相对优化量 s 为目标函数优化量与理论优化量的比值：

$$s = \frac{|f'(h) - f(h)|}{10}$$

通过 MATLAB 编程计算可得，由于长度范围在数量级 10^{-4} ，因此可以忽略不计，因此目前的循环遍历算法的步长是十分合理的。

②拟合曲线的误差分析

我们使用 MATLAB 软件对催化剂组合 A1, A6, A11, B2 以及 B6 做温度关于乙醇转化率及 C4 烯烃选择性的误差分析，由于篇幅限制，我们选择部分数据进行展示，其误差分析结果如下表所示：

表 6：误差分析

组合编号	温度	乙醇转化率			C4 烯烃选择性		
		真实值	模拟值	相对误差	真实值	模拟值	相对误差
A1	250	2.07	2.37	0.15	34.05	32.70	-0.039
	275	5.85	6.10	0.04	37.43	40.46	0.081
	300	14.97	12.63	-0.16	46.94	45.76	-0.025
	325	19.68	22.41	0.14	49.7	48.36	-0.027
	350	36.80	35.86	-0.03	47.21	48.05	0.018
B2	250	2.8	3.19	0.39	3.26	2.56	-0.215
	275	4.4	4.54	-0.21	4.97	5.73	0.152
	300	6.2	5.20	0.17	9.32	9.94	0.066
	325	16.2	17.68	0.09	22.88	21.85	-0.044
	350	45.1	44.65	-0.01	38.7	39.05	0.009

从数据表中可以看出，乙醇转化率和 C4 烯烃转化率的误相对差均较小，相较于乙醇转化率来说，C4 烯烃的选择性相对误差更小，因此我们可以认为曲线的拟合度较高，且温度—C4 烯烃选择性曲线拟合度更高。

5.3 问题二：基于多重共线性的多元回归模型

为了探究不同催化剂组合及温度对乙醇转化率及 C4 烯烃选择性大小的影响，首先我们需要分析不同催化剂组合的成分区别，因此我们选取指标衡量催化剂组合，同时计算相关性系数矩阵判断各指标间的相关性，为了建立多元线性回归方程，我们需要判断其共线性。为了消除指标间共线性的影响，首先建立基于主成分分析的多元回归模型，其次建立逐步多元回归模型，并对其结果进行对比，最后提出增加交互项多元回归的模型优化。



图 12：问题二流程图

5.3.1 相关性分析

为了研究这 7 个化学指标间的关系，我们需要对其进行相关性分析。假设此 7 个化学指标构成的指标向量为

$$X = (x_1, x_2, \dots, x_7)$$

其中， $x_1 \sim x_7$ 分别代表 Co 负载量、 HAP 含量、 Co/SiO_2 和 HAP 装料比、乙醇浓度、温度、乙醇转化率和 C_4 烯烃选择性。

我们构建各个变量的相关性系数矩阵：

$$R = XX^T = \begin{bmatrix} R_{11} & \dots & R_{71} \\ \vdots & \ddots & \vdots \\ R_{71} & \dots & R_{77} \end{bmatrix}$$

其中， $R_{ij} (i=1, 2, \dots, 7, j=1, 2, \dots, 7)$ 代表第 i 个指标和第 j 个指标间的相关性

5.3.2 相关性矩阵的求解

利用 SPSS 软件对 7 个指标进行相关性分析，得到相关性矩阵的具体数据如下：

表 1：相关性系数矩阵

	x_1	x_2	x_3	x_4	x_5	x_6	x_7
x_1	1	0.983	0.209	-0.299	-0.001	0.395	0.38
x_2	0.983	1	0.202	-0.293	-0.002	0.39	0.347
x_3	0.209	0.202	1	0.141	-0.014	0.042	-0.165
x_4	-0.299	-0.293	0.141	1	-0.038	-0.331	0.107
x_5	-0.001	-0.002	-0.014	-0.038	1	0.771	0.7
x_6	0.395	0.39	0.042	-0.331	0.771	1	0.735
x_7	0.38	0.347	-0.165	0.107	0.7	0.735	1

由表格可知，部分指标变量间，如 Co 负载量、 HAP 含量，存在较强的相关性，相关性系数达到 0.983，温度与乙醇转化率的相关性系数同样较高。因此可以使用多元线性回归模型对其进行回归拟合，但是，由于 Co 负载量、 HAP 含量间相关性系数较高，即其二者之间存在较强的相关性，若对此五个变量关于乙醇转化率和 C_4 烯烃选择性建立多元线性回归模型，会有较强的共线性，变量不显著，拟合结果较差，因此我们将此 5 个指标进行一定的转化，消除共线性，并在此基础上进行线性回归。

5.3.3 基于主成分分析的多元回归

由相关性结果可知，*Co* 负载量、*HAP* 含量间存在共线性，效果不显著。为消除变量间的共线性，我们通过主成分分析法，将自变量的 5 个指标划分为 2 个主成分，在此基础上对此 2 个主成分关于乙醇转化率和 C4 烯烃选择性建立多元线性回归模型。

● 主成分分析

主成分分析（Principal Component Analysis, PCA）是通过正交变换将一组可能存在相关性的变量转换为一组线性不相关的变量，转换后的这组变量叫主成分^[1]。求主成分的步骤^[2]如下：

- (1) 建立变量的相关系数矩阵，我们在上文中根据相关性系数分析已建立相关性系数矩阵

$$R = XX^T = \begin{bmatrix} R_{11} & \cdots & R_{71} \\ \vdots & \ddots & \vdots \\ R_{71} & \cdots & R_{77} \end{bmatrix}$$

其中， $R_{ij}(i=1,2,\cdots,7\ j=1,2,\cdots,7)$ 代表第*i*个指标和第*j*个指标间的相关性。

- (2) 我们假设有 *p* 个主成分，求 *R* 的特征根及相应的单位特征向量

$$a_1 = \begin{bmatrix} a_{11} \\ \vdots \\ a_{p1} \end{bmatrix} \cdots a_p = \begin{bmatrix} a_{1p} \\ \vdots \\ a_{pp} \end{bmatrix}$$

特征值为 $\lambda_i = a_{1i}X_1 + a_{2i}X_2 + \cdots + a_{pi}X_p\ i=1,2,\cdots,p$

- (3) 计算主成分贡献率与累计贡献率
贡献率：

$$\frac{\lambda_i}{\sum_{k=1}^p \lambda_k} \quad i=1,2,\cdots,p$$

累计贡献率：

$$\frac{\sum_{k=1}^i \lambda_k}{\sum_{k=1}^p \lambda_k} \quad i=1,2,\cdots,p$$

其中，一般选取累积贡献率大 85%~95%的特征值 $\lambda_1,\cdots,\lambda_m$ 所对应的第 1、第 2 至第*m*($m \leq p$)个主成分。

● 主成分分析的结果

我们使用 SPSS 软件对催化剂组合 5 个变量进行主成分分析，得到各个变量的特征根及累计贡献率如下表：

表 2：主成分结果

成分	特征根	累积贡献率
----	-----	-------

1	2.181	43.611%
2	1.145	66.518%
3	0.992	86.366%
4	0.664	99.653%
5	0.017	100%

一般选择贡献率大于 85%的主成分个数，但结合各个主成分间的多重共线性关系，取 3 个主成分时，主成分 1, 2, 3 存在多重共线性，因此我们选取 2 个主成分，主成分 1, 2 之间不存在多重共线性。

假设主成分 1 为 t_1 ，主成分 2 为 t_2 ，利用 SPSS 得到二者的主成分表达式为

$$t_1 = 0.447x_1 + 0.446x_2 + 0.13x_3 - 0.208x_4 + 0.004x_5$$

$$t_2 = 0.037x_1 + 0.036x_2 + 0.692x_3 - 0.584x_4 - 0.223x_5$$

其中， $x_1 \sim x_5$ 分别代表 Co 负载量、 HAP 含量、 Co/SiO_2 和 HAP 装料比、乙醇浓度、温度。

● 基本多元线性回归

多元线性回归方程的建立要求各指标间无共线性，而通过主成分分析模型确立的 2 个主成分之间无共线性，因此我们对自变量：主成分 1，主成分 2，因变量：乙醇转化率和 C4 烯烃选择性进行多元线性回归。

(1) 多元线性回归模型

涉及个自变量的多元线性回归模型可表示为

$$\begin{cases} y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_p x_p + \varepsilon \\ \varepsilon \sim N(0, \sigma^2) \end{cases}$$

其中， ε 为随机变量， $x_1 \sim x_5$ 分别代表 Co 负载量、 HAP 含量、 Co/SiO_2 和 HAP 装料比、乙醇浓度、温度，在此模型中取 5， y 代表乙醇转化率和 C4 烯烃转化率。

(2) 回归方程的显著性检验

当检验被解释变量 y 与一组解释变量 x_1, x_2, \dots, x_7 是否存在回归关系时，给出的零假设与备择假设分别是

$$H_0: \beta_1 = \beta_2 = \cdots = \beta_7 = 0$$

$$H_1: \beta_i \text{ 不全为 } 0, i = 1, 2, \dots, 7$$

首先构造 F 统计量，有

$$\sum (y - \bar{y})^2 = \sum (\hat{y} - \bar{y})^2 + \sum (y - \hat{y})^2$$

对总平方和 SST 进行分解，分解为回归平方和 SSR 和残差平方和 SSE 。本文采用 F 检验：检验回归平方和是否大于残差平方和，若成立，则说明该方程具有统计意义。

SST 具有 m 个自由度， SSR 具有 $n - m - 1$ 个自由度，其中， m, n 分别为自变量数量，样本数量。

在 H_0 成立的条件下有

$$F \sim F(m, n - m - 1)$$

设检验水平为 α ，则检验规则为

①若用样本计算的 $F \leq F_\alpha(m, n - m - 1)$ ，则接受 H_0

②若用样本计算的 $F > F_{\alpha}(m, n - m - 1)$, 则拒绝 H_0
拒绝 H_0 意味着一定存在解释变量与 y 存在回归关系。

● 基于主成分分析的线性回归模型求解

①我们使用统计软件 SPSS 得到 2 元线性回归方程为

$$y_1 = 0.27t_1 - 1.507t_2 - 94.485$$

其中, y_1 代表乙醇转化率, t_1 为主成分 1, t_2 为主成分 2。

$$y_2 = 0.148t_1 - 0.821t_2 - 47.042$$

其中, y_2 代表 C4 烯烃选择性。

②为了检验回归方程结果正确性, 我们利用统计软件 SPSS 对求解得到的两个 2 元线性回归方程分别进行显著性检验, 结果如下:

A) 判定系数

表 3: 判定系数

因变量	R	R^2	调整 R^2	标准的估计差
乙醇转化率	0.870	0.758	0.753	11.34
C4 烯烃选择性	0.801	0.642	0.635	8.173

通过上表可以得到, 多重判定系数 $R^2 > 0.74$, 调整后的 $R^2 > 0.63$, 回归方程的拟合精度较高。

B) 回归的显著性检验

表 4: 显著性检验

因变量		平方和	df	均方	F	Sig
乙醇的转化率	回归	4638.445	2	22313.9	1173.521	<0.001
	残差	12517.43	111	128.59		
	统计	58901.90	113			
C4 烯烃的选择性	回归	14330.95	2	6638.61	99.371	<0.001
	残差	6360.065	111	66.801		
	统计	20961.01	113			

由于 F 均较大, 所以回归方程较为显著, 即自变量与因变量之间存在较为明显的函数关系。

自变量为 2 个主成分时, 因变量为乙醇转化率均接受原假设, 用基于主成分分析的多元回归模型建立的自变量指标与因变量指标关系密切, 因此, 得到的 2 元线性方程是可行的。

5.3.4 多元逐步线性回归

为了消除共线性的影响, 我们采取逐步回归模型, 对自变量: Co 负载量、 HAP 含量、 Co/SiO_2 和 HAP 装料比、乙醇浓度、温度, 因变量乙醇转化率和 C4 烯烃选择性分别进行逐步线性回归。

我们通过逐步增加变量进行逐步线性回归, 具体步骤如下:

- Step 1: 建立多元线性回归模型
- Step 2: 进行回归系数显著性检验, 取最大概率 P_{\max}
- Step 3: 判断 P_{\max} 是否 ≤ 0.05 , 若满足则进入 Step 5, 若不满足, 则进入 Step 4
- Step 4: 则可接受 H_0 , 即这个指标与因变量线性关系不显著, 将指标剔除, 返回 Step 1
- Step 5: 则可拒绝 H_0 , 各指标与因变量线性关系显著, 输出方程, 结束。

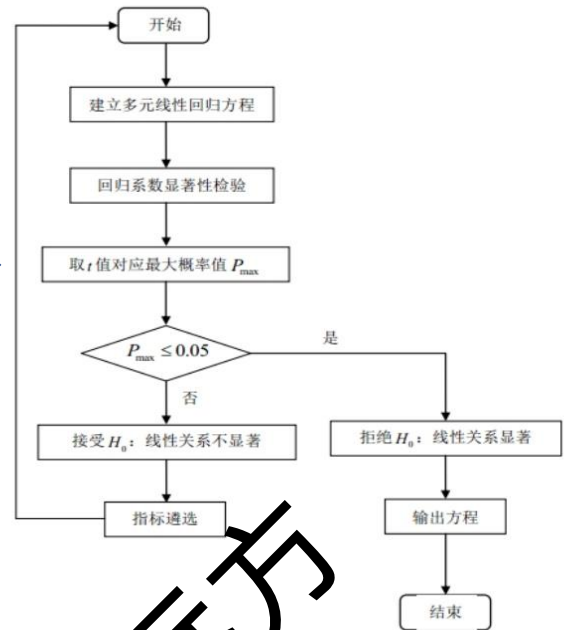


图 3-2 二分法

根据上述步骤, 首先对自变量: Co 负载量、 HAP 含量、 Co/SiO_2 和 HAP 装料比、乙醇浓度、温度, 因变量: 乙醇转化率, 不断添加变量, 逐步回归, 得到 5 个指标与乙醇转化率的关系, 其次对自变量: 负载量、含量、和装料比、乙醇浓度、温度, 因变量: C4 烯烃的选择性, 进行逐步回归, 最终得到这 5 个指标与 C4 烯烃的选择性的关系。

5.3.5 多元逐步回归的求解

我们对乙醇转化率和 C4 烯烃选择性的结果分别进行

● 乙醇转化率的结果

首先, 利用 SPSS 软件对自变量: Co 负载量、 HAP 含量、 Co/SiO_2 和 HAP 装料比、乙醇浓度、温度, 因变量: 乙醇转化率进行逐步回归, 最终经过筛选变量得出的结果如下表:

表 5: 逐步回归结果

总变量	Co 负载量、 HAP 含量、 Co/SiO_2 和 HAP 装料比、乙醇浓度、温度
保留变量	Co 负载量、乙醇浓度和温度
舍弃变量	HAP 含量、 Co/SiO_2 和 HAP 装料比

最终筛选和剔除后, 留下 Co 负载量、乙醇浓度和温度 3 个自变量, 逐步回归得到 5 元线性回归方程为

$$y_1 = 0.109x_1 - 8.833x_4 + 0.333x_5 - 80.637$$

其中, y_1 代表乙醇转化率, x_1 代表 Co 负载量, x_4 代表乙醇浓度, x_5 代表温度。

● C4 烯烃选择性的结果

利用 SPSS 软件, 逐步回归结果如下表所示:

表 6: 逐步回归结果

总变量	Co 负载量、 HAP 含量、 Co/SiO_2 和 HAP 装料比、乙醇浓度、温度
保留变量	Co 负载量、 Co/SiO_2 和 HAP 装料比、温度
舍弃变量	HAP 含量、乙醇浓度

最终筛选和剔除后，留下 Co 负载量、 Co/SiO_2 和 HAP 装料比、温度 3 个自变量，对比表 7 和表 8 可以发现， HAP 含量均舍弃， Co 负载量均保留，因相关性检验时，发现其二者的相关性较高，显著性差异小，存在共线性，因此二者不可同时存在，与逐步回归结果一致，十分合理。

通过逐步回归得到 5 元线性回归方程为

$$y_2 = 0.083x_1 - 2.86x_3 + 0.180x_5 - 44.512$$

其中， y_2 代表 C4 烯烃选择性， x_1 代表 Co 负载量， x_3 代表 Co/SiO_2 和 HAP 装料比， x_5 代表温度。

5.3.6 多元逐步回归显著性检验

为了检验回归方程结果正确性，我们利用统计软件 SPSS 求解得到的两个 5 元线性回归方程分别进行显著性检验，结果如下：

B) 判定系数

表 7：判定系数

因变量	R	R^2	调整 R^2	标准的估计差
乙醇转化率	0.877	0.787	0.782	10.667
C4 烯烃选择性	0.832	0.693	0.684	7.603

通过上表可以得到，多重判定系数 $R^2 > 0.69$ ，调整后的 $R^2 > 0.68$ ，回归方程的拟合精度较高。

B) 回归的显著性检验

表 8：显著性检验

因变量		平方和	df	均方	F	Sig
乙醇的转化率	回归	46384.47	3	15461.49	135.872	<0.001
	残差	12517.43	110	113.795		
	统计	58901.90	113			
C4 烯烃的选择性	回归	14330.95	3	4776.982	82.62	<0.001
	残差	6360.065	110	57.819		
	统计	20961.01	113			

由于 F 均较大，所以回归方程较为显著，即自变量与因变量之间存在较为明显的函数关系。

自变量为 Co 负载量、乙醇浓度和温度，因变量为乙醇转化率及自变量为 Co 负载量、 Co/SiO_2 和 HAP 装料比、温度，因变量为 C4 烯烃选择性时均接受原假设，用逐步分析法确立的自变量指标与因变量指标关系密切，因此，得到的 5 元线性方程是可行的。

5.3.7 模型的优化——增加交互项的逐步回归模型

在基于逐步线性回归模型的基础上，我们增加交互项对多元逐步线性回归模型拟合精度进行提高。在多元逐步线性回归已满足不存在多重共线性的基础上，其模型精度较低，其原因可能为我们仅考虑单一一次变量与因变量间关系，自变量数量较少，次数较低，不能很好地找到其与因变量间关系，故我们引入自变量间的二次交互项，增加自变量个数，提高自变量次数，此时在对所有一次、二次变量进行多元逐步回归。

增加二次交互项的逐步回归方程为

$$\begin{cases} y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_p x_p + \sum_{1 \leq i \leq j \leq p} x_i x_j + \varepsilon \\ \varepsilon \sim N(0, \sigma^2) \end{cases}$$

其中， ε 为随机变量，在此模型中 p 取 5， y 代表乙醇转化率和 C4 烯烃转化率。

接下来，重复模型 5.2.5 的逐步回归的步骤，依次对 5 个指标及其二次交互项与乙醇转化率和 C4 烯烃转化率进行多元逐步回归。

5.3.8 增加交互项的逐步回归模型的求解

使用统计软件 SPSS 进行加入二次交互项的逐步线性回归，分别分析因变量为乙醇转化率、C4 烯烃选择性的结果，其筛选后去掉共线性变量的结果如下：

● 乙醇转化率的结果

首先，利用 SPSS 软件对自变量： Co 负载量、 HAP 含量、 Co/SiO_2 和 HAP 装料比、乙醇浓度、温度及 15 个二次交互项，因变量：乙醇转化率及 C4 烯烃选择性，进行逐步回归，最终经过筛选变量得出的结果如下表：

表 9：逐步回归结果

总变量	因变量	Co 负载量、 HAP 含量、 Co/SiO_2 和 HAP 装料比、乙醇浓度、温度 $\sum_{1 \leq i \leq j \leq 5} x_i x_j$ (15 个二次交互项)
保留变量	乙醇转化率	Co 负载量*乙醇浓度、乙醇浓度*温度和温度二次项
	C4 烯烃转化率	Co 负载量*乙醇浓度、 Co 负载量*温度和温度二次项 Co/SiO_2 和 HAP 装料比*温度

进而得到如下回归方程：

$$y_1 = 0.001x_5 - 0.056x_4x_5 + 0.082x_1x_4 - 28.606$$

其中， y_1 代表乙醇转化率， x_1 代表 Co 负载量， x_4 代表乙醇浓度， x_5 代表温度。

$$y_2 = 0.001x_5^2 + 0.001x_1x_5 - 0.01x_3x_5 + 0.033x_1x_4 - 16.645$$

其中， y_2 代表乙醇转化率， x_1 代表 Co 负载量， x_3 代表 Co/SiO_2 和 HAP 装料比， x_4 代表乙醇浓度， x_5 代表温度。

5.3.9 增加二次交互项的回归模型的检验

我们分别对上述回归方程进行精度检验以及显著性检验。

A) 精度检验

我们使用 SPSS 进行逐步回归分析，在因变量为乙醇转化率时逐步回归的共有 5 步，因变量为 C4 烯烃的选择性时，逐步回归分析步骤为 4 步，其每步对应的拟合精度如下所示：

乙醇转化率					
	模型 1	模型 2	模型 3	模型 4	模型 5
R^2	0.61	0.789	0.826	0.844	0.844
调整后的 R^2	0.606	0.785	0.821	0.838	0.839
C4 烯烃选择性					
	模型 1	模型 2	模型 3	模型 4	
R^2	0.495	0.647	0.709	0.724	
调整后的 R^2	0.491	0.641	0.701	0.725	

所以，增加二次交互项后的拟合精度 $R_1^2 > 0.8$ ， $R_2^2 > 0.7$ ，拟合精度较高，因此，可以用增加二次项的回归方程。

B) 显著性检验

表 10: 显著性结果

因变量	F	Sig
乙醇的转化率	197.777	<0.001
C4 烯烃的选择性	75.337	<0.001

由于 F 均较大，所以回归方程较为显著，即自变量与因变量之间存在较为明显的函数关系。

5.3.10 模型的对比

● 基于主成分分析的回归模型与逐步回归模型的对比

我们从精度以及显著性两个角度进行对比，对比数据如下：

表 11: 对比结果

模型	R^2		调整后 R^2		F	
	主成分分析	逐步回归	主成分分析	逐步回归	主成分分析	逐步回归
乙醇的转化率	0.785	0.787	0.753	0.783	173.521	135.872
C4 烯烃的选择性	0.642	0.693	0.635	0.684	99.371	82.62

根据对比结果可知，对于拟合精度来说，多元逐步线性回归分析的拟合精度略好于基于主成分分析的回归模型，两组差别不太大；对于显著性而言，基于主成分分析的回归模型显著性高于逐步回归模型，但二者均具有较高显著性。

● 增加交互项模型前后对比

我们从精度以及显著性两个角度进行对比，对比数据如下：

表 12: 对比结果

	R^2		调整后 R^2		F	
增加交互项前后	前	后	前	后	前	后
乙醇的转化率	0.787	0.844	0.783	0.844	135.872	197.777
C4 烯烃的选择性	0.693	0.734	0.684	0.725	82.62	75.337

根据对比结果可知, 增加交互项后, 针对精度而言, 乙醇转化率的 R^2 从 0.787 提升到 0.844, 调整后的 R^2 从 0.783 提升到 0.844, C4 烯烃选择性具有相同规律, 拟合精度提升幅度较大, 因此, 增加交互项可以使拟合效果更好。对于显著性, 乙醇转化率对应的显著性提高, 但 C4 烯烃的选择性对应的转化率稍稍降低, 但仍接受原假设, 因此增加交互项后仍有较高的显著性。

故增加二次交互项后的模型得到优化, 更加符合问题二, 与实际结果相近。

5.4 问题三: 催化剂组合与温度的优化模型

问题三由 2 个递进的子问题构成, 其二者均为优化问题。首先我们建立 C4 烯烃收率与催化剂组合和温度二者的函数表达式, 其次, 建立 C4 烯烃收率最大的单目标优化模型。第二小问则在此模型的基础上, 修改约束条件, 建立模型。

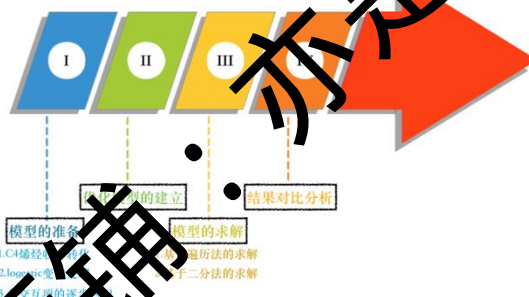


图 14: 问题三流程图

5.4.1 数据预处理

由题目所述, C4 烯烃的收率 w 计算公式为

$$w = y_1 y_2$$

其中, y_1 为乙醇转化率, y_2 为 C4 烯烃的选择性。

在实验过程中, 题目仅展示部分反应时间, 若考虑整个实验反应时间, 根据实验事实可知, 随着反应的不断进行, 反应物随时间的增加不断减少, 当催化剂消耗完毕时, 反应物含量不再改变, 趋于一稳定值, 而生成物会随着反应的进行不断增加, 由于催化剂含量有限度, 因此, 生成物含量也同样趋于一稳定值, 结合附件中数据可以大致发现, C4 烯烃收率满足 Logistic 曲线, 为了考虑实验的完整性, 同时便于分析, 我们使用 Logistic 变换, 将 C4 烯烃收率数据线性化。Logistic 变换为

令

$$Z = \ln\left(\frac{w}{5000 - w}\right)$$

其中, w 为 C4 烯烃的收率, Z 为变换后的线性 C4 烯烃收率数值。

5.4.2 增加交互项的多元逐步回归—建立 C4 烯烃收率函数表达式

为了探究催化剂组合与温度对 C4 烯烃收率的关系，我们沿用问题二中拟合程度最好的增加交互项的多元逐步回归模型，对自变量为催化剂组合（），因变量为 C4 烯烃收率，进行逐步回归，其回归方程为

$$\begin{cases} z = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_p x_p + \sum_{1 \leq i \leq j \leq p} x_i x_j + \varepsilon \\ \varepsilon \sim N(0, \sigma^2) \end{cases}$$

其中， ε 为随机变量，在此模型中 p 取 5， z 代表 *logistic* 变换后的 C4 烯烃的收率。

我们通过 SPSS 进行逐步回归，得到 5 元线性回归方程为

$$z = 0.035x_1 - 0.039x_2 + 0.816x_3 + 0.018x_5 - 0.29x_3^2 + 0.007x_2x_4 - 0.002x_4x_5 - 0.815$$

其中， $x_1 \sim x_5$ 分别代表 Co 负载量、 HAP 含量、 Co/SiO_2 和 HAP 表料比、乙醇浓度、温度。

代入 C4 烯烃收率 w 有

$$\ln(w/(5000 - w)) = 0.035x_1 - 0.039x_2 + 0.816x_3 + 0.018x_5 - 0.29x_3^2 + 0.007x_2x_4 - 0.002x_4x_5 - 0.815$$

经验证，该回归方程精度 R^2 达到 0.812，显著性 F 达到 65.543，我们认为回归方程的精度较高且具有显著性。因此，我们得到 C4 烯烃收率关于催化剂组合和温度的函数表达式。

5.4.3 基于 C4 烯烃收率最高的单目标优化模型

● 目标函数的确定

由于题目所述，在相同实验条件下，找到催化剂组合和温度的最优方案，使得 C4 烯烃收率最高，因此目标函数为

$$\max w$$

其中， w 为 C4 烯烃的收率。

● 约束条件的确定

为了使模型更加贴合实际，我们以附件 1 中所做的 21 组实验为基础，假设每一个自变量的下限为保证实验正常进行的最低要求，上限认为是该实验室所能提供的最高实验设施的要求，故建立如下约束条件：

约束条件一： Co 负载量约束

附件 1 中所给数据的负载量的下限为 10，上限为 200，因此我们认为实验中负载量应满足

$$10 \leq x_1 \leq 200$$

约束条件二：质量约束

在实验中，质量的下限为 10，上限为 200，我们仅考虑实验所给出的有效范围，因不知道超出范围外的情况，因此质量应满足

$$10 \leq x_2 \leq 200$$

约束条件三：装料比约束

根据实验实际情况，我们可得到装料比的下限为 0.5，上限为 5，同理，建立装料比约束为

$$0.5 \leq x_3 \leq 5$$

约束条件四：乙醇浓度约束

实验中乙醇浓度最低为 0.3，最高为 100，因此，乙醇浓度约束为

$$0.3 \leq x_4 \leq 2.1$$

约束条件五：实验温度约束

在附近 1 的 21 组实验中，实验温度为 250 度、275 度、350 度、375 度、425 度及 450 度，因此我们考虑温度的限制为

$$250 \leq x_5 \leq 450$$

由于 C4 烯烃收率与催化剂组合及温度存在函数关系式，且多元回归方程结果为优化模型的基础，因此我们建立基于 C4 烯烃收率最高的单目标优化模型：

$$\begin{aligned} & \max w \\ & s.t. \quad \left\{ \begin{array}{l} \ln(w/(w/5000)) = 0.035x_1 - 0.039x_2 + 0.816x_3 + 0.018x_5 \\ \quad - 0.19x_1^2 + 0.007x_2x_4 - 0.002x_4x_5 - 8.015 \\ 10 \leq x_1 \leq 200 \\ 10 \leq x_2 \leq 200 \\ 0.5 \leq x_3 \leq 5 \\ 0.3 \leq x_4 \leq 2.1 \\ 250 \leq x_5 \leq 450 \end{array} \right. \end{aligned}$$

5.4.4 单目标优化模型的求解

为了求解单目标优化模型，我们通过遍历法和二分法分别进行求解并对比分析二者结果。

● 遍历法

step 1: x_1 在区间[10,200]以步长为 5, x_2 在区间[10,200]以步长 5, x_3 在区间[0.5,5]以步长 0.1, x_4 在区间[0.3,2.1]以步长 0.1, x_5 在区间[250,450]以步长 5 依次进行遍历，在 MATLAB 中搜索结果如下：

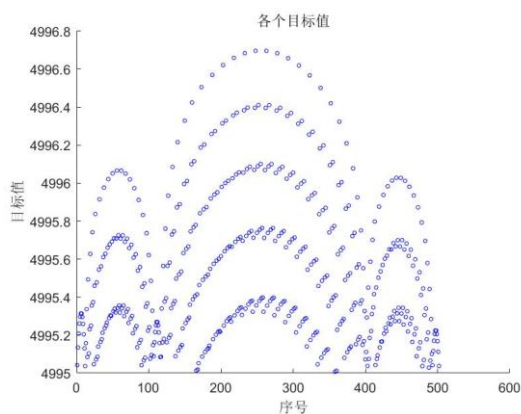


图 15: 遍历结果 1

step 2: x_1 在区间[190,200]以步长 1, x_2 在区间[10,20]以步长 1, x_3 在区间[2,3]以步长 0.01, x_4 在区间[0.3,2.1]以步长 0.01, x_5 在区间[440,450]以步长 1 依次进行遍历, 在 MATLAB 中搜索结果如下:

在每个子区间内均呈现先增大后减小趋势, 此后减小各个变量的搜索区间, 通过第一次遍历得到的最优解为 49.966969959%, 此时得到对应的各变量取值如下:

$$\begin{cases} x_1 = 200 \\ x_2 = 10 \\ x_3 = 2.1 \\ x_4 = 0.3 \\ x_5 = 450 \end{cases}$$

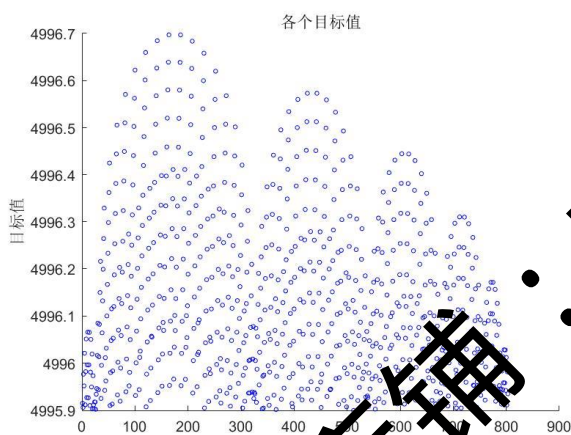


图 16: 遍历结果 2

step 3: x_1 在区间[199,200]以步长 0.1, x_2 在区间[10,11]以步长 0.1, x_3 在区间[2,3]以步长 0.001, x_4 在区间[0.3,0.4]以步长 0.01, x_5 在区间[499,450]以步长 0.1 依次进行遍历, 在 MATLAB 中搜索结果如下:

在每个子区间内均呈现先增大后减小趋势, 相较于第一次遍历结果, 区间变小, 搜索散点图更为密集, 此后减小各个变量的搜索区间, 通过第二次遍历得到的最优解为 49.966983985%, 此时得到对应的各变量取值如下:

$$\begin{cases} x_1 = 200 \\ x_2 = 10 \\ x_3 = 2.15 \\ x_4 = 0.3 \\ x_5 = 450 \end{cases}$$

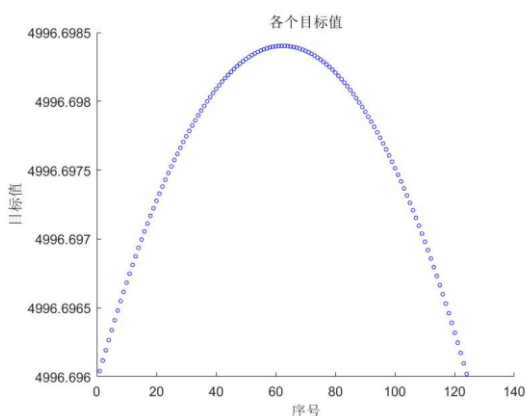


图 17: 遍历结果 3

最终遍历结果呈现抛物线趋势, 先增大后减小, 存在最大值, 此时的最大值为最终 C4 烯烃收率的最大值, 为 49.966983985%, 此时得到对应的各变量取值如下:

$$\begin{cases} x_1 = 200 \\ x_2 = 10 \\ x_3 = 2.147 \\ x_4 = 0.3 \\ x_5 = 450 \end{cases}$$

● 二分法

二分法求给定区间极值的步骤如下：

➤ 计算目标函数在有解区间端点处的值 $f(a), f(b)$

➤ 计算目标函数在区间中点处的值 $f(x_0)$

➤ 判断 $f(x_0)$ ：

若 $f(x_0) = 0$ ，即为 C4 烯烃的最大吸率

若 $f(x_0)f(a) < 0$ ，极值区间缩小为 $[a, x_0]$

若 $f(x_0)f(a) > 0$ ，极值区间缩小为 $[x_0, b]$

➤ 重复上述步骤，得到最大的 C4 烯烃收率

C_0 负载量 x_1	➡	有解区间 $[10, 200]$
HAP 质量 x_2	➡	有解区间 $[10, 200]$
C_0/SiO_2 和 HAP 装料比	➡	有解区间 $[0.5, 5]$
乙醇浓度	➡	有解区间 $[0.3, 2.1]$
温度	➡	有解区间 $[250, 450]$

我们使用 MATLAB 基于上述二分法求得的结果如下：

表 13: 基于二分法最优方案

C4 烯烃最大收率	x_1	x_2	x_3	x_4	x_5
49.9669%	199.9998	10.0002	2.1	0.3	449.9998

利用 MATLAB 做出各求解区间对应的 C4 烯烃收率，如下

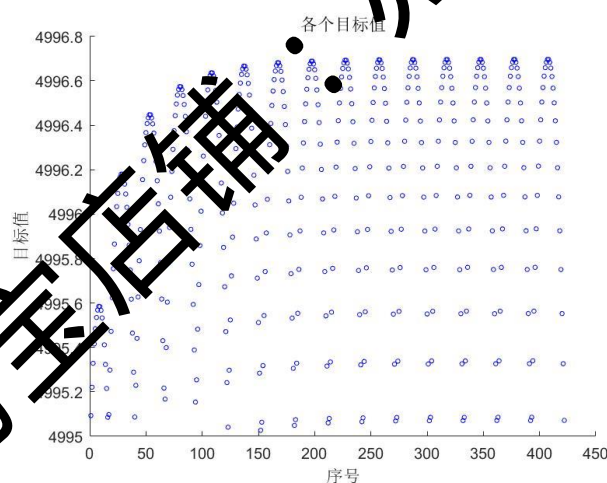


图 14: 二分法结果

由图可知，C4 烯烃收率关于求解区间呈震荡趋势，在每个子区间内为先增加后减少，存在区间极值。从总体趋势来看，区间极值呈现逐步上升并趋于稳定的趋势。最大值在后半区间取到。

5.4.5 加入温度限制的优化模型

问题三在求解第一个子问题的基础上进而增加温度限制，其他约束条件相同。因此，只需修改温度限制约束条件。

将约束条件五修改为

修改约束条件五：温度约束

由于问题考虑在实验温度不超过 350 度的情况下，找到催化剂组合与温度的最优

方案，因此温度限制变为

$$250 \leq x_5 \leq 350$$

由于目标仍是使 C4 烯烃收率达到最大，因此加入温度限制的单目标优化模型为

$$\begin{aligned} & \max w \\ & s.t. \left\{ \begin{array}{l} \ln(w/(w/5000)) = 0.035x_1 - 0.039x_2 + 0.816x_3 + 0.018x_5 \\ \quad - 0.19x_3^2 + 0.007x_2x_4 - 0.002x_4x_5 - 8.015 \\ 10 \leq x_1 \leq 200 \\ 10 \leq x_2 \leq 200 \\ 0.5 \leq x_3 \leq 5 \\ 0.3 \leq x_4 \leq 2.1 \\ 250 \leq x_5 \leq 350 \end{array} \right. \end{aligned}$$

5.4.6 加入温度限制的优化模型的求解

由于仅改变了 5.3.3 模型中的约束条件五，因此，为了沿用模型 5.3.3 的求解，我们仍使用遍历法和二分法对其进行求解，得到结果如下。

表 20：结果数据

方法	最大 C4 烯烃收率	x_1	x_2	x_3	x_4	x_5
遍历法	49.8124%	200	10	2.147	0.3	350
二分法	49.8123%	199.9998	10.0002	2.1	0.3	349.999

根据表格数据对比可知，遍历法结果与二分法基本一致，且取最优值时各变量值差别几乎没有，但考虑到精度，我们认为遍历法结果更接近最大值。

5.5 问题四：基于正交试验的方案确定

由于乙醇偶合制备 C4 烯烃的实验，从催化剂与温度等多因素探究 C4 烯烃收率的变化，因此可利用正交试验对组别进行设计。正交试验设计是研究多因素多水平的的设计方法，它是根据正交性从全面试验中挑选出部分有代表性的点进行试验，这些有代表性的点具备了“均匀分散，齐整可比”的特点，是一种高效率、快速、经济的实验设计方法[3]。正交试验设计法，可使用已经造好的正交表来安排试验并进行数据分析的一种方法，简单易行。

5.5.1 试验条件的确定

安排试验时，首先把所考察的每一个因子任意地对应于正交表的一列，然后把每列的数字“翻译”成所对应因子的水平^[5]。这样，每一行的各水平组合就构成了一个试验条件。在本实验中，共有 5 个因子，分别为：Co 负载量、HAP 含量、Co/SiO₂ 和 HAP 装料比、乙醇浓度和温度。同时综合考虑附件 1 中已完成的实验，选取 4 水平进行实验。如若按照传统的例举组合实验法，需进行 $4^5 = 1024$ 次实验，次数较多且容易产生误差，因此更适合利用正交试验进行分析。

5.5.2 水平试验点的选取

基于因子与水平个数的确定，需对 5 个因子 4 个水平各自进行取值。考虑到简单对比实验操作性强且极具代表性，而正交试验选点分布均匀。因此兼顾两种试验方

法的优点，从已进行的试验点中选择具有典型性、代表性的点，使试验点在试验范围内分布得很均匀，能反映全面情况。具体数值如下：

表 21：试验点数据

	Co 负载量	HAP 质量	Co/SiO_2 和 HAP 装料比	乙醇浓度	温度
水平 1	25	25	0.5	0.3	250
水平 2	50	50	1	0.9	300
水平 3	100	100	2	1.68	350
水平 4	200	200	5	2.1	400

5.5.3 基于正交表的方案确定

根据数据个数的选择，可考虑选用正交表 $L_{16}(4,5)$ 进行分析，利用 SPSS 软件，使用拟水平法确定正交试验方案，数据如下：

表 22：正交设计表

编号	Co 负载量	HAP 质量	Co/SiO_2 和 HAP 装料比	乙醇浓度	温度
1	25	25	0.5	0.3	250
2	25	50	1	0.9	300
3	25	100	2	1.68	350
4	25	200	5	2.1	400
5	50	25	1	1.68	400
6	50	50	0.5	2.1	350
7	50	100	5	0.3	300
8	50	200	2	0.9	250
9	100	25	2	2.1	300
10	100	50	5	1.68	250
11	100	100	0.5	0.9	400
12	100	200	1	0.3	350
13	200	25	5	0.9	350
14	200	50	2	0.3	500
15	200	100	1	2.1	250
16	200	200	0.5	1.68	300

如表所示， Co 负载量、 HAP 含量、 Co/SiO_2 和 HAP 装料比因子、乙醇浓度和温度可任意地对应于正交表的某五列，然后试验按行进行，每一行中各因素的水平组合就是每一次的试验条件，从上到下即为这个正交试验的方案。此时，已将水平试验的数据填入，即为理想情况下的正交实验方案。

5.5.4 正交试验方案的改进

为了增加 5 组实验，我们基于正交实验及模型三已经求解得到的最优催化剂组合和温度方案进行考虑。

(1) 筛选已进行的试验

为了节约实验用品及人力物力的浪费，我们不再对附件 1 中已经试验过的催化剂组合与温度方案重复实验，基于此想法，我们将筛选正交实验中已在附件 1 中做过的实验，去掉的方案编号为 6，11，16。

(2) 筛选不符合试验基本规则的试验

在筛选已做过实验中剩下的实验进行考虑，由于附件 1 中的实验具有参考价值，

分析各指标数据可以明显看到 Co 负载量、 HAP 含量在大多数实验中均相同，少部分实验其数值差别不大，因此我们通过此实验规律筛选正交试验中不符合实验规律的结果，其中有编号 2, 3, 4, 8, 9, 12, 13, 14。如编号 14 的试验， Co 负载量、 HAP 含量的比值相差近 4 倍，明显不符合实验规律。

(3) 增添最优试验方案

经过步骤(1)与步骤(2)的筛选，符合实验规律且不重复实验的编号有 1, 5, 7, 10, 15，同时考虑到乙醇偶合制备 C4 烯烃实验中，目的是为了获取 C4 烯烃，因此 C4 烯烃转化率同样为实验好坏的衡量标准。计算这 5 组试验的 C4 烯烃收率可以发现编号 10 的试验 C4 烯烃收率较低，实验效果不理想，同样为了验证模型三最优方案的实际试验结果，因此，我们将编号 10 的实验替换为模型三中的最优方案。综上所述，我们增加的 5 组实验如下：

实验 1：催化剂组合：

0.5mg 25wt% Co/SiO_2 - 25mg HAP - 乙醇浓度0.3ml/min
温度：250 摄氏度

实验 2：催化剂组合：

1mg 50wt% Co/SiO_2 - 25mg HAP - 乙醇浓度1.8ml/min
温度：400 摄氏度

实验 3：催化剂组合：

5mg 50wt% Co/SiO_2 - 100mg HAP - 乙醇浓度0.3ml/min
温度：300 摄氏度

实验 4：催化剂组合：

1mg 200wt% Co/SiO_2 - 100mg HAP - 乙醇浓度2.1ml/min
温度：250 摄氏度

实验 5：催化剂组合：

0.147mg 200wt% Co/SiO_2 - 10mg HAP - 乙醇浓度0.3ml/min
温度：350 摄氏度

六、模型的评价

6.1 模型的评价

本文的亮点之一是综合使用统计类模型，使得指标与数据更加科学合理。首先，综合利用聚类分析以及主成分分析对数据进行降维，深入探究不同指标间的关联程度。其次，在多元回归模型的基础上，着重考虑多重共线性，引入二次交互性进行逐步回归，不仅降低多重共线性，更提高了拟合程度。再者，考虑实际实验，利用 logistic 变换对数据进行处理，使其更具有现实意义。然后，综合利用二分法与遍历法算法对数值进行求解，速度快且精度高，相互对比照应使得结果更具有说服力。最后，本文在正交试验的基础上对实验进行一定的补充，不仅减少试验次数，更提高了实验的速度。

但是，本文建立的模型也有一定的局限性，具体体现在未考虑装料方式所带来的轻微影响，后续可对此缺陷进行相应的提升与改善。

6.2 模型的改进

在实际的化学实验中，不同的装料方式会对产物产生轻微的影响，并且实验室条件也并非理想的环境。因此，在我们的模型建立中忽略了二者的影响，后续可针对不同装料方式下的产物进行单独分析，以此使得模型更具有广泛性。

6.3 模型的推广

本文建立的乙醇偶合制备 C4 烯烃策略具有一定的实用性，可对 C4 烯烃的实验设计与准备提供一定的参考。该模型具有一定的应用前景，可应用于催化剂与乙醇低碳烯烃等化学实验。

七、参考文献

- [1] <https://www.zhihu.com/question/485822400> 2022.8.12
- [2] <https://baike.baidu.com/item/%E4%B8%BB%E6%88%90%E5%88%86%E5%88%86%E6%9E%90/829840> 2022.8.11
- [3] <https://zhuanlan.zhihu.com/p/63139206> 2022.8.11
- [4] 吕绍沛. 乙醇偶合制备丁醇及 C4 烯烃[D]. 辽宁大连理工大学, 2018
- [5] 姜启源, 谢金星, 叶俊. 数学模型(第五版)[M]. 北京: 高等教育出版社, 2018
- [6] <https://baike.baidu.com/item/正交试验/8699365?fr=aladdin> 2022.8.11
- [7] 董如何, 肖必华, 方永水. 正交试验设计的理论分析方法及应用[J]. 安徽建筑大学学报, 2004, 12(6):103-106.

附录

附录 1

介绍：遍历算法

第一次遍历：

clear;

X1 = 10 : 5 : 200; % Co 负载量

X2 = 10 : 5 : 200; % HAP 质量

X3 = 0.5 : 0.1 : 5; % 质量比

X4 = 0.3 : 0.1 : 2.1; % 乙醇浓度

X5 = 250 : 5 : 450; % 温度

maxq = 4995;

Z_arr = [];

Z_X1 = [];

Z_X2 = [];

Z_X3 = [];

Z_X4 = [];

Z_X5 = [];

n=0;

Index = [];

for i1=1:length(X1)

 x1 = X1(i1);

 for i2=1:length(X2)

 x2 = X2(i2);

 for i3=1:length(X3)

 x3 = X3(i3);

 for i4=1:length(X4)

 x4 = X4(i4);

 for i5=1:length(X5)

 x5 = X5(i5);

 z = 0.018*x5 + 0.035*x1 - 0.039*x2 - 0.19*x3*x3 + 0.007*x2*x4 - ...
 0.002*x4*x5 + 0.816*x3 - 8.015;

 zz = 5000*exp(z)/(1+exp(z));

 if (zz > maxq)

 n=n+1;

 Z_X1(n) = x1;

 Z_X2(n) = x2;

 Z_X3(n) = x3;

 Z_X4(n) = x4;

 Z_X5(n) = x5;

 Index(n) = n;


```

        Z_arr(n) = zz;
    end
end
end
end
end
end

[maxValue, index] = max(Z_arr);
scatter(Index, Z_arr, 8, 'b');
xlabel('序号');
ylabel('目标值');
title('各个目标值');
Z_X1(index)
Z_X2(index)
Z_X3(index)
Z_X4(index)
Z_X5(index)
第3次遍历:
clear;

X1 = 199 : 0.1 : 200;    % Co 负载量
X2 = 10 : 0.1 : 11;     % HAP 质量
X3 = 2 : 0.001 : 3;     % 质量比
X4 = 0.3 : 0.01 : 0.4;  % 乙醇浓度
X5 = 449 : 0.1 : 450;   % 温度

maxq = 4996.696;
Z_arr = [];

Z_X1 = [];
Z_X2 = [];
Z_X3 = [];
Z_X4 = [];
Z_X5 = [];

n=0;
Index = [];

for i1=1:length(X1)
    x1 = X1(i1);
    for i2=1:length(X2)
        x2 = X2(i2);
        for i3=1:length(X3)
            x3 = X3(i3);
            for i4=1:length(X4)
                x4 = X4(i4);

```

```

for i5=1:length(X5)
    x5 = X5(i5);

    z = 0.018*x5 + 0.035*x1 - 0.039*x2 - 0.19*x3*x3 + 0.007*x2*x4 - ...
        0.002*x4*x5 + 0.816*x3 - 8.015;
    zz = 5000*exp(z)/(1+exp(z));
    if (zz > maxq)
        n=n+1;
        Z_X1(n) = x1;
        Z_X2(n) = x2;
        Z_X3(n) = x3;
        Z_X4(n) = x4;
        Z_X5(n) = x5;
        Index(n) = n;
        Z_arr(n) = zz;
    end
end
end
end
end
end

[maxValue, index] = max(Z_arr);
scatter(Index, Z_arr, 8, 'b');
xlabel('序号');
ylabel('目标值');
title('各个目标值');
Z_X1(index)
Z_X2(index)
Z_X3(index)
Z_X4(index)
Z_X5(index)

```

附录 2

介绍： 二分法代码 MATLAB

```

clear;

maxq = 4960;
Z_arr = [];

Z_X1 = [];
Z_X2 = [];
Z_X3 = [];
Z_X4 = [];
Z_X5 = [];

n=0;

```

```

Index = [];

left1 = 10;
right1 = 200;
mid1 = 0;

left2 = 10;
right2 = 200;
mid2 = 0;

left4 = 0.3;
right4 = 2.1;
mid4 = 0;

left5 = 250;
right5 = 350;
mid5 = 0;

X3 = 0.5 : 0.1 : 5;    % 质量比

for i1=1:20
    mid1 = right1 - (right1-left1)/2;
    x1 = mid1;
    left1 = mid1;
    mid2 = left2 + (right2-left2)/2;
    x2 = mid2;
    right2 = mid2;
    mid4 = left4 + (right4-left4)/2;
    x4 = mid4;
    right4 = mid4;
    mid5 = right5 - (right5-left5)/2;
    x5 = mid5;
    left5 = mid5;
    for i3=1:length(X3)
        x3 = X3(i3);
        z = 0.018*x5 + 0.035*x1 - 0.039*x2 - 0.19*x3*x3 + 0.007*x2*x4 - ...
            0.002*x4*x5 + 0.816*x3 - 8.015;
        zz = 5000*exp(z)/(1+exp(z));
        if (zz > maxq)
            n=n+1;
            Z_X1(n) = x1;
            Z_X2(n) = x2;
            Z_X3(n) = x3;
            Z_X4(n) = x4;
            Z_X5(n) = x5;
            Index(n) = n;
            Z_arr(n) = zz;
        end
    end
end

```

```
end  
end
```

```
[maxValue, index] = max(Z_arr);  
scatter(Index, Z_arr, 8, 'b');  
xlabel('序号');  
ylabel('目标值');  
title('各个目标值');  
Z_X1(index)  
Z_X2(index)  
Z_X3(index)  
Z_X4(index)  
Z_X5(index)
```

附录 3

介绍： 模型三二分法 MATLAB

```
clear;  
  
maxq = 4995;  
Z_arr = [];  
  
Z_X1 = [];  
Z_X2 = [];  
Z_X3 = [];  
Z_X4 = [];  
Z_X5 = [];  
  
n=0;  
Index = [];  
  
left1 = 10;  
right1 = 200;  
mid1 = 0;  
  
left2 = 10;  
right2 = 200;  
mid2 = 0;  
  
left4 = 0.3;  
right4 = 2.1;  
mid4 = 0;  
  
left5 = 250;  
right5 = 450;
```

```

mid5 = 0;

X3 = 0.5 : 0.1 : 5;    % 质量比

for i1=1:20
    mid1 = right1 - (right1-left1)/2;
    x1 = mid1;
    left1 = mid1;
    mid2 = left2 + (right2-left2)/2;
    x2 = mid2;
    right2 = mid2;
    mid4 = left4 + (right4-left4)/2;
    x4 = mid4;
    right4 = mid4;
    mid5 = right5 - (right5-left5)/2;
    x5 = mid5;
    left5 = mid5;
    for i3=1:length(X3)
        x3 = X3(i3);
        z = 0.018*x5 + 0.035*x1 - 0.039*x2 - 0.19*x3*x3 - 0.007*x2*x4 - ...
            0.002*x4*x5 + 0.816*x3 - 8.015;
        zz = 5000*exp(z)/(1+exp(z));
        if (zz > maxq)
            n=n+1;
            Z_X1(n) = x1;
            Z_X2(n) = x2;
            Z_X3(n) = x3;
            Z_X4(n) = x4;
            Z_X5(n) = x5;
            Index(n) = n;
            Z_arr(n) = zz;
        end
    end
end
end

[maxValue, index] = max(Z_arr);
scatter(Index, Z_arr, 8, 'b');
xlabel('序号');
ylabel('目标值');
title('各个目标值');
Z_X1(index)
Z_X2(index)
Z_X3(index)
Z_X4(index)
Z_X5(index)

```