

基于遗传算法求解 C4 烯烃制备的最优策略

摘 要

本文针对乙醇偶合制备 C4 烯烃的问题，使用 Pearson 相关系数、控制变量法、BP 神经网络、遗传算法等方法，分别建立 Pearson 相关系数模型、基于控制变量的灵敏度模型和基于 BP 神经网络的遗传算法模型，通过 MATLAB 软件编程依次得到：乙醇转化率、C4 烯烃的选择性与温度之间的关系，给定 350℃ 实验组的结果分析；不同催化剂组合及温度对乙醇转化率以及 C4 烯烃选择性的影响程度；使 C4 烯烃收率尽可能高的全局最优解（催化剂与温度组合），以及在温度低于 350℃ 时的全局最优解；基于第二、第三问的结论设计了五次新实验。

针对问题一，通过查阅相关文献了解到乙醇偶合制备 C4 烯烃的化学反应过程及原理，对附件提供的实验数据展开分析和论证，选取了三次样条插值法进行处理。画出乙醇转化率、C4 烯烃的选择性与温度的三维折线图，观察到其基本呈线性相关，在通过正态性检验的前提下建立了 Pearson 相关系数模型对乙醇转化率、C4 烯烃的选择性和温度的关系进行量化分析，发现每种催化剂组合下的乙醇转化率、C4 烯烃的选择性与温度均有很强正相关性，在催化剂组合为 A3 的情况下，乙醇转化率、C4 烯烃的选择性与温度相关性最大，Pearson 相关系数的计算结果见附录。

针对问题二，首先进行相关数据的收集与整理，然后利用 MATLAB 软件构建了基于控制变量的灵敏度模型，并结合 EXCEL 制作得到不同催化剂组合及温度对乙醇转化率以及 C4 烯烃选择性大小影响的折线图，通过比较分析上述折线图与相关数据得到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，首先使用 MATLAB 软件构建了 BP 神经网络模型来计算 C4 烯烃的收率，然后对该神经网络模型进行验证，发现模型具有较高的准确性，然后构建基于 BP 神经网络的遗传算法模型进行搜索，最后得到全局最优解(温度, Co 负载量, Co/SiO₂ 的质量, HAP 的质量, Co/SiO₂ 和 HAP 装料比, 乙醇浓度): 422.1402℃, 0.5698wt%, 259.6742mg, 166.6059mg, 1.5586, 1.2210ml/min, 此时的 C4 烯烃收率最高: 52.33%, 并利用该模型搜索得到温度低于 350℃ 时的全局最优解(温度, Co 负载量, Co/SiO₂ 的质量, HAP 的质量, Co/SiO₂ 和 HAP 装料比, 乙醇浓度): 349.9780℃, 1.9669wt%, 217.2945mg, 177.0551mg, 1.2273, 1.6317ml/min, 此时的 C4 烯烃收率最高: 50.78%。

针对问题四，以 C4 烯烃的收率高为前提，以验证问题三模型及结果的准确性和可行性为目的设计 5 个实验方案，对附件 1 的实验数据分析，选择设计 2 个在催化剂组合 A1、A2 的条件下温度为 400℃ 的实验；选择设计 3 个催化剂组合方式及温度在第三问得到的全局最优解的条件下小幅度波动，将其代入第三问的模型来验证模型的准确性和可行性，最终设计方案见设计方案的选取表。

最后对模型进行了中肯的评价，并对模型的应用进行了推广。

关键词：Pearson 相关系数、控制变量、BP 神经网络、遗传算法

一、问题重述

1.1 问题背景

C4 烯烃是一种重要的化工原料，在医药领域具有极高的经济价值，在传统的 C4 烯烃生产过程中，化石能源扮演着极其重要的角色，但由于化石能源的不可再生性，且对环境污染严重，寻找 C4 烯烃新的制备方案已成为当下非常重要的方向，乙醇属于清洁的可再生能源，C4 烯烃可以通过乙醇脱氢、偶合、脱水、加氢制备，但该路线制备 C4 烯烃需要酸、碱活性位点的共同催化，如何选择合适的催化剂组合及温度使得乙醇的收率最大化的问题有着重要的研究意义^[1]。

1.2 问题重述

1. 依据附件一分别研究每种组合下乙醇转化率、C4 烯烃的选择性与温度的关系，并对 350°C 时给定的催化剂组合在一次实验不同时间的测试结果进行分析。
2. 研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。
3. 在相同实验条件下，如何选择催化剂组合与温度使得 C4 烯烃收率尽可能高。若温度低于 350°C，应如何选择催化剂组合与温度，使得 C4 烯烃收率尽可能高。
4. 如果再增加 5 次实验，应该如何设计，并给出详细理由。

二、问题分析

2.1 问题一的分析

题目要求分别研究乙醇转化率、C4 烯烃的选择性与温度的关系，考虑到附件数据较少，不妨采用三次样条插值对数据进行内插处理，首先对数据进行可视化，分别建立转化率、选择性与温度的三维可视图，可以观察到两个指标与温度具有明显的线性相关性，不妨采用 Pearson 相关系数描述各指标与温度的相关性大小。对于 350°C 条件下给定的测试结果，观察到曲线一直在波动，且 C4 烯烃选择性与其他选择性总是此消彼长，我们猜测 C4 烯烃与其它可能是由同一种中间产物转化而来，随着环境条件的变化，中间产物选择性的合成 C4 烯烃、其他。

2.2 问题二的分析

为探究不同催化剂组合及温度对乙醇转化率及 C4 烯烃选择性的影响，不妨采用控制变量的思想，从附件中筛选出合适的成分并绘制折线图，观察折线图的斜率大小，可以得出该成分对转化率、选择性的影响程度大小。另外我们还可以发现在其它实验条件相同的情况下，装填方式对最终的影响不大，400°C 情况下差距均未超过 5%。综合分析对比成分，最终可以得到催化剂组合中每种成分及温度对乙醇转化率、C4 烯烃选择性的影响情况。

2.3 问题三的分析

问题三属于典型的优化性问题，我们的目标是找到 C4 烯烃收率最高时对应的催化剂组合及温度，这就需要我们首先建立一个催化剂组合及温度与 C4

烯烃收率的对应关系，首先分析出对收率产生影响的自变量主要有温度、Co 负载量、Co/SiO₂ 的质量、HAP 的质量、乙醇浓度、催化剂质量比六种，接着考虑到各自变量之间具有复杂的非线性关系，我们选择通过 BP 神经网络建立自变量与收率的对应关系，调试模型参数，使模型误差尽可能小。接着可以利用训练好的 BP 神经网络搜寻最优解，由于因变量参数较多，通过遍历操作求最优解的时间复杂度太高，不妨采用遗传算法搜寻全局最优解，适应度函数即为训练好的 BP 神经网络，接着通过单亲遗传变异、精英遗传等操作不断迭代，最终即可得到较优的解。对于 350℃ 的约束条件，我们可以对加入种群的每个个体进行检查，若温度大于 350℃ 就筛去，直到种群达到上限，其余步骤与无约束时相同。

2.4 问题四的分析

题目要求再设计 5 次实验，并给出详细理由。我们需要根据问题二和问题三的结论，在 C4 烯烃的收率尽可能高的情况下对实验进行设计，并且我们希望设计的实验要与问题二和问题三的结论息息相关，可以根据增加的实验结果来对问题二和问题三的模型进行验证，判断模型的准确性和可行性。

问题分析流程图入下：

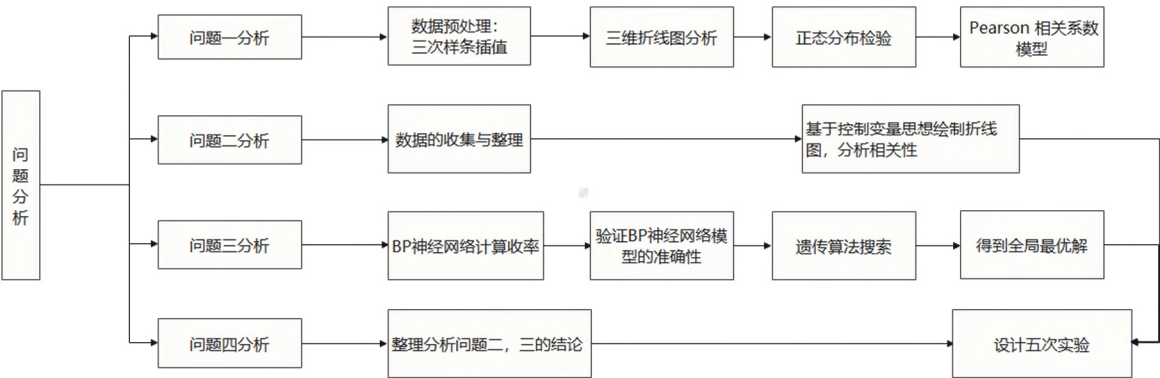


图 2.1：问题分析流程图

三、模型假设

1. 不考虑除附件范围之外的影响因素。
2. 假设所采用的数据具有较好的可信度和典型性。
3. 本文不讨论极端情况，即各实验数据的变化保持在合理的范围之内。
4. 假设实验期间不存在明显的外界干扰。
5. 假设一次反应期间温度保持不变

四、符号说明

关键符号	符号说明
r	Pearson 相关系数
l	神经网络隐层数量
P_{C_4}	C4 烯烃的收率
X	催化剂组合及温度向量集
$f(X)$	遗传算法适应度函数

这里只列出论文各部分通用符号，部分符号在首次引用时会进行说明。

五、模型的建立与求解

5.1.1 Pearson 相关系数模型的建立与求解

➤ 分析对象与指标

本问要求我们分别研究乙醇转化率、C4 烯烃的选择性与温度的关系，并分析在 350°C 时给定的某种催化剂组合的测试数据。查阅资料知乙醇制备 C4 烯烃反应需要经历脱水、偶合等过程，在催化剂 Co/SiO₂ 和 HAP 的协同作用下进行，在不同反应温度下催化剂的催化性能会有明显不同，且乙醇转化率，反应的主要产物也会发生改变。其中乙醇转化率是指单位时间内乙醇的单程转化率，其值为 $100\% \times (\text{乙醇进气量} - \text{乙醇剩余量}) / \text{乙醇进气量}$ ，C4 烯烃的选择性是指其在所有产物中的占比。

➤ 数据预处理与三维图分析

①由于实验数据量较少，数据量不足以支撑分析以及建立合理的模型，所以我们首先需要利用 MATLAB 软件对乙醇转化率和 C4 烯烃选择性的数据进行关于温度的三次样条插值（相关代码见附录），来模拟产生一些新的数据。

②结合附件所给数据以及上文经过插值所得数据，利用 MATLAB 做出乙醇转化率、C4 烯烃的选择性与温度的三维折线图（相关代码见附录）。

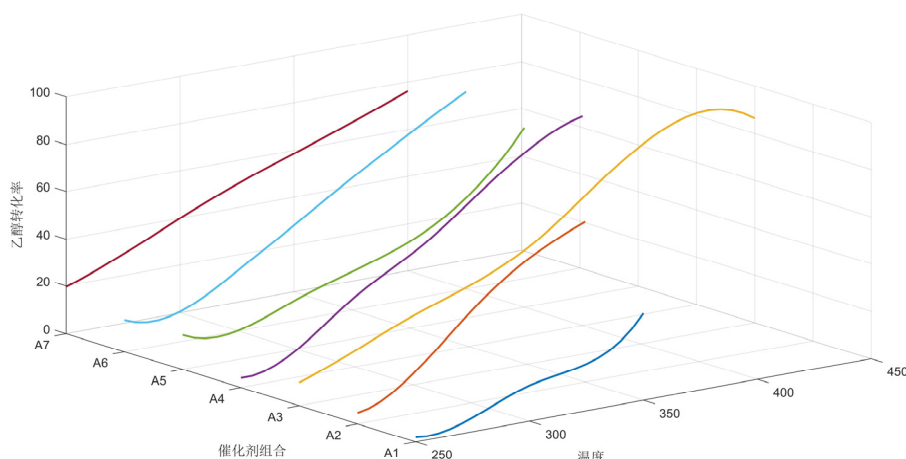


图 5.1：A1-A7 组实验乙醇的转化率与温度的关系图

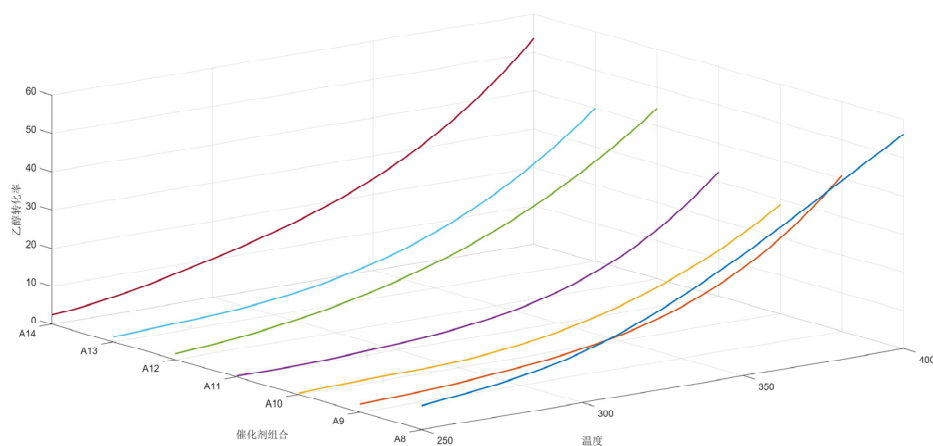


图 5.2: A8-A14 组实验乙醇的转化率与温度的关系图

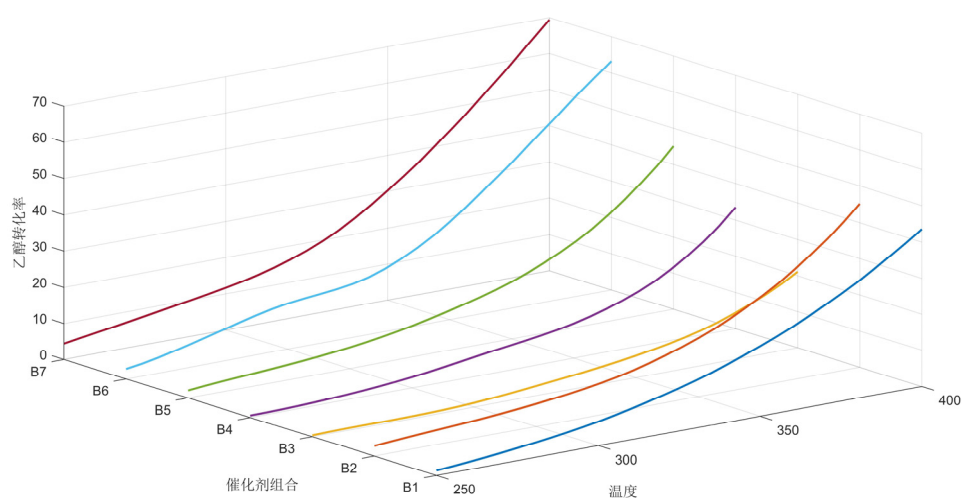


图 5.3: B1-B7 组实验乙醇的转化率与温度的关系图

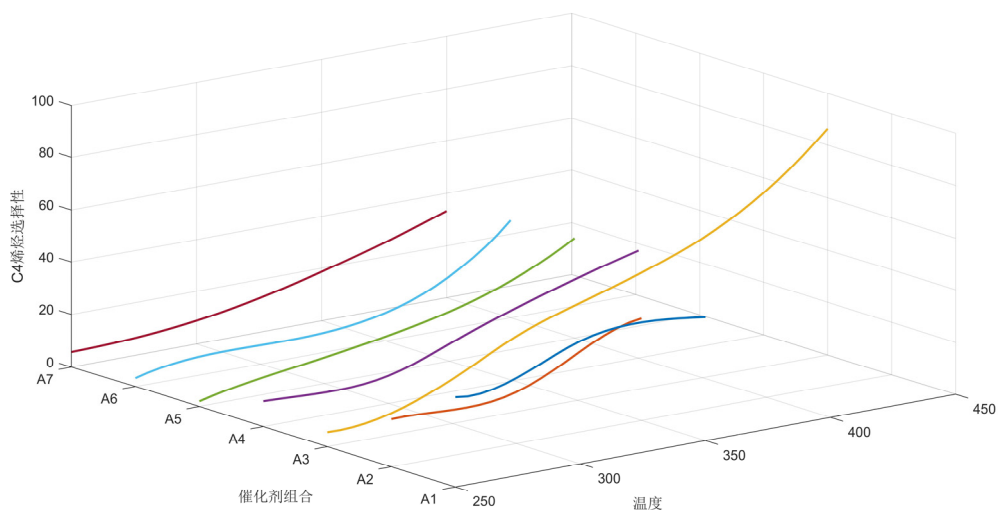


图 5.4: A1-A7 组实验 C4 烯烃的选择性与温度的关系图

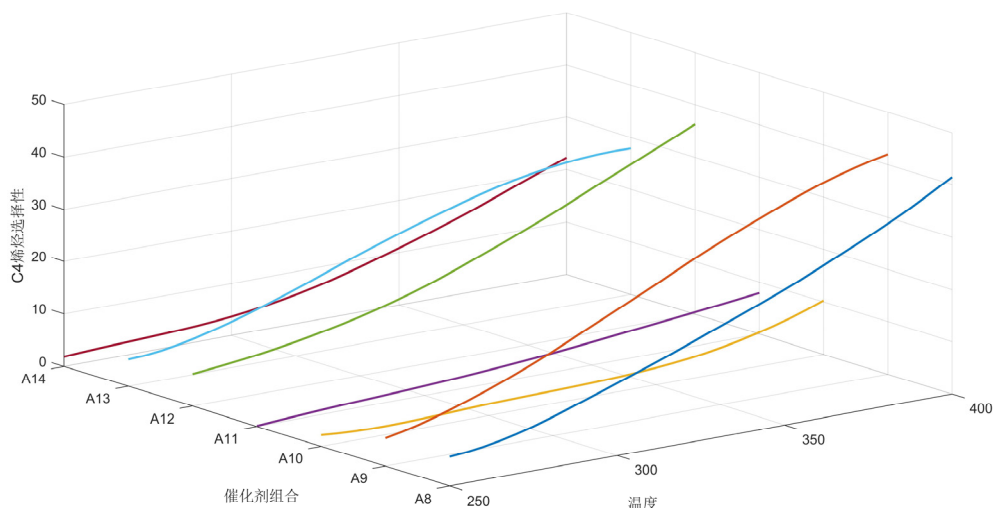


图 5.5: A8-A14 组实验 C4 烯烃的选择性与温度的关系图

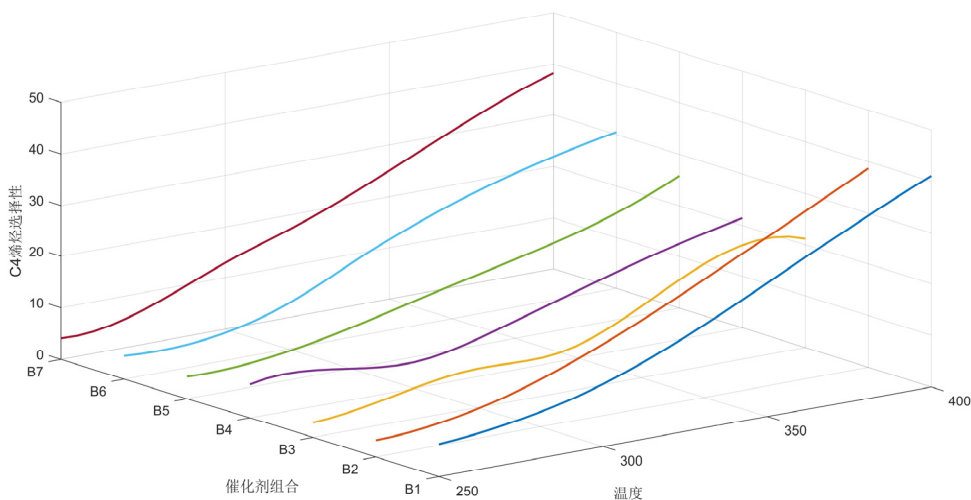


图 5.6: B1-B7 组实验 C4 烯烃的选择性与温度的关系图

首先观察乙醇转化率随温度的变化情况，对于 A1-A7 组，温度和乙醇转化率基本呈线性关系，存在特殊情况，A3 组在 400°C 之后乙醇转化率随温度上升而下降，A5、A6 组在 250°C 后短暂的时间段内，乙醇转化率随温度上升而下降，A8-A14 组、B1-B7 组均随温度上升而上升，且斜率逐渐增大；

然后观察 C4 烯烃选择性随温度的变化情况，A1-B7 大部分组中，C4 烯烃选择性随温度上升而上升，但 B3 组在 350°C 后 C4 烯烃选择性有略微下降。

接着我们对所有数据进行雅克-贝拉(Jarque-Bera)检验，除 A1 组乙醇转化率关于温度的数据在 90%的置信区间上通过了检验，其余组均在 95%的置信区间上通过了检验，

接着计算 Pearson 相关系数，计算公式如下：

$$r = \frac{\sum_{i=1}^n (X_i - \bar{X})(Y_i - \bar{Y})}{\sqrt{\sum_{i=1}^n (X_i - \bar{X})^2} \cdot \sqrt{\sum_{i=1}^n (Y_i - \bar{Y})^2}} \quad (5.1.1)$$

式中， \bar{X} 、 \bar{Y} 分别代表两组数据的平均值， n 代表样本个数，用 Excel 色阶可视化处理结果如下(详细表格及代码见附录)：

表 5.1：各指标与温度的 Pearson 相关系数表

	A1	A2	A3	B5	B6	B7
乙醇转化率	0.98183	0.99562	0.98815	0.92825	0.94846	0.94829
C4 烯烃选择性	0.90864	0.91063	0.99184	0.98846	0.98896	0.99742

综合上文的折线图以及 Pearson 相关系数分析表，我们可以得到以下结论：21 组实验的乙醇转化率和 C4 烯烃选择性都与温度具有显著的相关性，由此可知在乙醇偶合制备 C4 烯烃的反应过程中，乙醇的转化率和 C4 烯烃的选择性受温度的影响较大，且催化剂组合 A3 的乙醇转化率和 C4 烯烃的选择性受温度影响的程度最大，催化剂组合 A6 和催化剂组合 A10 的 C4 烯烃的选择性受温度影响的程度相对于其他催化剂组合较小，温度是影响制备反应结果的重要影响因素。

5.1.2 350°C 时给定催化剂组合的测试结果分析

由于直接观察表格不容易观察出选择性的变化趋势，我们对表格中的数据进行了可视化，如下图：

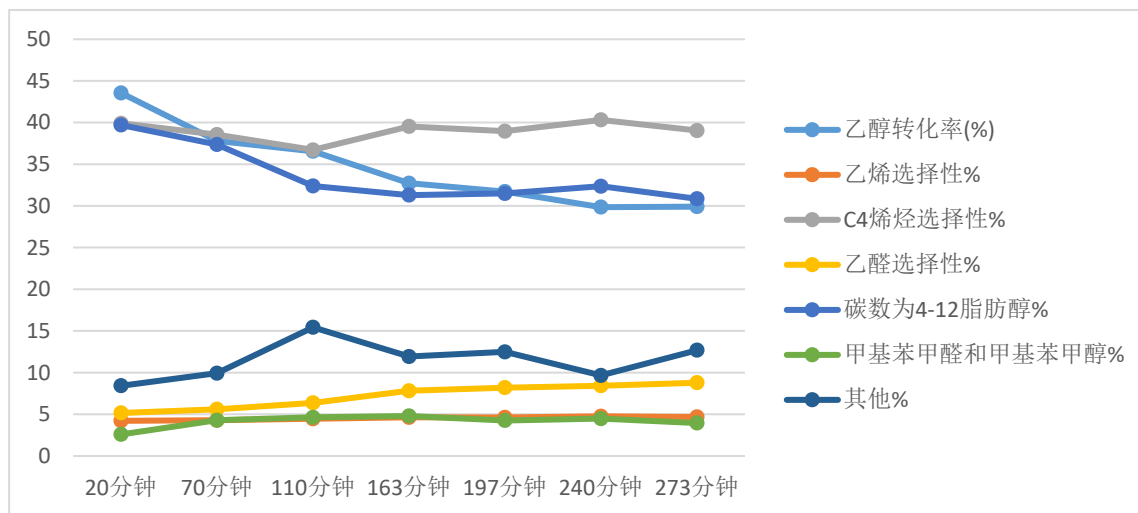


图 5.7：350°C 时给定的某种催化剂组合的测试数据

➤ 乙醇转化率变化规律的分析

观察上图，该实验为一次实验，每分钟通入固定数量的乙醇，随着反应时间的增长，产物不断生成，乙醇的浓度下降，从而导致乙醇的转化率和乙醇的反应速度下降，乙醇的反应速度下降导致乙醇浓度的下降速率减缓，故乙醇的转化率下降的速率也逐渐减缓，乙醇的通入速度和乙醇的反应速度逐渐趋于平衡，故乙醇的转化率最终趋于稳定，具体如下图所示：

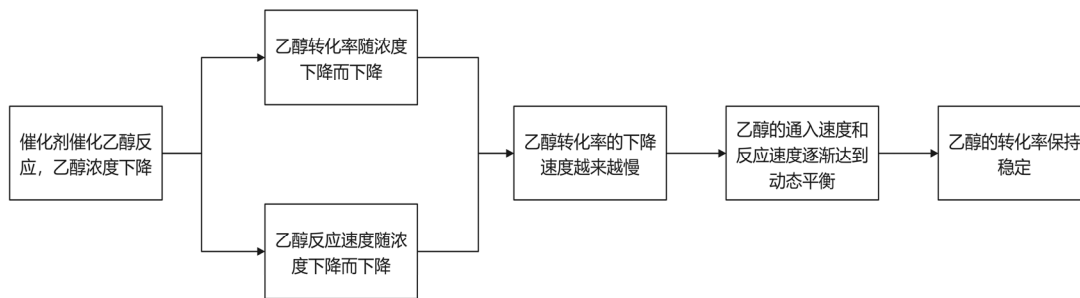


图 5.8：乙醇转化率的变化过程

➤ 产物选择性变化规律的分析

观察数据：在此实验条件下的主要产物是 C4 烯烃和碳数为 4-12 脂肪醇（以下简称高碳脂肪醇）。在时间为 20min-110min 之间时，C4 烯烃和高碳脂肪醇的选择性显著下降，其他的选择性显著上升；在时间为 110min-240min 之间时，C4 烯烃的选择性呈上升趋势，其他的选择性呈下降趋势，高碳脂肪醇的选择性无明显变化趋势；在时间为 240min-273min 之间时，C4 烯烃和高碳脂肪醇的选择性下降，其他的选择性上升。在整个时间段里，乙烯、甲基苯甲醛和甲基苯甲醇的选择性没有明显变化趋势，乙醛的选择性在缓慢上升，C4 烯烃的选择性与其他的选择性的变化趋势完全相反。

原因分析：结合折线图以及相关资料我们发现由于 C4 烯烃呈酸性且是主要的反应产物之一，C4 烯烃的大量存在会导致溶液呈酸性，而酸性条件会抑制乙醛转化生成 C4 烯烃和高碳脂肪醇，从而乙醛会更多的转化生成成为其他；随着反应进行，环境的酸性减弱，进行脱水缩合反应后又生成 C4 烯烃，这是一个酸碱性随着时间不断波动的过程，各生成物的选择性也随之波动，且波动程度不断减小，最终趋于稳定。

其中影响产物的选择性的变化的具体过程如下图所示：

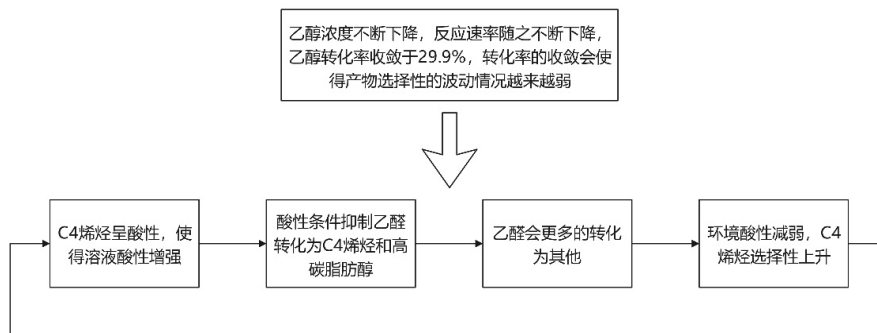


图 5.9：产物的选择性的变化过程

5.2.1 控制变量模型的建立

第二问要求分析不同催化剂组合及温度对乙醇转化率、C4 烯烃的选择性大小的影响，考虑到最终结果是由所有成分共同作用产生的，整体上无法做较为深入的分析，所以我们利用了实验中常见的控制变量的思想，分别对 Co 负载量、乙醇浓度、装料比、装料方式、温度、HAP、装料量对乙醇转化率及 C4 烯烃选择性的影响进行了分析。

首先收集整理原始数据，利用 EXCEL 做出各成分及温度与乙醇转化率以及 C4 烯烃选择性关系的折线图，通过观察折线图并结合相关文献分析总结出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。

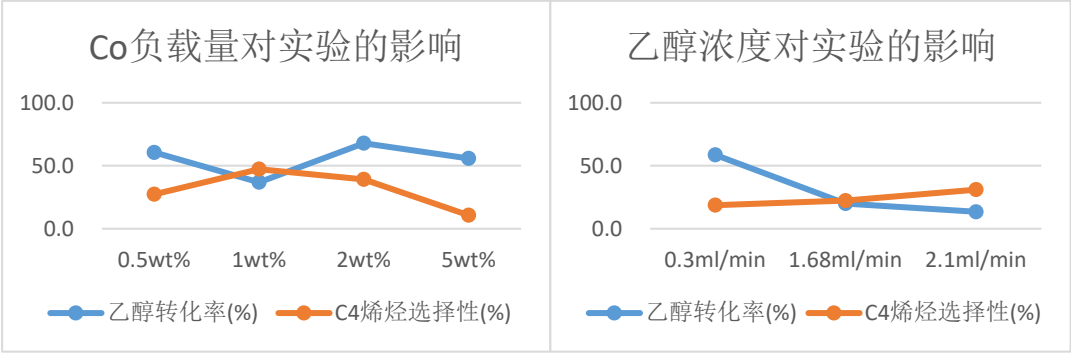


图 5.10: Co 负载量对实验的影响

图 5.11: 乙醇浓度对实验的影响

观察左图可知 Co 负载量从 0.5wt%-1.0wt%时，乙醇的转化率在降低，C4 烯烃的选择性在升高，从 1.0wt%-5wt%时，乙醇的转化率呈上升趋势，C4 烯烃的选择性下降；Co 的含量过多或过少都会产生不利影响，在 1wt%-2wt%之间较佳，观察右图可知随着乙醇浓度的上升，乙醇的转化率降低，C4 烯烃的选择性缓慢上升；

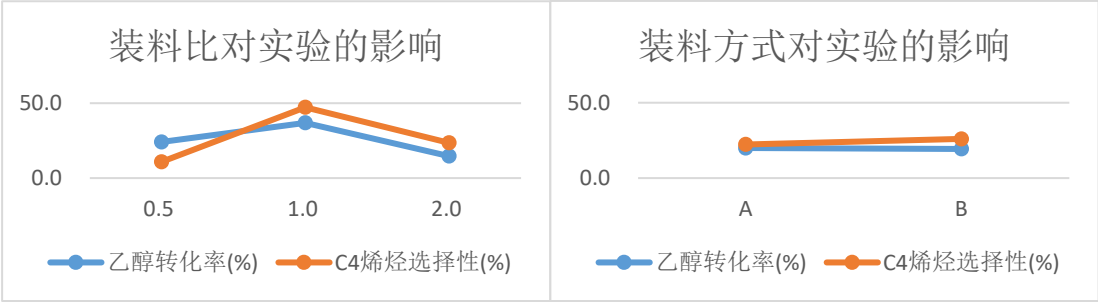


图 5.12: 装料比对实验的影响

图 5.13: 装料方式对实验的影响

观察左图可知装料比为 1:1 时，乙醇的转化率和 C4 烯烃的选择性最高；真实最高值应该在 1:2 到 2:1 之间。观察右图可知装料方式对乙醇的转化率和 C4 烯烃的选择性并没有太大的影响（这条结论会在后面确定神经网络训练集时用到）；

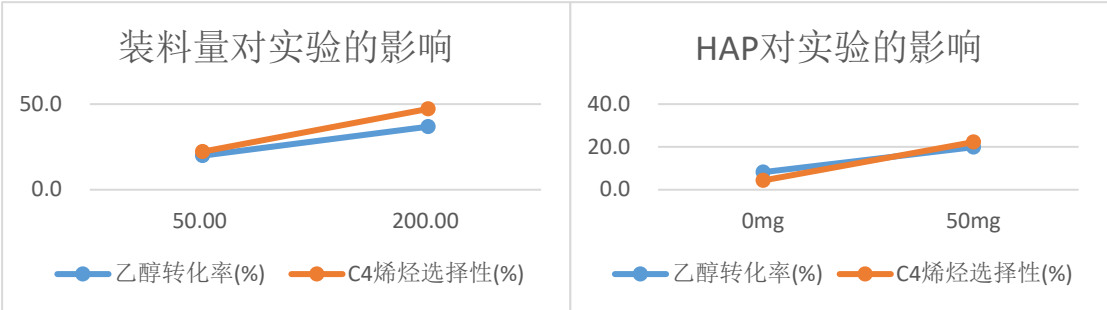


图 5.14: 装料量对实验的影响

图 5.15: HAP 对实验的影响

观察左图可知乙醇的转化率和 C4 烯烃的选择性会随装料量上升而上升；观察右图可知在无 HAP 参与的反应条件下，乙醇的转化率和 C4 烯烃的选择性

都很低，在有 HAP 参与的反应条件下，乙醇的转化率和 C4 烯烃的选择性会提高，故我们可以得出 HPA 会促进乙醇形成 C4 烯烃；

另外关于温度对乙醇转化率、C4 烯烃选择性的影响是：随着温度上升，乙醇转化率、C4 烯烃选择性不断上升，在第一问中已经讨论过这个问题，这里不再赘述。

5.3.1BP 神经网络模型的建立与求解

■ 模型的建立

BP 神经网络（Back-Propagation Network）又称反向传播神经网络，通过样本数据的训练，不断修正网络权值和阈值使误差函数沿负梯度方向下降，逼近期望输出。当一对学习模式提供给网络后，其神经元的激活值将从输入层经隐含层向输出层传播，在输出层的各神经元输出对应输入模式的网络相应。然后，按减少希望输出与实际输出误差的原则，从输出层经各隐含层，最后回到输入层逐层修正连接权。可以看出，这种修正从输出到输入逐层进行的，所以称它为“误差层修正各连接权”^[2]。

我们建立 BP 神经网络的具体流程图如下：

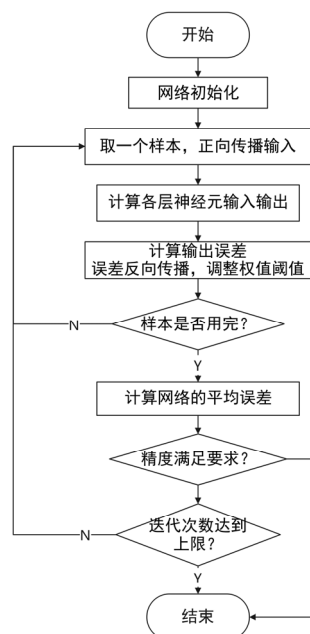


图 5.16：建立 BP 神经网络的流程图

➤ 数据整理：

我们先将不同催化剂组合及温度作为输入，将对乙醇转化率以及 C4 烯烃选择性的乘积收率作为输出，用 MATLAB 自带的 `premnmx()` 函数将这些数据归一化处理（通过第二问数据分析，我们发现装料方式对实验结果的影响不大，所以在本题中不再将其纳入影响范围考虑）。

➤ 网络结构设计

1. 输入输出层的设计

该模型由附件中的催化剂组合和温度作为输入，催化剂组合我们拆分为 Co 载重量、Co/SiO₂ 的质量、HAP 的质量、乙醇浓度、催化剂载重比，以 C4 烯烃收率为输出，所以输入层的节点数为 6，输出层的节点数为 1。

2. 隐层设计

在网络设计过程中，隐层神经元数的确定十分重要，隐层神经元个数过多，会加大网络计算量并容易产生过度拟合问题；神经元个数过少，则会影响网络性能，达不到预期效果。本文在选取隐层神经元个数的问题上参照了以下的经验公式：

$$l = \sqrt{n+m} + a \quad (5.3.1)$$

结合公式和我们的多次测试结果，我们最终选取隐层个数为 6。

■ 模型求解

➤ 输入数据

神经网络模型的建立需要大量的数据，少量数据建立的神经网络模型在精度上满足不了解题的需要，本文所用神经网络的数据部分来源于第一问的插值数据。

➤ 精度检验

为了对求解后的 BP 神经网络模型精度进行测试，我们取 40 为步长，取出所在为位置的数据添加到测试集，多神经网络模型进行多次训练，取效果最好的模型计算其平均绝对误差为 0.0022326，均方根误差为 0.0041709，绘制测试集期望值、预测值及其误差的图像如下（代码见附录）：

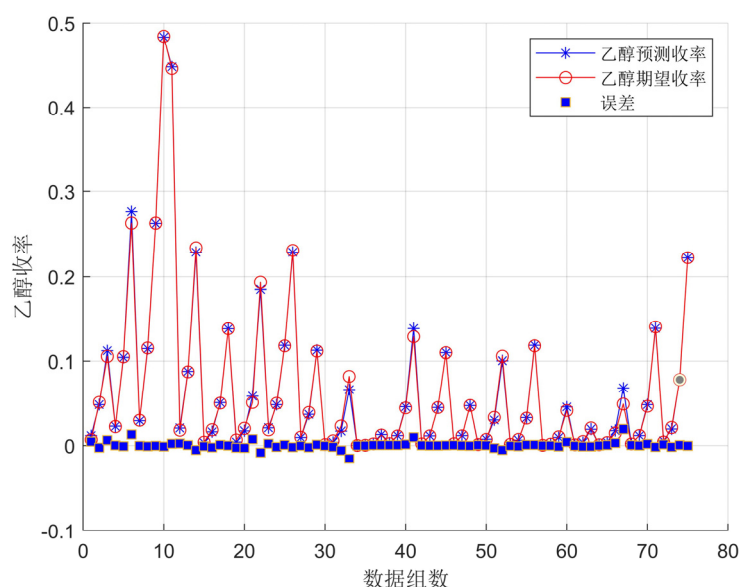


图 5.17：测试结果

观察上图可以看到，预测值和期望值基本吻合，误差基本保持在 0 刻度线附近，说明我们的模型效果良好。

■ 灵敏度分析

用控制变量的方法分别对温度，Co 载重量，装料比，乙醇浓度进行灵敏度分析，绘制曲线如下图（代码见附录）：

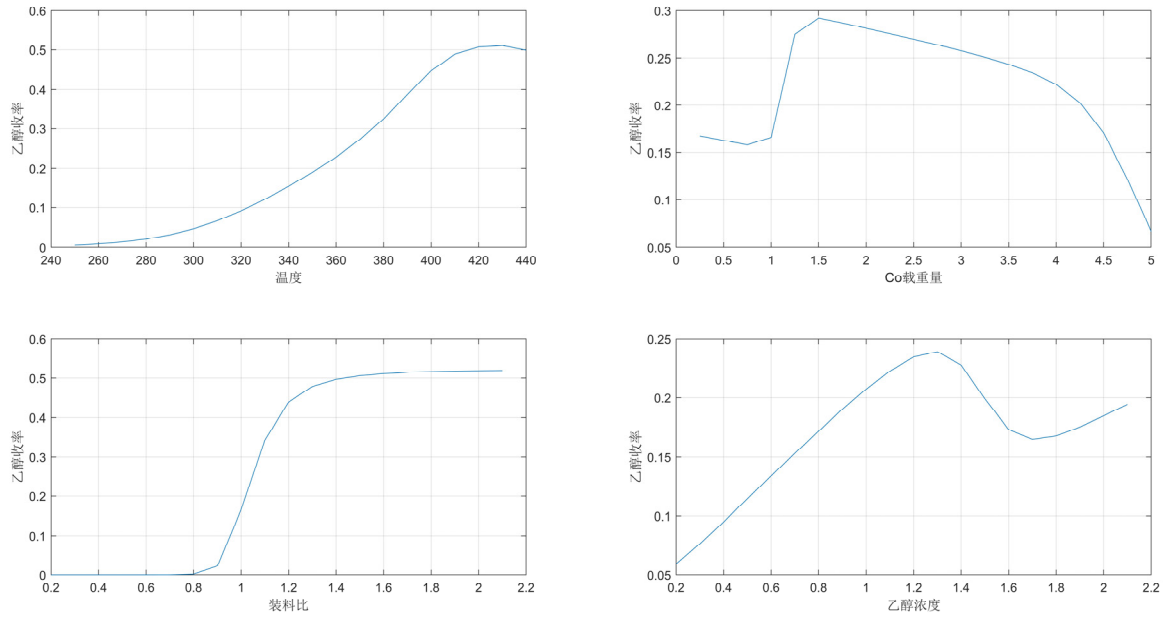


图 5.18：灵敏度分析

观察上图，可以发现乙醇收率的变化情况符合常理，这证明我们的神经网络是可靠的。

5.3.2 遗传算法模型的建立

■ 模型的建立

第三问要求我们找到最优组合使得 C4 烯烃的收率尽可能高，在前文建立的 BP 神经网络模型中建立了各种催化剂组合及温度与 C4 烯烃收率的一一对应关系，我们只需找到一个最优组合代入神经网络使得 C4 烯烃的收率最大即可。

目标函数为：

$$P_{C_4} = \max(f(X)) \quad (5.3.2)$$

决策向量为：

$$X = (x_1, x_2, x_3, x_4, x_5, x_6) \quad (5.3.3)$$

其中 P_{C_4} 代表 C4 烯烃的收率， $x_1 \sim x_6$ 依次代表温度、Co 负载量、Co/SiO₂ 的质量、HAP 的质量、乙醇浓度、催化剂质量比。其中 $f(X)$ 代表向 BP 神经网络输入 X 后输出的收率，其在遗传算法中也代表个体的适应度函数。

本文采用的遗传算法是单亲遗传算法，且没有对个体进行编码，每次迭代结束后，编码的目的是为了方便交叉变异，这里使用了一种扰动算子，对于每个个体的 6 个特征值按比例施加扰动，从而达到避免陷入局部最优的目的，另外，为了避免每次搜寻到的最优解又被扰动算子破坏，本文设置了精英遗传机制，每次迭代后遍历所有个体的适应度（即 C4 烯烃的收率），对种群内所有个体按适应度进行降序排列，取前十个个体为精英群体，每次迭代都原封不动的保留精英群体到下一代，下一代的其它个体都是对精英群体施加扰动算子后生成，不断循环迭代，最终排序后取种群第一个个体即为求得的全局最优解。遗传算法流程图如下：

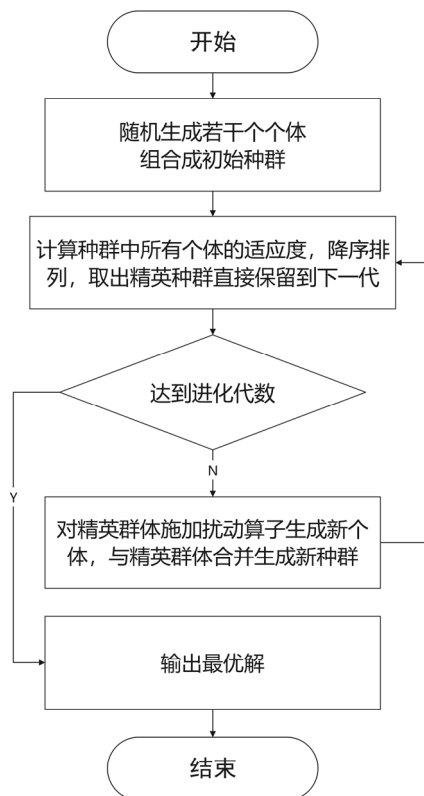


图 5.19: 遗传算法流程图

■ 模型的求解

编写 MATLAB 代码求解上述遗传算法, 得到无约束条件下最优解情况如下:

表 5.1: 无约束条件下最优解

C4 烯烃 收率	温度	Co 负载 量	Co/SiO ₂ 的质量	HAP 的 质量	乙醇浓度	催化剂质 量比
0.5233	422.1402	0.5698	259.6742	166.6059	1.2210	1.5586

对遗传算法生成的新个体进行检查, 当温度超过 350°C 时淘汰该个体, 继续对精英群体施加扰动算子生成新个体, 直到达到种群数量上限, 得到有约束条件下最优解情况如下:

表 5.2: 有约束条件下最优解

C4 烯烃 收率	温度	Co 负载 量	Co/SiO ₂ 的质量	HAP 的 质量	乙醇浓度	催化剂质 量比
0.5078	349.9780	1.9669	217.2945	177.0551	1.6317	1.2273

5.4.1 问题四模型的建立与求解

■ 化学实验方案设计的思路

观察附件 1 的性能数据表我们可以发现部分催化剂组合的实验不够充足, 根据以上问题的模型以及得到的结论我们可以设计实验来增加实验次数, 以提高模型的准确性, 也可以设计实验来验证模型的准确性。我们以 C4 烯烃收率最优的情况为目的来进行分析设计。

■ 5 个实验方案的选取

表 5.3: 实验设计表

	温度 (°C)	Co 负载 量(wt%)	Co/siO ₂ 的 质量(mg)	HAP 的 质量 (mg)	Co/siO ₂ 和 HAP 装料 比	乙醇浓度 (ml/min)
最优组	422	0.57	260	167	1.56	1.22
方案 1	400	1.00	200	200	1.00	1.68
方案 2	400	2.00	200	200	1.00	1.68
方案 3	430	0.62	253	170	1.49	1.30
方案 4	420	0.50	265	160	1.66	1.15
方案 5	425	0.53	255	165	1.55	1.20

■ 方案的设计理由

实验方案 1、2 是针对附件 1 的催化剂组合 A1、A2 进行设计的，因为在附件 1 的实验中我们可以发现 C4 烯烃收率在催化剂组合 A3 且温度为 400°C 时是最高的，而有些催化剂组合只做了在 350°C 以内的实验，故我们对 350°C 时 C4 烯烃的收率进行分析，得出催化剂组合 A2 的 C4 收率比催化剂组合 A3 的高，在 350°C 以内时催化剂组合 A1 的部分 C4 烯烃收率的值相对比催化剂组合 A3 高，故增加两组在催化剂组合 A1、A2 的条件下温度为 450°C 的实验。

实验方案 3、4、5 的设计是对催化剂的最优组合值取小幅度的波动。最后我们可以将得到的实验结果代入前三问的模型再次进行分析，也可以得出问题三得到的最优催化剂组合是否正确，对第三问的模型进行了可行性与准确性的检验。

六、模型的评价与推广

6.1 优缺点分析

优点：

1. 连续数据，正态分布利用 pearson 相关系数分析相对恰和且效率较高
2. BP 神经网络具有较强的非线性映射能力，能够通过学习自动提取输入、输出数据间的“合理规则”，然后将学习成果应用于新知识，即使系统受到局部损伤时还是可以正常工作的，具有一定的容错能力。
3. 结合遗传算法分析最优实验条件，在搜索过程中不容易陷入局部最优，即使在所定义的适应度函数是不连续的，非规则的或有噪音的情况下，也能以很大的概率找到全局最优解。

缺点：

1. BP 神经网络算法收敛速度较慢，且由于网络结构选择不一，可能造成网络性能低，容错性下降或造成网络不收敛，网络模型的逼近和推广能力与学习样本的典型性密切相关，而从问题中选取典型样本实例组成训练集比较困难。
2. 全局最优解的准确性完全依赖于神经网络模型的精确度。
3. 遗传算法在编码时容易出现不规范不准确的问题，因为可能需要考虑对不可行解采用阈值，工作量和求解时间也会大大增加，且容易出现过早收敛的问题。

6.2 模型的改进与推广

改进：由于 BP 神经网络模型的精确度直接影响了全局最优解的值，所以神经网络的精确度仍有待进一步提高。

推广：本文所用到的模型也可以用来研究生产规划、通行网络设置、背包问题、生命的遗传进化等问题。

七、参考文献

- [1] 吕绍沛.乙醇偶合制备丁醇及 C₄ 烯烃[D].大连理工大学,2018.
- [2] 司守奎,孙玺菁.数学建模算法与应用[M].3 版.北京:国防工业出版社,2021.4:432.

附 录

一、第一问源代码

三次样条插值的 MATLAB 源代码

```
clc,clear;
name = '插值数据及图表.xlsx';
% 乙醇转化率插值
a=xlsread(name,'乙醇转化率','G2:Q7');
a(4,:)=[];
x=[250 275 300 350 400];
new_x = linspace(250,400,7);
new_y = zeros(7,size(a,2));
for i=1:size(a,2)
    test = spline(x,a(:,i),new_x);
    new_y(:,i) = test';
end
test_y = new_y(4,:);
xlswrite(name,test_y,'乙醇转化率','G5');
a = xlsread(name,'乙醇转化率','D2:V7');
b = xlsread(name,'乙醇转化率','B2:C6');
x1 = [250 275 300 325 350 400];
x2 = [250 275 300 325 350 400 450];
new_x2 = linspace(250,450,201);
new_x1 = linspace(250,400,151);
new_y = zeros(151,size(a,2)-1);
for i=1:size(a,2)
    if i==1
        test = a(:,i);
        test = [test;86.4];
        test_y = spline(x2,test,new_x2);
        xlswrite(name,test_y,'转化率插值数据','D2');
    else
        new_y(1:151,i - 1) = spline(x1,a(:,i),new_x1);
    end
end
xlswrite(name,new_y,'转化率插值数据','E2');
xlswrite(name,new_x2,'转化率插值数据','A2');
x = [250 275 300 325 350];
new_x = linspace(250,350,101);
new_y = zeros(101,2);
new_y(:,1) = spline(x,b(:,1),new_x);
new_y(:,2) = spline(x,b(:,2),new_x);
xlswrite(name,new_y,'转化率插值数据','B2');

% C4 烯烃插值，与上部分代码基本相同
a=xlsread(name,'C4 烯烃选择性','G2:Q7');
a(4,:)=[];
```

```

x=[250 275 300 350 400];
new_x = linspace(250,400,7);
new_y = zeros(7,size(a,2));
for i=1:size(a,2)
    test = spline(x,a(:,i),new_x);
    new_y(:,i) = test';
end
test_y = new_y(4,:);
xlswrite(name,test_y,'C4 烯烃选择性','G5');
a = xlsread(name,'C4 烯烃选择性','D2:V7');
b = xlsread(name,'C4 烯烃选择性','B2:C6');
x1 = [250 275 300 325 350 400];
x2 = [250 275 300 325 350 400 450];
new_x2 = linspace(250,450,201);
new_x1 = linspace(250,400,151);
new_y = zeros(151,size(a,2)-1);
for i=1:size(a,2)
    if i==1
        test = a(:,i);
        test = [test;49.9];
        test_y = spline(x2,test,new_x2);
        xlswrite(name,test_y,'C4 烯烃选择性插值数据','D2');
    else
        new_y(1:151,i - 1) = spline(x1,a(:,i),new_x1);
    end
end
xlswrite(name,new_y,'C4 烯烃选择性插值数据','E2');
xlswrite(name,new_x2,'C4 烯烃选择性插值数据','A2');
x = [250 275 300 325 350];
new_x = linspace(250,350,101);
new_y = zeros(101,2);
new_y(:,1) = spline(x,b(:,1),new_x);
new_y(:,2) = spline(x,b(:,2),new_x);
xlswrite(name,new_y,'C4 烯烃选择性插值数据','B2');

```

计算 Pearson 相关系数的 MATLAB 源代码

```

clc,clear;
name = '插值数据及图表.xlsx';
sheet1 = '转化率插值数据';
sheet2 = 'C4 烯烃选择性插值数据';
sheet = sheet1; % 更改这里可以计算另一个相关系数
% 温度与乙醇选择性 pearson 相关系数的计算
a12 = xlsread(name,sheet,'B2:C102');
a3 = xlsread(name,sheet,'D2:D202');
a = xlsread(name,sheet,'E2:V400');
a1 = a12(:,1);
a2 = a12(:,2);
tem12 = 250:350;

```

```

tem3 = 250:450;
tem = 250:400;
% qqplot(a1)
% 进行正态分布检验
n = 21; % 数据的列数
H = zeros(1,n); % 初始化节省时间和消耗
P = zeros(1,n);
[H(1),P(1)] = jbtest(a1,0.05);
[H(2),P(2)] = jbtest(a2,0.05);
[H(3),P(3)] = jbtest(a3,0.05);
for i = 1:n-3
    [h,p] = jbtest(a(:,i),0.05);
    H(i+3)=h;
    P(i+3)=p;
end
disp(['在 95%的置信区间上有',num2str(sum(H)), '种催化剂组合通过了 JB 检验'])
% disp(P)
% 计算 pearson 相关系数
R = zeros(1,n);
p = zeros(1,n);
[test1,test2] = corrcoef([a1,tem12]);
R(1) = test1(1,2);
p(1) = test2(1,2);
[test1,test2] = corrcoef([a2,tem12]);
R(2) = test1(1,2);
p(2) = test2(1,2);
[test1,test2] = corrcoef([a3,tem3]);
R(3) = test1(1,2);
p(3) = test2(1,2);
for i = 1:n-3
    [test1,test2] = corrcoef([a(:,i),tem]);
    R(i+3) = test1(1,2);
    p(i+3) = test2(1,2);
end
disp('相关系数如下: ')
disp(R)
% disp('p 值 (双侧检验): ')
% disp(p)

```

三次样条插值的 MATLAB 源代码 2 (稀疏插值)

```

clc,clear;
name = '插值数据及图表.xlsx';
% 乙醇转化率插值
a=xlsread(name,'乙醇转化率','G2:Q7');
a(4,:)=[];
x=[250 275 300 350 400];
new_x = linspace(250,400,7);
new_y = zeros(7,size(a,2));

```

```

for i=1:size(a,2)
    test = spline(x,a(:,i),new_x);
    new_y(:,i) = test';
end
test_y = new_y(4,:);
xlswrite(name,test_y,'乙醇转化率','G5');
a = xlsread(name,'乙醇转化率','D2:V7');
b = xlsread(name,'乙醇转化率','B2:C6');
x1 = [250 275 300 325 350 400];
x2 = [250 275 300 325 350 400 450];
new_x2 = linspace(250,450,41);
new_x1 = linspace(250,400,31);
new_y = zeros(31,size(a,2)-1);
for i=1:size(a,2)
    if i==1
        test = a(:,i);
        test = [test;86.4];
        test_y = spline(x2,test,new_x2);
        xlswrite(name,test_y,'转化率插值数据 2','D2');
    else
        new_y(1:31,i - 1) = spline(x1,a(:,i),new_x1);
    end
end
xlswrite(name,new_y,'转化率插值数据 2','E2');
xlswrite(name,new_x2,'转化率插值数据 2','A2');
x = [250 275 300 325 350];
new_x = linspace(250,350,21);
new_y = zeros(21,2);
new_y(:,1) = spline(x,b(:,1),new_x);
new_y(:,2) = spline(x,b(:,2),new_x);
xlswrite(name,new_y,'转化率插值数据 2','B2');

% C4 烯烃插值， 与上部分代码基本相同
a=xlsread(name,'C4 烯烃选择性','G2:Q7');
a(4,:)=[];
x=[250 275 300 350 400];
new_x = linspace(250,400,7);
new_y = zeros(7,size(a,2));
for i=1:size(a,2)
    test = spline(x,a(:,i),new_x);
    new_y(:,i) = test';
end
test_y = new_y(4,:);
xlswrite(name,test_y,'C4 烯烃选择性','G5');
a = xlsread(name,'C4 烯烃选择性','D2:V7');
b = xlsread(name,'C4 烯烃选择性','B2:C6');
x1 = [250 275 300 325 350 400];
x2 = [250 275 300 325 350 400 450];

```



```

new_x2 = linspace(250,450,41);
new_x1 = linspace(250,400,31);
new_y = zeros(31,size(a,2)-1);
for i=1:size(a,2)
    if i==1
        test = a(:,i);
        test = [test;49.9];
        test_y = spline(x2,test,new_x2);
        xlswrite(name,test_y,'C4 烯烃选择性插值数据 2','D2');
    else
        new_y(1:31,i - 1) = spline(x1,a(:,i),new_x1);
    end
end
xlswrite(name,new_y,'C4 烯烃选择性插值数据 2','E2');
xlswrite(name,new_x2,'C4 烯烃选择性插值数据 2','A2');
x = [250 275 300 325 350];
new_x = linspace(250,350,21);
new_y = zeros(21,2);
new_y(:,1) = spline(x,b(:,1),new_x);
new_y(:,2) = spline(x,b(:,2),new_x);
xlswrite(name,new_y,'C4 烯烃选择性插值数据 2','B2');

```

画三维折线图的 MATLAB 源代码

```

clc,clear,close;
name = '插值数据及图表.xlsx';
sheet1 = '转化率插值数据 2';
sheet2 = 'C4 烯烃选择性插值数据 2';
zlab1 = '乙醇转化率';
zlab2 = 'C4 烯烃选择性';
% 更改下方 sheet、zlab 指向绘制不同三维图
sheet = sheet2;
zlab = zlab2;
% 不同催化剂组合下 C4 烯烃选择性与温度的三维图
a1 = xlsread(name,sheet,'B2:B22');
a2 = xlsread(name,sheet,'C2:C22');
a3 = xlsread(name,sheet,'D2:D42');
a = xlsread(name,sheet,'E2:V32');
n=21;
x1 = linspace(250,350,21);
x2 = x1;
x3 = linspace(250,450,41);
x = linspace(250,400,31);
figure(1)
surf([],[],[])
xlabel('温度')
ylabel('催化剂组合')
zlabel(zlab)
hold on;

```

```

plot3(x1,ones(1,21),a1);
plot3(x2,ones(1,21)*2,a2);
plot3(x3,ones(1,41)*3,a3);
for i=1:4
    plot3(x,ones(1,31)*(i+3),a(:,i));
end
yticks(1:7)
yticklabels({'A1','A2','A3','A4','A5','A6','A7'});
figure(2)
surf([],[],[])
xlabel('温度')
ylabel('催化剂组合')
zlabel(zlab)
hold on;
for i=1:7
    plot3(x,ones(1,31)*(i),a(:,i+4));
end
yticks(1:7)
yticklabels({'A8','A9','A10','A11','A12','A13','A14'});
figure(3)
surf([],[],[])
xlabel('温度')
ylabel('催化剂组合')
zlabel(zlab)
hold on;
for i=1:7
    plot3(x,ones(1,31)*(i),a(:,i+11));
end
yticks(1:7)
yticklabels({'B1','B2','B3','B4','B5','B6','B7'});

```

二、第二问源代码

训练神经网络的 MATLAB 源代码

```

%% 神经网络
clc,clear;
name = '神经网络训练集.xlsx';
a = xlsread(name,'sheet1','A2:I2971');
a = a';
in = a(1:6,:); out = a(9,:);
% 选出部分数据作为测试集
i = 1:40:2971;
in_test = in(:,i); out_test = out(i);
% in(:,i) = []; out(i) = [];
in_train = in; out_train = out;
% 归一化输入输出训练数据
[shuru, ps] = mapminmax(in_train);
[shuchu, ts] = mapminmax(out_train);
% 设置模型参数
a = 2; l = size(out_test, 2);

```

```

while a > 0.002
    net = newff(shuru,shuchu,6,{'tansig','tansig','tansig'},'trainlm');
    net.trainParam.goal = 0.00001; % 训练目标最小误差，这里设置为 0.00001
    net.trainParam.epochs = 1000; % 训练次数，这里设置为 10000 次
    net.trainParam.lr = 0.02; % 学习速率，这里设置为 0.02
    % 训练
    net = train(net, shuru, shuchu);
    % 测试
    load net
    shuru_test = mapminmax('apply', in_test, ps); % 归一化测试数据
    an = sim(net, shuru_test); % 预测测试数据
    shuchu_test = mapminmax('reverse', an, ts); % 恢复数据
    deltal = shuchu_test - out_test; % 计算误差
    a = sum(abs(detal)) / l;
end
% 绘图
figure; hold on; grid on;
plot(shuchu_test, 'b*-')
plot(out_test, 'ro-')
plot(detal, 'square', 'MarkerFaceColor', 'b')
legend('乙醇预测收率', '乙醇期望收率', '误差')
xlabel('数据组数')
ylabel('乙醇收率')
b = sqrt(detal * deltal' / l);
disp(['-----', num2str(l), '个测试样例误差如下-----'])
disp(['平均绝对误差为: ', num2str(a)])
disp(['均方根误差为: ', num2str(b)])

% 多次模拟，保存误差最小的模型

```

关于神经网络灵敏度分析的 MATLAB 源代码

```

%% 神经网络模型的灵敏度分析
clc,clear;
load ps
load ts
load net
name = '控制变量.xlsx';
sheet = {'温度','Co 载重量','装料比','乙醇浓度'};
% 关于温度的灵敏度分析
figure(1)
subplot(2,2,1)
a = xlsread(name, sheet{1}, 'A1:T6');
shuru_test = mapminmax('apply', a, ps);
an = sim(net, shuru_test); % 预测测试数据
shuchu_test = mapminmax('reverse', an, ts);
plot(a(1,:),shuchu_test)
hold on

```

```

grid on
xlabel(sheet{1})
ylabel('乙醇收率')
% 关于 Co 载重量的灵敏度分析
subplot(2,2,2)
a = xlsread(name, sheet{2}, 'A1:T6');
shuru_test = mapminmax('apply', a, ps);
an = sim(net, shuru_test); % 预测测试数据
shuchu_test = mapminmax('reverse', an, ts);
plot(a(2,:),shuchu_test)
hold on
grid on
xlabel(sheet{2})
ylabel('乙醇收率')
% 关于装料比的灵敏度分析
subplot(2,2,3)
a = xlsread(name, sheet{3}, 'A1:T6');
shuru_test = mapminmax('apply', a, ps);
an = sim(net, shuru_test); % 预测测试数据
shuchu_test = mapminmax('reverse', an, ts);
plot(a(6,:),shuchu_test)
hold on
grid on
xlabel(sheet{3})
ylabel('乙醇收率')

% 关于乙醇浓度的灵敏度分析
subplot(2,2,4)
a = xlsread(name, sheet{4}, 'A1:T6');
shuru_test = mapminmax('apply', a, ps);
an = sim(net, shuru_test); % 预测测试数据
shuchu_test = mapminmax('reverse', an, ts);
plot(a(5,:),shuchu_test)
hold on
grid on
xlabel(sheet{4})
ylabel('乙醇收率')

```

三、第三问源代码

遗传算法的“适应度函数”

```

function shuchu_test = f(a,net,ps,ts)
shuru_test = mapminmax('apply', a, ps);
an = sim(net, shuru_test); % 预测测试数据
shuchu_test = mapminmax('reverse', an, ts);
end

```

基于神经网络的遗传算法（无约束）

```

clc,clear;

```

```

load net
load ps
load ts
popsize = 50; % 种群规模
elite = 0.2; % 优势比例
lowerbound = [200 0 0 0 0];
upperbound = [500 6 300 300 3];
x = [400 1 200 200 0.9];
d = [10 0.2 10 10 0.1]; % 步长
% 初始化种群
pop = x;
for i=1:49
    delta = d.*(2*rand(1,5)-1);
    test = x + delta;
    pop = [pop;test];
end

% 开始迭代
maxiter = 5;
iter = 1;
while iter < maxiter
    alphi = zeros(1,50);
    for i = 1:50
        a = pop(i,:);
        a(6) = a(3)/a(4);
        alphi(i) = f(a',net,ps,ts);
    end
    [alphi_sort, index] = sort(alphi,'descend');
    pop = pop(index(1:popsize*elite),:);
    % 前五个个体各生成 8 个子代
    for i = 1:5
        for j = 1:8
            delta = d.*(2*rand(1,5)-1);
            test = pop(i,:) + delta;
            pop = [pop;test];
        end
    end
    iter = iter + 1;
end
a = pop(1,:);
a(6) = a(3)/a(4);
res = f(a',net,ps,ts)
a'

```

基于神经网络的遗传算法（有约束）

```

%% 基于神经网络的遗传算法（温度<350）
clc,clear;
load net
load ps

```

```

load ts
popsize = 50; % 种群规模
elite = 0.2; % 优势比例
lowerbound = [200 0 0 0 0];
upperbound = [500 6 300 300 3];
x = [345 2 200 200 1.68];
d = [10 0.2 10 10 0.1]; % 步长
% 初始化种群
pop = x;
while size(pop,1) < 50
    deltal = d.*(2*rand(1,5)-1);
    test = x + deltal;
    if test(1) < 350
        pop = [pop;test];
    end
end

% 开始迭代
maxiter = 4;
iter = 1;
while iter < maxiter
    alphi = [];
    for i = 1:50
        a = pop(i,:);
        a(6) = a(3)/a(4);
        alphi(i) = f(a',net,ps,ts);
    end
    [alphi_sort, index] = sort(alphi,'descend');
    pop = pop(index(1:popsize*elite),:);
    % 前五个每个生成 8 个子代
    for i = 1:5
        while size(pop,1) < popsize*elite+i*8
            deltal = d.*(2*rand(1,5)-1);
            test = pop(i,:) + deltal;
            if test(1) < 350
                pop = [pop;test];
            end
        end
    end
    iter = iter + 1;
end
a = pop(1,:);
a(6) = a(3)/a(4);
f(a',net,ps,ts)
a'

```


四、Pearson 相关系数表

乙醇转化率、C4 烯烃选择性与温度的 pearson 相关系数表

催化剂组合	乙醇转化率	C4 烯烃选择性
A1	0.981831	0.908644
A2	0.995619	0.910632
A3	0.988151	0.991835
A4	0.99856	0.971608
A5	0.96277	0.974952
A6	0.99299	0.836318
A7	0.998824	0.970953
A8	0.97786	0.994668
A9	0.916287	0.998156
A10	0.915261	0.859139
A11	0.895007	0.987838
A12	0.964852	0.984135
A13	0.934342	0.991464
A14	0.967087	0.957459
B1	0.963644	0.987669
B2	0.923623	0.988129
B3	0.90839	0.952887
B4	0.91713	0.904137
B5	0.928248	0.988455
B6	0.948462	0.98896
B7	0.94829	0.997424

五、支撑材料内容组成

文件夹	文件名	主要功能/用途
Code	p_1_1.m	三次样条插值
	p_1_2.m	Pearson 相关系数的计算
	p_1_3.m	三次样条插值 2（稀疏插值）
	p_1_4.m	三维折线图的绘制
	p_2_1.m	BP 神经网络模型的训练
	p_2_2.m	BP 神经网络模型的灵敏度分析
	p_3_1.m	遗传算法无约束条件下搜寻全局最优解
	p_3_2.m	遗传算法有约束条件下搜寻全局最优解
	f.m	遗传算法的适应度函数
Data	net.mat	训练好的 BP 神经网络模型
	ps.mat	归一化输入处理统一设置
	ts.mat	归一化输出处理统一设置
Img	P_1 三维图	问题一的 6 张三维折线图
	亿图源代码	亿图源文件
	BP 神经网络测试集误差分析图.jpg	BP 神经网络测试集误差分析图
	BP 神经网络流程图.jpg	BP 神经网络流程图
	产物选择性变化的分析.jpg	产物选择性变化的分析
	灵敏度分析.jpg	BP 神经网络灵敏度分析
	问题分析.png	问题分析流程图
	遗传算法流程图.jpg	遗传算法流程图
	乙醇转化率分析.jpg	乙醇转化率分析
Table	pearson 相关系数.xlsx	pearson 相关系数的计算结果
	插值数据及图表.xlsx	三次样条插值的结果
	控制变量.xlsx	灵敏度分析所用的数据
	神经网络训练集.xlsx	训练 BP 神经网络模型的数据