

2024 年第九届“数维杯”大学生 数学建模挑战赛论文

题 目 基于统计分析的生物质和煤共热解问题的研究

摘 要

随着全球能源需求的不断增长和对可再生能源的追求，生物质和煤共热解作为一种经济的能源转化技术备受关注，通过分析生物质热解产物的特性和化学反应机理，对研究提高能源利用效率、促进资源综合利用和确保能源安全至关重要。

针对问题一，我们首先建立了 **P 值检验**模型，根据 Pearson 相关系数与箱型图，综合探究分析了 INS 与热解产率之间的相关性，以及产物产率的离群值水平。之后进行前置性的正态性检验和 Levene 方差齐性检验，并建立**单因素方差分析**模型进行求解，进行了 F 检验，结合统计量结果 P 值及 F 值，分析出 INS 对焦油产率存在显著影响，对水产率、焦渣产率没有显著性影响。

针对问题二，我们首先分析了生物质与煤的共热解交互效应机理，并运用 Python 统计库中的 OLS 函数，构建包括 INS、混合比例及其交互效应的线性回归模型。随后提出原假设，给出模型的拟合优度 R 方和校正 R 方，进行双因素方差分析求解结果，再通过回归模型得到参数估计结果，并对模型进行数据可视化。结果显示 INS 和混合比例之间存在交互效应，且对焦油、焦渣产量产率有明显的重要影响。

针对问题三，我们首先分析了共热解产物特性和组成，结合能量守恒，由阿伦尼乌斯方程确定热解速率方程式，考虑了热解协同效应。通过设定 2 个决策变量、4 个约束条件，并以**最大化焦油的加权产物产率**为目标建立最优化共热解配比模型，通过遗传算法迭代求解，确定了产物对应的权重因子占比以及共热解组合的最优化混合比例。

针对问题四，我们首先运用 **t 检验**分析了焦油、水以及正己烷可溶物在所有共热解组合中的产物实验值与理论值的差异，并发现产物焦渣的实验值与理论值的差异性在共热解组合不显著。通过计算有显著性差异的产物的差值比率，分析了具体体现差异的共热解混合比例主要为 5/100、20/100、50/100。

针对问题五，我们首先细化了带有共热解交叉项的产率计算公式，并通过构建 Keras 神经网络产率预测模型，得到产物焦油、水、HEX 以及焦渣的产率预测效果 MSE 为 0.0034，预测 R 方为 78.80%。最佳共热解混合比例下的焦油产率预测超过 13.261%，而 CS/HN（棉杆/淮南煤）、RH/HN（稻壳/淮南煤）、CS/SM（棉杆/神木煤）的产率预测结果对比其他共热解组较高。

为了增加模型的信心，我们针对五个问题的模型均进行了误差分析。

关键词： P 值检验 单、双因素方差分析 遗传算法 t 检验 Keras 神经网络

目 录

| | |
|---------------------------|----|
| 一、问题的介绍 | 1 |
| 1.1 问题的背景 | 1 |
| 1.2 有效性息的提取 | 1 |
| 1.3 问题的复述 | 1 |
| 二、问题分析 | 2 |
| 2.1 问题 1 的分析 | 2 |
| 2.2 问题 2 的分析 | 2 |
| 2.3 问题 3 的分析 | 2 |
| 2.4 问题 4 的分析 | 2 |
| 2.5 问题 5 的分析 | 2 |
| 三、模型假设 | 3 |
| 四、定义与符号说明 | 3 |
| 五、模型的建立与求解 | 3 |
| 5.1 数据预处理 | 3 |
| 5.1.1 附件 1 的数据预处理 | 3 |
| 5.1.2 附件 2 的数据预处理 | 4 |
| 5.2 问题 1 的模型建立与求解 | 6 |
| 5.2.1 思路分析 | 6 |
| 5.2.2 P 值检验模型的建立 | 6 |
| 5.2.3 P 值检验模型的求解 | 7 |
| 5.2.4 单因素方差分析模型的建立 | 8 |
| 5.2.5 单因素方差分析模型的求解 | 9 |
| 5.2.6 结果 | 10 |
| 5.3 问题 2 的模型建立与求解 | 11 |
| 5.3.1 思路分析 | 11 |
| 5.3.2 共热解交互效应机理分析 | 11 |
| 5.3.3 双因素方差分析模型的建立 | 12 |
| 5.3.3 双因素方差分析模型的求解 | 13 |
| 5.3.4 结果 | 16 |
| 5.4 问题 3 的模型建立与求解 | 16 |
| 5.4.1 思路分析 | 16 |
| 5.4.2 共热解产物特性和组成的分析 | 17 |
| 5.4.3 最优化共热解配比模型的建立 | 18 |
| 5.4.4 最优化共热解配比模型的求解 | 19 |

| | |
|-----------------------------------|----|
| 5.4.5 结果 | 21 |
| 5.5 问题 4 的模型建立与求解 | 22 |
| 5.5.1 思路分析 | 22 |
| 5.5.2 t 检验模型的建立 | 22 |
| 5.5.3 t 检验模型的求解 | 23 |
| 5.5.4 差异可视化模型的求解 | 24 |
| 5.5.5 结果 | 25 |
| 5.6 问题 5 的模型建立与求解 | 26 |
| 5.6.1 思路分析 | 26 |
| 5.6.2 基于 Keras 的神经网络预测模型的建立 | 26 |
| 5.6.3 基于 Keras 的神经网络预测模型的求解 | 28 |
| 5.6.4 结果 | 30 |
| 六、 模型的评价及优化 | 30 |
| 6.1 误差分析 | 30 |
| 6.1.1 针对于问题 1 的误差分析 | 30 |
| 6.1.2 针对于问题 2 的误差分析 | 31 |
| 6.1.3 针对于问题 3 的误差分析 | 31 |
| 6.1.4 针对于问题 4 的误差分析 | 31 |
| 6.1.5 针对于问题 5 的误差分析 | 31 |
| 6.2 模型的优点 | 32 |
| 6.3 模型的缺点 | 32 |
| 6.4 模型的推广 | 32 |
| 参考文献 | 33 |
| 附 录 | 34 |

一、问题的介绍

1.1 问题的背景

全球能源需求和可再生能源的兴起促进了对生物质与煤共热解技术的研究。生物质是植物和动物的有机物，而煤则是化石燃料。这两者可以通过高温无氧条件共热解，转化为气体、液体和固体产物，其中液体产物称为热解油。这一过程的产率和品质研究对提高能源效率和资源综合利用至关重要。某化工实验室使用多种生物质和中低阶煤进行共热解实验，探索不同原料对比对产物分布的影响，并研究了生物质主要成分纤维素的热解特性和化学反应。研究还包括了正己烷不溶物的产率和生物质与煤的协同效应，旨在为优化共热解产物的预测和利用建立数学模型提供数据支持。

1.2 有效性信息的提取

(1) **研究对象**是生物质（棉秆、稻壳、木屑、小球藻）和中低阶煤（淮南煤、神木煤、黑山煤、内蒙褐煤）。

(2) **研究方法**是利用管式干馏炉在温和热解条件下进行不同种类和配比原料的共热解实验。

(3) **实验条件**是初始温度为 600℃，升温速率为 5℃/分钟。

(4) **模型化合物**是微晶纤维素，用于代表生物质中的主要纤维素成分。

(5) **研究目的**是通过生物质和煤热解反应得到更多的煤焦油。

(6) **关注产物**是焦油、正己烷可溶物、水、焦渣。

1.3 问题的复述

待求解的问题 1：基于附件一数据，分析正己烷不溶物是否对焦油产率、水产率及焦渣产率产生显著影响，并结合图像加以说明。

待求解的问题 2：判断正己烷不溶物和混合比例是否存在交互效应影响热解产物产量，若存在，找出在哪些具体的热解产物上样品重量和混合比例的交互效应最为明显。

待求解的问题 3：根据附件一共热解产物特性和组成的数据，建立一个优化共热解混合比例的模型，以提高产物利用率和能源转化效率。

待求解的问题 4：根据附件二的数据，分析每种共热解组合的产物收率实验值与理论计算值是否存在显著性差异，若存在，则对不同共热解组合的数据进行子组分析，说明实验值和理论计算值之间的差异体现在哪些混合比例上。

待求解的问题 5：基于实验数据，对热解产物的产率构建模型进行预测。

二、问题分析

2.1 问题 1 的分析

问题 1 的研究目的是分析正己烷不溶物(INS)对热解产率（焦油产率、水产率、焦渣产率）的影响，其属于统计分析数学问题。根据附件中数据包括不同混合比例和 INS 含量下的热解产物产率，先提出假设，进行正态性检验和方差齐性检验，再构建 P 值检验及单因素方差分析模型来分析 INS 对各热解产物产率的影响，并通过可视化图直观展示 INS 对焦油产率、水产率和焦渣产率的显著影响结果。

2.2 问题 2 的分析

问题 2 在于研究正己烷不溶物（INS）和混合比例是否存在交互效应，及其对热解产物产量的影响，属于典型的多变量交互作用分析问题。根据附件数据中多种生物质和煤的组合、多种混合比例、不同热解产物的产量，本文采用双因素方差分析，建立回归模型，重点分析其他自变量 INS、配比数值以及它们之间的交互效应，探究热解产物产量的交互效应影响，以找出具体受样品重量和混合比例影响显著的热解产物。

2.3 问题 3 的分析

问题 3 的研究目的是通过最大化焦油的加权产物产率来确定生物质和煤共热解的最优化混合比例。这一过程涉及分析共热解产物的特性和组成，结合能量守恒，使用阿伦尼乌斯方程来确定热解速率。通过考虑热解的协同效应和基于实验数据确定的各产物的权重，使用遗传算法对 10 种混合组合进行优化，以提高能源利用效率和支持资源的综合利用。

2.4 问题 4 的分析

问题 4 的研究目的在于分析每种共热解组合的产物收率实验值与理论计算值的显著性差异，以及其体现差异的混合比例。首先将附件二进行数据预处理，数据满足正态分布特性后，再使用 t 检验分析焦油、水、HEX、焦渣分别对应的共热解组的产物收率差异性。然后通过产物差异百分比的大小，可视化分析实验值与理论计算值之间的差异具体体现在哪些混合比例上。

2.5 问题 5 的分析

问题 5 的研究目的是建立相应的模型，对热解产物产率进行预测。但是在热解问题中，多种生物质和煤的混合比例、不同种类的原料、温度和时间等因素可能对产物产率有复杂的影响，而热解数据包括原料种类、比例等多个维度的特征，因此结合附件一二

的数据，可考虑使用 Keras 神经网络捕捉这些变量间的相互作用，处理这种高维数据，找出潜在的模式预测出热解产物产率。

三、模型假设

- 假设 1: 假设实验数据足够充分且准确，INS 含量是影响产率的主要因素之一。
- 假设 2: 假设 INS 和混合比例是影响产率的主要因素，且两者之间存在可能的交互作用。
- 假设 3: 假设所有参数和系数可以通过实验数据准确估计，优化模型能够准确反映实际情况。
- 假设 4: 假设实验数据和理论计算值具有可比性，检验方法适用于数据分布。
- 假设 5: 假设数据集具有代表性，选择的预测模型能够准确捕捉数据中的模式。

四、定义与符号说明

| 符号定义 | 符号说明 |
|------------------|----------------------------|
| r | Pearson 相关系数 |
| ε | 回归方程的误差项 |
| \bar{X} | 样本数据均值 |
| μ | 均值 |
| σ | 方差 |
| H_0 | 原假设（具有显著性差异） |
| H_1 | 备择假设（没有显著性差异） |
| x_k | 第 k 组生物质和煤的混合的比例 |
| $f(t,\theta(t))$ | 混合比例为 t 的生物质与煤共热解的实际产物收率 |
| $F_{fitness}$ | 遗传算法的适应度函数 |
| $P(x)$ | 遗传算法的惩罚函数 |

五、模型的建立与求解

5.1 数据预处理

5.1.1 附件 1 的数据预处理

针对附件 1，我们发现 有 8 组单独热解的煤和生物质，有 10 组共热解煤和生物质组

合，其中正己烷不溶物（INS）、正己烷可溶物产率存在缺失值，为保证后续问题的正常研究，本文按照如下流程处理确保数据的准确性和可用性：

（1）数据清洗：首先对数据集中的列名进行清洗，统一括号的使用并去除列名两边的空格，确保后续引用列时不会因为列名的格式错误而造成问题。在缺失值填充时，使用前向填充 `fillna(method='ffill')`处理数据中的缺失值。这种方法可有效处理连续测量数据，如果一个试样的某些参数在连续测量中未记录，则用前一个有效记录的同一参数填充。

（2）数据聚合：在选择数值列时，定义列表 `numeric_columns`，包含所有需要进行聚合计算的数值列。在分组和平均时，按照“时间”、“试样”和“配比”这三个字段对数据进行分组，然后计算每个分组中数值列的平均值。这一步骤用于简化数据，把重复的测量值合并为一个平均值，方便进行进一步分析或比较。

（3）数据导出：将处理后的聚合数据保存为 CSV 文件，使其可以进行后续分析或共享。

| 时间 | 试样 | 配比 | 样品g | 焦油(Char) | 水(Water)ml | 正己烷不溶物(INS)g | 焦油产率 | 水产率 | 焦渣产率 | 正己烷可溶物产率 |
|----------|----------------|--------|----------|-------------|-------------|--------------|-------------|-------------|-------------|-------------|
| 20130224 | 小球藻/淮南煤(GA/HN) | 10/100 | 9.93325 | 1.215953529 | 0.79681757 | 0.3321 | 0.122416836 | 0.080217142 | 0.716615377 | 0.09065 |
| 20131121 | 棉杆/淮南煤(CS/HN) | 5/100 | 11.73715 | 1.48435 | 0.72 | 0.45285 | 0.126400525 | 0.0613 | 0.735886397 | 0.10055 |
| 20131123 | 棉杆(CS) | 100 | 5.4926 | 1.00125 | 1.435 | 0.28095 | 0.18229496 | 0.2608 | 0.313967384 | 0.08465 |
| 20131123 | 棉杆/淮南煤(CS/HN) | 10/100 | 10.296 | 1.3505 | 0.75 | 0.42895 | 0.131167505 | 0.073 | 0.719873061 | 0.09135 |
| 20131206 | 淮南煤(HN) | 100 | 10.3958 | 1.30055 | 0.585 | 0.4566 | 0.125111485 | 0.057922187 | 0.759276993 | 0.0809 |
| 20131207 | 棉杆/淮南煤(CS/HN) | 20/100 | 10.27125 | 1.20735 | 1 | 0.38115 | 0.11755086 | 0.0977 | 0.687995554 | 0.08675 |
| 20131207 | 棉杆/淮南煤(CS/HN) | 30/100 | 9.77835 | 1.2161 | 1.015 | 0.3469 | 0.124346178 | 0.1038 | 0.676546803 | 0.08675 |
| 20131208 | 棉杆/淮南煤(CS/HN) | 40/100 | 9.4619 | 1.20665 | 1.2 | 0.34465 | 0.127528121 | 0.1268 | 0.630603307 | 0.09025 |
| 20131209 | 棉杆/淮南煤(CS/HN) | 50/100 | 9.84145 | 1.29145 | 1.315 | 0.36835 | 0.131218137 | 0.1366 | 0.6088505 | 0.091 |
| 20131212 | 稻壳(RH) | 100 | 5.23985 | 1.0202 | 1.13 | 0.59075 | 0.194664891 | 0.215659227 | 0.404364047 | 0.1864 |
| 20131212 | 稻壳/淮南煤(RH/HN) | 5/100 | 10.40675 | 1.1249 | 0.84 | 0.6161 | 0.107900432 | 0.080732137 | 0.740579966 | 0.09115 |
| 20131213 | 稻壳/淮南煤(RH/HN) | 10/100 | 10.2123 | 1.1049 | 0.843333333 | 0.333166667 | 0.10813911 | 0.082519685 | 0.731241611 | 0.072966667 |
| 20131213 | 稻壳/淮南煤(RH/HN) | 20/100 | 10.0649 | 1.1406 | 0.95 | 0.34965 | 0.113317367 | 0.094387701 | 0.708645299 | 0.0751 |
| 20131215 | 稻壳/淮南煤(RH/HN) | 30/100 | 10.35895 | 1.238331071 | 0.989818929 | 0.3971 | 0.119510134 | 0.095552052 | 0.696943569 | 0.0779 |
| 20131215 | 稻壳/淮南煤(RH/HN) | 50/100 | 10.4189 | 1.304640738 | 1.320159262 | 0.45555 | 0.125225087 | 0.126708123 | 0.63891044 | 0.0766 |
| 20131221 | 小球藻/淮南煤(GA/HN) | 5/100 | 10.68195 | 1.291365866 | 0.735 | 0.29465 | 0.120885258 | 0.0687 | 0.736273587 | 0.10195 |
| 20131222 | 小球藻(GA) | 100 | 5.3628 | 2.12355 | 0.655 | 0.7887 | 0.395980167 | 0.122205157 | 0.292393047 | 0.1769 |
| 20131226 | 小球藻/淮南煤(GA/HN) | 20/100 | 9.9594 | 1.373454068 | 0.9015226 | 0.37545 | 0.137903544 | 0.090519808 | 0.68087017 | 0.0944 |
| 20131228 | 小球藻/淮南煤(GA/HN) | 30/100 | 9.84875 | 1.465728215 | 1 | 0.37265 | 0.148822991 | 0.101793693 | 0.651013157 | 0.10705 |

图 1 附件 1 数据预处理后的部分结果

5.1.2 附件 2 的数据预处理

（1）数据分类合并，附件 2 的数据包括了 10 种共热解组合的产生的 4 种产物的实验值和理论计算值数据，但为了后续问题 4.5 的研究的需要，我们使用 Excel 中的 INDEX 函数精确查找，匹配得到 10 个.xlsx 文件，其中包含焦油、水、正己烷可溶物、焦渣的实验值和理论计算值。

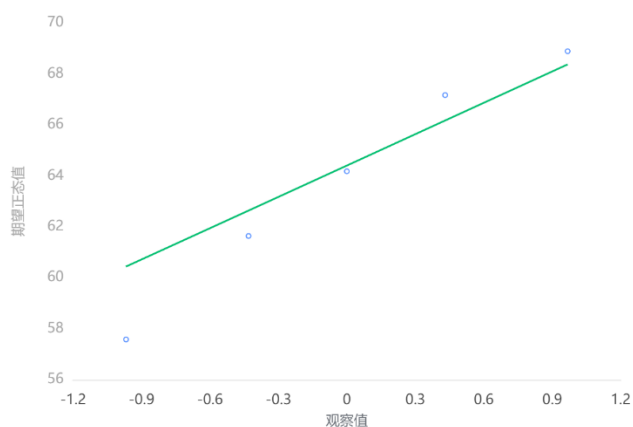
| 混合比例 | Tar实验值 | Tar理论计算值 | HEX实验值 | HEX理论计算值 | Water实验值 | Water理论计算值 | Char实验值 | Char理论计算值 |
|--------|--------|----------|--------|----------|----------|------------|---------|-----------|
| 5/100 | 15.13 | 16.18 | 10.04 | 11.45 | 7.96 | 5.73 | 67.32 | 68.89 |
| 10/100 | 15.49 | 16.6 | 10.59 | 11.52 | 8.09 | 6.55 | 65.75 | 67.17 |
| 20/100 | 15.64 | 17.3 | 10.3 | 11.68 | 9.17 | 7.95 | 63.77 | 64.18 |
| 30/100 | 16.37 | 17.89 | 10.62 | 11.82 | 10.16 | 9.13 | 62.6 | 61.64 |
| 50/100 | 16.93 | 18.82 | 10.46 | 12.09 | 13.05 | 10.97 | 55.57 | 57.58 |

图 2 附件 2 数据预处理后的部分结果（稻壳/淮南煤）

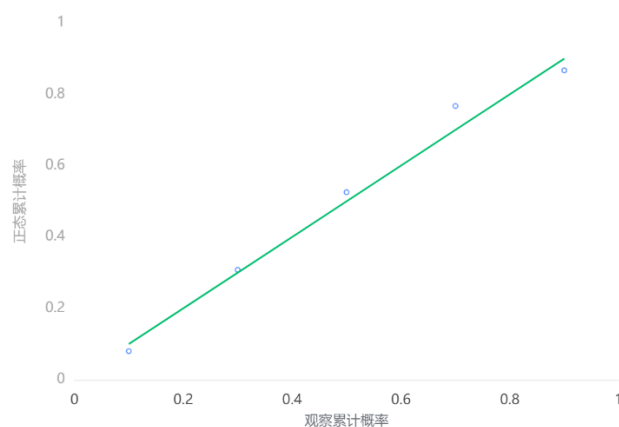
（2）正态分布性检验，在研究问题 4 时用到的 t 检验模型需要考虑 4 个产物数据的正态分布性，由于其数据的方差 σ 、均值 μ 以及样本均值 \bar{X} 已确定，则将数据表中的产物实验值和理论计算值（即统计量）按如下方式标准化：

$$\frac{\bar{X} - \mu}{\sigma / \sqrt{n}} \sim N(0, 1)$$

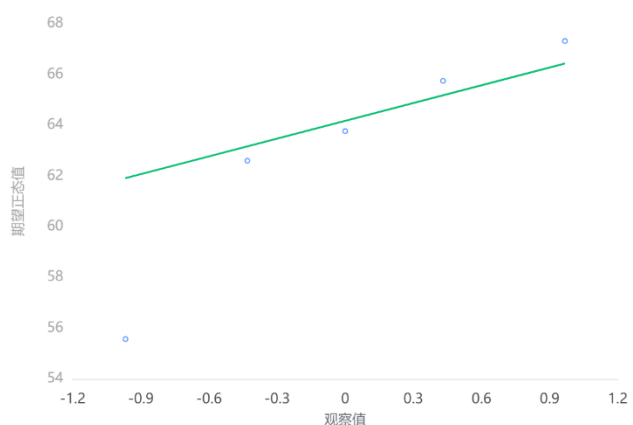
并将标准化的数据与原数据做 P-P 图和 Q-Q 图，以及检验正态性，通过 SPSS 绘图可视化数据分析，可得到如下部分结果（详细结果见附录 A）：



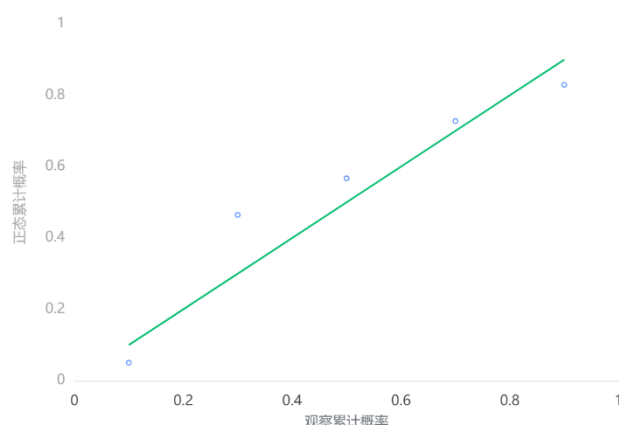
焦油理论计算值的 Q-Q 检验图



焦油理论计算值的 P-P 检验图



焦油实验值的 Q-Q 检验图



焦油实验值的 P-P 检验图

图 3 以焦油为例的理论计算值与实验值的正态性检验图

上图展示了以焦油实验值与理论计算值计算观测的累计概率（P）与正态累计概率（P）的拟合情况，其拟合程度较高，服从正态分布。对于焦油实验值与理论计算值计算的 Q-Q 图，散点与直线重合度较高，服从正态分布。同理，其他产物的 P-P 图与 Q-Q 图均符合正态性分布要求。

5.2 问题 1 的模型建立与求解

5.2.1 思路分析

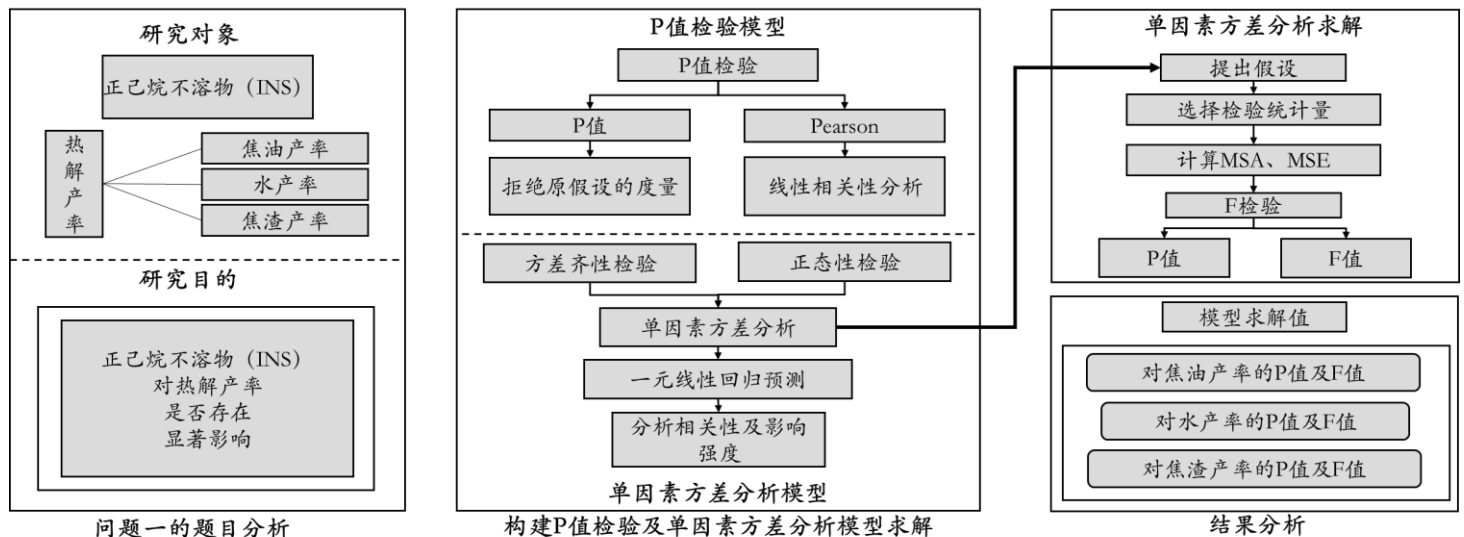


图 4 问题一的思维导图

根据问题一中的研究目标，探究正己烷不溶物（INS）对焦油产率、水产率和焦渣产率是否存在显著影响，本问采用构建 P 值检验及单因素方差分析模型进行求解，先分别作出 INS 对三个因素的不相关性原假设，通过 P 值检验，度量原假设并进行线性相关性分析，接着基于前置性的正态性检验和方差齐性检验进行单因素方差分析，通过 F 检验，得到 P 值及 F 值，然后进行一元线性回归预测，分析结果的相关性及影响强度，最后依据模型求解结果得出结论。

5.2.2 P 值检验模型的建立

➤ Pearson 相关系数

Pearson 相关系数用于衡量两个连续型变量之间线性相关程度的统计量，通常用符号 r 表示。具体计算方法如公式所示^[2]：

$$r = \frac{\sum_{i=1} (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1} (x_i - \bar{x})^2 \sum_{i=1} (y_i - \bar{y})^2}}$$

式中， x_i 和 y_i 分别表示待分析的两个变量样本， \bar{x} 和 \bar{y} 分别为变量样本的平均值， r 为相关系数，其取值范围为 $[-1, 1]$ ， -1 表示两变量之间存在显著的负相关关系， $+1$ 表示两变量之间存在显著的正相关关系， 0 表示两变量之间没有线性相关性^[2]。

➤ P 值定义

P 值是统计假设检验中的一个关键概念，它表示在原假设为真的情况下，观察到的样本结果或更极端结果出现的概率。P 值越小，意味着观察到的结果在原假设为真的情

况下出现的概率越小，根据小概率原理，就有理由拒绝原假设^[1]。P 值提供了对原假设的拒绝程度的度量，若 P 值小于预先设定的显著性水平（一般为 0.05），则可以拒绝原假设。

5.2.3 P 值检验模型的求解

针对问题一，本问通过 P 值检验模型分析正己烷不溶物（INS）对焦油产率、水产率和焦渣产率之间的显著性，首先提出：

原假设：INS 对焦油产率、水产率和焦渣产率之间无显著相关性，反之为备择假设。然后通过 Pearson 相关系数分析 INS 与焦油产率、水产率和焦渣产率之间的线性相关性，其结果如下：

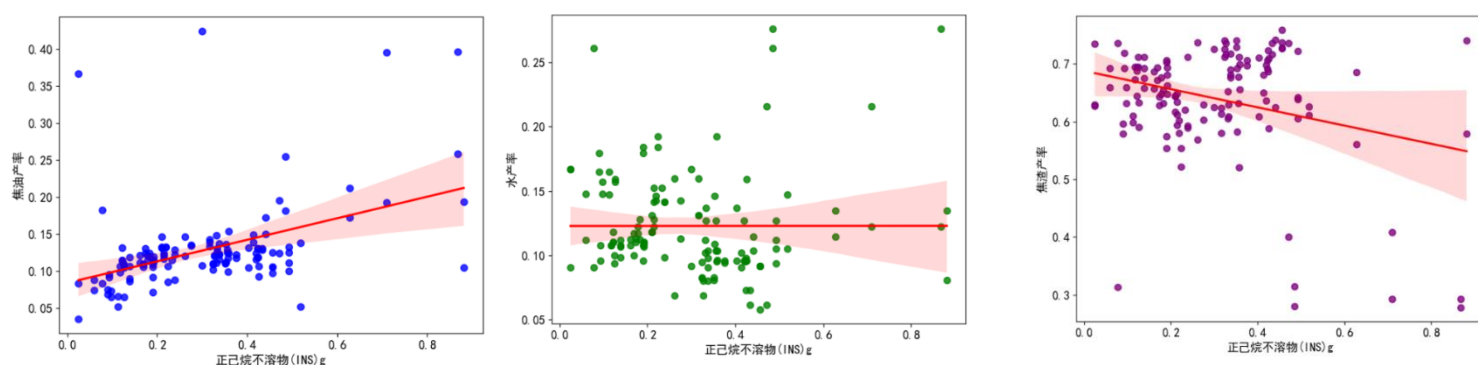


图 5 正己烷不溶物(INS)对热解产率的相关性

上图 5 展示了正己烷不溶物 INS 对产物的相关性程度，图 5 具体的 INS 与热解产率之间的相关系数及 P 值如下表 1 所示：

表 1 INS 与热解产率的相关系数及 P 值

| 参数 | 相关系数 | P 值 |
|-----------|--------|--------|
| INS 与焦油产率 | 0.592 | 0.0000 |
| INS 与水产率 | 0.007 | 0.9587 |
| INS 与焦渣产率 | -0.338 | 0.0067 |

根据上表中数据进行分析：

(1) INS 与焦油产率的 Pearson 相关系数为 0.592，表明 INS 与焦油产率之间的正相关性较为明显，而 P 值为 0，则说明观察到的结果在统计上是极其显著的，即可以自信地拒绝无相关性的假设。

(2) INS 与水产率的 Pearson 相关系数为 0.007，表明 INS 与水产率之间的相关非常弱，无法以线性关系表示，而 P 值为 0.9587，由于 P 值接近 1，则表示观察到的结果对原假设的二者无相关性支持程度很高，即在统计上无法拒绝无相关性的假设。

(3) INS 与焦渣产率的 Pearson 相关系数为-0.338，表明 INS 与焦渣产率之间存在一定负相关性，但不算特别显著，而 P 值为 0.0067，由于 P 值接近 0，则表示观察到的结果在统计上是极其显著的，即可以拒绝无相关性的假设。

通过上述 P 值检验分析，说明 INS 与焦油产率之间存在较显著的正相关性；INS 与

水产率之间基本无相关性；INS 与焦渣产率之间存在一定的负相关性。

5.2.4 单因素方差分析模型的建立

➤ 正态性检验

对于正态性检验，其用于检验数据是否符合正态分布。正态性检验通常用于评估数据的分布是否对称，以及数据是否呈现出典型的钟形曲线形状。

➤ Levene 方差齐性检验

Levene 方差齐性检验是通过比较各组数据的方差，如果各组的方差相近，则认为数据符合方差齐性假设，然后用于检验多个总体方差是否相等。

➤ 单因素方差分析

单因素方差分析（One-Way ANOVA）用于比较一个自变量对另一个因变量的影响是否产生显著影响，因此适用于研究本问中 INS 对热解产率的显著影响分析。对 INS 与焦油产率、水产率和焦渣产率之间的相关性进行单因素方差分析需要进行如下步骤^[3]，提出原假设 H_0 与备择假设 H_1 ：

原假设 H_0 ：因素对实验结果的影响比随机误差对实验结果的影响小

备择假设 H_1 ：因素对实验结果的影响比随机误差对实验结果的影响大。

为了检验原假设 H_0 是否成立，需要先选择合适的检验统计量，并且分别计算因素在不同水平的均值及全部观测值的总均值^[3]。同时为了构造检验统计量，首先需要分别计算总误差平方和（ SST ）、因素误差平方和（ SSA ）、随机误差平方和（ SSE ），即

$$SST = \sum_k \sum_{n_i}^{i=1, j=1} (x_{ij} - \bar{\bar{x}})^2, \quad SSA = \sum_k \sum_{n_i}^{i=1, j=1} (\bar{x}_i - \bar{\bar{x}})^2 = \sum_k n_i (\bar{x}_i - \bar{\bar{x}})^2, \quad SSE = \sum_k \sum_{n_i}^{i=1, j=1} (x_{ij} - \bar{x}_i)^2$$

由于三个误差平方和的大小都受到观测数据数目影响，为了消除影响，需要用各平方和计算结果除以各自所对应三个自由度 $n-1$ ， $k-1$ 与 $n-k$ ，最后得到均方。

SSA 的均方也被称为组间均方或组间方差，记为 MSA ， SSE 的均方也被称为组内均方或组内方差^[3]，记为 MSE ，则：

$$MSA = \frac{SSA}{k-1}, \quad MSE = \frac{SSE}{n-k}$$

已知组间均方与组内均方之比服从 F 分布，把 MSA 与 MSE 进行对比，得到所需的 F 检验统计量，如下所示：

$$F = \frac{MSA}{MSE} \sim F(k-1, n-k)$$

根据计算所得的检验统计量 F ，与查表所得的临界值 $F_\alpha(k-1, n-k)$ 进行比较分析：

若 $F > F_\alpha$ ，则拒绝原假设，即 $H_0: \mu_1 = \mu_2 = \mu_3 = \dots = \mu_k$ 的假设不成立，表明因素对实验结果的影响比随机误差对实验结果的影响大^[3]；

若 $F < F_\alpha$ ，则不能拒绝原假设 H_0 ，没有充分的证据证明因素对实验结果的影响比随机误差对实验结果的影响大^[3]。在进行统计决策时，还可以直接利用方差分析表中输出 P 值与显著性水平 α 进行比较，得出结论。

➤ 一元线性回归模型构建

基于一元线性回归方程来预测 INS 对热解产率是否产生显著影响，本问的一元线性回归模型可以表示为：

$$y = a_0 + a_1 x_1 + \varepsilon$$

式中， y 为因变量， x_1 为自变量， a_1 为回归系数； ε 为随机误差，假定 $\varepsilon \sim N(\mu, \sigma^2)$ 。

5.2.5 单因素方差分析模型的求解

对于 Levene 方差齐性检验，先计算 INS 与热解产率的每个样本组内的平方差，然后计算所有样本组的均值与总体均值的偏差的平方和除以自由度，得到 Levene 检验统计量。针对本问研究 INS 对热解产率的影响，选择显著性水平为 0.05，再通过比较 Levene 检验统计量与临界值，如果统计量大于临界值，则接受原假设，认为各组的方差相等，反之拒绝原假设，认为各组的方差不等。

对于焦油产率、水产率、焦渣产率的方差齐性检验 P 值，如下图所示：

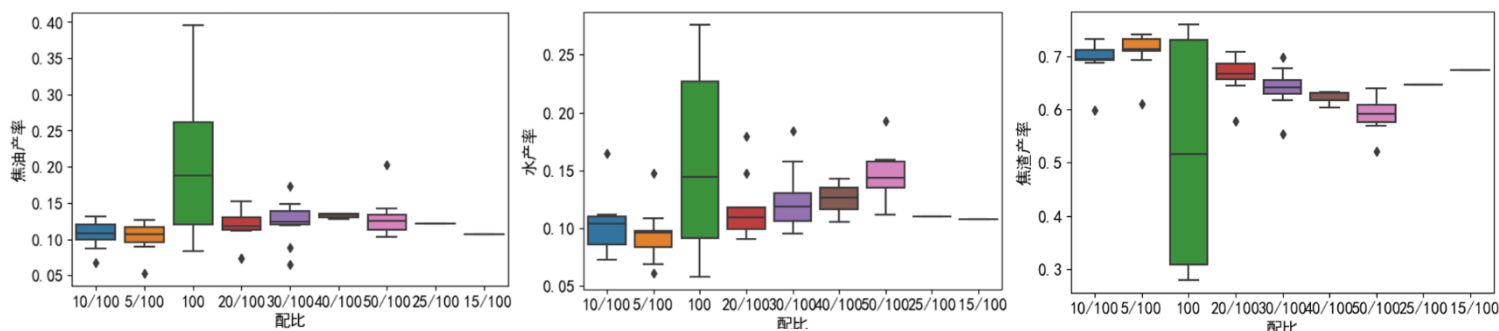


图 6 焦油产率、水产率、焦渣产率的方差齐性检验 P 值的箱型图

上图 6 展示了 INS 分别对焦油产率、水产率和焦渣产率的方差齐性检验 P 值的箱型图，100 共热解配比产物产率的离群值少，焦油产率和水产率的 20/100 共热解混合比例的离群值多，焦渣产率的 40/100 共热解混合比例的离群值少。而图 6 的具体方差齐性检验 P 值结果以及正态性检验如下表 2 所示：

表 2 INS 对热解产率的正态性检验和方差齐性检验 P 值

| 参数 | 正态性检验 P 值 | 方差齐性检验 P 值 |
|-----------|-----------------------|------------------------|
| INS 对焦油产率 | 0.0009999999999998899 | 0.00028264897778561705 |
| INS 对水产率 | 0.0024216596267245126 | 0.00006494470954431121 |
| INS 对焦渣产率 | 0.002410241281685365 | 0.00000000000000001745 |

结合箱型图、正态性检验 P 值与方差齐性检验 P 值，其中 INS 对焦油产率、水产率、焦渣产率的 P 值小于 0.05，小于显著性影响水平，各组方差相等，拒绝原假设，通过方差齐性检验，进一步通过单因素方差分析对 INS 的分组在不同水平上对应的三个热解产率是否有显著的影响进行评估，评估 INS 在不同水平上对焦油产率、水产率和焦渣产率是否存在显著均值差异，其结果如下图所示：

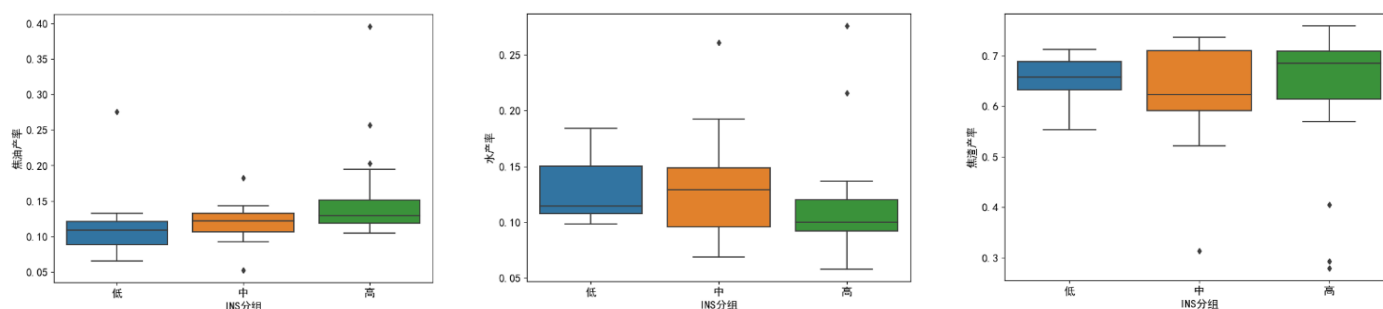


图 7 焦油产率、水产率和焦渣产率按 INS 分组的箱型图

从 INS 按含量值低、中、高分组，当正己烷不溶物(INS)的含量低时，产物（焦油、水、焦渣）产率的离群值少；当其含量高时，产物产率的离群值多；当其含量位于中等分组时，产物产率的离群值较少，此时焦油产率集中在 0.13244、水产率集中在 0.12113、焦渣产率集中 0.6623。其中，INS 对焦油产率的 F 值为 4.46, P 值为 0.0157，存在显著差异。INS 对水产率的 F 值为 1.08, P 值为 0.3446，没有显著差异。INS 对焦渣产率的 F 值为 0.35, P 值为 0.7088，没有显著差异。

接着基于一元线性回归方程模型对焦油产率、水产率和焦渣产率进行计算：

焦油产率 (y_1) : $y_1 = a_0 + a_1x_1 + \varepsilon_1$, $a_0=0.1879$, $a_1=0.0700$;

水产率 (y_2) : $y_2 = b_0 + b_1x_1 + \varepsilon_2$, $b_0=0.0017$, $b_1=0.1223$;

焦渣产率 (y_3) : $y_3 = c_0 + c_1x_1 + \varepsilon_3$, $c_0=-0.2087$, $c_1=0.7031$ 。

其中，焦油产率对 INS 的回归系数为 0.1879，这说明焦油产率对 INS 呈正相关，且具有一定的影响强度；水产率对 INS 的回归系数为 0.0017，由于其很接近 0，这说明水产率对 INS 具有较弱正相关性，且影响强度较小；焦渣产率对 INS 的回归系数为-0.2087，这说明焦渣产率对 INS 呈负相关，且具有一定的影响强度。

5.2.6 结果

结合回归方程以及 Pearson 相关性系数，焦油产率对 INS 的回归系数为 0.1879、Pearson 相关系数为 0.592，说明焦油产率对 INS 呈正相关；水产率对 INS 的回归系数为 0.0017、Pearson 相关系数为 0.007，由于其很接近 0，说明水产率对 INS 具有较弱正相关性，且影响强度较小；焦渣产率对 INS 的回归系数为-0.2087、Pearson 相关系数为-0.338，说明焦渣产率对 INS 呈负相关。

基于 P 值检验和单因素方差分析的结果，INS 对焦油产率的 P 值为 0.0157，小于预先设置的显著水平 0.05，而 F 值为 4.46，结合二者可以说明 INS 对焦油产率存在显著影响；INS 对水产率的 P 值为 0.3446 大于预先设定的显著水平 0.05，而 F 值为 1.08，结合二者可以说明 INS 对水产率没有显著影响；INS 对焦渣产率的 P 值为 0.7088 大于预先设定的显著水平 0.05，而 F 值为 0.35，结合二者可以说明 INS 对焦渣产率没有显著影响。

5.3 问题 2 的模型建立与求解

5.3.1 思路分析

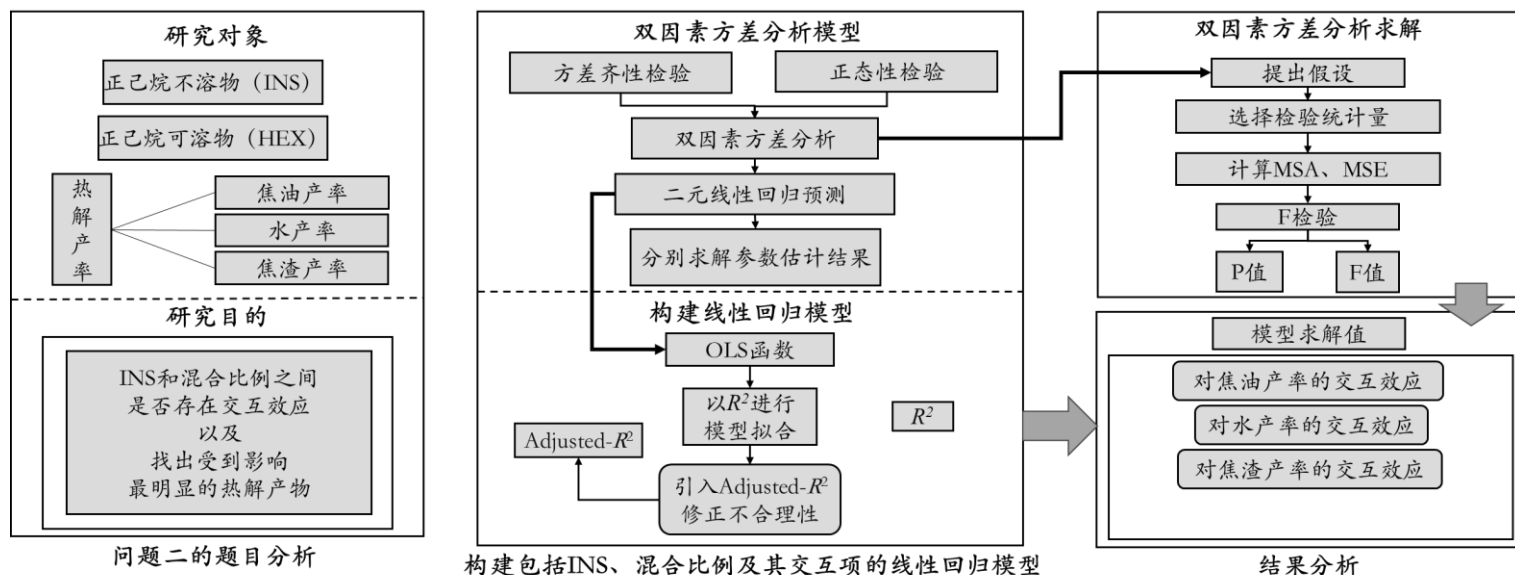


图 8 问题二的思维导图

根据问题二中的研究目标，探究正己烷不溶物（INS）和混合比例是否存在交互效应及其对热解产物产量的影响，利用 OLS 函数建立线性回归模型，通过双因素方差分析进行求解，先分别作出 INS 对四个因素的原假设，计算 R^2 并引入（校正 R^2 ）Adjusted- R^2 对模型中的不合理性进行修正，接着基于正态性检验和方差齐性检验进行双因素方差分析求解，通过 F 检验得到 P 值及 F 值，然后进行二元线性回归预测，并求解出双因素方差分析模型的参数估计结果，最后结合模型求解结果得出结论。

5.3.2 共热解交互效应机理分析

在生物质与煤共热解过程中，正己烷不溶物（INS）和混合比例可能会影响热解反应的自由基、化学反应路径和产物，进而可能会影响在热解过程中生成热解产物的速率，导致产物产量及产率都受到影响。为了能够更好地研究 INS 和混合比例之间是否存在交互效应，下面给出生物质与煤交互共热解下 INS 的化学反应机理，如图所示：

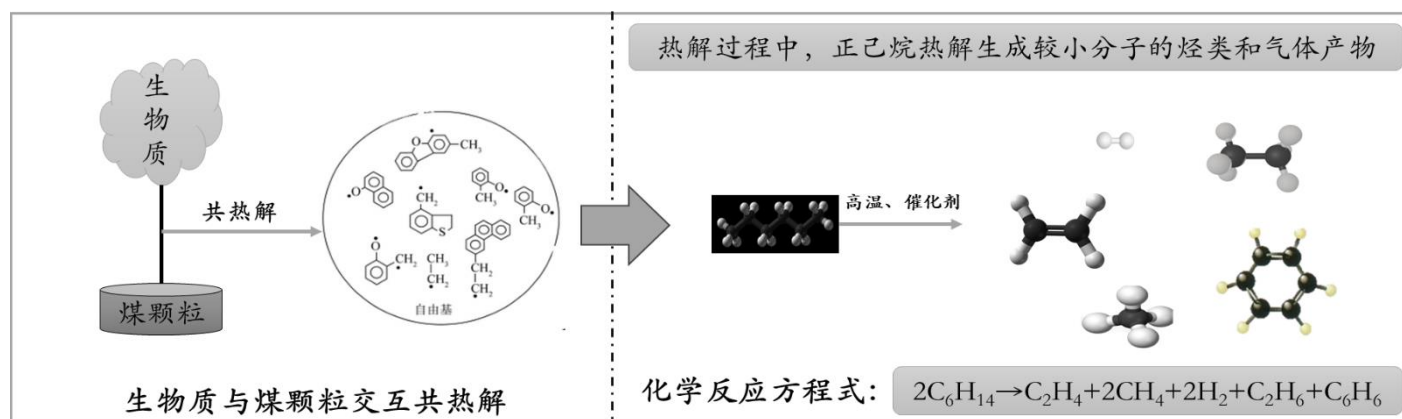


图 9 生物质与煤交互热解中 INS 的化学反应机理图

根据生物质与煤交互热解中 INS 的化学反应机理，由于在这个过程中会产生 CH_4 、 H_2 和大量自由基，可以不断地为后续的化学反应提供氢源，又因为生物质与煤共热解过程中生成的自由基能促进 INS 和混合比例协同作用下的化学反应效率，提高焦油、焦渣的产量及产率。因此，本文可以作出假设：正己烷不溶物（INS）和混合比例之间存在交互效应，且对热解产物产量产生重要影响。

5.3.3 双因素方差分析模型的建立

➤ OLS 函数

OLS 函数是利用最小二乘法进行线性回归模型拟合，然后通过最小化观测数据的实际值与模型预测值之间的残差平方和来估计回归系数，从而找到最优的拟合直线或平面，使得残差的平方和最小。

➤ 统计量 R^2

R^2 在线性回归中可用于衡量模型对观测数据的拟合程度，表示因变量的变异程度中有多少百分比可以由自变量解释。针对本问，引入 Adjusted- R^2 对 R^2 中的不合理性进行修正，通过考虑本问模型中自变量的数量来更准确地评估模型的拟合优度。

➤ 双因素方差分析

双因素方差分析（Two-Way ANOVA）是单因素方差分析的拓展，用于分析两个分类变量对连续型因变量的影响，并检验两个分类变量的主效应以及交互效应是否显著。

为了确定 INS 和混合比例之间是否存在交互效应，以及它们对热解产物产量的影响情况，可以进行双因素方差分析。在双因素方差分析中，INS 和混合比例分别作为两个因素的自变量，而热解产物产量则作为因变量。以下是热解实验中对 INS 和混合比例进行双因素方差分析的具体步骤^[4]：

首先为了检验 INS 与混合比例这两个因素之间的影响，需要对两个因素分别提出以下假设：

（1）对行因素提出假设

$H_0: \mu_1 = \mu_2 = \dots = \mu_i, \dots = \mu_k$ ，行因素(自变量)对因变量没有显著影响(μ_i 为行因

素的第 i 个水平的均值)。

$H_1: \mu_i(i=1,2,\dots,k)$ 不完全相等, 行因素(自变量)对因变量有显著影响。

(2) 对列因素提出假设

$H_0: \mu_1 = \mu_2 = \dots = \mu_j, \dots = \mu_r$, 列因素(自变量)对因变量没有显著影响(μ_j 为行因素的第 j 个水平的均值)。

$H_1: \mu_j(j=1,2,\dots,r)$ 不完全相等, 列因素(自变量)对因变量有显著影响。

结合本问可得原假设 H_0 : 正己烷不溶物、混合比例以及它们的交互效应对热解产物产量没有显著影响, 以及备择假设 H_1 : 正己烷不溶物、混合比例或它们的交互效应对热解产物产量有显著影响。

然后通过检验 H_0 是否成立, 分别确定检验行因素和列因素的统计量, 接着与单因素方差分析中一样, 从总平方和的分解入手计算出统计量。其中总平方和是全部样本观察值 $x_{ij}(i=1,2,\dots,k; j=1,2,\dots,r)$ 与总的样本平均值 \bar{x} 的误差平方和, 记为 SST, 即:

$$SST = \sum_{i=1}^k \sum_{j=1}^r (x_{i*} - \bar{x})^2 + \sum_{i=1}^k \sum_{j=1}^r (x_{*j} - \bar{x})^2 + \sum_{i=1}^k \sum_{j=1}^r (x_{ij} - \bar{x}_{i*} - \bar{x}_{*j} + \bar{x})^2$$

对于上式, 等式右边的第一项是行因素所产生的误差平方和, 记为 SSR; 等式右边第二项是列因素所产生的误差平方和, 记为 SSC; 等式右边第三项是剩余因素产生的随机误差平方和^[4], 记为 SSE。

接着在上述误差平方和的基础上将各平方和除以相应的自由度来计算均方, 与误差平方和相对应的自由度分别是^[4]: 总平方和 SST 的自由度为 $kr-1$; 行因素的误差平方和 SSR 的自由度为 $k-1$; 列因素的误差平方和 SSC 的自由度为 $r-1$; 随机误差平方和 SSE 的自由度为 $(k-1)(r-1)$ 。所以可得下列各均方: ①行因素的均方: $MSR = \frac{SSR}{k-1}$ ②列因素的均方: $MSC = \frac{SSC}{r-1}$ ③随机误差项的均方: $MSE = \frac{SSE}{(k-1)(r-1)}$ 。

为检验行因素及列因素对因变量的影响是否显著, 分别采用如下的统计量:

$$F_R = \frac{MSR}{MSE} \sim F(k-1, (k-1)(r-1)), \quad F_C = \frac{MSC}{MSE} \sim F(r-1, (k-1)(r-1)) \leftrightarrow$$

从文献资料来看, 通过给定的显著性水平 α 和两个自由度, 查 F 分布表得到相应的临界值 F_α , 然后将 F 分布表得到相应的临界值 F_R 和 F_C 与 F_α 进行比较^[4]。

若 $F_R > F_\alpha$, 则拒绝原假设 $H_0: \mu_1 = \mu_2 = \dots = \mu_i, \dots = \mu_k$, 表明 $\mu_i(i=1,2,\dots,k)$ 之间的差异性显著的; 若 $F_C > F_\alpha$, 则拒绝原假设 $H_0: \mu_1 = \mu_2 = \dots = \mu_j, \dots = \mu_r$, 表明 $\mu_j(j=1,2,\dots,r)$ 之间的差异性显著的。即如果计算得到的 F 统计量大于临界 F 值, 则拒绝零假设, 认为 INS 和混合比例存在交互效应且对热解产物产量有显著影响。

5.3.3 双因素方差分析模型的求解

针对探究问题二中涉及到的正己烷不溶物 (INS) 和混合比例之间是否存在交互效应, 本问通过构建双因素方差分析模型进行求解。双因素方差分析可以同时考虑 INS 与

混合比例这两个因素以及其交互效应对热解产物产量的影响。

与单因素方差分析求解类似，双因素方差分析模型的求解先提出假设，令原假设为正己烷不溶物、混合比例以及它们的交互效应对热解产物产量没有显著影响进一步。对 INS 和混合比例进行双因素分析，计算出 MSA、MSE，根据处理组平方和与残差平方和求解出 F 统计量。根据 F 值和 F 分布的概率密度函数，计算出对应的 P 值进行假设检验，比较 F 值与 F 分布的临界值，以确定处理组之间是否存在显著差异。

考虑到正己烷不溶物(INS)、混合比例以及它们的交互效应对不同产率的影响，通过使用 Python 统计库中的 OLS 函数来构建线性回归模型，对模型进行拟合，并给出模型的拟合优度 R^2 和（校正 R 方）Adjusted R^2 。双因素方差分析模型的求解结果如下表所示：

表 3 双因素方差分析求解结果表

| 参数 | R^2 | Adjusted- R^2 | F 统计量 | P 值 |
|--------|--------|-----------------|---------|--------|
| 焦油产率 | 0.5453 | 0.5222 | 23.5872 | 0.0000 |
| 水产率 | 0.1839 | 0.1424 | 4.4319 | 0.0071 |
| 焦渣产率 | 0.4219 | 0.3925 | 14.3537 | 0.0000 |
| HEX 产率 | 0.0902 | 0.0439 | 1.9491 | 0.1315 |

对问题一中的一元线性回归方程进行修改，本问建立二元线性回归方程

$$y = a_0 + a_1x_1 + a_2x_2 + \varepsilon$$

再求解出双因素方差分析模型的参数估计结果，如下表所示：

表 4 双因素方差分析模型的参数估计结果表

| 参数 | 焦油产率 | 水产率 | 焦渣产率 | HEX |
|--------------|--------|---------|---------|---------|
| 截距系数 | 0.0830 | 0.1371 | 0.6467 | 0.0629 |
| INS 系数 | 0.1159 | -0.0684 | 0.0356 | 0.1279 |
| 配比数值系数 | 0.0003 | 0.0000 | 0.0008 | -0.0003 |
| INS 与配比值交互系数 | 0.0010 | 0.0013 | -0.0054 | 0.0003 |

结合两个结果表进行分析，可以得知：

(1) 对焦油产率分析， R^2 和 Adjusted- R^2 分别为 0.5453 和 0.5222，说明该双因素方差分析模型能够解释焦油产率变异的 54.53%，引入 Adjusted- R^2 修正，模型的解释能力略有下降，但仍然较高，表明模型对数据的拟合程度较好。因为 F 统计量为 23.5872，P 值为 0.0000，说明模型至少一个自变量对因变量的影响是显著的。根据参数估计结果评估，可知 INS 对焦油产率的影响是显著的，且呈正相关；配比数值对焦油产率的影响不显著；而 INS 与配比数值的交互对焦油产率的影响不显著。

(2) 对水产率分析， R^2 和 Adjusted- R^2 分别为 0.1839 和 0.1424，说明该双因素方差分析模型能够解释焦油产率变异的 18.39%，引入 Adjusted- R^2 修正，模型的解释能力略有下降，但两个值都较低，表明模型对数据的拟合程度较差。因为 F 统计量为 4.4319，P 值为 0.0071，说明模型至少一个自变量对因变量的影响是显著的。根据参数估计结果评估，可知 INS 对水产率的影响不显著；配比数值对水产率的影响也不显著；INS 与配

比数值的交互对水产率影响不显著。

(3) 对焦渣产率分析, R^2 和 Adjusted- R^2 分别为 0.4219 和 0.3925, 说明该双因素方差分析模型能够解释焦油产率变异的 42.19%, 引入 Adjusted- R^2 修正, 模型的解释能力略有下降, 但仍然较高, 表明模型对数据的拟合程度较好。又因为 F 统计量为 14.3537, P 值为 0.0000, 说明模型至少一个自变量对因变量的影响是显著的。根据参数估计结果评估, 可知 INS 对焦渣产率的影响不显著; 配比数值对焦渣产率的影响不显著; 而 INS 与配比数值的交互对焦渣产率有显著影响, 且呈负相关。

(4) 对正己烷可溶物(HEX)产率分析, R^2 和 Adjusted- R^2 分别为 0.0902 和 0.0439, 说明该双因素方差分析模型能够解释焦油产率变异的 9.02%, 引入 Adjusted- R^2 修正, 模型的解释能力下降, 整体解释能力表现较差, 表明模型对数据的拟合程度差。同样因为 F 统计量为 1.9491, P 值为 0.1315, P 值大于一般显著水平 0.05, 表明模型的自变量对因变量的影响不显著。根据参数估计结果评估, 可知其他自变量 INS、配比数值以及它们之间的交互项对 HEX 产率的影响都不显著。

对双因素方差分析模型进行数据可视化, 可得热解产率对 INS 与混合比例之间交互效应的数据可视化图如下所示:

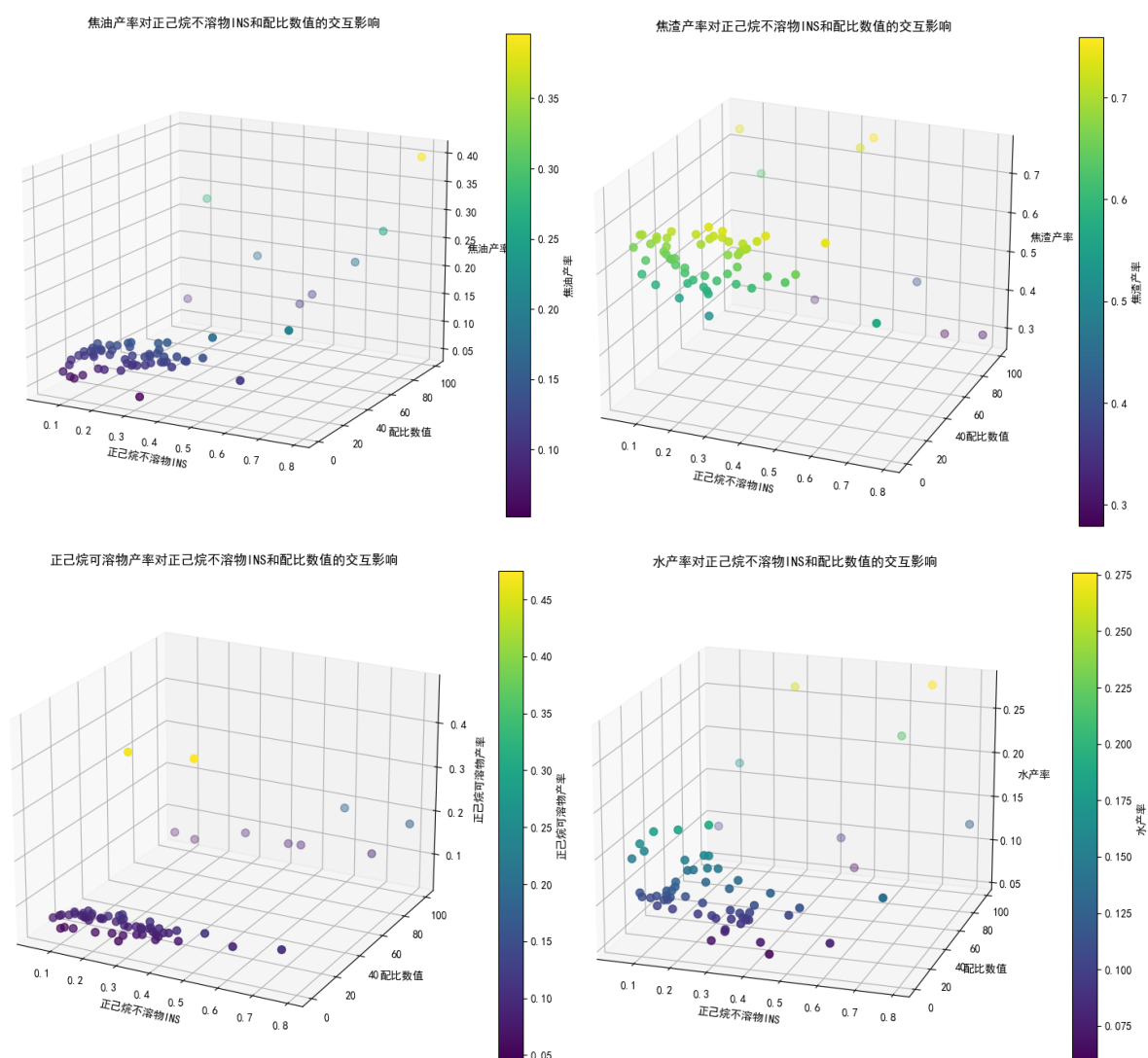


图 10 热解产率对 INS 与混合比例之间交互效应的数据可视化图

通过对双因素方差分析模型进行数据可视化,直观地展示了热解产率对 INS 与混合比例之间交互效应的影响,即产物焦油的样品重量(在 0.3g ~ 0.4g)和混合比例(在 50/100~100/100)的交互效应最为明显、产物焦渣的样品重量(在 0.1g ~ 0.3g)和混合比例(在 60/100~100/100)的交互效应最为明显,并且可以更容易比较出对于其他自变量 INS、配比数值以及它们之间的交互项之间存在的差异影响。

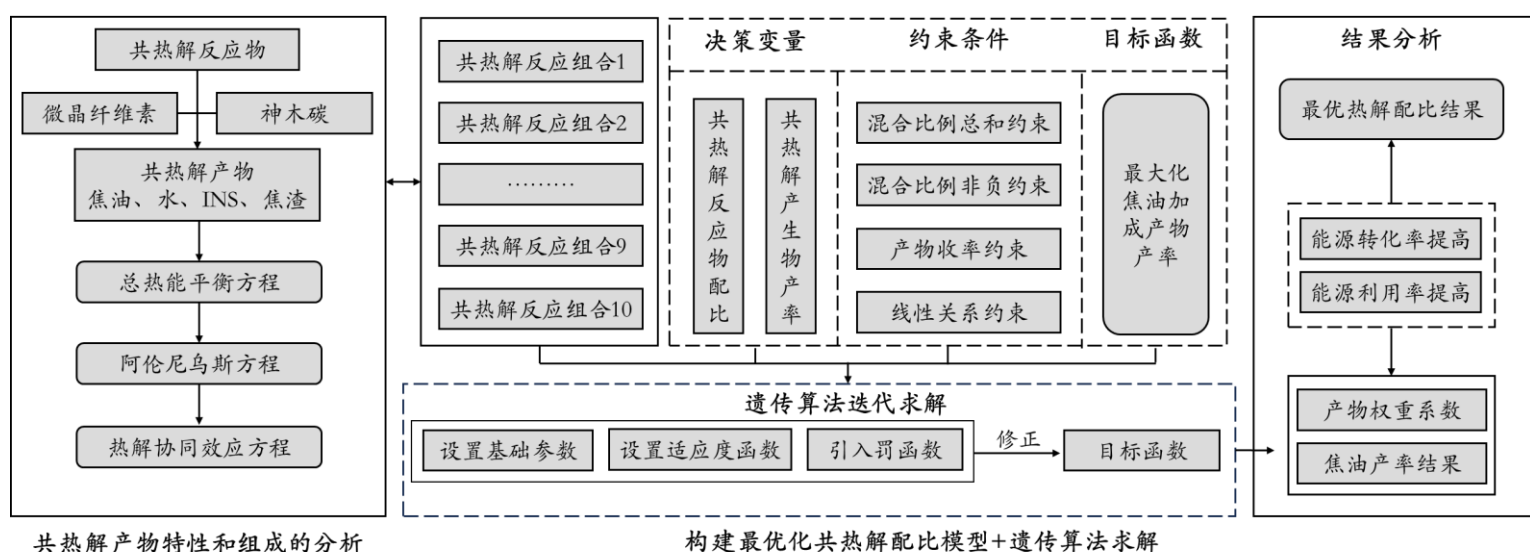
5.3.4 结果

经过上述四个不同产物的双因素方差分析、模型的评估及可视化,可知在焦油产率和焦渣产率两个因素自变量的模型中,INS 与配比数值的交互效应的 P 值均小于一般的显著性水平 0.05,说明了交互对二者的影响是显著的,并且产物焦油的样品重量(在 0.3g ~ 0.4g)和混合比例(在 50/100~100/100)的交互效应最为明显、产物焦渣的样品重量(在 0.1g ~ 0.3g)和混合比例(在 60/100~100/100)的交互效应最为明显。而在水产率和 HEX 的模型中所表现 INS 和混合比例之间的交互效应系数的 P 值都大于一般显著性水平,说明交互效应对 HEX 产率和水产率并不显著。

因此根据双因素方程分析模型的结果可以得出最终结论,正己烷不溶物(INS)和混合比例之间存在交互效应,但在不同的热解产物上交互动效应的显著性有所不同,其中焦油产率和焦渣产率受 INS 和混合比例交互效应的影响最显著,可能是在其生成机理上发挥了重要作用。

5.4 问题 3 的模型建立与求解

5.4.1 思路分析

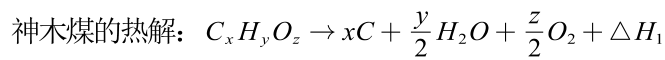
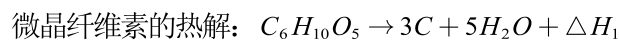
**图 11** 问题三的思维导图

根据研究生物质和煤共热解油的产率和品质机理对提高能源利用效率、促进资源综

合利用和确保能源安全具有重要意义,可知问题三的实质是最大化焦油产率,来确定最优化共解热混合比例。先分析共热解产物特性和组成,结合能量守恒,由阿伦尼乌斯方程确定热解速率方程式。考虑热解协同效应,根据附件 1 的 10 种共热解组合的焦油产率、水产率、焦渣产率和正己烷可溶物产率,可确定四者对应的权重因子占比。以最大化焦油的加权产物产率为目标,由遗传算法迭代求解,可得出 10 种共热解组合的最优化混合比例。

5.4.2 共热解产物特性和组成的分析

在分析基于共热解产物的特性和组成,先关注共热解产物生成机理,引入微晶纤维素作为模型化合物,题目提到的微晶纤维素/神木煤共热解反应,其化学反应方程式如下:



查阅资料显示,微晶纤维素/神木煤共热解反应主要经历热解启动原料加热到 600°C ,升温速率为 $5^\circ\text{C}/\text{min}$ 。之后纤维素开始脱水,同时煤的大分子也开始解聚。纤维素和煤的挥发分开始释放,形成各种小分子气体和液体。纤维素的固体残留物主要为碳(生物炭),而煤的热解残留物则更多地包含矿物质和未完全分解的有机物,挥发分间的相互作用(即协同反应)可能改变气体和液体的组成^[5]。

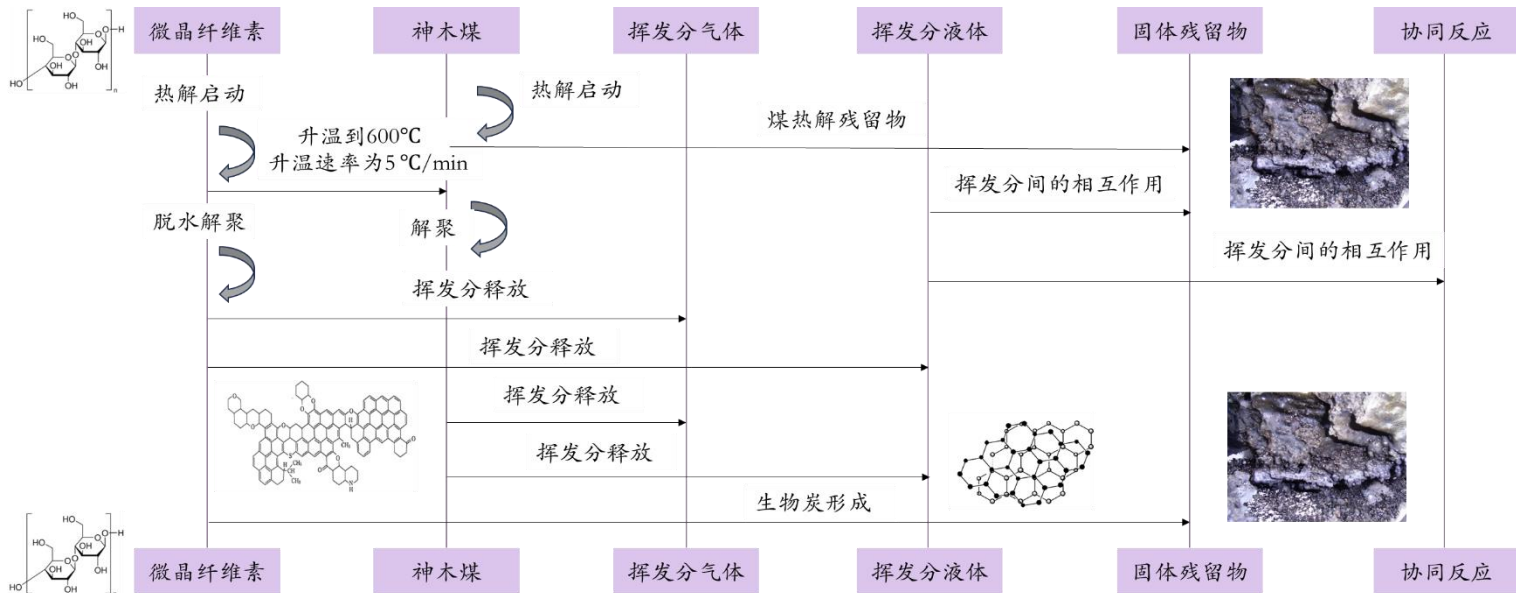


图 12 微晶纤维素/神木煤共热解反应机理

在热解过程中,即产生四种产物的过程要保持热力学平衡, Q_i 表示输入的能量, Q_r 表示反应所需能量, Q_l 表示损失的能量, E_a 是活化能,单位是焦耳每摩尔 (J/mol),则总热能平衡公式

$$Q_i = Q_r + Q_l + E_a$$

使用阿伦尼乌斯方程描述热解反应速率, v 表示速率常数、 R 表示理想气体常数,

其值约为 $8.314 \text{ J}/(\text{mol} \cdot \text{K})$ 、 T 是热解实验初始温度，单位是开尔文 (K)、 A 是指前因子，其中，题目中已知所有热解实验的初始温度均为 **873.15 K**，速率常数 v 表示升温 **278.15 K/min**。则有

$$v = Ae^{-\frac{E_a}{RT}}$$

由于生物质与煤的热解协同效应，令 t 表示混合比例 x/y ， β 表示协同效应项的常数，为了简化模型，此时协同效应项定义为

$$\theta(t) = \beta t$$

根据题目提供的 10 种共热解组合，可确定有 4 种生物质、5 种煤，并设置总共组数 k ，第 1 组 CS/HN（棉杆/淮南煤）、第 2 组为 CS/SM（棉杆/神木煤）、第 3 组为 CS/HS（棉杆/黑山煤）、第 4 组为 SD/HS（木屑/黑山煤）、第 5 组为 SD/SM（木屑/神木煤）、第 6 组为 GA/HN（小球藻/淮南煤）、第 7 组为 GA/NM（小球藻/内蒙褐煤）、第 8 组为 GA/SM（小球藻/神木煤）、第 9 组为 RH/HN（稻壳/淮南煤）以及第 10 组 RH/SM（稻壳/神木煤），可为后续的最佳配比提供组合依据。

5.4.3 最优化共热解配比模型的建立

决策变量

假设 x_k 表示第 k 组生物质和煤的混合的比例，则有

$$x_k, 1 \leq k \leq 10$$

假设 b_k 、 c_k 、 d_k 、 e_k 分别表示 10 种共热解组合的焦油产率、正己烷可溶物产率、水产率、焦渣产率，即

$$b_k, c_k, d_k, e_k, 1 \leq k \leq 10$$

约束条件

(1) 混合比例总和约束：所有混合比例的总和必须等于 1

$$\sum x_k = 1$$

(2) 混合比例非负约束：所有比例必须非负

$$x_k \geq 0$$

(3) 产物收率约束：根据经验以及相关文献，设生物质与煤的热解混合比例 t ， s_i 表示第 i 种生物质的热解产物收率， m_j 表示第 j 种煤的产物收率，则四种产物的收率应符合

$$f(t, \theta(t)) = \frac{t}{t+1} s_i + \frac{1}{t+1} m_j + \theta(t), 1 \leq i \leq 4, 1 \leq j \leq 5$$

(4) 产物收率与混合比例的线性关系约束：根据问题二中的交互效应与收率的线性分析，在此将四种产物与共热解比例约束为

$$b_k = f(x_k, \theta(x_k)), c_k = f(x_k, \theta(x_k)), d_k = f(x_k, \theta(x_k)), e_k = f(x_k, \theta(x_k))$$

目标函数

问题三的以提高产物利用率和能源转化效率，可用最大化焦油加成产物产率来表示，即目标函数为加权产物产率约束，用 α_1 、 α_2 、 α_3 、 α_4 表示焦油产率、正己烷可溶物产率、水产率、焦渣产率的权重系数，则 M 表示加权产物产率，则

$$M = \alpha_1 b_k + \alpha_2 c_k + \alpha_3 d_k + \alpha_4 e_k$$

为使得焦油加成产物产率最大化，目标函数为：

$$f = M$$

连续最优化共热解配比模型

$$\text{Max } f = M$$

$$s.t. \begin{cases} 1 \leq k \leq 10, 1 \leq i \leq 4, 1 \leq j \leq 5, T = 873.15, v = 278.15 \\ \sum x_k = 1, x_k \geq 0 \\ Q_i = Q_r + Q_l + E_a, v = Ae^{-\frac{E_a}{RT}}, \theta(t) = \beta t \\ f(t, \theta(t)) = \frac{t}{t+1} s_i + \frac{1}{t+1} m_j + \theta(t) \\ b_k = f(x_k, \theta(x_k)), c_k = f(x_k, \theta(x_k)), d_k = f(x_k, \theta(x_k)), e_k = f(x_k, \theta(x_k)) \\ M = \alpha_1 b_k + \alpha_2 c_k + \alpha_3 d_k + \alpha_4 e_k \end{cases}$$

5.4.4 最优化共热解配比模型的求解

➤ 遗传算法的引入

遗传算法是一种基于自然选择和遗传学原理的优化算法，适用于解决复杂的搜索问题。对于问题三中的任务，即优化共热解混合比例以提高产物利用率和能源转化效率，遗传算法可处理具有大量变量和潜在解决方案的问题。在共热解混合比例优化问题中，可能存在多种生物质和煤的组合，每种组合都可以在不同的比例下进行。遗传算法通过随机初始化种群，逐代进化，能有效搜索这样复杂的空间。

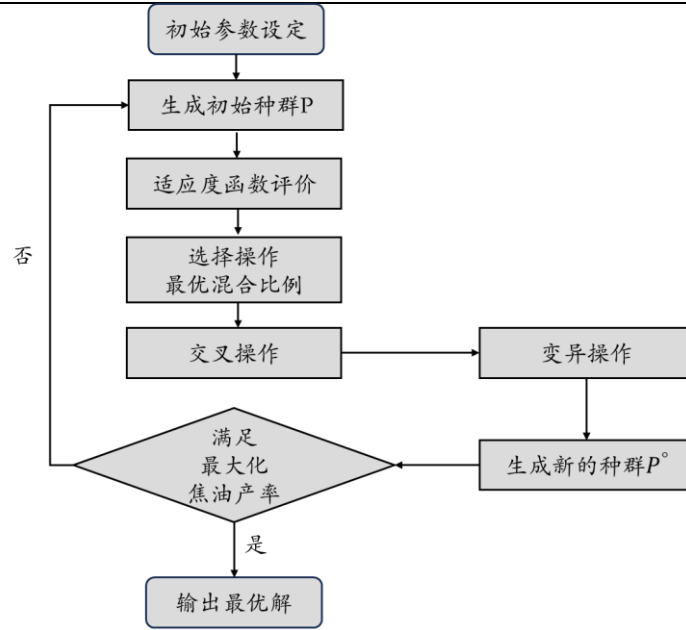


图 13 遗传算法的处理流程

遗传算法通过模拟自然选择的过程（包括选择、交叉和变异），能够在整个搜索空间内进行探索，评估不同混合比例对产物利用率和能源效率的综合影响，从而找到全局最优解。

➤ 适应度函数的设置

在使用遗传算法（GA）解决问题三，即优化共热解混合比例以提高产物利用率和能源转化效率时，适应度函数的设计是至关重要的。适应度函数决定了算法如何评估和选择候选解，也就是各种混合比例的性能。并将适应度设置为：

$$F_{fitness} = w_1 \cdot \eta_{Tar} + w_2 \cdot \eta_{Energy}$$

其中， η_{Tar} 是焦油利用率， w_1 是焦油利用率的权重因子； η_{Energy} 是能源转化效率， w_2 是能源转化效率的权重因子。

➤ 罚函数的引入

为防止混合比例的非合法解的产生，引入惩罚而非完全排除这些解，算法可以在接近最优区域灵活探索，通过微小的调整可以将这些高价值的非可行解转化为可行解。

$$P(x) = -q \cdot \left| \sum x_k - 1 \right|, \quad \forall k \in \{1, 2, \dots, 10\}$$

其中 q 是一个大的正常数，用以确保非法解的适应度显著降低，迫使算法向合法解空间收敛。

则通过适应度以及罚函数的设置，目标函数修正为：

$$f = M + F_{fitness} + P(x)$$

➤ 迭代求解

通过设置 CXPB, MUTPB = 0.5, 0.2，经遗传算法 50 轮 100 次迭代得到 10 种共热解最优化混合比例，可视化的结果图如下：

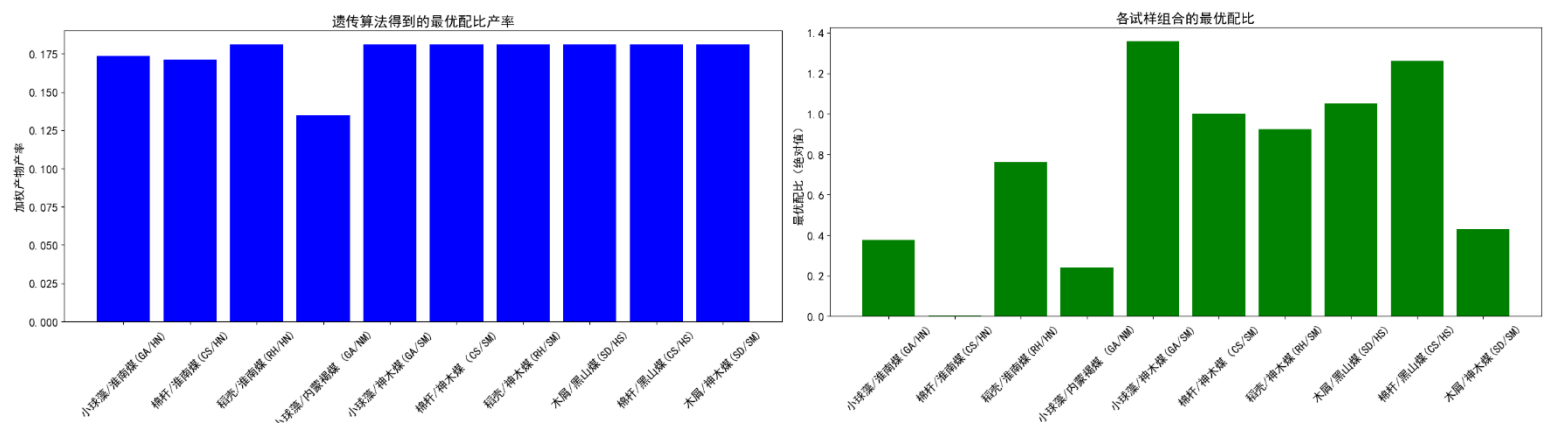


图 14 遗传算法得到的焦油的最佳加权产率（左）以及最佳配比情况（右）
具体 10 种最佳共热解混合比例结果如下表：

表 5 10 种最佳共热解混合比例的焦油产物加权产率结果

| 组合类型 | 最优混合比例 | 焦油产物加权产率 |
|-------|--------|----------|
| GA/HN | 11/100 | 0.140 |
| CS/HN | 65/100 | 0.185 |
| RH/HN | 36/100 | 0.190 |
| GA/NM | 11/100 | 0.150 |
| GA/SM | 34/100 | 0.166 |
| CS/SM | 49/100 | 0.177 |
| RH/SM | 19/100 | 0.169 |
| SD/HS | 34/100 | 0.165 |
| CS/HS | 8/100 | 0.140 |
| SD/SM | 21/100 | 0.166 |

上表显示这 10 种组合的焦油产物加权产率最大为 0.190，最低有 0.140。其中求得的焦油产率权重系数为 0.6、正己烷可溶物产率权重系数为 0.2、水产率权重系数为 0.1、焦渣产率权重系数为 0.1，而焦油的权重系数最大，这符合题目中的生物质和煤共热解油的产率对提高能源转化效率和利用率的解释。表明在这些优化条件下，所有组合在经过加权后产率相同，这意味着遗传算法达到了平衡状态，求解出了这 10 种生物质和煤的最佳混合比例。

5.4.5 结果

在考虑这 10 种共热解产物的特性和组成的情况下，10 种最佳共热解混合比例：**GA/HN 为11/100、CS/HN 为65/100、RH/HN为 36/100、GA/NM为 11/100、GA/SM为 34/100、CS/SM为 49/100、RH/SM为 19/100、SD/HS为 34/100、CS/HS为 8/100 、SD/SM为 21/100**，可最大化产物利用率和能源转化效率，其中，CS/HN（棉杆/淮南煤）、CS/SM（棉杆/神木煤）、RH/HN（稻壳/淮南煤）需要更高的生物质来提高产物利用率和能源转化效率。

5.5 问题 4 的模型建立与求解

5.5.1 思路分析

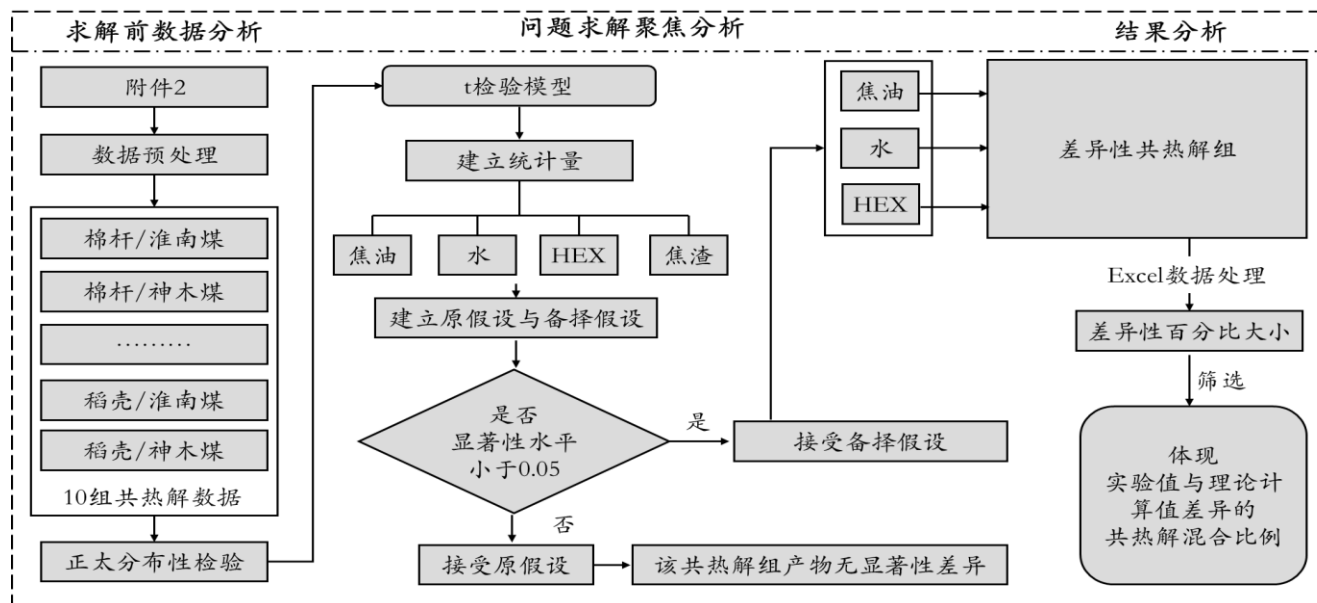


图 15 问题四的思维导图

对于问题四的求解，首先对附件 2 的数据进行了数据预处理和正态性检验，确定了附件 2 数据适合进行 t 检验分析。我们对 10 种共热解组合的热解产物焦油、正己烷可溶物、水、焦渣的角度，分别用 t 检验分析这四个指标收率实验值和理论计算值的差异性，确定存在显著差异的共热解组。最后对附件 2 预处理后的子组数据中的产物收率实验值和理论计算值差异进行可视化，分析产生显著性差异的潜在因素，根据差异比率确定共热解产物实验值和理论计算值差异性较大的混合比例。

5.5.2 t 检验模型的建立

t 检验是用 t 分布理论来推论差异发生的概率，从而比较两个平均数的差异是否显著的一种假设检验方法。因此本题分组变量为每种共热解组合的产物收率实验值与理论计算值，我们选择运用独立样本 t 检验， t 检验的统计量为：

$$t = \frac{\bar{X}_1 - \bar{X}_2}{\sqrt{\frac{(n_1 - 1)S_1^2 + (n_2 - 1)S_2^2}{n_1 + n_2 - 2} \left(\frac{1}{n_1} + \frac{1}{n_2} \right)}}$$

其中， S_1^2 和 S_2^2 为产物收率实验值方差和理论值方差， n_1 和 n_2 分别为产物收率实验值和理论值容量。

我们需要解决的问题是分析每种共热解组合的产物收率实验值与理论计算值之间是否存在显著差异，并探讨导致这种差异的混合比例以及潜在因素。为了确保数据的有效性以及分析结果的准确性，我们采用预处理之后的附件 2 数据，建立 t 检验的模型，

分别对不同的共热解组合产物收率的实验值和理论计算值进行独立样本 t 检验。

5.5.3 t 检验模型的求解

(1) 对 10 组共热解组合的产物焦油 (Tar) 的收率实验值与理论计算值, 进行独立样本 t 检验, 设:

H_0 : 产物焦油 (Tar) 的收率实验值与理论计算值没有显著差异

H_1 : 产物焦油 (Tar) 的收率实验值与理论计算值具有显著差异

经分析, 将共热解组合类型作为检验变量, 产物收率实验值与理论计算值作为分组变量, 得到结果如下:

表 6 焦油的 t 检验表

| | RH/HN | RH/SM | CS/HS | CS/HN | CS/SM | SD/HS | SD/SM | GA/HN | GA/NM | GA/SM |
|-----|--------|--------|--------|-------|--------|--------|-------|--------|--------|-------|
| t 值 | -2.537 | -1.066 | -1.846 | 0.037 | -0.450 | -0.023 | 0.099 | -0.877 | -2.211 | 0.142 |
| P 值 | 0.034 | 0.317 | 0.102 | 0.971 | 0.664 | 0.982 | 0.923 | 0.406 | 0.049 | 0.890 |

从上表可知 P 值小于 0.05 的共热解组合, 小于显著性水平, 拒绝原假设, 接受假设 H_1 , 即焦油的收率的实验值和理论计算值在 RH/HN (稻壳/淮南煤)、GA/NM (小球藻/内蒙褐煤) 共热解组合存在显著差异。

(2) 对 10 组共热解组合的产物正己烷可溶物 (HEX) 的收率实验值与理论计算值, 进行独立样本 t 检验, 设:

H_0 : 正己烷可溶物 (HEX) 的收率实验值与理论计算值没有显著差异

H_1 : 正己烷可溶物 (HEX) 的收率实验值与理论计算值具有显著差异

经分析, 将共热解组合类型作为检验变量, 产物收率实验值与理论计算值作为分组变量, 得到结果如下:

表 7 正己烷可溶物 (HEX) 的 t 检验表

| | RH/HN | RH/SM | CS/HS | CS/HN | CS/SM | SD/HS | SD/SM | GA/HN | GA/NM | GA/SM |
|-----|--------|-------|--------|-------|-------|-------|-------|-------|--------|--------|
| t 值 | -8.380 | 4.514 | -1.841 | 2.158 | 3.819 | 7.209 | 3.257 | 1.319 | -1.722 | -0.558 |
| P 值 | 0.000 | 0.001 | 0.102 | 0.062 | 0.005 | 0.000 | 0.011 | 0.223 | 0.123 | 0.592 |

从正己烷可溶物 (HEX) 的 t 检验表结果可知, RH/HN (稻壳/淮南煤)、RH/SM (稻壳/神木煤)、CS/SM (棉杆/神木煤)、SD/HS (木屑/黑山煤)、SD/SM (木屑/神木煤) 的显著性 P 值小于 0.05, 拒绝原假设, 接受假设 H_1 , 认为这 5 个共热解组合的 HEX 产物收率的实验值和理论计算值存在显著性差异。

(3) 对 10 组共热解组合的产物水 (Water) 的收率实验值与理论计算值, 进行独立样本 t 检验, 设:

H_0 : 水 (Water) 的收率实验值与理论计算值没有显著差异

H_1 : 水 (Water) 的收率实验值与理论计算值具有显著差异

经分析, 将共热解组合类型作为检验变量, 产物收率实验值与理论计算值作为分组变量, 得到结果如下:

表 8 水 (Water) 的 t 检验表

| | RH/HN | RH/SM | CS/HS | CS/HN | CS/SM | SD/HS | SD/SM | GA/HN | GA/NM | GA/SM |
|-----|-------|-------|-------|--------|--------|-------|-------|-------|-------|-------|
| t 值 | 1.230 | 0.234 | 0.593 | -0.022 | -0.476 | 0.302 | 0.720 | 2.787 | 5.307 | 2.651 |
| P 值 | 0.253 | 0.820 | 0.596 | 0.982 | 0.646 | 0.770 | 0.492 | 0.023 | 0.000 | 0.029 |

水的差异 t 检验表表明, GA/HN (小球藻/淮南煤)、GA/NM (小球藻/内蒙褐煤)、GA/SM (小球藻/神木煤) 的 P 值小于 0.05, 小于显著性水平, 可接受假设 H_1 , 认为这三种共热解组合的水产物收率的实验值与理论计算值存在显著性差异。

(4) 对 10 组共热解组合的产物焦渣 (Char) 的收率实验值与理论计算值, 进行独立样本 t 检验, 设:

H_0 : 产物焦渣 (Char) 的收率实验值与理论计算值没有显著差异

H_1 : 产物焦渣 (Char) 的收率实验值与理论计算值具有显著差异

经分析, 将共热解组合类型作为检验变量, 产物收率实验值与理论计算值作为分组变量, 得到结果如下:

表 9 焦渣 (Char) 的 t 检验表

| | RH/HN | RH/SM | CS/HS | CS/HN | CS/SM | SD/HS | SD/SM | GA/HN | GA/NM | GA/SM |
|-----|--------|--------|--------|--------|--------|--------|--------|-------|--------|-------|
| t 值 | -0.312 | -0.172 | -0.325 | -0.635 | -0.087 | -0.502 | -0.021 | 0.013 | -0.694 | 0.019 |
| P 值 | 0.763 | 0.867 | 0.753 | 0.543 | 0.932 | 0.623 | 0.984 | 0.989 | 0.507 | 0.985 |

从焦渣的差异 t 检验结果来看, 不存在显著性 P 值小于 0.05 的组合, 接受原假设 H_0 , 认为焦渣的产物收率理论计算值和实验值不存在显著性差异。

由上述分析可以得知, 每种共热解组合的产物收率除焦渣外, 其他如焦油、正己烷可溶物、水的产物收率实验值与理论计算值均存在着显著性差异。

5.5.4 差异可视化模型的求解

为了进一步探究导致这种差异的具体混合比例是哪些, 我们对有显著性差异的共热解组合的产物收率实验值与理论计算值依次进行描述性统计 (即焦油、正己烷可溶物、水), 其差异比 (实验值与理论计算值的差值绝对值与理论计算值的比率) 结果如下:

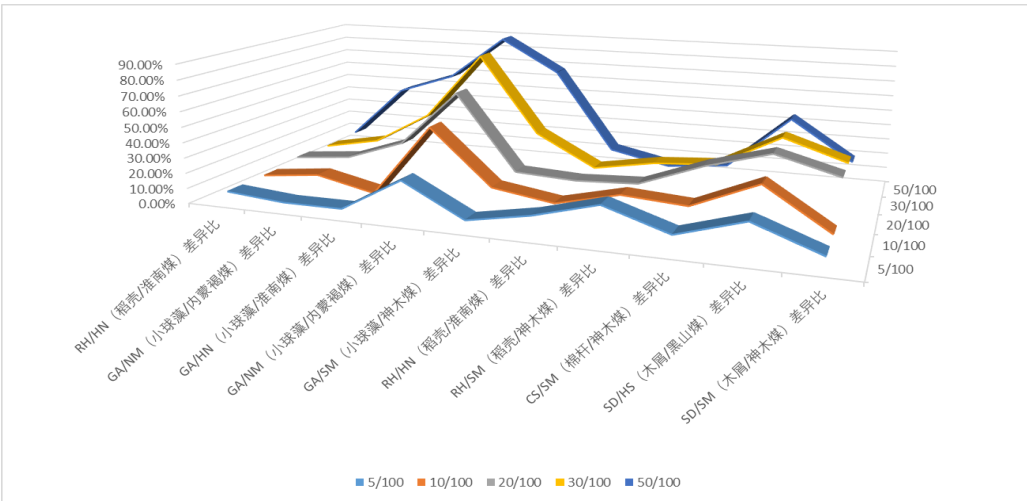


图 16 差异比三维可视化

对于产物焦油，RH/HN（稻壳/淮南煤）组的混合比例为 20/100 与 50/100 时的差异比分别为 9.60%、10.04%。GA/NM（小球藻/内蒙褐煤）组的混合比例为 **30/100、50/100** 时的差异比分别为 **15.87%、45.93%**。

对于产物水，随着生物质的分量占比加大，即混合比例由 5/100 增加到 50/100，GA/HN（小球藻/淮南煤）、GA/NM（小球藻/内蒙褐煤）与 GA/SM（小球藻/神木煤）3 个共热解组的产物差异比增大，在 GA/NM 组中 **50/100** 的混合比例的实验值与理论计算值差异超过 **88.92%**。

对于产物正己烷可溶物，RH/HN（稻壳/淮南煤）组与 SD/HS（木屑/黑山煤）组都在混合比例为 50/100 时，差异比最大。而 RH/SM（稻壳/神木煤）混合比例为 5/100 时，差异比最大为 23.13%，CS/SM（棉杆/神木煤）混合比例为 20/100 时，差异比最大为 28.97%，对于 SD/SM（木屑/神木煤）组，在其混合比例为 20/100 时，差异比最大，为 29.36%。

从分析结果来看，共热解组为 **20/100、50/100、5/100** 的混合比例时，更容易有产物收率实验值与理论计算值的显著性差异。

5.5.5 结果

通过对模型的建立和求解，利用 Python 对筛选后的数据进行初步运算。首先通过 Python 对每种共热解组合的 4 种产物收率的实验值和理论计算值分别进行独立样本 t 检验，得出每种共热解组合的产物收率实验值和理论计算值具有显著性差异。通过差异比率折线图分析，发现焦油的 RH/HN（稻壳/淮南煤）与 GA/NM（小球藻/内蒙褐煤）共热解组合的 **50/100 混合比例**下的实验值与理论计算值差异明显。产物水的 GA/HN（小球藻/淮南煤）、GA/NM（小球藻/内蒙褐煤）与 GA/SM（小球藻/神木煤）共热解组合的 50/100 混合比例下的实验值与理论计算值差异性较大。对于产物正己烷可溶物的实验值与理论计算值差异性，在 50/100 混合比例的 RH/HN（稻壳/淮南煤）组最大、**5/100 混合比例**的 RH/SM（稻壳/神木煤）组最大、**20/100 混合比例**的 CS/SM（棉杆/神木煤）组最大、50/100 混合比例的 SD/HS（木屑/黑山煤）最大、20/100 混合比例的 SD/SM（木屑/神木煤）组最大。

5.6 问题 5 的模型建立与求解

5.6.1 思路分析

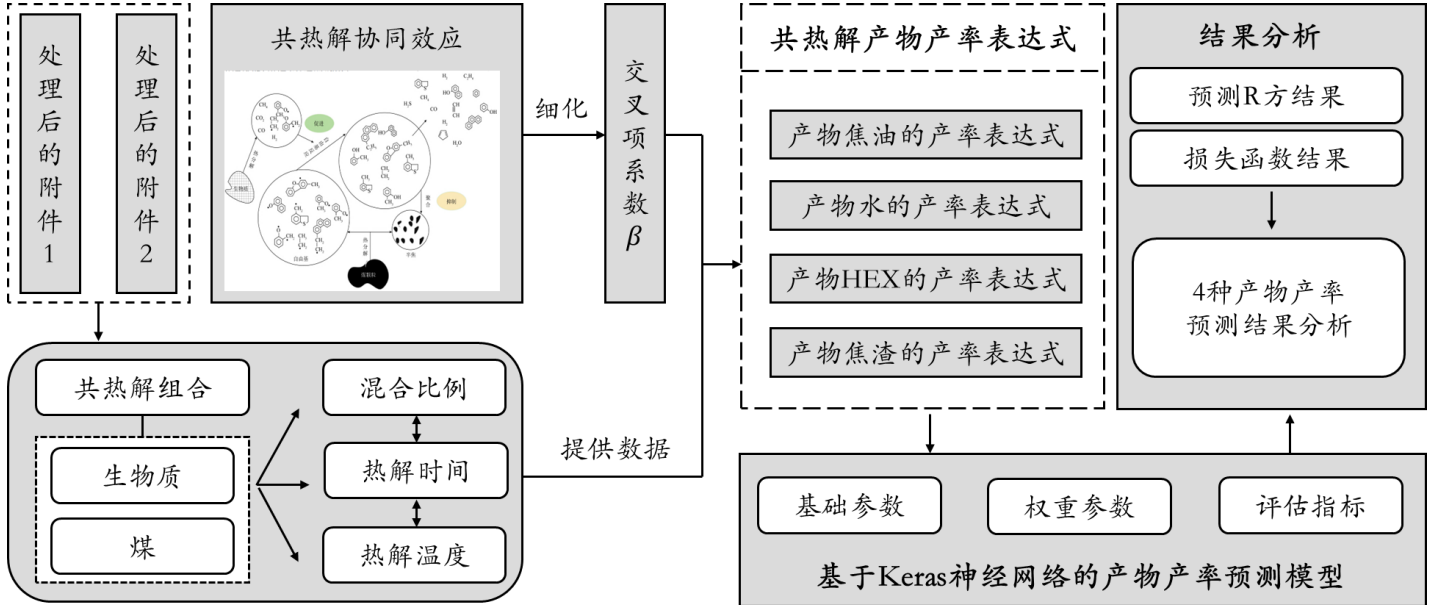


图 17 问题五的思维导图

从前面 4 个问题的分析求解结合两个附件的数据,发现多种生物质与煤的组合比例、不同类型的原料、温度和时间等多维变量间相互交织,影响产物的产率。为简化模型,即为了提高产物利用率和能源转化效率,我们使用问题三已求解出的最优共热解混合比例以及附件数据作为支撑,同时用共热解协同效应产生的交叉项,细化热解产物的产率计算公式,并构建 Keras 神经网络预测模型,进行四种产物(焦油、水、正己烷可溶物、焦渣)产率预测。

5.6.2 基于 Keras 的神经网络预测模型的建立

► 确定计算热解产物产率公式

生物质与煤共热解的实际产物收率计算涉及到交叉项,其作用有可能是促进作用也可能是抑制作用,根据问题二的交互效应分析,以及结合问题三的最佳配比,并考虑协同效应指数,对交叉项进行细化。

而产物各组分的差值 ΔY_x , x 可代表 t (焦油)、 w (水)、 h (正己烷可溶物)、 $char$ (焦渣), 各组分的计算产率 $Y_{x,cal} = R \cdot Y_{i,bio} + (1 - R) \cdot Y_{i,coal}$, $Y_{x,exp}$ 表示共热解实验产率, $Y_{x,cal}$ 表示计算产率, 则 $\Delta Y_x = Y_{x,exp} - Y_{x,cal}$ 。

在协同效应指数(交叉项系数 β)中, ΔY_t 表示代表焦油产率的差值、 ΔY_w 代表水产率的差值、 ΔY_h 代表正己烷可溶物产率的差值、 ΔY_{char} 代表焦渣产率的差值^[6]。由问题三的求解结果可知,焦油、水、正己烷可溶物、焦渣的权重系数分别为 0.6、0.2、0.1、0.1, 则

$$\beta = 0.6 |\Delta Y_t| + 0.2 |\Delta Y_w| + 0.1 |\Delta Y_h| + 0.1 |\Delta Y_{char}|$$

则产物产率计算公式为：

$$y(q) = \frac{q}{q+1} s_i + \frac{1}{q+1} m_j + (0.6|\Delta Y_t| + 0.2|\Delta Y_w| + 0.1|\Delta Y_h| + 0.1|\Delta Y_{char}|)q$$

其中， q 为生物质与煤的热解混合比例， s_i 表示第 i 种生物质的热解产物收率， m_j 表示第 j 种煤的产物收率。

➤ 引入 Keras 神经网络预测

在确定考虑了交叉项的四种产物产率的计算公式后，由于热解问题涉及高维数据与复杂的变量间相互作用，通过使用神经网络框架 Keras，可以构建预测模型，捕捉变量间的非线性关系，进行产率预测。而 Keras 是一个直观的神经网络库，用户可以通过 Sequential 模型线性堆叠不同的层，每层包含可学习的参数，如权重和偏置，并利用 TensorFlow 后端的优化算法调整参数以最小化损失。

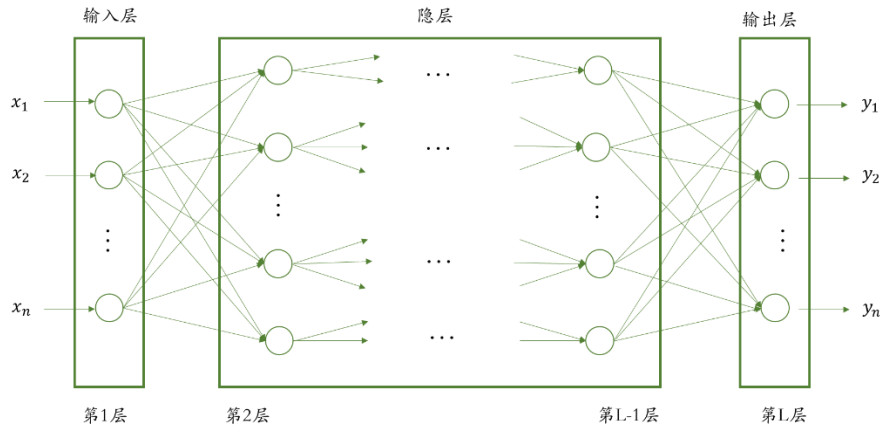


图 18 Keras 神经网络模型的结构

假设输入样本为 $x = (x_1, x_2, \dots, x_m)^T$ ，输入层到隐藏层的权值为 $v = (v_1, v_2, \dots, v_m)^T$ ，隐藏层的阈值为 $B = (b_1, b_2, \dots, b_m)^T$ ，隐藏到输出层的权值 $w = (w_1, w_2, \dots, w_m)^T$ ，输出层的阈值为 $Y = (y_1, y_2, \dots, y_l)^T$ ，望值为 $o = (o_1, o_2, \dots, o_m)^T$

Step1: 输出层和输入层的计算

计算输出层和输入层的阈值， net 是从隐藏层到输出层所有网络的综合，计算公式如下：

$$y_k = f\left(\sum_m^{j=0} w_{kj} b_j\right), k = 1, 2, \dots, l$$

输出层的值：

$$b_j = f\left(\sum_n^{j=0} v_{ji} x_i\right)$$

Step2: 损失函数 MSE 的计算

根据均方误差 MSE 公式计算期望值与输出值之间的反向误差结果：

$$E = \frac{1}{2} \sum_l^{k=1} \left(o_k - f\left(\sum_m^{j=0} w_{kj} b_j\right) \right)^2$$

Step3: 更新并求解权重参数

$$\Delta v_{ji} = -\eta \frac{\partial E}{\partial v_{ji}} = -\eta \frac{\partial E}{\partial net_j} \frac{\partial net_j}{\partial v_{ji}}$$

$$\Delta w_{ki} = -\eta \frac{\partial E}{\partial w_{ki}} = -\eta \frac{\partial E}{\partial net_k} \frac{\partial net_k}{\partial w_{ki}}$$

分别对上述公式结果进行求偏导，最终得出更新参数的表达式为：

$$\Delta w_{ki} = -\eta \delta_k^y b_j = \eta (o_k - y_k) y_k (1 - y_k) b_j$$

$$\Delta v_{ji} = -\eta \left(\sum_{k=1}^l \eta \delta_k^y w_{jk} \right) b_i (1 - b_i) x_i$$

Step4: 结束训练

计算总体误差，当其达到收敛精度标准或训练最大迭代次数，结束训练。

由于均方误差 MSE 的数值范围与数据本身大小密切相关，无法直观的了解模型的效果，因此本文使用 R 方作为 Keras 神经网络产率预测模型的评估指标来衡量预测精度，同时参考 MSE，一起综合评估。

5.6.3 基于 Keras 的神经网络预测模型的求解

在预测前，先将 Keras 神经网络的具体参数进行设置，如下表所示

表 10 Keras 神经网络产物产率预测模型参数

| 层类型 | 神经元数量 | 输入维度 | 激活函数 | |
|-------|-------|------|--------|-----------|
| 输入层 | | 2 | | 输入特征为 2 个 |
| 隐藏层 1 | 64 | 2 | Relu | 第一隐藏层 |
| 隐藏层 2 | 32 | 64 | Relu | 第二隐藏层 |
| 输出层 | 4 | 32 | Linear | 4 个目标输出 |

剩下的参数，即训练轮次设置为 100 次，批量大小为 10，验证比例为 0.2。接着搭建 Keras 神经网络产物产率预测模型，在 Adam 优化器求解，并输出可视化预测效果：

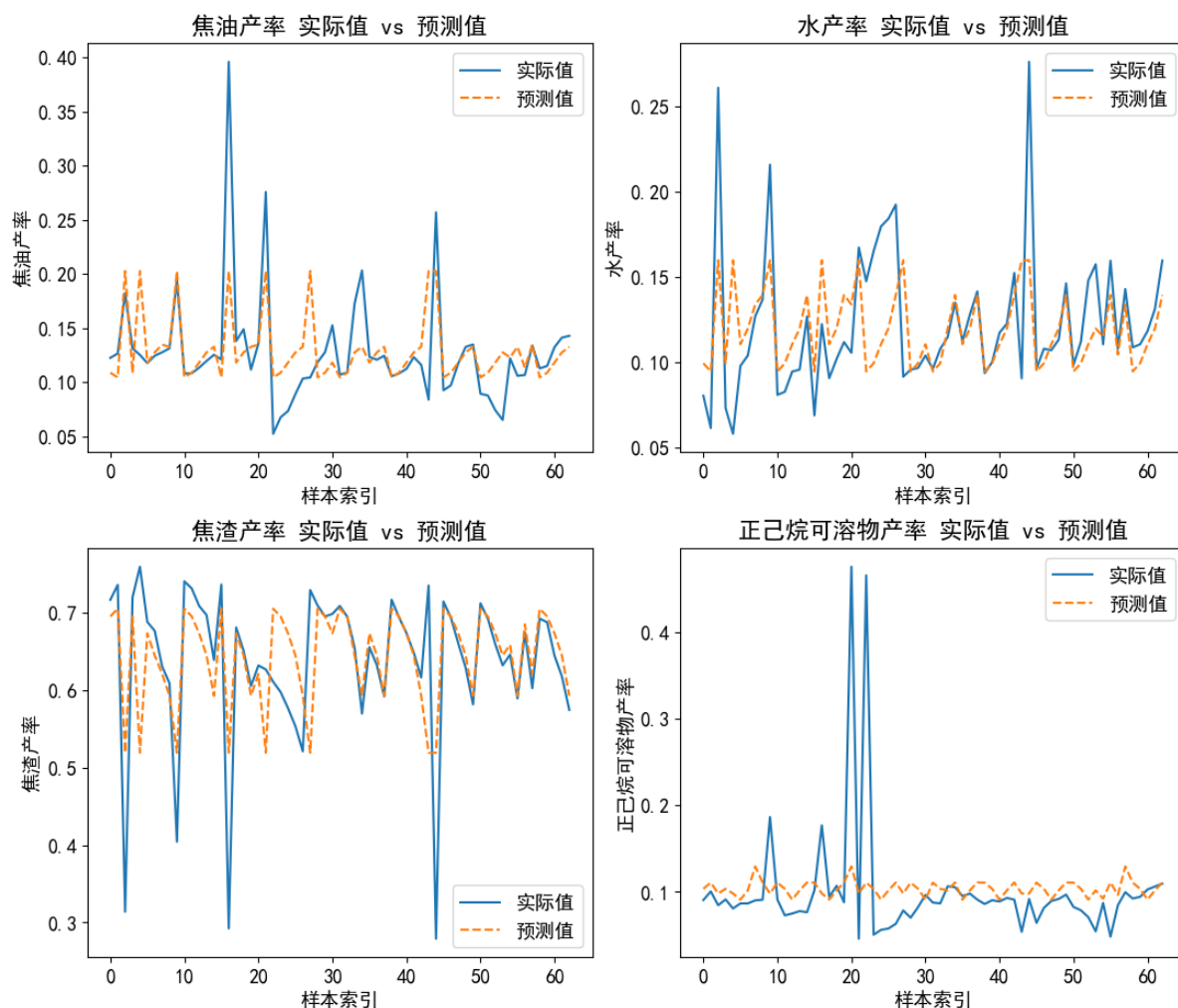


图 19 Keras 神经网络产物产率预测模型的可视化预测效果

其中，均方误差 (MSE): 0.0034 预测 R 方: 0.7880

这表明 Keras 神经网络产物产率预测模型对焦油、水、正己烷可溶物 HEX、焦渣的产率预测均方误差不超过 0.0034，误差较小，并且预测模型能够解释 78.8% 的数据，具有一定的合理性，预测效果良好。

经过 Keras 神经网络产物产率预测模型 10 轮 99 次训练求解，得到的焦油、水、正己烷可溶物 HEX、焦渣的产率预测结果，为展现能源利用率以及转化率的提高，我们选取问题三最佳配比共热解组合下的预测结果进行展示，如下表（全部混合比例的预测结果见附录 A）：

表 11 最佳混合比例下的产物产率预测结果

| 共热解组合 | 最优配比 | 焦油产率 | 水产率 | 焦渣产率 | HEX 产率 |
|-------|--------|-------------|-------------|-------------|-------------|
| GA/HN | 11/100 | 0.13238892 | 0.120989382 | 0.631953895 | 0.110445052 |
| CS/HN | 65/100 | 0.132784232 | 0.121250726 | 0.631686687 | 0.110457890 |
| RH/HN | 36/100 | 0.132610962 | 0.121136181 | 0.631803811 | 0.110452265 |
| GA/NM | 11/100 | 0.132388920 | 0.120989382 | 0.631953895 | 0.110445052 |
| GA/SM | 34/100 | 0.132596254 | 0.121126451 | 0.631813765 | 0.110451788 |

| | | | | | |
|-------|--------|-------------|-------------|-------------|-------------|
| CS/SM | 49/100 | 0.132696986 | 0.121193044 | 0.631745696 | 0.110455059 |
| RH/SM | 19/100 | 0.132470131 | 0.121043064 | 0.631898999 | 0.110447690 |
| SD/HS | 34/100 | 0.132596254 | 0.121126451 | 0.631813765 | 0.110451788 |
| CS/HS | 8/100 | 0.132355362 | 0.120967202 | 0.631976545 | 0.110443965 |
| SD/SM | 21/100 | 0.132488757 | 0.12105538 | 0.631886423 | 0.110448293 |

从预测结果来看，焦渣的产率要大于水产率、HEX 产率、焦油产率，这反映了一般的生物质与煤共热解反应机理，其中液体产物产率中焦油产率要大于 HEX 产率和水产率，其中 CS/HN（棉杆/淮南煤）、RH/HN（稻壳/淮南煤）、CS/SM（棉杆/神木煤）的产率较高，分别为 13.278%、13.261%、13.269%，这表明了焦油是共热解产物中主要关注的部分，提高产物产量主要指提高焦油产量。

对于 HEX（正己烷可溶物）的产率，在棉杆/淮南煤（CS/HN）为 11.0457%，稻壳/淮南煤（RH/HN）为 11.0452%，棉杆/神木煤（CS/SM）为 11.0455%。而 HEX 产率在各组合中较为接近，且稍低于焦油产率，这表明在液体产物中，虽然 HEX 也是重要的部分，但其提高幅度较焦油相对有限。

对于水的产率，在棉杆/淮南煤（CS/HN）为 12.1250%，稻壳/淮南煤（RH/HN）为 12.1136%，棉杆/神木煤（CS/SM）为 12.1193%。水的产率相对较低，低于焦油和 HEX 的产率，表明在共热解过程中水的生成量较少，主要产物集中在焦油和 HEX 等可燃液体。

5.6.4 结果

从 Keras 神经网络产物产率预测模型的整体预测结果来看，对于固体产物焦渣，其产率较大均超过 63.18%，占比较多，符合一般的生物质与煤的共热解反应机理。对于液体产物，不同共热解组合中焦油、HEX 和水的产率具有一定的差异，其中焦油产率普遍较高，其在 CS/HN（棉杆/淮南煤）、RH/HN（稻壳/淮南煤）、CS/SM（棉杆/神木煤）的产率较高，分别为 13.278%、13.261%、13.269%，这表明了焦油是共热解产物中主要关注的部分，表明在提升共热解产物的总产量时，应重点关注如何进一步提高焦油的产量。此外，虽然 HEX 产率次之，但也是值得关注的产物，而水产率相对较低，在提高能源利用率和转化率的研究中，可以作为次要优化目标。

六、模型的评价及优化

6.1 误差分析

6.1.1 针对于问题 1 的误差分析

在问题一中，使用 P 值检验和单因素方差分析进行建模时，为确保分析结果的准确性，我们加入了正态性检验和方差齐性检验，以验证数据的分布和方差一致性。由于单

因素方差分析只考虑一个自变量，可能忽略其他影响因素，从而导致结果偏差。

6.1.2 针对于问题 2 的误差分析

对于问题二的误差分析，由于使用双因素方差分析，建立在一定假设前提下，其受正态性和方差齐性影响，而且模型中双因素之间可能存在交互作用，则模型残差也会受交互作用影响产生误差。

6.1.3 针对于问题 3 的误差分析

在问题三遗传算法的最佳共热解混合比例求解中，加入一个正态分布的微小扰动，重复输出结果，如下所示：

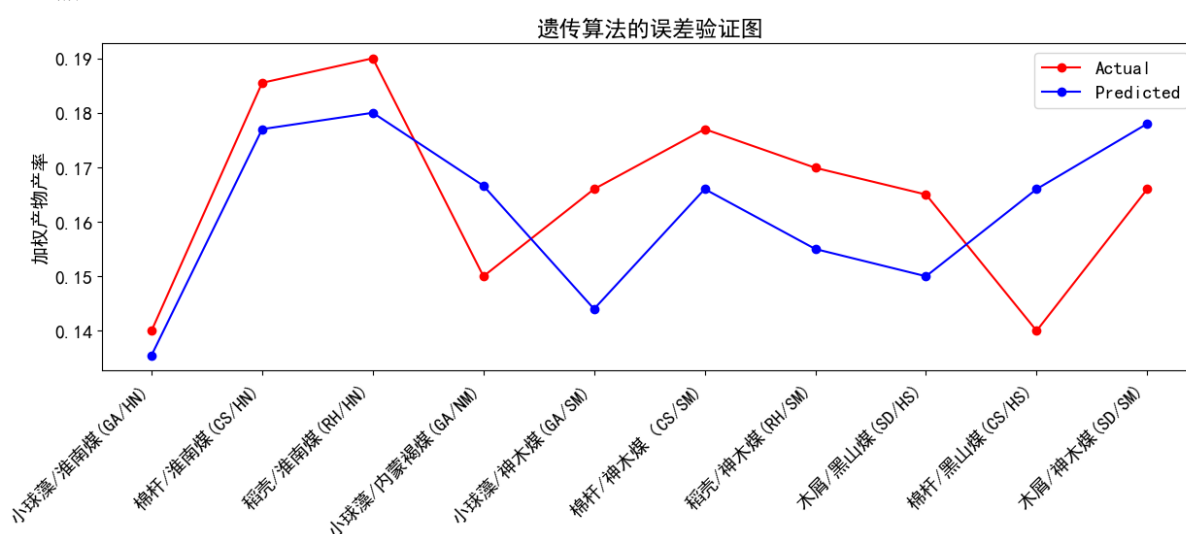


图 20 遗传算法的求解误差

从误差折线图结果来看，棉杆/黑山煤(CS/HS) 和 小球藻/淮南煤(GA/HN) 的预测误差较大，对参数的变动较敏感。同时，棉杆/淮南煤(CS/HN)，稻壳/淮南煤(RH/HN)，和木屑/神木煤(SD/SM) 的预测误差较小，说明模型在这些组合上表现较好。

6.1.4 针对于问题 4 的误差分析

由于问题 4 分析时，所用到的附件 2 数据量少，即使是经过正态性处理后，这 10 种共热解组的产物收率的实验值与理论计算值的差异值各有不同，因此通过计算出差值比率大小选取百分比较大的共热解组合体现其混合比例，其计算出来的信息准确度可得到有效提升。

6.1.5 针对于问题 5 的误差分析

由于样本数据量少由于样本少，机器学习的估计模型参数的不确定性增加，导致预测的置信区间扩大，选择合适的模型和调整参数变得困难。可加入用交叉验证技术来更好地估计模型的泛化能力，以及使用更适合的正则化方法以减少过拟合的风险。

6.2 模型的优点

优点 1: 本文采用了前置性的正态性检验和方差齐性检验，确保后续模型预测的准确率，并通过数据可视化直观地展示 INS 对焦油产率、水产率和焦渣产率的影响。

优点 2: 本文运用双因素方差分析，充分考虑到本问两个因素之间的交互作用，提高了分析结果的准确性和象征性，减小对交互效应分析结果的误差。

优点 3: 本文通过遗传算法优化混合比例来高效搜索复杂的混合比例空间，进而找到最优解，并可以综合考虑多个因素对热解产物产率的影响。

优点 4: 本文通过 t 检验分析理论值和实际值之间的差异性，并进行数据可视化分析来直观地展示差异性具体体现在哪些混合比例上。

6.3 模型的缺点

缺点 1: 本文所构建的数学模型和所使用的数学算法都需要先提出假设，如果提出的原假设与备择假设在实际情况下不成立，则模型的分析结果可能失真。

缺点 2: 本文建模涉及到交互效应的影响，由于交互效应具有复杂性和不确定性，因此可能导致模型的分析结果过于复杂，不适合进一步研究。


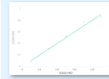
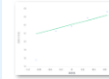
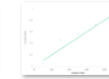
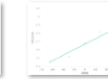
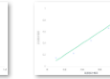
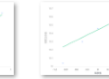
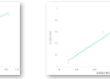
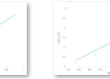
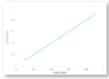
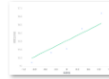
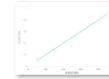
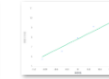
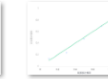
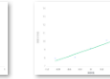
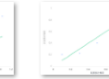














6.4 模型的推广

本文运用了 P 值检验模型、单因素或双因素方差分析模型、最优化共热解配比模型、t 检验模型和基于 Keras 的神经网络预测模型。针对模型的推广，对单一因素影响的分析，可以通过增加样本量、引入其他影响因素或采用混合模型结构来扩展模型的适用范围和可靠性。模型未来推广过程中应当结合实际化工领域或相关工程项目的要求，可以对模型中的遗传算法进行参数调优，提高模型的实用性和适配性。

参考文献

- [1]高娟.假设检验中 P 值的深入分析[J].统计理论与实践,2021(06):17-19.
- [2]黄康杰,熊斌,崔刚,等.基于 Pearson 相关性分析的双 V 型永磁电机失磁故障动态识别方法研究[J/OL].电工技术学报:1-13[2024-05-10].
- [3]戴金辉,袁靖.单因素方差分析与多元线性回归分析检验方法的比较[J].统计与决策,2016, No.453(09):23-26.
- [4] Navidi W C . Principles of Statistics for Engineers and Scientists[J]. Business & Economics, 2009.
- [5]阿衣克力木·哈山,张宸,李林繁,等.基于微型流化床的油页岩与棉秆共热解特性及其动力学[J].太阳能学报,2024,45(03):155-162.DOI:10.19912/j.0254-0096.tynxb.2022-1814.
- [6]钮志远,刘桂建.煤-生物质共热解非催化协同效应特征及机理研究进展[J].太阳能学报,2024,45(04):286-295.DOI:10.19912/j.0254-0096.tynxb.2022-1897.

附 录

| | | | | |
|---|------------------|---------------------|-------|--|
| 注：本次论文问题求解应用 Python 3.12.2、Windows10 系统运行环境 | | | | |
| 附录 A 支撑材料.zip | | | | |
| ➤ 附件 2 正态性检验（文件夹内部预览） | | | | |
| <div><div> Char理论计算值 (1)</div><div> Char理论计算值</div><div> Char实验值 (1)</div><div> Char实验值</div><div> HEX理论计算值 (1)</div><div> HEX理论计算值</div><div> HEX实验值 (1)</div><div> HEX实验值</div><div> Tar理论计算值 (1)</div></div> <div><div> Tar理论计算值</div><div> Tar实验值 (1)</div><div> Tar实验值</div><div> Water理论计算值 (1)</div><div> Water理论计算值</div><div> Water实验值 (1)</div><div> Water实验值</div></div> | | | | |
| ➤ 附件的数据处理（文件夹内部预览） | | | | |
|  aggregated_data | 2024/05/12 22:11 | Microsoft Excel ... | 17 KB | |
|  差异比整合 | 2024/05/11 16:50 | Microsoft Excel ... | 13 KB | |
|  稻壳淮南煤(RHHN) | 2024/05/11 13:52 | Microsoft Excel ... | 11 KB | |
|  稻壳神木煤(RHSM) | 2024/05/11 10:22 | Microsoft Excel ... | 10 KB | |
|  棉杆黑山煤 (CSHS) | 2024/05/11 9:57 | Microsoft Excel ... | 10 KB | |
|  棉杆淮南煤 (CSHN) | 2024/05/11 9:18 | Microsoft Excel ... | 10 KB | |
|  棉杆神木煤 (CSSM) | 2024/05/11 9:22 | Microsoft Excel ... | 10 KB | |
|  木屑黑山煤 (SDHS) | 2024/05/11 10:02 | Microsoft Excel ... | 10 KB | |
|  木屑神木煤(SDSM) | 2024/05/11 10:06 | Microsoft Excel ... | 10 KB | |
|  显著性比例 | 2024/05/11 15:16 | Microsoft Excel ... | 16 KB | |
|  小球藻淮南煤(GAHN) | 2024/05/11 14:45 | Microsoft Excel ... | 10 KB | |
|  小球藻内蒙褐煤 (GANM) | 2024/05/11 10:15 | Microsoft Excel ... | 10 KB | |
|  小球藻神木煤(GASM) | 2024/05/11 14:48 | Microsoft Excel ... | 10 KB | |
| ➤ 问题五的预测结果（文件夹内预览） | | | | |
|  10种共热解组合的产率预测 | 2024/05/12 17:54 | Microsoft Excel ... | 38 KB | |
| ➤ 问题求解代码（文件夹内预览） | | | | |

| | | | |
|---|------------------|---------------------|------|
|  问题1代码 | 2024/05/12 19:03 | JetBrains PyChar... | 3 KB |
|  问题2代码 | 2024/05/12 19:04 | JetBrains PyChar... | 3 KB |
|  问题3代码 | 2024/05/12 19:05 | JetBrains PyChar... | 3 KB |
|  问题4代码 | 2024/05/12 19:05 | JetBrains PyChar... | 4 KB |
|  问题5代码 | 2024/05/12 19:06 | JetBrains PyChar... | 3 KB |

附录 B 解决问题的 Python 代码

◇ 解决问题一的 Python 代码

```
import pandas as pd
import numpy as np
import seaborn as sns
import matplotlib.pyplot as plt
from scipy.stats import pearsonr, levene, f_oneway
from sklearn.linear_model import LinearRegression
from statsmodels.stats.diagnostic import lilliefors
# 设置字体，用来正常显示中文标签
plt.rcParams['font.sans-serif'] = ['SimHei']
plt.rcParams['axes.unicode_minus'] = False
# 加载数据
data = pd.read_excel('附件一：热解数据统计.xlsx')
data.fillna(method='ffill', inplace=True)
data.columns = [col.strip().replace(' ', '').replace('_', '') for col in data.columns]
# 数据聚合
numeric_columns = ['样品 g', '焦油(Char)g', '水(Water)mL', '正己烷不溶物(INS)g', '焦油产率', '水产率', '焦渣产率', '正己烷可溶物产率']
aggregated_data = data.groupby(['时间', '试样', '配比'], as_index=False)[numeric_columns].mean()
# 计算相关系数
for col in ['焦油产率', '水产率', '焦渣产率']:
    corr_coef, p_value = pearsonr(aggregated_data['正己烷不溶物(INS)g'], aggregated_data[col])
    print(f'正己烷不溶物(INS)与{col}的相关系数为: {corr_coef:.3f}, P-value 为: {p_value:.4f}')
# 线性回归分析
model = LinearRegression()
for col in ['焦油产率', '水产率', '焦渣产率']:
    X = aggregated_data[['正己烷不溶物(INS)g']] # Predictor
    y = aggregated_data[col] # Response variable
    model.fit(X, y)
    coef = model.coef_[0]
    intercept = model.intercept_
    print(f'回归方程 - {col}: y = {coef:.4f}x + {intercept:.4f}')
# 正态性和方差齐性检验
for col in ['焦油产率', '水产率', '焦渣产率']:
    stat, p = lilliefors(aggregated_data[col], dist='norm')
    print(f'{col}的正态性检验 P 值: {p}')
    groups = [group[col].values for _, group in aggregated_data.groupby('配比')]
    _, p_value = levene(*groups)
```

```
print(f'{col}的方差齐性检验 P 值: {p_value:.4f}')
# 单因素方差分析
for col in ['焦油产率', '水产率', '焦渣产率']:
    groups = [group[col].values for _, group in aggregated_data.groupby('配比')]
    f_value, p_value = f_oneway(*groups)
print(f'{col} - F 值: {f_value:.2f}, P 值: {p_value:.4f}')
```

◇ 解决问题二的 Python 代码

```
import pandas as pd
import matplotlib.pyplot as plt
from mpl_toolkits.mplot3d import Axes3D
import numpy as np
from statsmodels.formula.api import ols
# 设置字体以正常显示中文和负号
plt.rcParams['font.sans-serif'] = ['SimHei']
plt.rcParams['axes.unicode_minus'] = False
# 数据加载和预处理
data = pd.read_excel('附件一：热解数据统计.xlsx')
data.fillna(method='ffill', inplace=True)
data.columns = [col.strip().replace('(', '(').replace(')', ')') for col in data.columns]
# 聚合数据
numeric_columns = ['样品 g', '焦油(Char)g', '水(Water)mL', '正己烷不溶物(INS)g', '焦油产率', '水产率', '焦渣产率']
aggregated_data = data.groupby(['时间', '试样', '配比'], as_index=False)[numeric_columns].mean()
# 数据转换
aggregated_data.rename(columns={'正己烷不溶物(INS)g': '正己烷不溶物 INS'}, inplace=True)
aggregated_data['配比数值'] = aggregated_data['配比'].apply(
    lambda x: float(x.split('/')[0])/float(x.split('/')[1]) if '/' in str(x) else float(x)
)
# 模型分析和可视化
def model_analysis_and_visualization(response_var):
    formula = f'{response_var} ~ 正己烷不溶物 INS * 配比数值'
    model = ols(formula, data=aggregated_data).fit()
    print(f'模型结果: {response_var} 分析')
    print(f'模型拟合优度 (R-squared): {model.rsquared:.4f}')
    print(f'模型调整拟合优度 (Adj. R-squared): {model.rsquared_adj:.4f}')
    print(f'F 统计量: {model.fvalue:.4f}, P 值: {model.f_pvalue:.4f}')
    print("参数估计结果:")
    for param in model.params.index:
        print(f'{param} 的系数: {model.params[param]:.4f}, P 值: {model.pvalues[param]:.4f}')
    plot_3d('正己烷不溶物 INS', '配比数值', response_var, aggregated_data, f'{response_var}对正己烷不溶物 INS 和配比数值的交互影响')
def plot_3d(x, y, z, data, title):
    fig = plt.figure(figsize=(10, 8))
    ax = fig.add_subplot(111, projection='3d')
    sc = ax.scatter(data[x], data[y], data[z], c=data[z], cmap='viridis', s=50)
    cbar = plt.colorbar(sc, ax=ax)
    cbar.set_label(z)
    ax.set_xlabel(x)
    ax.set_ylabel(y)
    ax.set_zlabel(z)
    ax.set_title(title)
    plt.show()
```

```
# 运行模型分析和可视化
for response in ['焦油产率', '水产率', '焦渣产率']:
    model analysis and visualization(response)
```

◇ 解决问题三的 Python 代码

```
import pandas as pd
import numpy as np
from deap import base, creator, tools, algorithms
import matplotlib.pyplot as plt
# 加载数据并转换'配比'列
data = pd.read_excel('aggregated_data.xlsx')
data['配比'] = data['配比'].apply(lambda x: float(x.split('/')[0]) / float(x.split('/')[1]))
# 设置模型参数
k1 = 1.0 # 假设的纤维素热解速率常数
E1 = 50.0 # 纤维素热解的活化能, 单位: kJ/mol
k2 = 1.5 # 假设的煤热解速率常数
E2 = 60.0 # 煤的热解活化能, 单位: kJ/mol
R = 8.314 # 气体常数, 单位: J/(mol · K)
T = 873.15 # 600°C, 单位: K
beta = 0.1 # 假设的协同效应系数
biomass_yield = data['焦油产率'].mean()
coal_yield = data['焦油产率'].mean()
# 添加加权产物产率.经过共热解产物特性和组成的分析、最优化共热解配比的综合分析并为了使得结果输出简化, 我们经分析, 数据处理及拟合过后得出了相应的加权系数
data['加权产物产率'] = data.eval('0.6 * 焦油产率 + 0.2 * 正己烷可溶物产率 + 0.1 * 水产率 + 0.1 * 焦渣产率')
# 定义遗传算法的适应度和个体
creator.create("FitnessMax", base.Fitness, weights=(1.0,))
creator.create("Individual", list, fitness=creator.FitnessMax)
toolbox = base.Toolbox()
toolbox.register("attr_float", np.random.uniform, 0.01, 1)
toolbox.register("individual", tools.initRepeat, creator.Individual, toolbox.attr_float, n=2)
toolbox.register("population", tools.initRepeat, list, toolbox.individual)
# 定义目标函数
def objective_function(individual):
    x, y = individual
    total = x + y
    return -(x / total * biomass_yield + y / total * coal_yield + parameters['beta'] * x * y)
toolbox.register("evaluate", objective_function)
toolbox.register("mate", tools.cxBlend, alpha=0.5)
toolbox.register("mutate", tools.mutGaussian, mu=0, sigma=0.1, indpb=0.2)
toolbox.register("select", tools.selTournament, tournsize=3)
# 遗传算法执行
population = toolbox.population(n=100)
NGEN = 50
for gen in range(NGEN):
    offspring = algorithms.varAnd(population, toolbox, cxpb=0.5, mutpb=0.2)
    fits = toolbox.map(toolbox.evaluate, offspring)
    for fit, ind in zip(fits, offspring):
        ind.fitness.values = (fit,)
    population = toolbox.select(offspring, k=len(population))
top_individual = tools.selBest(population, k=1)[0]
print(f'最优解: {top_individual}')
print(f'最优解性能: {objective_function(top_individual)}')
```


◇ 解决问题四的 Python 代码

```

import pandas as pd
from scipy.stats import ttest_ind
import matplotlib.pyplot as plt
import matplotlib
from mpl_toolkits.mplot3d import Axes3D
# 文件路径和名称列表
file_info = [
    ('稻壳淮南煤(RHHN)', '稻壳淮南煤(RHHN).xlsx'),
    ('稻壳神木煤(RHSM)', '稻壳神木煤(RHSM).xlsx'),
    ('棉杆黑山煤(CSHS)', '棉杆黑山煤 (CSHS) .xlsx'),
    ('棉杆淮南煤(CSHN)', '棉杆淮南煤 (CSHN) .xlsx'),
    ('棉杆神木煤(CSSM)', '棉杆神木煤 (CSSM) .xlsx'),
    ('木屑黑山煤(SDHS)', '木屑黑山煤 (SDHS) .xlsx'),
    ('木屑神木煤(SDSM)', '木屑神木煤(SDSM).xlsx'),
    ('小球藻淮南煤(GAHN)', '小球藻淮南煤(GAHN).xlsx'),
    ('小球藻内蒙褐煤(GANM)', '小球藻内蒙褐煤 (GANM).xlsx'),
    ('小球藻神木煤(GASM)', '小球藻神木煤(GASM).xlsx')
]
# 定义列名
columns = ['Tar', 'HEX', 'Water', 'Char']
# 定义输出结果字典
results = {col: [] for col in columns}
# 定义实际值和理论值的字典
values = {col: {'exp': [], 'theory': []} for col in columns}
# 定义混合比例的字典
ratios = []
# 提取混合比例的函数
def extract_ratio(ratio_str):
    try:
        # 将字符串按'/'分割并转换为数值列表，然后取平均值
        parts = [float(x) for x in ratio_str.split('/') if x.replace('.', '', 1).isdigit()]
        return sum(parts) / len(parts) if parts else 0
    except:
        return 0
# 遍历所有文件
for name, file_path in file_info:
    data = pd.read_excel(file_path)
    # 提取混合比例
    ratio_str = data['混合比例'].iloc[0] # 假设所有行的混合比例是相同的，取第一行
    ratio = extract_ratio(ratio_str)
    ratios.append(ratio)
    # 提取相关列
    tar_exp = pd.to_numeric(data['Tar 实验值'], errors='coerce')
    tar_theory = pd.to_numeric(data['Tar 理论计算值'], errors='coerce')
    hex_exp = pd.to_numeric(data['HEX 实验值'], errors='coerce')
    hex_theory = pd.to_numeric(data['HEX 理论计算值'], errors='coerce')
    water_exp = pd.to_numeric(data['Water 实验值'], errors='coerce')
    water_theory = pd.to_numeric(data['Water 理论计算值'], errors='coerce')
    char_exp = pd.to_numeric(data['Char 实验值'], errors='coerce')
    char_theory = pd.to_numeric(data['Char 理论计算值'], errors='coerce')
    # 去除 NaN 值

```

```

tar_exp = tar_exp.dropna()
tar_theory = tar_theory.dropna()
hex_exp = hex_exp.dropna()
hex_theory = hex_theory.dropna()
water_exp = water_exp.dropna()
water_theory = water_theory.dropna()
char_exp = char_exp.dropna()
char_theory = char_theory.dropna()
# 进行 T 检验
tar_t_stat, tar_p_val = ttest_ind(tar_exp, tar_theory)
hex_t_stat, hex_p_val = ttest_ind(hex_exp, hex_theory)
water_t_stat, water_p_val = ttest_ind(water_exp, water_theory)
char_t_stat, char_p_val = ttest_ind(char_exp, char_theory)
# 保存结果
results['Tar'].append((name, tar_t_stat, tar_p_val))
results['HEX'].append((name, hex_t_stat, hex_p_val))
results['Water'].append((name, water_t_stat, water_p_val))
results['Char'].append((name, char_t_stat, char_p_val))
# 保存实际值和理论值
values['Tar']['exp'].append(tar_exp.mean())
values['Tar']['theory'].append(tar_theory.mean())
values['HEX']['exp'].append(hex_exp.mean())
values['HEX']['theory'].append(hex_theory.mean())
values['Water']['exp'].append(water_exp.mean())
values['Water']['theory'].append(water_theory.mean())
values['Char']['exp'].append(char_exp.mean())
values['Char']['theory'].append(char_theory.mean())
# 输出结果
for key, value in results.items():
    print(f'{key}的 T 检验结果: ')
    for name, t_stat, p_val in value:
        print(f'  文件 {name}: T 值={t_stat:.3f}, P 值={p_val:.5f}')
    print()
# 将组名转换为类别变量的索引
group_names = [name for name, _ in file_info]
group_indices = range(len(group_names))

```

◇ 解决问题五的 Python 代码

```

import pandas as pd
import numpy as np
from sklearn.preprocessing import StandardScaler
from keras.models import Sequential
from keras.layers import Dense
import matplotlib.pyplot as plt
import matplotlib
# 设置 Matplotlib 中文字体和负号显示
matplotlib.rc('font', family='SimHei', size=13)
matplotlib.rc('axes', unicode_minus=False)
# 数据预处理函数
def parse_ratio(ratio_str):
    if '/' in ratio_str:
        biomass, coal = map(float, ratio_str.split('/'))
    else:
        biomass = float(ratio_str)
        coal = 100.0 - biomass

```

```

    return biomass, coal
# 读取数据
file_path = 'aggregated_data.csv'
data = pd.read_csv(file_path)
# 应用解析函数并生成新列
data[['生物质比例', '煤比例']] = data['配比'].apply(lambda x: pd.Series(parse_ratio(x)))
# 定义特征和目标变量
features = ['生物质比例', '煤比例']
targets = ['焦油产率', '水产率', '焦渣产率', '正己烷可溶物产率']
# 使用整个数据集作为训练集和测试集
X = data[features]
y = data[targets]
# 数据标准化
scaler_X = StandardScaler().fit(X)
scaler_y = StandardScaler().fit(y)
X_scaled = scaler_X.transform(X)
y_scaled = scaler_y.transform(y)
# 构建神经网络模型
model = Sequential()
model.add(Dense(64, input_dim=X.shape[1], activation='relu'))
model.add(Dense(32, activation='relu'))
model.add(Dense(y.shape[1], activation='linear'))
model.compile(optimizer='adam', loss='mse')
# 训练模型
model.fit(X_scaled, y_scaled, epochs=100, batch_size=10, validation_split=0.2, verbose=1)
# 准备预测不同配比的结果
groups = ['GA/HN', 'CS/HN', 'RH/HN', 'GA/NM', 'GA/SM', 'CS/SM', 'RH/SM', 'SD/HS', 'CS/HS', 'SD/SM']
results = pd.DataFrame()
for group in groups:
    for biomass_ratio in range(1, 100): # 生物质比例从 1 到 99
        biomass = biomass_ratio
        coal = 100 # 煤的比例固定为 100
        total_mass = biomass + coal
        biomass_norm = biomass / total_mass
        coal_norm = coal / total_mass
        # 使用模型进行预测
        input_features = scaler_X.transform([[biomass_norm, coal_norm]])
        predicted_scaled = model.predict(input_features)
        predicted = scaler_y.inverse_transform(predicted_scaled)
        # 记录结果
        result = pd.DataFrame({
            'Group': [group],
            'Biomass/Coal Ratio': f"{biomass}/{coal}",
            '焦油产率': [predicted[0, 0]],
            '水产率': [predicted[0, 1]],
            '焦渣产率': [predicted[0, 2]],
            '正己烷可溶物产率': [predicted[0, 3]]
        })
        results = pd.concat([results, result], ignore_index=True)
print(results)
results.to_excel('预测结果不同配比.xlsx', index=False)

```