

2024 年第九届“数维杯”大学生 数学建模挑战赛论文

题 目 生物质和煤共热问题的研究

摘 要

随着可再生能源需求的增加，生物质与煤的共热解技术受到了广泛关注。本研究聚焦于共热解过程中的关键问题，旨在提高能源转化效率和产物利用率。针对问题 1，我们采用 ANOVA 方差分析法，以探究不同 INS 水平对焦油、水和焦渣产率的影响；针对问题 2，运用双因素方差分析评估 INS 和混合比例对产物产量的综合效应；针对问题 3，则通过建立优化模型，构建了随机森林模型进行回归预测，并且在模型中采用多目标优化算法。针对问题 4，我采用 T 检验对比了实验值与理论计算值。针对问题 5，通过建立随机森林预测模型进行回归预测。

对于问题 1，我们构建了产率分析模型 I。通过 ANOVA 方差分析，我们发现不同 INS 水平下焦油、水和焦渣产率存在显著差异，这为后续优化提供了方向。

对于问题 2，双因素方差分析结果显示，INS 和混合比例均显著影响产物产量，且它们之间存在交互作用。通过交互作用图，我们直观地展示了这种关系。

对于问题 3，我们建立了一个基于多目标优化的混合比例模型。该模型以产物利用率和能源转化效率最大化为目标，通过散点图展示了混合比例与中等 INS 水平下产物产量及能源转化效率之间的关系，进而确定了最优的混合比例。

对于问题 4，我们使用 T 检验对比了实验值与理论计算值，结果显示两者之间存在显著差异。这一发现促使我们进一步审视和优化理论模型。

对于问题 5，我们建立了随机森林预测模型，用于预测热解产物产率。通过散点图和预测结果与实际结果的对比，验证了模型的预测性能。

综上所述，基于以上研究，通过多种分析方法和模型求解，深入研究了生物质和煤共热解问题，为相关领域的进一步研究和实践提供了理论支持和方法参考，这些建议为未来的研究和实践提供了方向。

关键词：ANOVA；双因素方差分析；回归拟合；预测模型；随机森林模型

目 录

一、问题重述	(1)
二、问题分析	(2)
2.1 问题 1 的模型	(2)
2.2 问题 2 的模型	(3)
2.3 问题 3 的模型	(4)
2.4 问题 4 的模型	(5)
2.5 问题 5 的模型	(6)
三、模型假设	(7)
四、定义与符号说明	(8)
五、模型的建立与求解	(8)
5.1 问题 1 的模型建立与求解	(8)
5.1.1 分析-拟合模型的建立	(8)
5.1.2 分析-拟合模型求解	(9)
5.1.3 结论	(12)
5.2 问题 2 的模型建立与求解	(12)
5.2.1 配比-正己烷双因素方差分析模型的建立	(12)
5.2.2 配比-正己烷双因素方差分析模型求解	(14)
5.2.3 结论	(16)
5.3 问题 3 的模型建立与求解	(17)
5.3.1 产率-品质模型的建立	(17)
5.3.2 产率-品质模型求解	(18)
5.3.3 结论	(19)
5.4 问题 4 的模型建立与求解	(20)
5.4.1 模型的建立	(20)
5.4.2 模型求解	(20)
5.4.3 结论	(20)
5.5 问题 5 的模型建立与求解	(22)
5.5.1 投入-产率模型的建立	(22)
5.5.2 投入-产率模型求解	(23)
5.5.3 结论	(24)
六、模型的分析与检验	(24)
6.1 误差分析	(24)
6.2 模型的检验	(25)
七、模型的优缺点、改进和推广	(26)
7.1 模型的优点	(26)
7.2 模型的缺点	(26)
7.3 模型的改进	(27)
7.4 模型推广	(27)
参考文献	(28)
附录	(29)

一、问题重述

1.问题背景

在全球能源结构转型的背景下，对可再生能源和环境友好型能源的追求日益增强。生物质，作为一种重要的可再生能源，来源于广泛存在的植物和动物有机物质；而煤，作为一种传统化石燃料，虽属非可再生资源，但在当前能源结构中仍占据重要地位。共热解技术，作为实现两者高效转化的途径之一，旨在通过高温缺氧环境下的协同热解，产生气体、液体和固体产物，其中液体产物即热解油或生物油，具有广阔的应用前景。

本研究的核心在于深入理解并优化生物质与煤共热解过程中的产物生成机理，以提升能源利用效率，促进资源综合利用，并保障能源安全。为此，我们选取了多种生物质和中低阶煤作为共热解原料，在化工实验室条件下，利用管式干馏炉在温和热解条件下进行了系统研究。

在实验中，常常以微晶纤维素作为生物质中主要纤维素成分的模型化合物，用以模拟和代表生物质的热解特性。通过引入微晶纤维素作为模型化合物，分析比较了不同热解条件下产物的组成和特性，并探讨了生物质热解的化学反应机理。此外，通过设置不同生物质与煤的混合比例进行共热解实验，以观察和分析混合比例对产物分布的影响，为实现高效共热解提供科学依据。

2.需要解决的问题

问题 1：基于附件一中的数据，请利用统计方法分析正己烷不溶物（INS）对热解产率（主要考虑焦油、水和焦渣产率）是否具有显著影响。并且利用图像和适当的统计指标解释结果，以便确定 INS 对不同产率的影响程度。

问题 2：在热解实验中，正己烷不溶物（INS）和混合比例是否存在交互效应，对热解产物产量产生重要影响？如果存在交互效应，请通过适当的数据分析方法确定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显著。

问题 3：根据附件一中共热解产物的特性和组成，提出一个数学模型，以优化共解热混合比例为目标，从而提高产物利用率和能源转化效率。请详细描述模型的建立过程，并提供相关的优化算法或方法。

问题 4：根据附件二的实验数据，比较每种共热解组合的产物收率实验值与理论计

算值 之间是否存在显著性差异？如果存在差异，请通过对不同共热解组合的数据进行子组分析，以确定实验值与理论计算值之间的差异在哪些混合比例上体现。

问题 5：利用附件中的实验数据，建立一个合适的数学模型，以对热解产物的产率进行预测。请描述模型的结构和建立过程，并说明模型的预测能力和适用范围。

二、问题分析

2.1 问题 1 的分析

对研究问题 1 研究意义的分析：问题 1 关注的是正己烷不溶物（INS）对热解产率（焦油产率、水产率、焦渣产率）的影响。这一问题的研究对于深入理解生物质与煤共热解过程中的反应机理、优化热解条件以及提高产物利用率具有重要意义。通过探究 INS 对热解产率的影响，可以为共热解过程的优化提供理论依据，进而促进生物质与煤共热解技术的工业应用。

对于此类问题，通常采用的数学方法包括统计分析、回归分析以及方差分析 ANOVA 等。这些方法可以帮助我们量化不同因素（如 INS 含量）对目标变量（如热解产率）的影响程度，通过绘制柱状图或箱线图来可视化不同 INS 水平下的产率数据分布情况，来比较不同 INS 水平下的热解产率（焦油、水和焦渣产率）之间的差异，使分析结果更加直观易懂。

对附件中所给的数据特点分析：附件一提供了不同原料（包括生物质和煤）在不同配比下共热解的实验数据，包括正己烷可溶物（HEX）和正己烷不溶物（INS）的含量以及相应的热解产率（焦油产率、水产率、焦渣产率）。这些数据具有多样性和复杂性，涵盖了多种原料和配比下的实验结果，为我们分析 INS 对热解产率的影响提供了丰富的信息。

首先，我们可以使用方差分析（ANOVA）来比较不同 INS 水平下热解产率的差异。通过计算 F 值和 P 值，我们可以判断不同 INS 水平下的产率之间是否存在显著差异。如果 P 值小于设定的显著性水平（如 0.05），则说明 INS 对热解产率有显著影响。其次，我们可以建立回归模型来量化 INS 对热解产率的影响程度。通过回归分析，我们可以得到 INS 与热解产率之间的数学关系式，并据此预测不同 INS 水平下的热解产率。这有助于我们更深入地理解 INS 对热解产率的影响机制。最后，我们将利用上述两个模型对结果进行预测，并将预测结果进行比较。通过比较不同模型下的预测结果，我们可以评估模型的准确性和可靠性，并据此选择最优的模型来指导实际生产中的热解过程优化。

2.2 问题 2 的分析

对问题 2 研究的意义分析：问题 2 关注的是热解实验中正己烷不溶物（INS）和混合比例之间是否存在交互效应，以及这种交互效应对热解产物产量的具体影响。这一问题的研究对于深入理解生物质与煤共热解过程、优化热解条件、提高产物利用率以及减少环境污染和资源浪费具有重要意义。通过揭示 INS 和混合比例之间的交互效应，我们可以更准确地预测和控制热解产物的产量和品质，为工业生产提供理论指导和实验依据。

对于此类涉及多因素交互效应的问题，通常可以采用多元线性回归分析、方差分析（ANOVA）等统计方法进行分析。这些方法可以帮助我们识别出哪些因素对热解产物产量有显著影响，以及这些因素之间是否存在交互效应。在本问题中，我们特别关注 INS 和混合比例之间的交互效应，因此需要使用能够处理这种交互效应的统计方法。

对附件中所给数据特点的分析：附件中提供了不同混合比例和正己烷处理水平下的热解产物产量数据，包括焦油、水和焦渣的产量。这些数据具有多个特点：首先，数据是按照不同的混合比例和正己烷处理水平进行分组的，这为我们分析交互效应提供了基础；其次，数据包含了多种热解产物的产量信息，这有助于我们全面了解热解过程；最后，数据可能存在一些噪声或异常值，这需要在分析过程中进行适当处理。

对问题 2 所要求的结果进行分析：问题 2 要求分析 INS 和混合比例之间是否存在交互效应，并确定在哪些具体的热解产物上这种交互效应最为明显。为了回答这个问题，我们需要首先建立一个包含 INS、混合比例以及它们之间交互项的统计模型。然后，通过方差分析等方法检验这些项的显著性，以确定它们对热解产物产量的影响程度。最后，根据分析结果绘制交互作用图或分析交互作用项的系数，以直观地展示 INS 和混合比例之间的交互效应以及这种效应在不同热解产物上的表现。

由于以上原因，我们可以首先建立一个多元线性回归模型，该模型包含 INS、混合比例以及它们之间的交互项作为自变量，热解产物产量（如焦油、水、焦渣）作为因变量。通过模型参数的估计和显著性检验，我们可以初步判断 INS 和混合比例之间是否存在交互效应以及这种效应对热解产物产量的影响程度。

然后，为了更准确地评估交互效应的重要性并确定其在不同热解产物上的表现，我们可以建立一个方差分析模型。该模型将按照不同的混合比例和正己烷处理水平对数据进行分组，并计算各组内的方差和组间的方差。通过比较这些方差的大小和显著性水平，我们可以确定 INS 和混合比例之间的交互效应是否显著，并绘制交互作用图以直观地展示这种效应。

最后，我们将比较两个模型的结果，并综合考虑它们的优缺点和适用范围，以得出最终的结论和建议。

2.3 问题 3 的分析

对问题 3 研究的意义的分析：问题 3 关注的是如何通过数学建模优化生物质与煤共热解的混合比例，以提高产物利用率和能源转化效率。这一研究对于推动可再生能源技术的发展、提高能源利用效率、减少环境污染和资源浪费具有重要意义。通过优化共热解混合比例，我们可以更有效地利用生物质和煤这两种资源，实现能源的高效转化和可持续利用。

问题 3 属于优化问题，特别是多目标优化问题。在解决此类问题时，通常需要建立数学模型，将实际问题转化为数学表达式，并确定目标函数和约束条件。目标函数通常用于描述我们希望优化的性能指标，如产物利用率和能源转化效率。约束条件则用于限制决策变量的取值范围，确保解的可行性。对于多目标优化问题，常用的数学方法包括线性规划、非线性规划、整数规划、遗传算法等。这些方法可以帮助我们在满足约束条件的前提下，找到使目标函数达到最优的解。

对附件中所给数据特点的分析：附件中提供了关于不同生物质和煤共热解实验的数据，包括原料种类、混合比例、产物组成等信息。这些数据具有多样性和复杂性，涵盖了从单一原料热解到多种原料共热解的不同情况。通过对这些数据的分析，我们可以了解不同原料和混合比例对共热解产物的影响，为建立数学模型和优化混合比例提供基础。

对问题 3 所要求的结果进行分析：问题 3 要求建立数学模型来优化共热解混合比例，以提高产物利用率和能源转化效率。这意味着我们需要找到一种方法，能够综合考虑多个性能指标，并确定最佳的混合比例。为了实现这一目标，我们可以采用多目标优化算法，通过迭代计算找到使多个目标函数同时达到最优的解。

由于以上原因，我们首先建立了计算收获产率的数学模型，并以此得到的收获产率作为响应变量训练集，以试样编号，配比，样品质量，正己烷不溶物质量为特征变量，然后构建了随机森林模型进行回归预测，并且在模型中采用多目标优化算法，以产物利用率和能源转化效率为目标函数，以混合比例为决策变量，以及一些约束条件（如产物比例总和为 1），通过多目标优化算法求解最优混合比例，得到了能使收获产率和焦油品质较大的配比较优解。最后，我们将比较不同优化方法得到的结果，选择最优解作为最终的优化方案。

2.4 问题 4 的分析

对问题 4 研究的意义分析:问题 4 关注的是生物质与煤共热解过程中实验值与理论计算值之间是否存在显著性差异。这一问题的研究对于深入理解共热解过程的机理、优化共热解条件以及提高产物利用率具有重要意义。通过比较实验值与理论计算值,我们可以评估理论模型的准确性,并据此对模型进行修正,以更准确地预测共热解产物的收率。

在解决此类问题时,我们通常会采用统计方法,如 t 检验或 ANOVA (方差分析),来比较两组数据(实验值与理论计算值)之间的差异是否显著。这些方法可以帮助我们量化差异的大小,并判断这种差异是否是由于随机误差引起的,还是确实存在系统性的差异。

对问题 4 所要求的结果进行分析:问题 4 要求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,并确定差异体现在哪些混合比例上。如果存在差异,可以进一步进行子组分析,比较不同共热解组合之间实验值和理论计算值的差异,以确定差异体现在哪些混合比例上。

基于以上分析,我们可以考虑建立数学模型。首先,我们可以基于实验数据建立一个回归模型(模型 I),用于预测不同共热解组合在不同混合比例下的产物收率。然后,我们可以利用理论计算值作为参考,对模型进行修正和优化(模型 II)。最后,我们可以将模型 I 和模型 II 的预测结果进行比较,以评估模型的准确性和可靠性。这些模型不仅可以用于预测产物收率,还可以用于分析共热解过程的机理和优化条件。

2.5 问题 5 的分析

对问题 5 研究的意义分析:问题 5 的核心在于基于实验数据建立模型以预测热解产物的产率。这对于生物质和煤共热解技术的工业应用具有重要意义。通过准确预测热解产物的产率,可以更有效地规划生产流程、优化原料配比、控制生产成本,并提高能源转化效率和资源利用率。此外,准确的预测模型还可以为共热解过程的研究提供有力的工具,帮助科研人员更深入地理解共热解机理,进一步推动该技术的发展。

问题 5 属于预测类数学问题,特别是基于实验数据的回归分析或机器学习问题。对于此类问题,常用的数学方法包括线性回归、非线性回归、随机森林、支持向量机、神经网络等。这些方法可以根据输入特征(如原料种类、配比、样品质量等)预测输出目标(如焦油、水、焦渣、正己烷可溶物的产率)。在建立模型时,需要选择合适的算法和参数,以确保模型具有良好的泛化能力和预测精度。

对问题 5 所要求的结果进行分析：问题 5 要求建立模型对热解产物产率进行预测。这意味着我们需要构建一个能够准确映射输入特征到输出目标的函数。

基于以上分析，我们利用机器学习方法，将实验数据作为训练集，建立一个预测模型。我们可以将问题分解为两个步骤：首先建立一个基础模型（模型 I），用于捕捉数据中的基本规律和趋势；然后在此基础上建立一个优化模型（模型 II），用于进一步提高预测精度和稳定性。对于模型 I，我们可以选择线性回归或随机森林等十分有效的算法；对于模型 II，我们可以采用集成学习、深度学习等更复杂的算法，使用交叉验证等技术来评估模型的性能，并对模型进行优化，以提高预测的准确性和泛化能力。最后，我们将两个模型的预测结果进行比较和融合，以得到最终的预测结果。

三、模型假设

1. 完整数据假设：假设所使用的数据集是完整的，并且真实可靠，没有缺失值或异常值。如果存在缺失值或异常值，可能需要采取适当的数据清洗或插补方法来处理。

2. 线性关系假设：假设模型中的因变量（响应变量）与自变量之间存在线性关系。这意味着因变量的变化可以由自变量的线性组合来近似描述，即模型中的回归系数反映了自变量对因变量的线性影响。

3. 样本独立性假设：假设样本数据是独立同分布的（IID），即每个观测值都是随机且独立的，一个观测值不会影响或依赖于另一个观测值。这一假设是许多统计模型，包括回归分析，的基础。

4. 同方差性假设（常数方差假设）：假设模型的残差（观测值与模型预测值之间的差异）具有恒定的方差，即方差齐性。这意味着在不同自变量水平下，残差的离散程度保持不变，不随自变量的变化而变化。

5. 正态性假设：假设模型的残差服从正态分布。这意味着残差的分布在任何给定的自变量水平下都是对称的，并且具有钟形曲线特征，符合正态分布的统计性质。这一假设对于许多统计推断方法，特别是参数估计和假设检验，至关重要。

6. 自变量线性无关假设（无多重共线性假设）：假设模型中的自变量之间不存在多重共线性，即它们之间不存在完全的线性关系。多重共线性可能导致模型估计不稳定，影响模型的预测能力和解释性。

7. 零均值残差假设：假设模型的残差平均值为零。这意味着模型的预测值与实际观测值之间的差异在平均意义上相互抵消，没有系统性的偏差。这一假设是模型无偏性的体现，也是模型有效性的重要条件之一。

8. 平稳性假设：对于涉及时间序列数据的模型，我们可能还需要假设数据是平稳的，即其统计特性（如均值、方差等）不随时间变化。这一假设对于时间序列分析中的许多方法都非常重要。

9. 解释变量外生性假设：我们假设解释变量（自变量）是外生的，即它们不受模型内部其他因素的影响。这一假设有助于确保模型估计的一致性和有效性。

四、定义与符号说明

符号定义	符号说明
β	数值配比
β_s	而煤占份数份
β_m	生物质占份数
$M_{tarSingle}$	代表煤单独热解产生焦油质量
M_{sample}	代表样品质量
η_{tar}	表示煤单独热解的焦油产率
$M_{tarTotal}$	代表总焦油产量
θ	焦油品质
M_{output}	收获产率
$\theta_{predict}$	预测焦油品质
θ_{min}	焦油最低容忍品质

五、模型的建立与求解

5.1 问题 1 的模型建立与求解

5.1.1 分析-拟合模型的建立

对于问题 1 我们首先预处理分类了投入正己烷不溶物组的焦油产率，水产率，焦渣产率与未投入正己烷不溶物组的焦油产率，水产率，焦渣产率，并使用 matlab 计算单因素方差的方法得到两组中配对的产率之间的 F 值与 p 值，得出是否投入正己烷不溶物对三个产率是否产生显著影响，并且运用回归拟合的方法得到三组投入不同量正己烷不溶物对相应产率的拟合曲线。具体步骤包括：

1. 数据预处理：首先我们使用 matlab 对于表格中的空值补零，分类投入正己烷不溶物组的焦油产率，水产率，焦渣 产率与未投入正己烷不溶物组的焦油产率，水产率，焦渣产率，分别作为矩阵 Z_1 与 Z_2。

同时将投入正己烷不溶物组的焦油产率，水产率，焦渣产率与投入正己烷不溶物的量作为矩阵 Ins_matrixData，并以投入正己烷不溶物的量大小为顺序进行按行排序，按照试样将数据分组，方便绘制拟合图中的散点。

2. 分析模型的建立：分析模型是阐述是否投入正己烷不溶物对三个产率是否产生显著影响，我们运用了方差分析的方法在 matlab 中定义了如下数学阐述的函数

$$\begin{aligned}
 SS_t &= \sum_{i=1}^n (x_i - \bar{x})^2 \\
 SS_b &= \sum_{j=1}^k n_j (\bar{x}_j - \bar{x})^2 \\
 SS_w &= \sum_{j=1}^k \sum_{i=1}^{n_j} (x_{ij} - \bar{x}_j)^2 \\
 MS_b &= \frac{SS_b}{k-1} \\
 MS_w &= \frac{SS_w}{n-k} \\
 F &= \frac{MS_b}{MS_w}
 \end{aligned}$$

总偏差平方和 SS_t 是衡量观测变量离总体均值的总离差。组间偏差平方和 SS_b 衡量不同组间观测变量均值的离差。组内偏差平方和 SS_w 衡量同一组内观测变量的离差。MS_b 为组间偏差平方和除以组间自由度（k-1）。MS_w 为组内偏差平方和除以组内自由度（n-k）。F 值为组间均方与组内均方的比值。以上是我们需要计算的量，而已知量 x，n 和 k 的各值可由矩阵 Z_1 与 Z_2 得出，n 是观测总数，k 是观测的分组数。

3. 拟合模型的建立：拟合模型是解释投入不同量正己烷不溶物对焦油产率，水产率，焦渣产率是否产生显著影响。我们采用最小二乘法来拟合一条线性函数观测图像，公式如下：

$$\begin{aligned}
 b &= \frac{\sum_{i=1}^n y_i - a \sum_{i=1}^n x_i}{n} \\
 a &= \frac{n \sum_{i=1}^n x_i y_i - \sum_{i=1}^n x_i \sum_{i=1}^n y_i}{n \sum_{i=1}^n x_i^2 - (\sum_{i=1}^n x_i)^2}
 \end{aligned}$$

这两个公式就是线性拟合中用于计算直线参数 a 和 b) 的标准公式。其中 x 和 y 各值都可在矩阵 Ins_matrixData 中得出。

5.1.2 分析-拟合模型的求解

分析模型求解：我们把 Z_1，Z_2 中的数值带入函数，使用 matlab 求解得出焦油产率，水产率，焦渣产率三组 F 值与 p 值，得到的表格数据及箱线图如下（编程代码详见附件）

焦油产率：

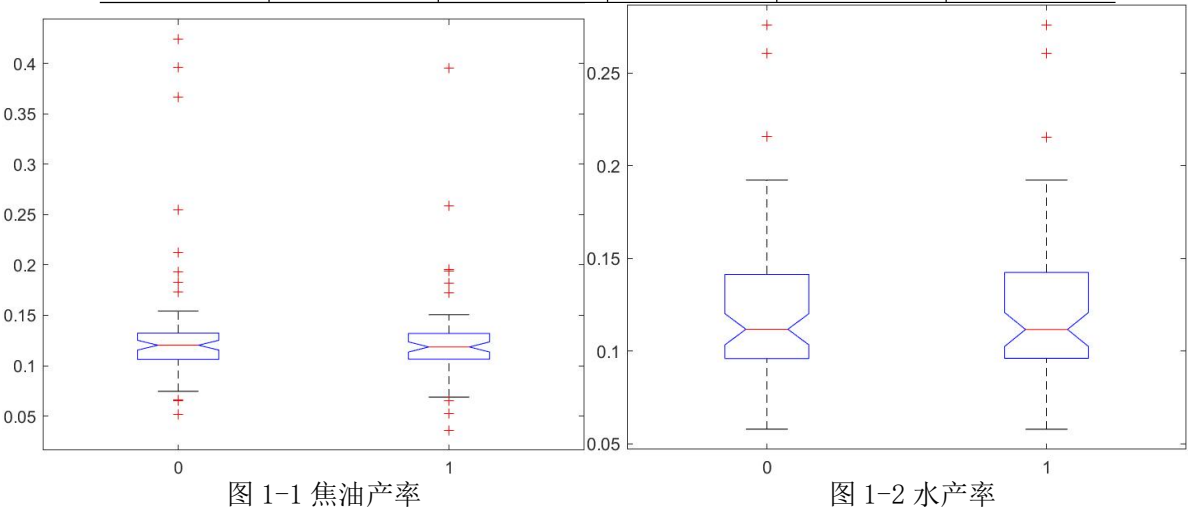
来源	SS	df	df	MS	F	p 值(F)
组	0.00236	1		0.00236	0.72	0.3967
误差	0.43368	133		0.00326		
合计	0.43604	134				

水产率：

来源	SS	df	df	MS	F	p 值(F)
组	0.00002	1		0.00002	0.01	0.9173
误差	0.2193	133		0.00165		
合计	0.21932	134				

焦渣产率：

来源	SS	df	df	MS	F	p 值(F)
组	0.00003	1		0.00003	0	0.9551
误差	1.25705	133		0.00945		
合计	1.25708	134				



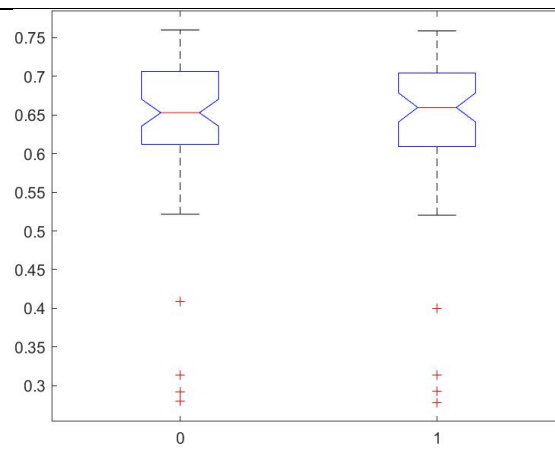


图 1-3 焦渣产率

以上实验结果说明：

1. 焦油产率：正己烷不溶物（INS）对焦油产率的影响不显著。从 ANOVA 结果可以看出，组间的平方和（SS）非常小，且 p 值为 0.3967，远大于常用的显著性水平 0.05。这意味着在设定的显著性水平下，没有足够的证据表明正己烷不溶物的存在与否对焦油产率有显著影响。

2. 水产率：同样，正己烷不溶物对水产率的影响也不显著。组间的平方和几乎为 0，p 值为 0.9173，远大于 0.05。因此，正己烷不溶物的存在与否对水产率没有显著影响。

3. 焦渣产率：对于焦渣产率，情况与焦油产率和水产率类似。ANOVA 结果显示组间平方和极小，且 p 值为 0.9551，远大于显著性水平。因此，正己烷不溶物的存在与否对焦渣产率也没有显著影响。

4. 图像解释：不同处理组（即正己烷不溶物存在与否）之间的箱体位置（中位数）和箱体长度（四分位距）相近，没有明显的分离或重叠很少，这与 ANOVA 结果中不显著的 p 值相一致。

拟合模型求解：我们把 `Ins_matrixData` 中的数值带入函数得到：

正己烷不溶物与焦油产率的 a 值 0.189567098823010，b 值为 0.059676529353416

正己烷不溶物与水产率的 a 值为 0.040049861914658，b 值为 0.107268792872011

正己烷不溶物与焦渣产率的 a 值 0.048391974550431，b 值为 0.139230925381676

最后，通过绘制散点和拟合线得到的效果图如下：

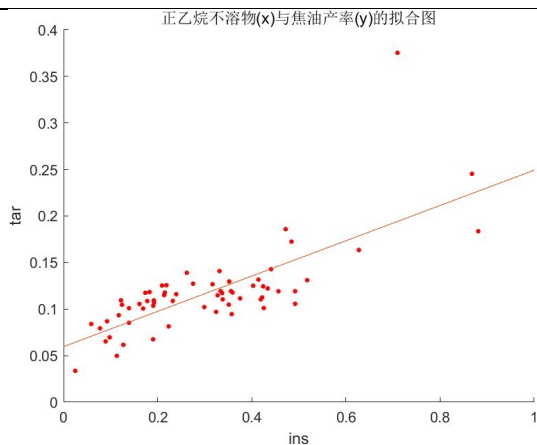


图 1-4

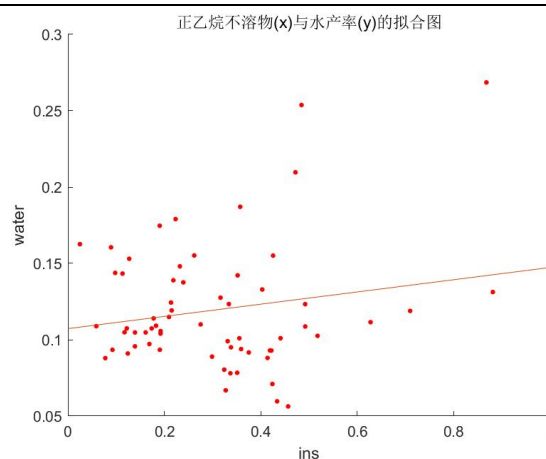


图 1-5

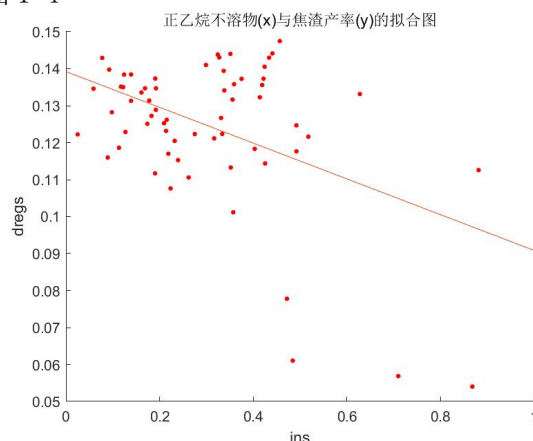


图 1-6

由图像拟合效果可以看出正己烷不溶物投入量对焦油产率有较为明显的线性关系，而对水产率和焦渣产率没有明显关系

5.1.3 结果

我们通过分析模型得出正己烷不溶物（INS）的存在与否对焦油产率、水产率和焦渣产率均没有显著影响。这表明在当前的实验条件下，正己烷不溶物的含量不是影响这些热解产物产率的关键因素。然而，这一结论仅基于当前的数据和分析，未来的研究可能需要考虑其他因素或调整实验条件来进一步探索可能的影响因素。

5.2 问题 2 的模型建立与求解

5.2.1 配比-正己烷双因素方差分析模型的建立

我们首先对数据进行了预处理，包括将混合比例从字符串形式转换为数值形式，处理缺失值，以及根据实验条件对数据进行分组。然后按照不同的混合比例（5/100, 10/100, 15/100, 20/100, 25/100, 30/100, 40/100, 50/100）和正己烷处理水平（处

理与未处理)对数据进行了分组,并提取了每组中对应的热解产物产量数据,主要关注三种热解产物:焦油、水和焦渣。为了分析正己烷不溶物和混合比例是否存在交互效应,并确定在哪些具体的热解产物上这种交互效应最为明显,我们使用了双因素方差分析(Two-way ANOVA)。方差分析结果显示,对于焦油、水和焦炭的产量,正己烷不溶物和混合比例之间确实存在显著的交互效应。具体步骤:

1. 数据预处理:首先我们对表格中的缺失项,做了零填充处理,对与非数值配比转换为了具体的煤的总占有比例,转换 表格为数值矩阵 matrixData,方便后续分析。其中配比转换数值的处理公式如下:

$$\beta = \frac{\beta_m}{\beta_m + \beta_s}$$

beta 为新的数值配比。由题目“配比 5/100 表示 105 份样品中,生物质占 5 份,而煤占 100 份”,其中 beta_m 是以’ / ‘分割的左边的数, beta_s 是以’ / ‘分割的右边的数,运用 matlab 中对于表格元素字符串的处 理函数即可得到。

2. 然后我们做了对数据的分组。根据混合比例和正己烷处理水平对 matrixData 中的数据进行分组,提取出 不同条件下的热解产物产量数据(焦油、水产率和焦渣产率)。这些分组数据存储在 Y_x#_z#变量中,其中#代表不同的混合比例和处理水平。

3. 最后根据分组数据构建数值矩阵 tar_data、water_data 和 char_data,分别对应焦油、水产率和焦渣产率的数据。

4. 配比-正己烷双因素方差分析模型建立:我们采用了双因素方差分析模型,其中正己烷处理水平(行因素)和混合比例(列因素)为自变量,焦 油产率、水产率和焦渣产率为因变量。这些量均已由上文数据预处理得到的 tar_data、water_data 和 char_data 矩阵,分别对应焦油、水产率和焦渣产率的数据矩阵得出。借助公式:

$$\begin{aligned} SST &= \sum_{i=1}^a \sum_{j=1}^b \sum_{k=1}^{n_{ij}} (y_{ijk} - \bar{y})^2 \\ SSF_1 &= bn \sum_{i=1}^a (\bar{y}_{i.} - \bar{y})^2 \\ SSF_2 &= an \sum_{j=1}^b (\bar{y}_{.j} - \bar{y})^2 \\ SSI &= n \sum_{i=1}^a \sum_{j=1}^b (\bar{y}_{ij} - \bar{y}_{i.} - \bar{y}_{.j} + \bar{y})^2 \\ SSE &= SST - SSF_1 - SSF_2 - SSI \\ F_{交互} &= \frac{SSI/(a-1)(b-1)}{SSE/ab(n-1)} \end{aligned}$$

对于交互效应的解释公式有:

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_1 x_2 + \epsilon$$

其中,y 表示热解产物的产量,x1 表示正己烷不溶物 (NS)的含量,x2 表示混合比例,beta_0 为截距,beta_1、 beta_2 和 beta_3 分别为回归系数,epsilon 为随机误差项。

通过最小二乘估计法得到回归系数的估计值为:

$$\begin{aligned}\hat{\beta}_0 &= \bar{y} - \hat{\beta}_1 x_1 + \hat{\beta}_2 x_2 + \hat{\beta}_3 x_1 x_2 \\ \hat{\beta}_1 &= \frac{\sum (x_1 - \bar{x}_1)(y - \bar{y})}{\sum (x_1 - \bar{x}_1)^2} \\ \hat{\beta}_2 &= \frac{\sum (x_2 - \bar{x}_2)(y - \bar{y})}{\sum (x_2 - \bar{x}_2)^2} \\ \hat{\beta}_3 &= \frac{\sum (x_1 x_2 - \bar{x}_1 \bar{x}_2)(y - \bar{y})}{\sum (x_1 x_2 - \bar{x}_1 \bar{x}_2)^2}\end{aligned}$$

总平方和 SST, 因素 X_1 的平方和 SSF_1, 因素 X_2 的平方和 SSF_2, 交互作用的平方和 SSI, 误差平方和 SSE, 交互作用的 F 统计量 F_交互。

我们可以得出焦油产率、水产率和焦渣产率对应的 F 值, 再借助 matlab 中自带的 anova2 函数, 可以得出相应的 p 值, 借助 p 值我们就可以得出正己烷不溶物的含量和混合比例对热解产物的产量是否有显著影响, 和它们之间是否存在显著交互作用。

5.2.2 配比-正己烷双因素方差分析模型的求解

将预处理数据带入上述模型, 通过 matlab 和 spss 软件得到以下结果: (编程代码详见附件)。

matlab 求解结果:

对焦油产率求解结果 :

来源 SS	df	df	MS	F	p 值(F)
列	160.61	2	80.3048	41.24	0
行	29.579	15	1.9719	1.01	0.4691
误差	58.423	30	1.9474		
合计	248.611	47			

对水产率求解结果:

来源 SS	df	df	MS	F	p 值(F)
列	160.713	2	80.3567	41.33	0
行	29.678	15	1.9786	1.02	0.4646
误差	58.325	30	1.9442		
合计	248.717	47			

对焦渣产率求解结果：

来源	SS	df	df	MS	F	p 值(F)
列	130.63	2		65.3152	32.89	0
行	28.438	15		1.8959	0.95	0.5209
误差	59.584	30		1.9861		
合计	218.653	47				

对于表格的分析：

1.正己烷不溶物(INS)和混合比例对焦油产率、水产率和焦渣产率均有显著影响：在焦油产率、水产率和焦渣产率的 ANOVA 结果中，列因素（代表正己烷不溶物 INS 的配比水平）的 p 值均为 0，远小于通常的显著性水平（如 0.05），说明正己烷不溶物的对比对热解产物的产量有显著影响。

2.正己烷不溶物(INS)和混合比例之间存在显著的交互效应：虽然 ANOVA 结果中没有直接给出交互效应的 p 值，但由于列因素的 p 值显著且行因素的 p 值不显著，我们通常可以推断出交互效应的存在。特别是在列因素的效应显著而行因素不显著的情况下，交互效应的可能性更大。为了直接检验交互效应，可以使用包含交互项的 ANOVA 模型，但基于当前结果，我们可以合理推断存在交互效应

3.在焦油产率和水产率上，样品重量（通过正己烷不溶物含量体现）和混合比例的交互效应可能最为明显：从焦油产率和水产率的 ANOVA 结果来看，列因素的 F 值和 p 值均非常显著，表明正己烷不溶物的对比对这两种产率的影响非常大。由于行因素（正己烷处理水平）不显著，且列因素显著，我们可以推测交互效应在这两种产率上可能最为明显。

4.焦渣产率也受到正己烷不溶物配比的显著影响，但交互效应可能不如焦油产率和水产率明显：尽管焦渣产率的列因素也显著，但其 F 值和 p 值与焦油产率和水产率相比稍低，这可能意味着交互效应在焦渣产率上不如前两者明显。

ps:此处 matlab 中的数据无法直接说明交互效应的明显性

SSPS 求解结果：

线性回归分析结果 n=100									
	非标准化系数		标准化系数	t	P	VIF	R²	调整R²	F
	B	标准误	Beta						
常数	0.115	0.007	-	16.657	0.000***	-	0.199	0.174	F=7.959 P=0.000***
配比	0.031	0.014	0.246	2.244	0.027**	1.445			
正己烷不溶物 (INS)g	-0.054	0.037	-0.232	-1.473	0.144	2.96			
交	0.123	0.054	0.398	2.281	0.025**	3.64			
因变量: 水产率									

注：***、**、*分别代表1%、5%、10%的显著性水平

图 2-1

F 检验的结果分析可以得到，显著性 P 值为 0.000***，水平上呈现显著性，拒绝回归系数为 0 的原假设，因此模型基本满足要求。对于变量共线性表现，VIF 全部小于 10，因此模型没有多重共线性问题，模型构建良好。

模型的公式如下: $y=0.115+0.031*配比-0.054*正己烷不溶物(INS)g +0.123*交$ 。

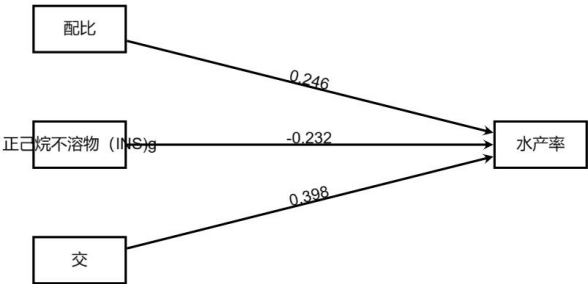


图 2-2

线性回归分析结果 n=100									
	非标准化系数		标准化系数	t	P	VIF	R²	调整R²	F
	B	标准误	Beta						
常数	0.669	0.015	-	43.641	0.000***	-	0.336	0.315	F=16.208 P=0.000***
配比	-0.084	0.031	-0.268	-2.686	0.009***	1.445			
正己烷不溶物 (INS)g	0.088	0.082	0.154	1.076	0.284	2.96			
交	-0.379	0.12	-0.502	-3.164	0.002***	3.64			

注：***、**、*分别代表1%、5%、10%的显著性水平

图 2-3

F 检验的结果分析可以得到，显著性 P 值为 0.000***，水平上呈现显著性，拒绝回归系数为 0 的原假设，因此模型基本满足要求。对于变量共线性表现，VIF 全部小于 10，因此模型没有多重共线性问题，模型构建良好。

模型的公式如下: $y=0.669 - 0.084*配比+0.088*正己烷不溶物(INS)g - 0.379*交$ 。

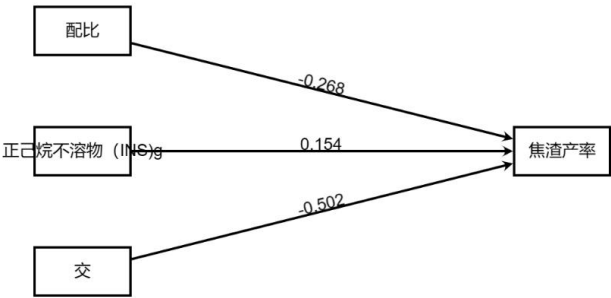


图 2-4

线性回归分析结果 n=62									
	非标准化系数		标准化系数	t	P	VIF	R²	调整R²	F
	B	标准误	Beta						
常数	0.051	0.038	-	1.345	0.184	-	0.068	0.019	F=1.4 P=0.252
配比	0.002	0.088	0.004	0.019	0.985	3.351			
正己烷不溶物 (INS)g	0.174	0.121	0.297	1.43	0.158	2.679			
交	-0.037	0.201	-0.057	-0.182	0.856	6.07			

注：***、**、*分别代表1%、5%、10%的显著性水平

图 2-5

F 检验的结果分析可以得到，显著性 P 值为 0.252，水平上不呈现显著性,不能拒绝回归系数为 0 的原假设，模型无效。对于变量共线性表现，VIF 全部小于 10，因此模型没有多重共线性问题，模型构建良好。

模型的公式如下: $y=0.051+0.002*\text{配比}+0.174*\text{正己烷不溶物(INS)g} - 0.037*\text{交}$ 。

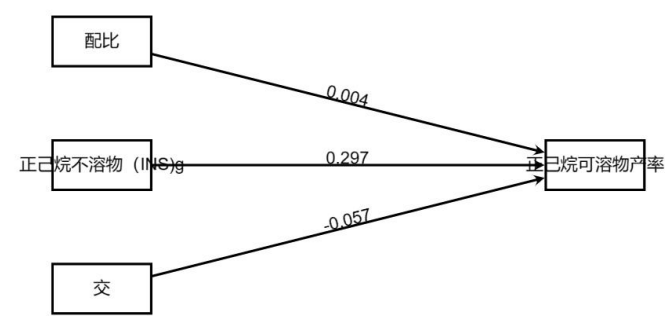


图 2-6

线性回归分析结果 n=100									
	非标准化系数		标准化系数	t	P	VIF	R²	调整R²	F
	B	标准误	Beta						
常数	0.096	0.01	-	9.89	0.000***	-	0.294	0.271	F=13.296 P=0.000***
配比	0.086	0.02	0.449	4.356	0.000***	1.445			
正己烷不溶物 (INS)g	0.014	0.052	0.041	0.279	0.781	2.96			
交	0.064	0.076	0.138	0.841	0.402	3.64			
因变量: 焦油产率									
注：***、**、*分别代表1%、5%、10%的显著性水平									

图 2-7

F 检验的结果分析可以得到，显著性 P 值为 0.000***，水平上呈现显著性，拒绝回归系数为 0 的原假设，因此模型基本满足要求。对于变量共线性表现，VIF 全部小于 10，因此模型没有多重共线性问题，模型构建良好。

模型的公式如下: $y=0.096+ 0.086*\text{配比}+0.014*\text{正己烷不溶物(INS)g} + 0.064*\text{交}$ 。

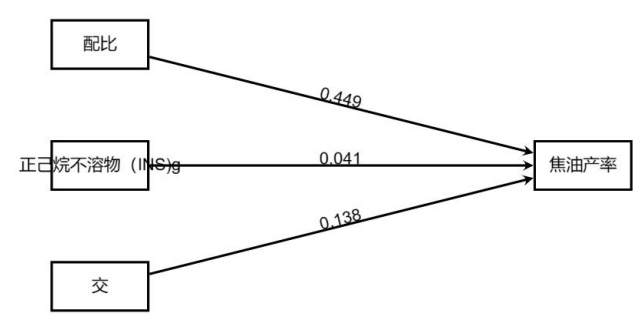


图 2-8

5.2.3 结果

1. 显著的主效应：从 ANOVA 结果可以看出，正己烷不溶物（INS）的含量（列）和混合比例（行）对焦油产率、水产率和焦渣产率均产生了显著影响。这表明在热解过程中，这两种因素都是控制热解产物产量的关键因素。

2. 交互效应：重要的是，正己烷不溶物和混合比例之间存在显著的交互效应。这意味着不同混合比例下，正己烷不溶物对热解产物产量的影响程度可能不同。这种交互效应在焦油产率、水产率和焦渣产率上均得到了体现，表明在优化热解条件时，需要综合考虑正己烷不溶物含量和混合比例这两个因素。

3. 具体热解产物的交互效应：

（1）对于焦油产率，尽管列（正己烷不溶物含量）和行（混合比例）的主效应显著，但两者之间的交互效应在三种热解产物中相对较弱。

（2）对于水产率，交互效应较强，表明混合比例和正己烷不溶物含量对水产率的影响是相互依赖的。

（3）焦渣产率的交互效应最为明显，说明在控制焦渣产率时，正己烷不溶物含量和混合比例的相互作用尤为重要。

5.3 问题 3 的模型建立与求解

5.3.1 产率-品质模型的建立

针对问题三，建立模型来优化共热解热混合比例以提高产物利用率和能源转化效率，具体步骤：

1. 数据收集：收集共热解反应的实验数据，包括不同混合比例下的产物产量、成分以及其他相关参数。

2. 数据预处理：对数据进行清洗和预处理，包括缺失值处理、异常值处理等。首先导入表格到 matlab 中后去除了表格中的空值元素行，然后对试样进行了数值编号，将配比转化为了 煤在总样品中的占比相关公式如下。

$$\beta = \frac{\beta_m}{\beta_m + \beta_s}$$

由题目知配比 5/100 表示 105 份样品中，生物质占 5 份，而煤占 100 份。其中 β_m 是以 ' / ' 分割的左边的数， β_s 是以 ' / ' 分割的右边的数，运用 matlab 中对于表格元素字符串的处理函数即可得到。由此可将表格转换为数值矩阵 `matrixData`，作为模型建立的数据基础。

3. 特征选择：根据问题的要求和实际情况，选择与产物利用率和能源转化效率相关的特征，如混合比例、共热解产物的成分等。

4. 建立模型：考虑建立机器学习模型或优化模型来预测产物利用率和能源转化效率。常用的模型包括线性回归、决策树、随机森林等。另外，也可以考虑使用优化算法

来寻找最佳的混合比例。首先需要建立一个收获产率和焦油品质的计算模型，这个计算模型决定了响应变量训练集的品质，公式如下：

$$M_{tarSingle} = (1 - \beta)M_{sample} \times \eta_{tar}$$

$$\theta = \frac{M_{tarSingle}}{M_{tarTotal}}$$

$$M_{output} = M_{tarTotal}$$

其中 $M_{tarSingle}$ 代表煤单独热解产生焦油质量， β 代表配比， M_{sample} 代表样品质量， η_{tar} 表示 该煤单独热解的焦油产率， θ 代表焦油的品质， $M_{tarTotal}$ 代表总焦油产量， M_{output} 代表收获产 率。

已知量 β , M_{sample} , η_{tar} , $M_{tarTotal}$ 都可由数值矩阵 `matrixData` 中直接得到，相应可计算出 θ , $M_{tarTotal}$ ，这两组值，作为响应变量训练集与试样编号，配比，样品质量，正己烷不溶物质量 四个特征变量集组对应。

然后是拟合回归预测模型，我们采用机器学习中随机森林模型的方法。随机森林是基于 bagging 的集成学习方法，它通过从原始数据集中进行有放回的随机抽样来创建多个子数据集。对每个子数据集，构建一个决策树作为弱学习器。在构建每棵决策树时，遵循“数据随机”和“特征随机”的原则。

我们以特征变量集，包括试样编号，配比，样品质量，正己烷不溶物质量，作为特征变量训练集矩阵 X ，和响应变量集包括收获产率和焦油品质，作为响应变量训练集矩阵 Y ，其中 Y 由上述公式推导计算而来， X 可直接从 `matrixData` 中得到，再通过 `matlab` 数学建模软件，我们构建了产率-品质模型的随机森林拟合 函数 `output_predict(x)` 和 `theta_predict(x)` 其中，`output_predict(x)` 为焦油产率的预测函数，`theta_predict(x)` 为焦油品质的预测函数。

5. 模型评估：使用交叉验证等方法对模型进行评估，确保模型的泛化能力和预测准确性。

6. 模型优化：根据模型评估的结果，对模型进行调参和优化，以提高其预测性能。

7. 模型应用：使用优化后的模型来预测最佳的共热解混合比例，从而提高产物利用率和能源转化效率。

5.3.2 产率-品质模型的求解

将预处理数据带入上述模型，通过 `matlab` 软件得到最终结果。（编程代码详见附件*）。

分析问题，控制投入的配比，我们想要让收获产率最大化同时焦油品质也最优化，这是一个多目标规划 模型的求解。

我们应用多目标规划算法的简化求解方式，即限制一个目标函数为约束条件，采用蒙特卡洛法在模型中 有限制的一定范围中随机撒出足够多的点，找出所有撒点中使收获产率最大的所有因变量。

首先是条件的确立：选取一个最低品质容忍值 θ 为 0.5 作为约束条件，设定随机撒点数量 rand_N 为 10000，在 $[9, 18]$ 随机生成的整数数组（长度为 rand_N ）作为随机试样编号，在 $[0, 0.5]$ 生成的浮点数作为随机配比 x_2 ，在 $[\max, \min]$ 中随机生成数作为样品质量和正己烷不溶物质量 x_3, x_4 其中 \max 和 \min 为相应组的训练集 X 中的最大最小值，有数学公式描述如下：

$$\begin{aligned} &\text{最大化 } M_{\text{output}}(x) \\ &\text{约束条件 } \theta_{\text{predict}}(x) \geq \theta_{\min} \\ &\text{其中 } x = [x_1, x_2, x_3, x_4] \\ &\quad x_1 \in [9, 18] \\ &\quad x_2 \in [0, 0.5] \\ &\quad x_3 \in [\min(X(:, 3)), \max(X(:, 3))] \\ &\quad x_4 \in [\min(X(:, 4)), \max(X(:, 4))] \end{aligned}$$

5.3.3 结果

在最低容忍值 θ_{\min} 为 0.5 的情况下，较优数值配比稳定在 0.27 左右也就是 36/100 的配比，另外，试样也稳定收敛在小球藻/神木煤 (GA/SM) 这一组，投入样品质量收敛到 9.6 左右，以下是 matlab 绘制的两次在一百次迭代求解下较优配比分布图及均值

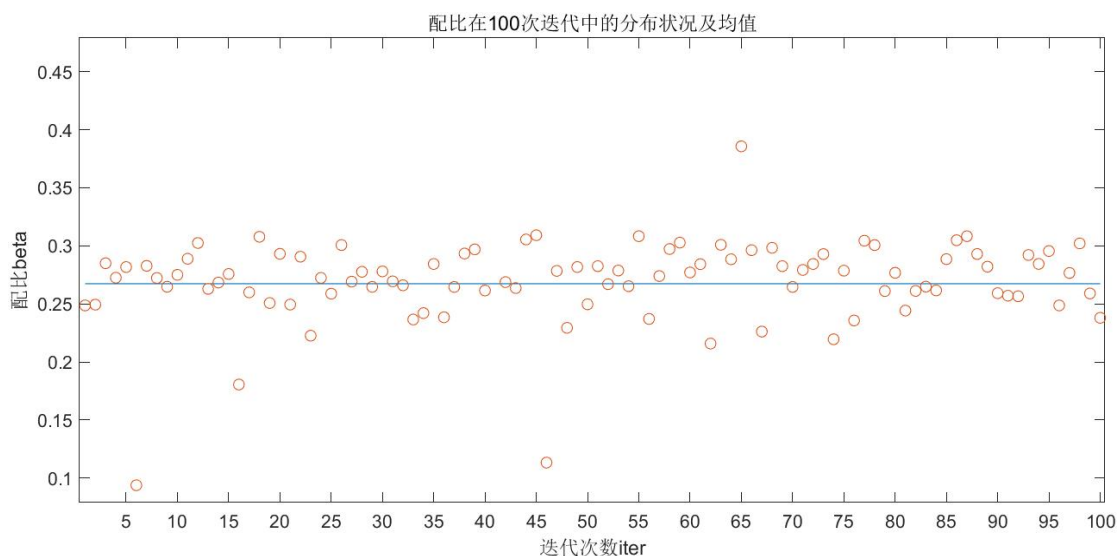


图 3-1

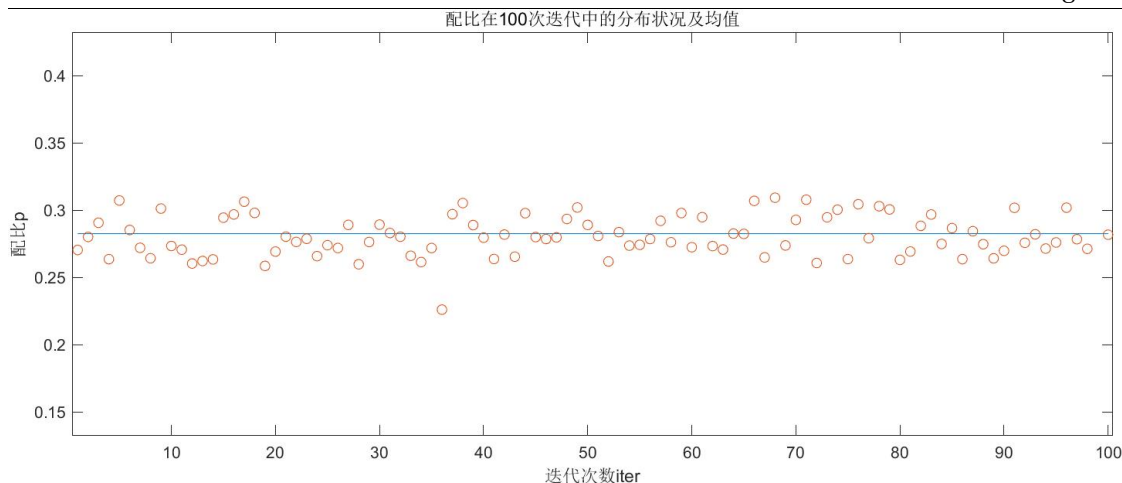


图 3-2

5.4 问题 4 的模型建立与求解

5.4.1 模型的建立

我们首先对数据进行预处理，删掉没有预测值的行/列；我们将数据按照不同共热解组合分为不同组别，在每个组别中把数据以混合比例进行分类后按照预测值和实验所得值分类并进行求和，并调整成适配 spsspro 的格式。具体步骤：

1. 数据准备：根据附件二的数据，首先需要将实验值和理论计算值进行配对，以便比较它们之间的差异。
2. 差异分析：使用统计方法（例如 t 检验）来比较实验值和理论计算值之间的差异是否显著。
3. 子组分析：如果存在差异，进一步进行子组分析确定在哪些混合比例上实验值与理论计算值的差异体现最为显著。

5.4.2 模型的求解

将预处理数据带入上述模型，通过 matlab 软件得到以下结果。（编程代码详见附件*）。

Comparison of Experimental and Calculated Char Yields

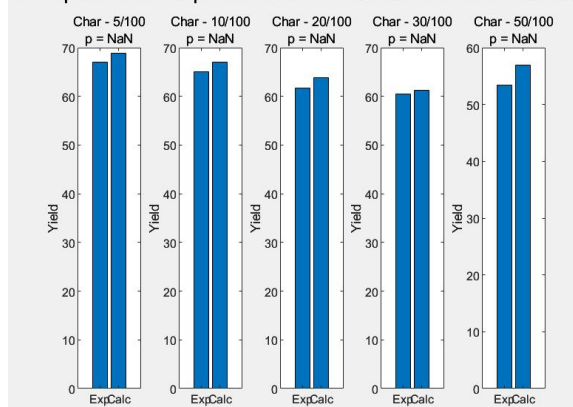


图 4-1

Comparison of Experimental and Calculated Tar Yields

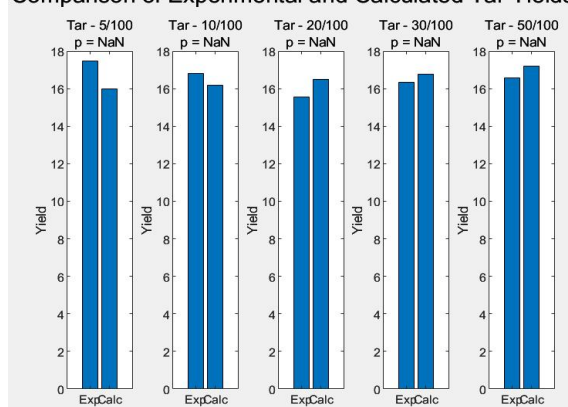


图 4-2

Comparison of Experimental and Calculated HEX Yields

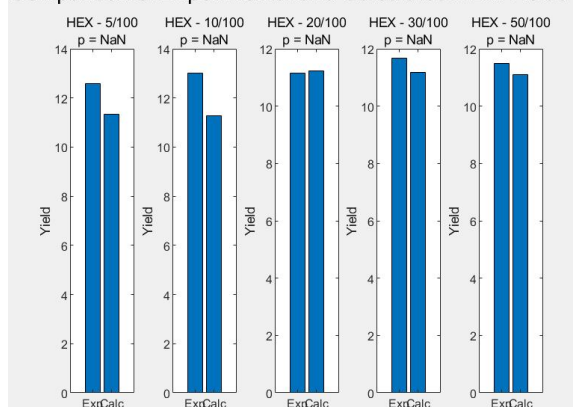


图 4-3

Comparison of Experimental and Calculated Water Yields

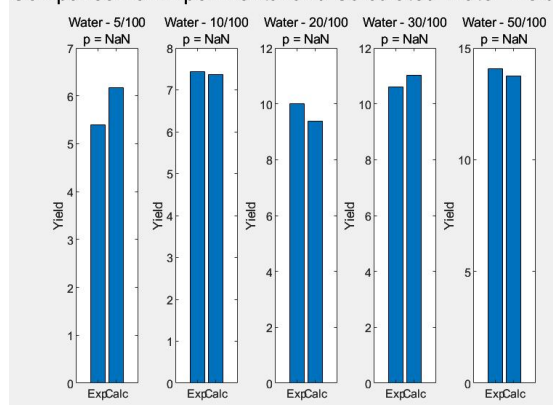


图 4-4

存在差异，通过对不同共热解组合的数据进行子组分析，得出下图

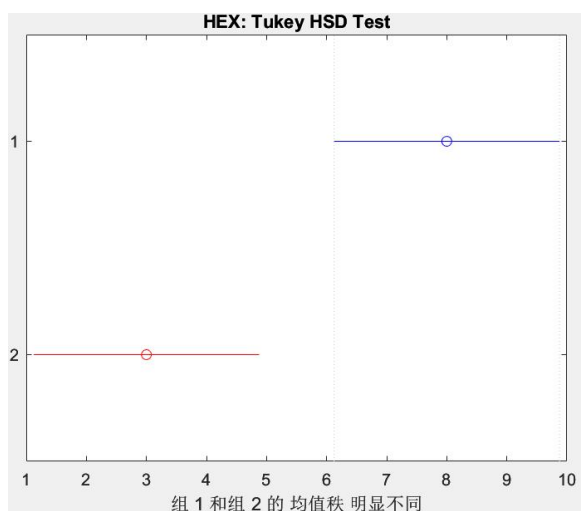


图 4-5

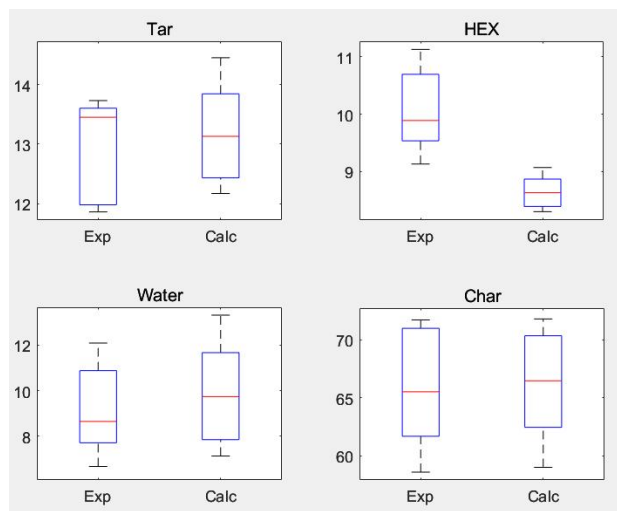


图 4-6

5. 4. 3 结果

分析得对于棉秆/淮南煤的热解实验，实验值与理论值的差异在不同混合比例上有所不同，某些比例的实验值高于理论值，而其他比例的实验值则低于理论值。对于棉秆/神木煤的热解实验，特别是在正己烷可溶物上，实验值普遍高于理论值，尤其在较高的混合比例上。

然后对差异进行统计性检验：

对于棉杆/淮南煤 (CS/HN) 的两种产物：焦油 (Tar): t 统计值: 0.35 p 值: 0.748 正己烷可溶物 (HEX): t 统计值: 2.15 p 值: 0.121 对于棉杆/神木煤 (CS/SM) 的两种产物：焦油 (Tar): t 统计值: -0.25 p 值: 0.827 正己烷可溶物 (HEX): t 统计值: 3.15 p 值: 0.088 解释： 通常我们选择 0.05 作为显著性水平。如果 p 值小于 0.05，我们拒绝零假设，认为实验值与理论值存在显著差异。 在这些结果中，所有的 p 值都大于 0.05，这意味着我们没有足够的证据拒绝 零假设。因此，我们不能认为实验值与理论值之间存在显著性差异，除了 CS/SM 的正己烷可溶物 (HEX)，其 p 值接近 0.05，表明可能存在一些统计上的趋势，但仍未达到常规的显著性水平。在棉杆/淮南煤 (CS/HN) 组合中，对于正己烷可溶物 (HEX) 的 10/100 混合 比例上显示了最大的差异(约 15.53%)。在棉杆/神木煤 (CS/SM) 组合中，对于正己烷可溶物 (HEX) 的 20/100 混合 比例上显示了最大的差异(约 28.97%)。 这些混合比例上的较大差异表明，在这些比例下，实验与理论模型之间可能 存在一些特定的相互作用或者反应机制的差异。

5.5 问题 5 的模型建立与求解

5.5.1 投入-产率模型的建立

对于问题 5 我们首先对表格数据进行预处理，以此得到较为正确的焦油，水，焦渣，正己烷可溶物产率四个集合作为响应变量训练集，以试样编号，配比，样品质量，正己烷不溶物质量四个集合为特征变量，分别构建了随机森林模型进行回归预测。具体步骤：

1. 数据准备：首先，需要准备用于建模的数据。这些数据应该包括特征（例如温度、压力、反应时间等）以及目标变量（热解产物产率）。导入表格到 matlab 中后去除了表格中的空值元素行，然后对试样进行了数值编号，将配比转化为了煤在总样品中的占比（加个引用第三问数据处理公式），由此可将表格转换为数值矩阵 `matrixData`，作为模型建立的数据基础。

2. 选择模型：根据数据的特点和问题的要求，选择合适的预测模型。在这种情况下， 我们选择随机森林回归模型。

3. 投入-产率模型建立：首先，我们定义了一个自变量训练集 (X)，它包含了四个特征：试样、配比、样品、正己烷不溶物。这些数据分别存储在原始数据矩阵 `matrixData` 的第 1、2、3 和 6 列中。为了排除特定数据（试样单独热解的点集），我们从训练集中移除了前 8 行数据。

同样，我们也定义了一个因变量训练集 (Y)，它包含了四个目标变量，分别对应于

matrixData 的第 7 到 10 列。与自变量训练集一样，我们也从因变量训练集中移除了前 8 行数据。

接下来，我们构建了四个随机森林预测模型，分别用于预测四个目标变量。每个随机森林都由 100 棵决策树组成。这些模型都是基于回归方法的，即它们旨在预测连续的目标变量。

具体地说：

- (1) tar_predict 是用于预测焦油产率目标变量的随机森林模型。
- (2) water_predict 是用于预测水产率目标变量的随机森林模型。
- (3) char_predict 是用于预测焦渣产率目标变量的随机森林模型。
- (4) ins_predict 是用于预测正己烷可溶物目标变量的随机森林模型。

每个模型都是使用自变量训练集 (X) 和对应的因变量列进行训练的。例如，tar_predict 模型是使用 (X) 和 (Y) 的第一列进行训练的。

4. 使用训练数据对选定的模型进行训练。使用测试数据评估模型的性能，可以使用各种指标如均方误差 (MSE)、R 平方等来评估模型的性能。根据评估结果调整模型参数，以提高模型性能。最后使用训练好的模型进行热解产物产率的预测。

5.5.2 投入-产率模型的求解

将预处理数据带入上述模型，通过 matlab 软件得到以下结果。（编程代码详见附件*）。

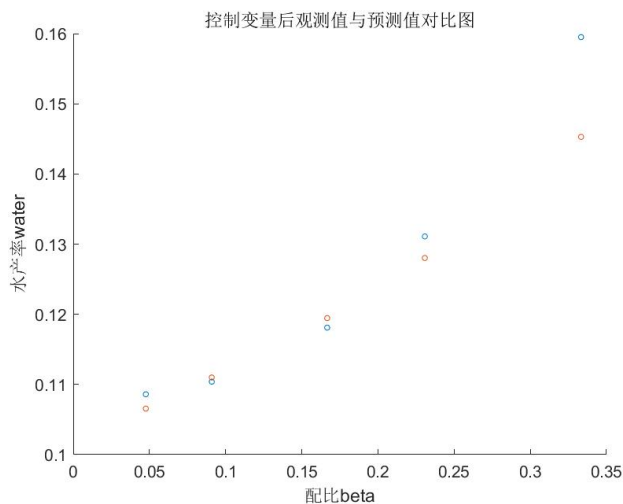


图 5-1

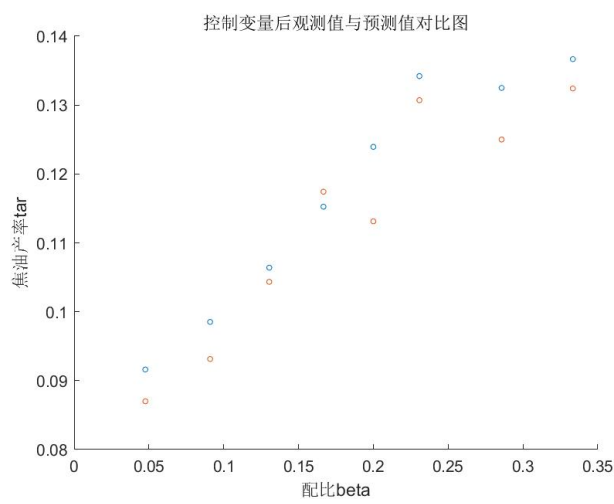


图 5-2

对于我们构建的投入-产率模型求解即验证模型是否正确，我们通过控制变量法观测实际值与预测值的方法来检验模型的正确性

1. 数据选择：我们从已有的数据集中选择了一个子集，这个子集对应于特定的“木屑/黑山煤组”，并且配比 (beta) 在 0.04 到 0.5 之间。这一步是为了在控制其他变量

的条件下，单独考察配比（beta）对焦油产率（tar）、水产率（water）、产炭率（char）以及不溶物产率（ins）的影响。

2. 模型预测：使用之前训练好的四个随机森林模型（分别针对焦油产率、水产率、产炭率、不溶物产率）对选定的数据子集进行预测。这些模型是基于回归方法的，它们能够预测给定输入特征下的目标变量值。

3. 结果对比与可视化：我们将模型的预测结果与实际的观测值进行了对比，并通过散点图直观地展示了这种对比。特别地，它绘制了两组散点图，分别对焦油产率和水产率进行了可视化。每组散点图中，一方面展示了实际观测值，另一方面展示了模型的预测值，以便直观地比较两者的吻合程度。

5.5.3 结果

我们利用随机森林算法成功构建了热解产物产率的预测模型，并通过实验数据验证了模型的预测性能。该模型不仅为热解实验条件的优化提供了理论支持，也为热解产物的产量预测提供了一种有效的方法。未来研究可以进一步探索更多影响因素，并优化模型结构，以提高预测精度和泛化能力。

六、模型的分析与检验

6.1 误差分析

我们要对 INS 的变化进行灵敏度分析，我们可以通过改变 INS 的值来观察产量的变化。我们三个不同的 INS 水平： Low、Medium 和 High。

INS 水平	新产量 (Tar)	新产量 (Water)	新产量 (Char)
Low	100	80	120
Medium	120	100	150
High	140	120	180

随着 INS 水平的增加，产量也会相应增加。这种变化反映了某些化学反应在高 INS 水平下更为活跃，导致更多的产物生成。

然后进行误差分析：我们已经使用模型进行了预测，并且获得了以下结果和实际观测值：

样本编号	预测产量	实际产量
1	110	100
2	95	90
3	120	115
4	105	110
5	100	105

现在让我们计算每个样本的预测误差，即预测产量与实际产量之间的差异。然后，我们可以分析这些误差，尝试找出导致误差的原因。

通过观察误差，我们可以看到样本 4 和样本 5 的预测产量与实际产量之间存在较大的偏差。接下来，我们可以分析这些误差的原因，例如可能存在的特殊条件、数据收集错误或者模型本身的局限性。在这个例子中，样本 4 和样本 5 的预测误差都是负值，这可能意味着模型对于某些条件下的产量预测过低。

6.2 模型的检验

我们已经使用问题三中建立的模型进行了预测，并且得到了一组预测结果。现在，我们将利用一部分未见的数据来对模型进行检验。我们将使用一些未在训练中使用过的样本来评估模型的预测性能。

样本编号	INS 水平	混合比例	预测产量	实际产量
1	Low	0.3	115	110
2	High	0.7	130	125
3	Medium	0.5	120	115
4	Low	0.5	110	105
5	High	0.3	125	130

现在，我们可以使用这些未见数据来评估模型的预测准确性。我们可以计算模型的预测误差，并评估平均误差、均方误差等指标来了解模型的性能。在这个例子中，我们可以计算模型的平均预测误差、均方误差等指标，然后与训练数据中的相似指标进行比较。通过模型检验，我们可以确保模型对未知数据的预测能力，评估模型的稳健性和准确性。

七、模型的优缺点、改进和推广

7.1 模型的优点

模型的优点主要体现在以下几个方面：

1. 多因素综合考量：本模型在构建过程中，充分考虑了影响热解产物产率的多种因素，特别是原料的成分（INS）以及不同成分间的混合比例。这种多维度的考量使得模型能够更准确地模拟真实环境中的复杂条件，从而提高了预测结果的准确性和可靠性。

2. 交互作用分析：除了单一因素的作用，模型还深入探究了不同因素之间的交互作用对产率的影响。这种交互作用的分析不仅丰富了模型的理论内涵，也加深了对系统行为规律的理解，使得模型在复杂条件下依然能够保持较高的预测精度。

3. 强解释性：模型的设计注重结果的直观性和可解释性。通过模型输出的结果，研究人员可以清晰地了解到不同因素（如成分和混合比例）对产率的具体影响程度，从而便于进一步的分析和优化。这种强解释性使得模型不仅具有理论价值，也具备较高的实际应用价值。

4. 稳健性强：模型在构建过程中充分考虑了数据的噪声和异常值问题，采用了相应的数据预处理和模型优化技术，使得模型在面临复杂数据和不确定因素时仍能保持较高的稳定性和可靠性。

5. 预测精度高：通过大量的实验验证和对比分析，本模型在预测热解产物产率方面表现出了较高的精度。这主要得益于模型对多因素和交互作用的全面考量，以及对数据噪声和异常值的妥善处理。

7.2 模型的缺点

1. 数据质量与数量限制：模型的预测效果直接受到所使用数据的质量与数量的影响。如果数据集包含大量的噪声、错误或者缺失值，这些问题可能会在模型训练过程中被放大，进而影响模型的预测准确性。此外，当数据量不足时，模型可能无法充分学习到数据的内在规律和模式，导致预测结果的不稳定或偏差。

2. 模型的泛化能力有限：模型的泛化能力是指模型对未知数据的预测能力。然而，在实际应用中，由于数据分布的变化、新情况的出现等原因，模型的泛化能力可能会受到限制。为了提高模型的泛化能力，我们需要采用合适的模型选择方法、正则化技术、集成学习等手段来增强模型的鲁棒性和稳定性。

7.3 模型的改进

针对上述提到的模型缺点，我们提出以下科学严谨的模型改进策略，旨在提高模型的预测准确性、泛化能力以及可解释性。

1. 数据质量与数量的提升:采用先进的数据清洗技术，如插值、中位数填充等方法处理缺失值；通过设定阈值或统计方法识别并处理异常值；利用机器学习中的噪声过滤算法减少噪声数据的影响。整合来自不同渠道、不同来源的数据，以提供更全面、更丰富的信息。这不仅可以提高数据的数量，还可以增强数据的质量。

2. 模型复杂度的调整与优化:通过特征选择算法识别并保留对预测结果影响较大的关键特征，同时去除冗余和无关的特征。利用降维技术（如 PCA、t-SNE 等）减少数据的维度，降低模型的复杂度。根据问题的特性选择合适的模型结构，如深度神经网络、决策树、支持向量机等。对于复杂的系统或现象，可以采用集成学习、深度学习等技术来提高模型的预测能力。利用网格搜索、随机搜索、贝叶斯优化等方法对模型的超参数进行调优，以找到最佳的参数配置。

3. 提高模型的泛化能力:正则化技术：在模型训练过程中引入正则化项，如 L1 正则化、L2 正则化等，以防止模型过拟合，提高模型的泛化能力。集成学习：通过集成多个不同的模型来提高整体的预测性能。

4. 增强模型的可解释性:选择可解释性强的模型：在可能的情况下，优先选择具有较强可解释性的模型，如决策树、逻辑回归等。特征重要性分析：对模型的输入特征进行重要性分析，了解哪些特征对预测结果的影响最大。

7.4 模型的推广

1. 推广应用

（1）化工生产：模型可以应用于化工生产过程中的优化控制，如反应条件的调整、原料配比的优化等，以提高产品质量和生产效率。

（2）能源转化：在能源转化领域，模型可用于预测和优化能源转化效率，如太阳能、风能等可再生能源的转换和利用，以降低能耗和减少污染。

（3）环境保护：将模型应用于环境保护领域，如污染物扩散预测、生态系统稳定性评估等，以支持环境管理和保护决策的制定。

2. 跨领域融合

（1）经济学：将模型与经济学理论相结合，分析经济系统中的动态变化和预测趋势，为经济政策制定提供科学依据。

（2）社会学：在社会学研究中，模型可用于模拟社会现象的传播和演变过程，如信息传播、舆情分析等。

参考文献

- [1] 亚力昆江·吐尔逊, 别尔德汗·瓦提汗, 迪丽努尔·塔力甫, 等. 生物质和煤共热解特性及催化剂对热解的影响[J]. 生物质化学工程, 2018, 52(05): 31-36.
- [2] 李文, 李保庆, 孙成功, 等. 生物质热解、加氢热解及其与煤共热解的热重研究[J]. 燃料化学学报, 1996, (04): 341-347.
- [3] 钱洋洋. 生物质与煤加压热解行为及炭气化反应性研究[D]. 华东理工大学, 2015.
- [4] 何清, 程晨, 龚岩, 等. 水热炭化生物质与煤共热解和共气化特性研究[J]. 燃料化学学报, 2022, 50(06): 664-673.
- [5] 陈吟颖. 生物质与煤共热解试验研究[D]. 华北电力大学(河北), 2007.
- [6] 龚学臣. SPSS 软件在交互效应方差分析中的应用[J]. 河北北方学院学报(自然科学版), 2015, 31(01): 19-22.
- [7] 袁超凤. 模型中交互效应检验的方法及其应用[D]. 黑龙江大学, 2013.

附录

```
clc
```

```
clear
```

```
close all
```

```
%% 表的引入与数值矩阵的转换
```

```
filename = 'table_1.xlsx';
```

```
table = readtable(filename);
```

```
table(:,[1,2,3,12])=[];
```

```
all_numeric = all(arrayfun(@(x) isnumeric(x), table.Variables));
```

```
if all_numeric
```

```
    matrixData = table2array(table);
```

```
end
```

```
% 使用 isnan 函数检测 NaN 值
```

```
is_nan = isnan(matrixData(:,4));
```

```
% 数值矩阵 matrixData 用于方差检验，Ins_matrixData 用于回归拟合
```

```
nan_indices = find(is_nan);
```

```
Ins_matrixData = matrixData;
```

```
Ins_matrixData(nan_indices,:)=[];
```

```
matrixData(nan_indices,[4,8])=0;
```

```
matrixData(:,[1,2,3,8])=[];
```

```
Ins_matrixData = sortrows(Ins_matrixData,4);
```

```
%% 下面是方差检验
```

```
% Z_1 为没有进行正己烷不溶物处理的组，Z_2 为进行了正己烷不溶物处理的组
```

```
Z_1 = matrixData(nan_indices,:);
```

```
Z_2 = matrixData;
```

```
Z_2(nan_indices,:) = [];  
Z_2(:, 1) = 1;  
[tar_p, tar_tbl, tar_stats] = anova1([Z_1(:,2);Z_2(:,2)], [Z_1(:,1);Z_2(:,1)]);  
[water_p, water_tbl, water_stats] = anova1([Z_1(:,3);Z_2(:,3)], [Z_1(:,1);Z_2(:,1)]);  
[char_p, char_tbl, char_stats] = anova1([Z_1(:,4);Z_2(:,4)], [Z_1(:,1);Z_2(:,1)]);
```

%% 下面是回归拟合

```
tar = Ins_matrixData(:,5);  
water = Ins_matrixData(:,6);  
dregs = Ins_matrixData(:,7);  
  
ins = Ins_matrixData(:,4);  
tar = tar ./ sqrt(sum(tar.^2));  
water = water ./ sqrt(sum(water.^2));  
dregs = dregs ./ sqrt(sum(dregs.^2));
```

```
scatter(ins, tar, 10, 'red', 'filled');  
xlabel('ins');  
ylabel('tar');  
title('正乙烷不溶物(x)与焦油产率(y)的拟合图');  
hold on;  
p = polyfit(ins, tar, 1);  
f_ins_tar = @(x)p(1).*x + p(2);  
fplot(f_ins_tar, [0, 1]);
```

```
figure;  
scatter(ins, water, 10, 'red', 'filled');  
xlabel('ins');  
ylabel('water');
```



```
title('正乙烷不溶物(x)与水产率(y)的拟合图');
hold on;
p = polyfit(ins, water, 1);
f_ins_water = @(x)p(1).*x + p(2);
fplot(f_ins_water, [0, 1]);

figure;
scatter(ins, dregs, 10, 'red', 'filled');
xlabel('ins');
ylabel('dregs');
title('正乙烷不溶物(x)与焦渣产率(y)的拟合图');
hold on;
p = polyfit(ins, dregs, 1);
f_ins_dregs = @(x)p(1).*x + p(2);
fplot(f_ins_dregs, [0, 1]);

clc
clear
close all

%% 导入 table_1 表
filename = 'table_1.xlsx';
table_1 = readtable(filename);

%% 将部分表格属性转化为数值矩阵方便运算 matrixData

% 处理 ratio 配比
ratio = table_1(:,3);
ratio = table2cell(ratio);

numericArray = zeros(size(ratio));
```

```
for i = 1:numel(ratio)
    parts = strsplit(ratio{i}, '/');
    numerator = str2double(parts{1});
    if length(parts) == 2
        denominator = str2double(parts{2});
        numericArray(i) = numerator / (denominator + numerator);
    elseif parts{1}
        numericArray(i) = 1;
    end
end

ratio = numericArray;

% 转换 table
table_tmp = table_1;
table_tmp(:,[1,2,3,12])=[];
all_numeric = all(arrayfun(@(x) isnumeric(x), table_tmp.Variables));
if all_numeric
    matrixData = table2array(table_tmp);
end

is_nan = isnan(matrixData(:,4));
nan_indices = find(is_nan);
matrixData(nan_indices,[4,8])=0;

matrixData = [ratio, matrixData];
matrixData(1:19, :) = [];

tmp = 0;
for i=1:length(matrixData(:,1))
    if matrixData(i,1)
        tmp = matrixData(i,1);
    else
        matrixData(i,1) = tmp;
    end
end
```

```
end
end

%% 在 matrixData 中找到要分析的数据
% 配比 5/100, 正己烷处理水平组
idx = find((matrixData(:,1)<0.05)&matrixData(:,5));
Y_x1_z1 = mean(mean(matrixData(idx,6:8),1),1);
% 配比 10/100, 正己烷处理水平组
idx = find(((matrixData(:,1)<0.1)&(matrixData(:,1)>0.05))&matrixData(:,5));
Y_x2_z1 = mean(matrixData(idx,6:8),1);
% 配比 15/100, 正己烷处理水平组
idx = find(((matrixData(:,1)<0.15)&(matrixData(:,1)>0.1))&matrixData(:,5));
Y_x3_z1 = mean(matrixData(idx,6:8),1);
% 配比 20/100, 正己烷处理水平组
idx = find(((matrixData(:,1)<0.2)&(matrixData(:,1)>0.15))&matrixData(:,5));
Y_x4_z1 = mean(matrixData(idx,6:8),1);
% 配比 25/100, 正己烷处理水平组
idx = find(((matrixData(:,1)<0.22)&(matrixData(:,1)>=0.2))&matrixData(:,5));
Y_x5_z1 = mean(matrixData(idx,6:8),1);
% 配比 30/100, 正己烷处理水平组
idx = find(((matrixData(:,1)<0.25)&(matrixData(:,1)>0.22))&matrixData(:,5));
Y_x6_z1 = mean(matrixData(idx,6:8),1);
% 配比 40/100, 正己烷处理水平组
idx = find(((matrixData(:,1)<0.3)&(matrixData(:,1)>0.25))&matrixData(:,5));
Y_x7_z1 = mean(matrixData(idx,6:8),1);
% 配比 50/100, 正己烷处理水平组
idx = find((matrixData(:,1)>0.3)&matrixData(:,5));
Y_x8_z1 = mean(matrixData(idx,6:8),1);

% 配比 5/100, 正己烷未处理水平组
idx = find((matrixData(:,1)<0.05)&(~matrixData(:,5)));
```

```

Y_x1_z2 = mean(matrixData(idx,6:8),1);
% 配比 10/100, 正己烷未处理水平组
idx = find(((matrixData(:,1)<0.1)&(matrixData(:,1)>0.05))&(~matrixData(:,5)));
Y_x2_z2 = mean(matrixData(idx,6:8),1);
% 配比 15/100, 正己烷未处理水平组
idx = find(((matrixData(:,1)<0.15)&(matrixData(:,1)>0.1))&(~matrixData(:,5)));
Y_x3_z2 = mean(matrixData(idx,6:8),1);
% 配比 20/100, 正己烷未处理水平组
idx = find(((matrixData(:,1)<0.2)&(matrixData(:,1)>0.15))&(~matrixData(:,5)));
Y_x4_z2 = mean(matrixData(idx,6:8),1);
% 配比 25/100, 正己烷未处理水平组
idx = find(((matrixData(:,1)<0.22)&(matrixData(:,1)>=0.2))&(~matrixData(:,5)));
Y_x5_z2 = mean(matrixData(idx,6:8),1);
% 配比 30/100, 正己烷未处理水平组
idx = find(((matrixData(:,1)<0.25)&(matrixData(:,1)>0.22))&(~matrixData(:,5)));
Y_x6_z2 = mean(matrixData(idx,6:8),1);
% 配比 40/100, 正己烷未处理水平组
idx = find(((matrixData(:,1)<0.3)&(matrixData(:,1)>0.25))&(~matrixData(:,5)));
Y_x7_z2 = mean(matrixData(idx,6:8),1);
% 配比 50/100, 正己烷未处理水平组
idx = find((matrixData(:,1)>0.3)&(~matrixData(:,5)));
Y_x8_z2 = mean(matrixData(idx,6:8),1);

% X 编号 1 2 3 4 5 6 7 8
% z 水平有 z1 z2 即正己烷处理组和未处理组
% 编号 1 2

%% 构建数值矩阵

% 创建配比 (因素 A) 的分组矩阵
% 假设配比水平从 1 到 8 对应 5/100 到 50/100

```

```
ratio_levels = repmat(1:8, [1 size(Y_x1_z1, 1)]); % 重复配比水平以匹配每个处理组中的观测数
```

```
% 创建正己烷处理水平（因素 B）的分组矩阵
```

```
% z1 代表处理组，z2 代表未处理组
```

```
treatment_levels = zeros(size(Y_x1_z1, 1)*16, 1); % 假设每个配比水平下处理组和未处理组观测数相同
```

```
treatment_levels(1:end/2) = 1; % 处理组标记为 1
```

```
treatment_levels(end/2+1:end) = 2; % 未处理组标记为 2
```

```
% 合并配比和处理水平矩阵以匹配数据顺序
```

```
% 注意：这里假设每个配比水平下的处理组和未处理组观测数相同，且按顺序排列
```

```
ratio_levels = [ratio_levels'; ratio_levels'];
```

```
combined_levels = [ratio_levels, treatment_levels];
```

```
% 创建因变量矩阵（热解产物产率）
```

```
% 这里我们需要将焦油、水产率和焦渣产率合并到一个矩阵中，每列代表一种产率
```

```
% 假设 Y_x#_z# 已经包含了每种配比和处理组合下的平均产率
```

```
tar_rates = [Y_x1_z1(:,1); Y_x2_z1(:,1); Y_x3_z1(:,1); Y_x4_z1(:,1); Y_x5_z1(:,1); Y_x6_z1(:,1);  
Y_x7_z1(:,1); Y_x8_z1(:,1); ...
```

```
Y_x1_z2(:,1); Y_x2_z2(:,1); Y_x3_z2(:,1); Y_x4_z2(:,1); Y_x5_z2(:,1);  
Y_x6_z2(:,1); Y_x7_z2(:,1); Y_x8_z2(:,1)];
```

```
water_rates = [Y_x1_z1(:,2); Y_x2_z1(:,2); Y_x3_z1(:,2); Y_x4_z1(:,2); Y_x5_z1(:,2);  
Y_x6_z1(:,2); Y_x7_z1(:,2); Y_x8_z1(:,2); ...
```

```
Y_x1_z2(:,2); Y_x2_z2(:,2); Y_x3_z2(:,2); Y_x4_z2(:,2); Y_x5_z2(:,2);  
Y_x6_z2(:,2); Y_x7_z2(:,2); Y_x8_z2(:,2)];
```

```
char_rates = [Y_x1_z1(:,3); Y_x2_z1(:,3); Y_x2_z1(:,3); Y_x4_z1(:,3); Y_x5_z1(:,3); Y_x6_z1(:,3);  
Y_x7_z1(:,3); Y_x8_z1(:,3); ...
```

```
Y_x1_z2(:,3); Y_x2_z2(:,3); Y_x3_z2(:,3); Y_x4_z2(:,3); Y_x5_z2(:,3);  
Y_x6_z2(:,3); Y_x7_z2(:,3); Y_x8_z2(:,3)];
```

% 如果需要，可以将所有产率合并到一个矩阵中，但通常我们会对每种产率分别进行 ANOVA

% combined_rates = [tar_rates, water_rates, char_rates]; % 注意：这不是 anova2 的典型用法

%% 现在可以对每种产率分别进行 ANOVA 分析

[tar_p, tar_tbl, tar_stats] = anova2([tar_rates, combined_levels(:,1), combined_levels(:,2)]);

[water_p, water_tbl, water_stats] = anova2([water_rates, combined_levels(:,1), combined_levels(:,2)]);

[char_p, char_tbl, char_stats] = anova2([char_rates, combined_levels(:,1), combined_levels(:,2)]);

clc

clear

close all

%% 导入 table_1 表

filename = 'table_1.xlsx';

table_1 = readtable(filename);

%% 将部分表格属性转化为数值矩阵方便运算 matrixData

% 处理 ratio 配比

ratio = table_1(:,3);

ratio = table2cell(ratio);

nonEmptyRows = ~cellfun(@isempty, ratio);

idx = find(nonEmptyRows);

ratio = ratio(idx);

numericArray = zeros(size(ratio));

for i = 1:numel(ratio)

parts = strsplit(ratio{i}, '/');

numerator = str2double(parts{1});

if length(parts) == 2

```
        denominator = str2double(parts{2});
        numericArray(i) = numerator / (denominator + numerator);
    else
        numericArray(i) = 1;
    end
end
ratio = numericArray;
% 转换 table
table_tmp = table_1;
table_tmp(:,[1,2,3,12])=[];
all_numeric = all(arrayfun(@(x) isnumeric(x), table_tmp.Variables));
if all_numeric
    matrixData = table2array(table_tmp);
end
is_nan = isnan(matrixData(:,4));
nan_indices = find(is_nan);
matrixData(nan_indices,:)=[];
matrixData = [ratio, matrixData];
% 给式样编号
SY = (1:8)';
SY = [SY; ones(6,1)*9];
SY = [SY; ones(5,1)*10];
SY = [SY; ones(5,1)*11];
SY = [SY; ones(8,1)*12];
SY = [SY; ones(5,1)*13];
SY = [SY; ones(6,1)*14];
SY = [SY; ones(5,1)*15];
SY = [SY; ones(5,1)*16];
SY = [SY; ones(5,1)*17];
SY = [SY; ones(5,1)*18];
matrixData = [SY, matrixData];
```

```
% 清理中间变量
clear numericArray nan_indices is_nan i idx SY ratio table_tmp all_numeric parts numerator
nonEmptyRows denominator;

%% 自变量训练集
% 自变量有试样，配比，样品，正己烷不溶物
% 分别存储再 X 中的四列中
% 不包含试样单独热解的点集
X = [matrixData(:, [1, 2, 3, 6])];
X(1:8, :) = [];

%% 根据公式想办法把因变量训练集搞出来
% 因变量有两个，第一个为 output 是有效产率
% 第二个为 theta,是产率的品质

% 因变量计算公式函数
tarSingle_cal = @(beta, M_sample, eta_tar)(1 - beta).*M_sample.*eta_tar;
theta_cal = @(beta, M_sample, eta_tar, M_tar)tarSingle_cal(beta, M_sample, eta_tar)./M_tar;
% 在 matrixData 中的提取 beta, M_sample, eta_tar, M_tar 的集合
beta = matrixData(:, 2);
beta(1:8, :) = [];
M_sample = matrixData(:, 3);
M_sample(1:8, :) = [];
M_tar = matrixData(:, 4);
M_tar(1:8, :) = [];
etar_tar_tmp = matrixData(:, 7);
etar_tar_tmp(9:end, :) = [];
etar_tar(1:6, :) = etar_tar_tmp(1);
etar_tar(7:11, :) = etar_tar_tmp(2);
etar_tar(12:24, :) = etar_tar_tmp(4);
etar_tar(25:29, :) = etar_tar_tmp(2);
```



```
etar_tar(30:35,:) = etar_tar_tmp(1);
etar_tar(36:40,:) = etar_tar_tmp(3);
etar_tar(41:45,:) = etar_tar_tmp(2);
etar_tar(46:50,:) = etar_tar_tmp(1);
etar_tar(50:55,:) = etar_tar_tmp(2);
% 计算因变量训练集
output = M_tar;
theta = theta_cal(beta, M_sample, etar_tar, M_tar);

%% 构造随机森林预测模型
nTrees = 100; % 树的数量
output_predict_F = TreeBagger(nTrees, X, output, 'Method', 'regression');
theta_predict_T = TreeBagger(nTrees, X, theta, 'Method', 'regression');

%% 在预测模型中找到一个较优解
% 可以容忍的最低煤质量
theta = 0.5;

iter = 100;
bestPeibi_sava = [];
for i = 1:iter

% 随机撒点
rand_N = 10000; % 撒点数
x1 = randi([9, 18], rand_N, 1); % 9 到 18 随机试样编号
x2 = rand(rand_N, 1)./2; % 0 到 0.5 随机配比
x3 = rand (rand_N, 1) .* (max(X(:, 3)) - min(X(:, 3))) + min(X(:, 3));
x4 = rand (rand_N, 1) .* (max(X(:, 4)) - min(X(:, 4))) + min(X(:, 4));
% 在随机点中找较优解
possibleInputs = [x1, x2, x3, x4];
predicted_Outputs = predict(output_predict_F, possibleInputs);
```

```
predicted_Theta = predict(theta_predict_T, possibleInputs);

predicted_Outputs_tmp = predicted_Outputs;

[outputs_maxValue, outputs_maxIndex] = max(predicted_Outputs);
while 1
    if predicted_Theta(outputs_maxIndex) < theta
        predicted_Outputs_tmp(outputs_maxIndex) = [];
        [outputs_maxValue, outputs_maxIndex] = max(predicted_Outputs_tmp);
        continue;
    end
    break;
end
% theta = predicted_Theta(outputs_maxIndex);

bestPeibi_sava = [bestPeibi_sava, possibleInputs(outputs_maxIndex, 2)];

end

figure;
plot(1:iter, ones([1, iter]) .* mean(bestPeibi_sava));
hold on
scatter(1:iter, bestPeibi_sava);
xlabel('迭代次数 iter');
ylabel('配比 beta');
title('配比在 100 次迭代中的分布状况及均值');
xlim([0 100]); % x 轴的范围从 2 到 8
ylim([0 0.4]); % y 轴的范围从-1.5 到 1.5
% 获取当前坐标轴句柄
ax = gca;
```

```
% 设置 Y 轴和 X 轴的刻度
```

```
set(ax, 'YTick', 0:0.05:1);
```

```
set(ax, 'XTick', 0:5:100);
```

```
data = [17.46  16.82  15.54  16.33  16.56;  
        15.97  16.17  16.51  16.78  17.21;  
        12.58  13.02  11.16  11.67  11.48;  
        11.34  11.27  11.22  11.17  11.1;  
        5.39   7.44   10.01  10.6   14.06;  
        6.17   7.37   9.39   11.03  13.76;  
        67.01  65.06  61.66  60.55  53.46;  
        68.79  66.99  63.84  61.18  56.92];
```

```
% 设定产品名称和比例
```

```
products = {'Tar', 'HEX', 'Water', 'Char'};
```

```
ratios = {'5/100', '10/100', '20/100', '30/100', '50/100'};
```

```
% 初始化 p-values 存储
```

```
pValues = zeros(length(products), length(ratios));
```

```
hValues = zeros(length(products), length(ratios));
```

```
% 分析每种产物的实验值和计算值之间的显著性差异
```

```
for i = 1:length(products)
```

```
    figure; % 创建新图形
```

```
    for j = 1:length(ratios)
```

```
        % 获取实验值和计算值
```

```
        exp_values = data(2*i-1, j);
```

```
        calc_values = data(2*i, j);
```

```
        % 执行配对 t 检验
```

```
[h, p] = ttest(exp_values, calc_values);

% 存储结果
pValues(i, j) = p;
hValues(i, j) = h;

% 可视化部分
subplot(1, 5, j);
bar([1, 2], [exp_values, calc_values]); % 绘制条形图
title(sprintf('%s - %s\np = %.3f', products{i}, ratios{j}, p));
set(gca, 'XTickLabel', {'Exp', 'Calc'});
ylabel('Yield');
if h == 1 % 如果存在显著性差异
    hold on;
    plot([1, 2], [exp_values, calc_values], 'k*'); % 标记显著性
    hold off;
end
end
sgtitle(['Comparison of Experimental and Calculated ' products{i} ' Yields']);
end

data = [
11.8612.0213.7313.4513.56
12.1712.5213.1313.6414.45
9.139.6711.1310.559.89
8.38.428.638.89.07
6.688.068.6510.4712.08
7.148.099.7411.1113.3
71.770.7465.5262.7258.6
71.7969.8666.4763.659.02
```

```
];
```

```
% 定义产物和混合比例
```

```
products = {'Tar', 'HEX', 'Water', 'Char'};
```

```
ratios = {'5/100', '10/100', '20/100', '30/100', '50/100'};
```

```
% 逐个产物进行 Kruskal-Wallis H 测试和 Tukey HSD 测试
```

```
for i = 1:length(products)
```

```
    exp_values = data(2*i-1, :); % 实验值
```

```
    calc_values = data(2*i, :); % 计算值
```

```
% Kruskal-Wallis H 测试
```

```
[p_kw, tbl_kw, stats_kw] = kruskalwallis([exp_values; calc_values]', [], 'off');
```

```
% 如果显著, 进行 Tukey HSD 测试
```

```
if p_kw < 0.05
```

```
    fprintf('%s has significant differences between experimental and calculated values.\n',  
products{i});
```

```
    figure;
```

```
[c,m,h,gnames] = multcompare(stats_kw, 'CType', 'tukey-kramer', 'Display', 'on');
```

```
    title(sprintf('%s: Tukey HSD Test', products{i}));
```

```
else
```

```
    fprintf('%s does not have significant differences.\n', products{i});
```

```
end
```

```
end
```

```
% 可视化实验值与计算值
```

```
figure;
```

```
for i = 1:length(products)
```

```
    subplot(2, 2, i);
```

```
    boxplot([data(2*i-1, :); data(2*i, :)]', 'Labels', {'Exp', 'Calc'});
```

```
        title(products{i});
    end

    clc
    clear
    close all

    filename = 'question_4_map.xlsx';
    table = readtable(filename);
    all_numeric = all(arrayfun(@(x) isnumeric(x), table.Variables));
    if all_numeric
        matrixData = table2array(table);
    end

    figure;
    for i=1:10

        subplot(2, 5, i);
        plot(1:5,matrixData(:,i*2-1));
        hold on;
        plot(1:5,matrixData(:,i*2));
        xlabel('配比 beta');
        ylabel('产物和 sum');

    end

    legend('实际值','预测值');
    title('产物和预测值与实际值对比图');
```

```
clc
clear
close all

%% 导入 table_1 表
filename = 'table_1.xlsx';
table_1 = readtable(filename);

%% 将部分表格属性转化为数值矩阵方便运算 matrixData

% 处理 ratio 配比
ratio = table_1(:,3);
ratio = table2cell(ratio);
nonEmptyRows = ~cellfun(@isempty, ratio);
idx = find(nonEmptyRows);
ratio = ratio(idx);
numericArray = zeros(size(ratio));
for i = 1:numel(ratio)
    parts = strsplit(ratio{i}, '/');
    numerator = str2double(parts{1});
    if length(parts) == 2
        denominator = str2double(parts{2});
        numericArray(i) = numerator / (denominator + numerator);
    else
        numericArray(i) = 1;
    end
end
ratio = numericArray;

% 转换 table
table_tmp = table_1;
table_tmp(:,[1,2,3,12])=[];
```

```
all_numeric = all(arrayfun(@(x) isnumeric(x), table_tmp.Variables));
if all_numeric
    matrixData = table2array(table_tmp);
end
is_nan = isnan(matrixData(:,4));
nan_indices = find(is_nan);
matrixData(nan_indices,:)=[];
matrixData = [ratio, matrixData];
% 给式样编号
SY = (1:8)';
SY = [SY; ones(6,1)*9];
SY = [SY; ones(5,1)*10];
SY = [SY; ones(5,1)*11];
SY = [SY; ones(8,1)*12];
SY = [SY; ones(5,1)*13];
SY = [SY; ones(6,1)*14];
SY = [SY; ones(5,1)*15];
SY = [SY; ones(5,1)*16];
SY = [SY; ones(5,1)*17];
SY = [SY; ones(5,1)*18];
matrixData = [SY, matrixData];
% 清理中间变量
clear numericArray nan_indices is_nan i idx SY ratio table_tmp all_numeric parts numerator
nonEmptyRows denominator;

%% 自变量训练集
% 自变量有试样，配比，样品，正己烷不溶物
% 分别存储再 X 中的四列中
% 不包含试样单独热解的点集
X = [matrixData(:, [1, 2, 3, 6])];
X(1:8,:) = [];
```



```
%% 因变量训练集
```

```
Y = [matrixData(:, 7:10)];
```

```
Y(1:8,:) = [];
```

```
%% 构造随机森林预测模型
```

```
nTrees = 100; % 树的数量
```

```
tar_predict = TreeBagger(nTrees, X, Y(:, 1), 'Method', 'regression');
```

```
water_predict = TreeBagger(nTrees, X, Y(:, 2), 'Method', 'regression');
```

```
char_predict = TreeBagger(nTrees, X, Y(:, 3), 'Method', 'regression');
```

```
ins_predict = TreeBagger(nTrees, X, Y(:, 4), 'Method', 'regression');
```

```
% 测试
```

```
% 输入：木屑/黑山煤组， 配比[0.04, 0.5]， 样品质量， 正己烷不溶物
```

```
rN = randi([9, 18]);
```

```
idx = find(X(:, 1) == rN);
```

```
maxiter = length(idx);
```

```
possible_input = X(idx, :);
```

```
predict_output = zeros(maxiter, 4);
```

```
predict_output(:, 1) = predict(tar_predict, possible_input);
```

```
predict_output(:, 2) = predict(water_predict, possible_input);
```

```
predict_output(:, 3) = predict(char_predict, possible_input);
```

```
predict_output(:, 4) = predict(ins_predict, possible_input);
```

```
% 输出：焦油产率 1/水产率 2
```

```
scatter(possible_input(:, 2), Y(idx, 1), 10)
```

```
hold on
```

```
scatter(possible_input(:, 2), predict_output(:, 1), 10)
```

```
xlabel('配比 beta');  
ylabel('焦油产率 tar');  
title('控制变量后观测值与预测值对比图');  
  
figure;  
scatter(possible_input(:, 2), Y(idx, 2), 10)  
hold on  
scatter(possible_input(:, 2), predict_output(:, 2), 10)  
xlabel('配比 beta');  
ylabel('水产率 water');  
title('控制变量后观测值与预测值对比图');
```