

中国研究生创新实践系列大赛
“华为杯”第十七届中国研究生
数学建模竞赛

学 校 浙江理工大学

参赛队号 B20103380015

1.李昂

队员姓名 2.卜天

3.林佳佳

中国研究生创新实践系列大赛

“华为杯”第十七届中国研究生

数学建模竞赛

题 目 基于数据挖掘降低汽油精制过程辛烷值损失

摘 要：

汽油的辛烷值（RON）是衡量汽油在气缸内抗爆震燃烧能力的数字指标，值高表示其抗爆性能好。某企业汽油产品辛烷值损失有较大优化空间，急需数据挖掘技术（统计、机器学习等方法）解决辛烷值损失问题。

针对问题一，要按“样本确定方法”对 285 号和 313 号样本进行处理。我们对于某个变量数值全部缺失进行了删除，对于某个变量数值部分缺失的取均值代替。根据拉依达准则剔除异常值，其中利用了贝塞尔公式得到标准误差 σ 。样本 285 中没有发现异常值，样本 313 中发现 53 个异常值，进行了剔除。最后取六点到九点的操作数据的平均值作为对应辛烷值八点的操作变量数据。

针对问题二，要从 367 个变量中寻找影响产品辛烷值的三十个以下主要变量。首先我们进行了散点图分析，观察每个变量和产品辛烷值之间是否存在线性关系，对于能观察到线性关系的变量均予保留。对于不能观察到线性关系以及不能通过非线性变化转化成线性关系的变量进行筛出。接着进行斯皮尔曼相关分析，对在 99% 的置信区间上显著的变量进行相关系数的排名，剔除了相关系数小于 0.3 的变量，得到了 53 个变量。然后对这些变量运用 SPSS 软件进行系统聚类，得到聚类谱系图，从 23 类每类里选出相关系数较高的变量。最后广泛查阅相关文献，综合确定了 29 个主要变量。

针对问题三，根据确定的 29 个主要变量，要通过数据挖掘建立辛烷值损失预测模型，并进行验证。由于炼油工艺过程的复杂性和设备的多样性，它们的操作变量之间具有高度非线性的关系。并且传统的数据关联模型中变量比较少，急需优化。我们分别建立了 BP 神经网络模型和多元非线性回归模型进行辛烷值损失预测。

我们设计了 3 个三层全连接的神经网络模型，将 20 个重要变量全部输入，将 9 个非重要变量分批次输入到模型中。模型的输入层为 23 个神经元，模型预测的结果由三个模型预测值取均值得到。模型一中输入的三个非重要变量得到的均方误差最低，接近 0.6；模型二与模型三得到的均方误差接近 3，在实验的接受范围内。取均值后我们将 9 个非重要变量共同计入考虑，得到的预测值经过均值处理后更加平稳、准确。

建立多元非线性回归模型过程中，我们将 325 个样本分成拟合样本和预测样本。借助 SPSS 软件不断提高模型的拟合优度，得到了辛烷值损失预测函数。接着将 50 个预测样本代入函数，与实际辛烷值损失比较，辛烷值损失预测值与实际值之间的差值在 0.3 以内，精确度较高。

针对问题四，要求保证产品硫含量不大于 $5 \mu\text{g/g}$ 的前提下，得到样本对应的主要变量优化的操作条件，使辛烷值损失降幅大于 30%。由于问题有约束条件和目标，以及由问题三得到的多元非线性回归函数，我们建立了非线性规划模型。目标函数是最小化辛烷损失值，决策变量是 18 个主要操作变量，约束条件是硫含量小于等于 $5 \mu\text{g/g}$ 和每个变量的可操作

范围，还要满足损失降幅 30%以上。我们针对一些样本给出了具体的操作优化方案。

针对问题五，要求用图形展示 133 号样本主要操作变量优化调整过程中对应的汽油辛烷值和硫含量的变化轨迹。结论有反应压力、氢油比、催化汽油进装置总流量、反应器质量空速、氮气进装置流量和产品辛烷值的量化关系，以及主要变量和产品硫含量的量化关系。

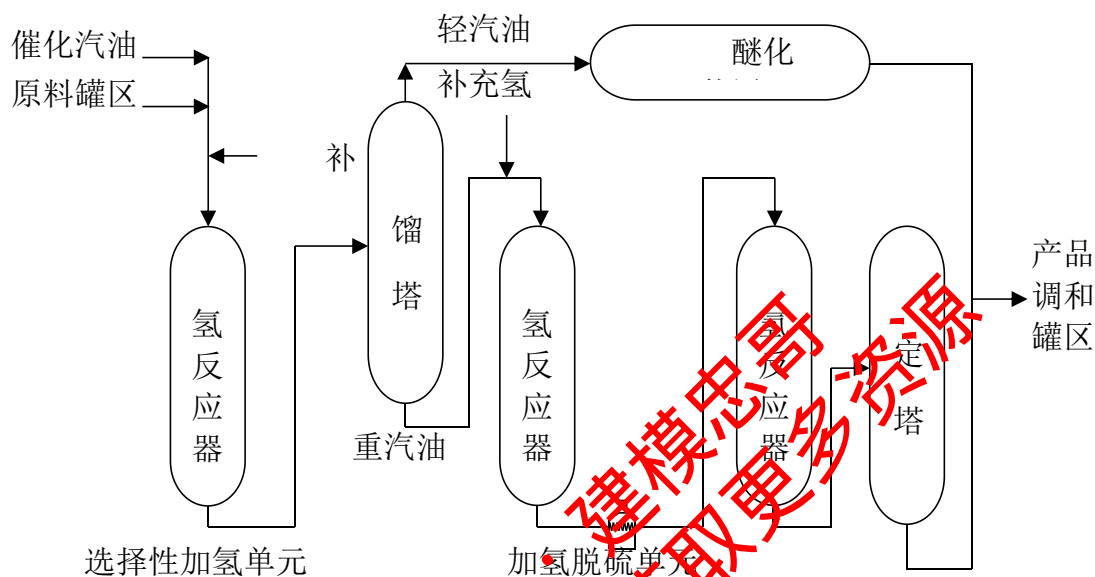
关键词：汽油辛烷值损失、数据挖掘、Spearman 相关系数、系统聚类、BP 神经网络模型、relu 函数、多元非线性回归模型、非线性规划模型

公众号关注：建模忠哥
获取更多资源

一、问题重述

1.1 问题背景

汽油是小型车辆的主要燃料，它燃烧产生的尾气排放会对大气环境造成重大影响。汽油清洁化的重点是降低汽油中的硫、烯烃含量，并且尽量保持其辛烷值。我国原油对外依存度超过。为了有效利用重油资源，需要进行催化裂化处理。所以为了满足汽油质量要求，必须对催化裂化汽油进行精制处理。下图是催化汽油加氢脱硫装置工艺流程图。



图

1 催化汽油加氢脱硫装置工艺流程

汽油的辛烷值（RON）是衡量汽油在气缸内抗爆震燃烧能力的数字指标，值高表示其抗爆性能好。辛烷值每降低1个单位，相当于损失约150元/吨。化工过程的建模一般是通过数据关联或机理建模的方法实现。

某石化企业的催化裂化汽油精制脱硫装置积累了大量历史数据，汽油产品辛烷值损失平均为1.37个单位，但同类装置的最小损失值只有0.6个单位。因此有较大优化空间。问题要求利用数据挖掘技术解决建模问题。数据挖掘指从大量数据中通过算法搜索隐藏在其中的信息的过程。数据挖掘和计算机科学有关，通过统计、机器学习和模式识别等方法来实现目标。建模目标要求产品硫含量不大于 $5\mu\text{g/g}$ 的前提下，尽量降低汽油辛烷值损失在30%以上。

1.2 待解决的问题

问题一：数据预处理，参考近4年的工业数据（附件一）的预处理结果，依“样本确定方法”（附件二）对285号和313号数据样本进行预处理。

问题二：根据附件一，为了建立降低辛烷值损失模型，需要对367个变量进行降维处理，选取30个以下主要变量。

问题三：建立辛烷值（RON）损失预测模型，并进行验证。

问题四：要求在保证产品硫含量不大于 $5\mu\text{g/g}$ 的前提下，进行辛烷值（RON）损失预测模型主要变量操作方案的优化；

问题五：对133号样本以图形展示其主要操作变量优化调整过程中对应的汽油辛烷值

和硫含量的变化轨迹。

二、模型假设

假设一：假设附件中的数据真实有效。

假设二：假设除了附件中的变量，没有别的变量会对产品中的辛烷值 RON 和产品中的硫有影响。

假设三：所有样本在同一个操作环节被加工的设备以及加工方式都相同。

假设四：假设该石化企业的催化裂化汽油都来自中东地区。

三、符号说明和名词解释

3.1 符号说明

符号	说明
RON	辛烷值
δ	拟合误差
ρ	相关系数
R^2	拟合优度
VLB	下界
WUB	上界
X	n 元向量
x_i	主要操作变量
$G(X)$	非线性函数组成的向量

3.2 名词解释

氢油比是指进料汽油和装置循环氢的比值。

质量空速在 S Zorb 装置中是指进料量与反应器内藏量的比值。

斯皮尔曼相关系数是用来描述两个样本的数据是否在一条直线上，也就是描述两个样本之间是否具有相关性，以及两样本之间相关强度的指标。

BP 神经网络是基于误差反向传播算法的若干层前向神经网络，代表神经网络里最精华的一部分。

非线性规划是求解目标函数或约束条件中有一个或几个非线性函数的最优化问题的方法。

四、问题一数据预处理

参照近 4 年 325 个样本数据的预处理结果，按“样本确定方法”对 285 号和 313 号数据样本进行预处理。操作变量数据来自于实时数据库，原料、产品和催化剂数据来自于 LIMS 实验数据库，其中原料及产品的辛烷值是重要的建模变量。

表 1 附件三样本操作变量缺失情况表

样本编号	原始数据时间	缺失情况
285 号	2017 年 7 月 17 日 6 点至 2017 年 7 月 17 日 8 点	有 11 列操作变量全部缺失。
313 号	2017 年 5 月 15 日 6 点至 2017 年 5 月 15 日 8 点	有 8 列操作变量全部缺失， 有 5 列操作变量部分缺失。

针对部分缺失的数据，我们用该列数据的均值进行代替。针对全部缺失的数据，我们将此类点删除。

下表是 285 号样本部分变量描述性统计

表 2 285 号样本部分操作变量描述性统计

操作变量	样本数	最小值	最大值	均值	标准差
氢油比	40	0.2733002	0.2734981	0.2733991	0.00005931
反应过滤器压差	40	24.17817	24.23831	24.208241	0.01802913
还原器压力	40	2.528462	2.529279	2.52887043	0.00024475
还原器流化氢气流量	40	855.5165	856.2485	855.882523	0.2193988
反应器上部温度	40	421.5055	421.5132	421.509325	0.0023060

利用 Spss 软件进行描述性统计，可以得到 285 号样本的描述统计。通过观察可以发现氢油比、还原器压力的均值比较低，氢油比的均值仅有 0.2，而还原器压力以及反应堆上部温度的均值非常高，还原器流化氢气流量的均值达到 855。本实验的数据指标有 367 个，对于观察法来说很难实现，我们需要在数据的预处理步骤对异常值、空缺值以及无效值进行筛查。

异常值检验：在汽油辛烷值的样本数据中，我们发现了一些异常值，例如存在样本值远远大于样本均值的现象，异常值检验是剔除那些过大、过小，或因外界因素而不符合规律的样本值，所以在数据预处理计算，我们利用异常值检验来剔除异常值。

我们根据拉依达准则来去除异常值，拉依达准则将随机误差进行了保留，目的在于去除粗大误差。在拉依达准则规范中，关键在于计算标准误差，这里的标准误差通过贝塞尔

公式得到：

$$\sigma = \left[\frac{1}{n-1} \sum_{i=1}^n v_i^2 \right]^{1/2} = \left\{ \left[\sum_{i=1}^n x_i^2 - \left(\sum_{i=1}^n x_i \right)^2 / n \right] / (n-1) \right\}^{1/2}$$

具体步骤如下流程图：

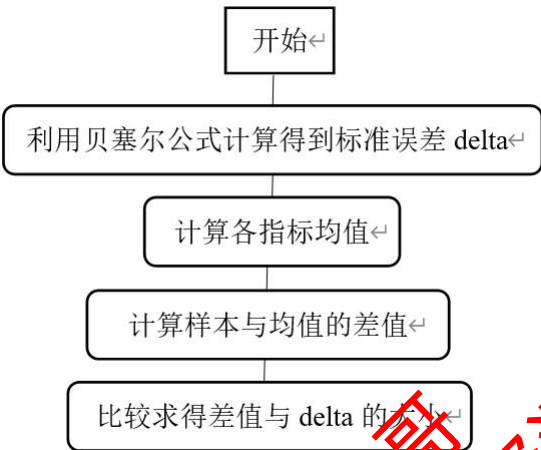


图 1 异常值检验流程图

下面两个表格分别是样本 285 和样本 313 部分变量的标准误差。

表 3 样本 285 部分变量的标准误差

氢油比	反应过 滤器压 差	还原器 压力	还原器 流化氢 气流量	反应器 上部温 度	反应器 底部温 度
5.9314E-05	0.018029	0.000245	0.219399	0.002306	0.001778

表 4 样本 313 部分变量的标准误差

氢油比	反应过 滤器压 差	还原器 压力	还原器 流化氢 气流量	反应器 上部温 度	反应器 底部温 度
5.9314E-05	0.018029	0.000245	0.219399	0.002306	0.001773

经过异常值检验后，我们在样本 285 数据中没有发现异常值，而在样本 313 数据中，共发现了 53 个异常值，说明样本 285 相较样本 313 而言标准很多。对于样本 313 的异常值我们进行了验证，我们通过手工操作就能发现 313 样本存在很多远大于均值的量，这表明我们的异常值剔除是非常必要的。将 285 每个原始操作变量 6 点到 8 点的数值求均值，作为 285 样本 8 点的各个变量值。部分结果如下表。

表 5 285 样本预处理后的部分结果

氢油比	反应过 滤器压 差	还原器 压力	还原器 流化氢 气流量	反应器 上部温 度	反应器 底部温 度
0.2734	24.2082	2.5288	855.8825	421.5093	421.1962

以上结果和附件一中的数据基本一致，有些是小数点后 3 位的差异。将 313 每个原始

操作变量 6 点到 8 点的数值求均值，作为 313 样本 8 点的各个变量值。部分结果如下表。

表 6 313 样本预处理后的部分结果

氢油比	反应过 滤器压 差	还原器 压力	还原器 流化氢 气流量	反应器 上部温 度	反应器 底部温 度
0.26193	17.1834	2.4170	849.228	424.968	424.522

以上结果除了异常值较多的变量有细微差异，变量值和附件一中的值基本一致。

五、问题二的分析与求解

5.1. 问题分析

问题二要求寻找建模主要变量。由于模型涉及 7 个原料性质、2 个产品性质、2 个待生吸附剂性质、2 个再生吸附剂性质以及 354 个操作变量，需要根据 325 个样本数据，筛选出建模主要变量。

通过观察发现，有一部分自变量对辛烷值的优化起到至关重要的作用，我们的目标就是找到这些在催化裂化过程中，即施工工艺的包括反应器和稳定塔中，影响产品辛烷值的三十个以下主要指标。我们运用相关性分析，系统聚类等方法，结合文献综合确定建模主要变量。

5.2 问题求解

相关性系数是研究两个变量的相似于线性关系的强度的指标，通过相关性分析以及显著性检验我们能够确定两个变量之间具有线性相关关系的强度。在 325 个样本中，我们针对因变量（产品辛烷值 RON），对指标进行了相关性分析。

5.2.1 散点图分析

我们以 2017 年至 2020 年的数据作为样本，将产品辛烷值作为对比变量，将重要的位点指标（以原料——饱和烃、原料——芳烃、原料——密度为例）与产品辛烷值的关系用散点图进行分析，观察位点指标与辛烷值之间是否存在线性关系，部分结果如下。

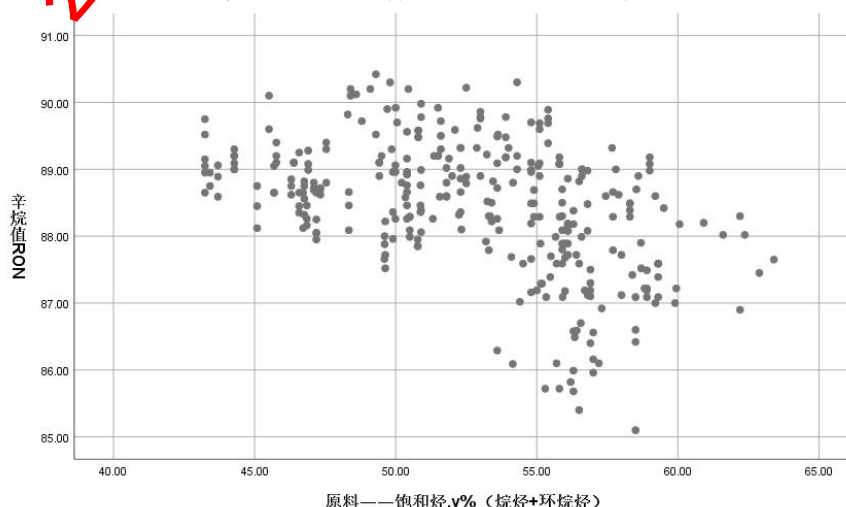


图 2 原料——饱和烃与产品辛烷值散点图

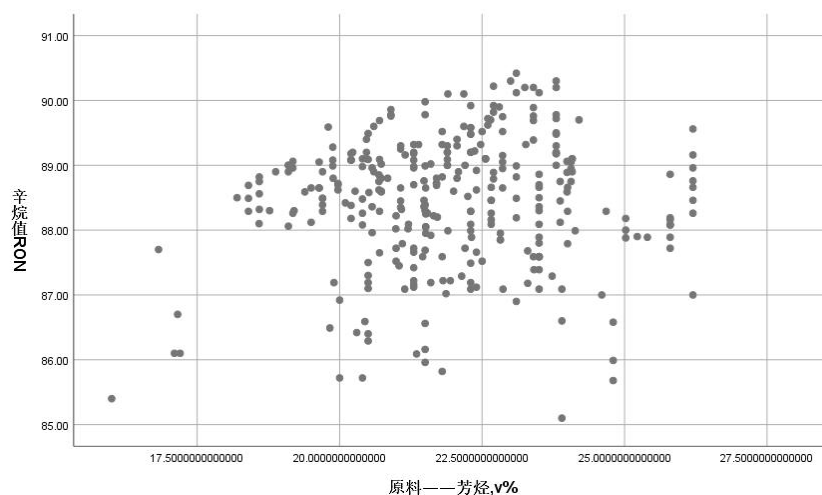


图 3 原料——芳烃与产品辛烷值散点图

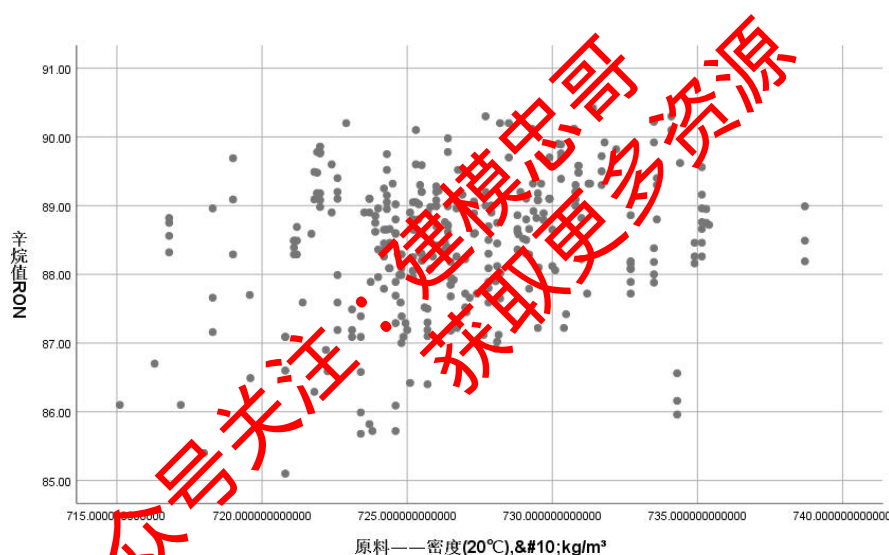


图 4 原料——密度与产品辛烷值散点图

观察上图可以发现，辛烷值与原料——饱和烃有较强的线性关系，二者之间存在一定的负相关关系，随着原料——饱和烃的增长，辛烷值逐渐减少。辛烷值与原料——芳烃不明显，但仍然可以发现一定的相关趋势。原料——密度辛烷值的线性关系也比较明显，可以发现一定的正相关的趋势。对于这些能够观察出线性关系的指标，我们均予保留。

除了标准线性关系以外，我们也发现了一些对于不存在线性关系以及不能通过非线性变化转化成线性关系的位点指标，以 D121 顶去放火炬流量为例，该类指标与辛烷值的这散点图如下：

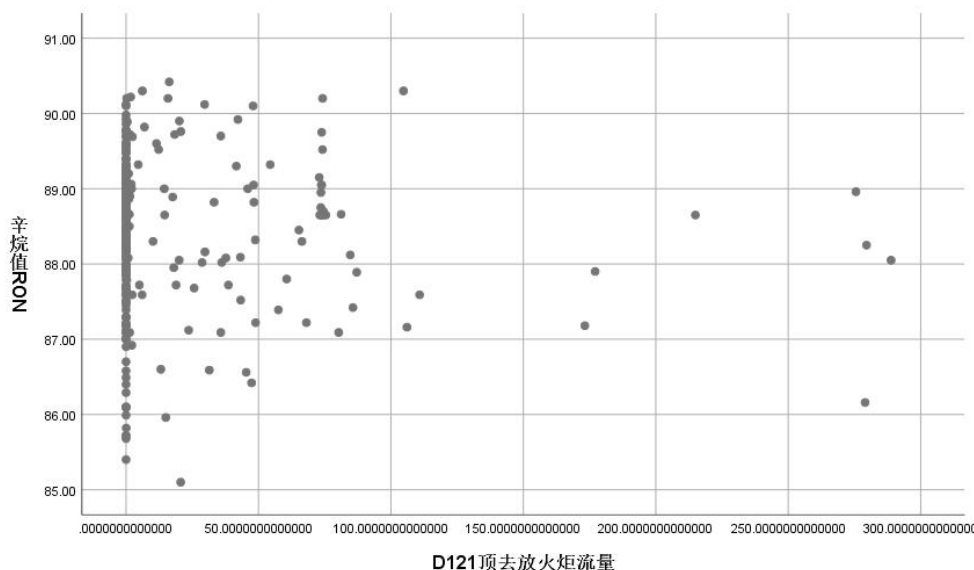


图 5 D121 顶去放火炬流量与产品辛烷值的散点图

根据观察可知, D121 顶去放火炬流量与辛烷值的散点图呈现与 y 轴垂直的式样, 这表明 D121 顶去放火炬流量的变化并不能直接影响辛烷值的变化, 从图中其他的点上也不能过观察出指数、对数、幂次的非线性关系。所以我们将这类与辛烷值关系都较少的变量进行筛出, 共筛选出 D121 顶去放火炬流量、污油出装置、低压热氮气压力、HY2533 手操器、火炬罐 D-206 液位等五个指标, 指标的散点图一并附录中。

5.2.2 斯皮尔曼 (Spearman) 相关系数

当样本不符合正态分布时，通常使用斯皮尔曼相关系数来做相关性分析。斯皮尔曼相关系数是用来描述两个样本的数据是否在同一条直线上，也就是描述两个样本之间是否具有相关性，以及两样本之间相关强度的指标，当相关系数的绝对值越接近于 1 时，就代表两个样本越相关。斯皮尔曼相关系数的公式如下：

$$\rho = \frac{\sum_i (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_i (x_i - \bar{x})^2 \sum_i (y_i - \bar{y})^2}}$$

其中 ρ 表示相关系数,当 ρ 值的绝对值越接近于1时,就表示两个样本之间越线性相关。用 Matlab 得到的各个变量和产品辛烷值的相关性系数如下表所示:

表 7 斯皮尔曼部分相关系数表

硫含量,μg/g,	饱和 烃,%(体积分数)	烯烃,%(体积分数)	芳烃,%(体积分数)	D105 压力	溴 值,gBr/100g
0.4473**	0.9727**	-0.4877**	0.4080**	0.0662	-0.0829
硫含量,%(质量分数)	碳含量,%(质量分数)	硫含量,%(质量分数)	氢油比	反应过滤器 压差	还原器压力
0.0253	0.0891	-0.0108	-0.1827**	-0.1076	0.0128

反应器顶部 压力	反应器顶底 压差	反应过滤器 压差	反吹氢气温 度	D105 温度	...
0.0460	-0.1251*	0.1785**	0.2326**	-0.1881**	...

注: **为在 0.01 水平上显著(双尾), *为在 0.05 水平上显著(双尾)

由上表可知, 数据中存在大量在 99%的置信区间上显著的指标, 经过统计有 193 个指标显示为显著, 这可能时样本量比较大的原因, 对于大量的显著指标我们也要客观的分析, 不能一味地处理为自变量。如显著相关的指标中硫含量、饱和烃、烯烃、芳烃等指标, 相关系数确实比较大, 达到中相关性和高相关性, 这说明这些指标确实在一定程度上与辛烷值线性相关。但同样显著相关的指标 D105 温度、反应器顶底压差、反应器底部温度、还原器流化氢气流量等指标的相关性系数比较低, 对于这些指标我们需要计算考量, 不能一概而论。

为了防止在置信区间上显著, 但相关系数较低的指标过多, 我们进行了进一步的筛选。我们对显著的相关性系数进行了数值上的排名, 排名结果如下图所示。



图 6 相关系数排名后部分结果图

我们将相关系数大于 0.3 的指标进行保留, 将相关系数小于 0.3 的指标进行筛选。最后共得到了 53 个筛选后的变量。

5.2.3 系统聚类

我们将 53 个变量进行系统聚类。系统聚类的流程如下图:

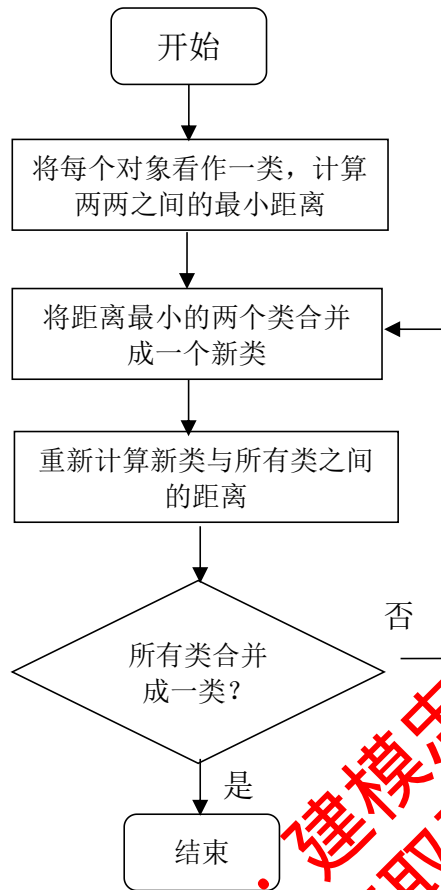


图7 系统聚类流程图

我们将这 53 个指标运用 SPSS 软件进行系统聚类，得到如下聚类谱系图。方法使用平均联接（组间），重新标度的距离聚类组合。

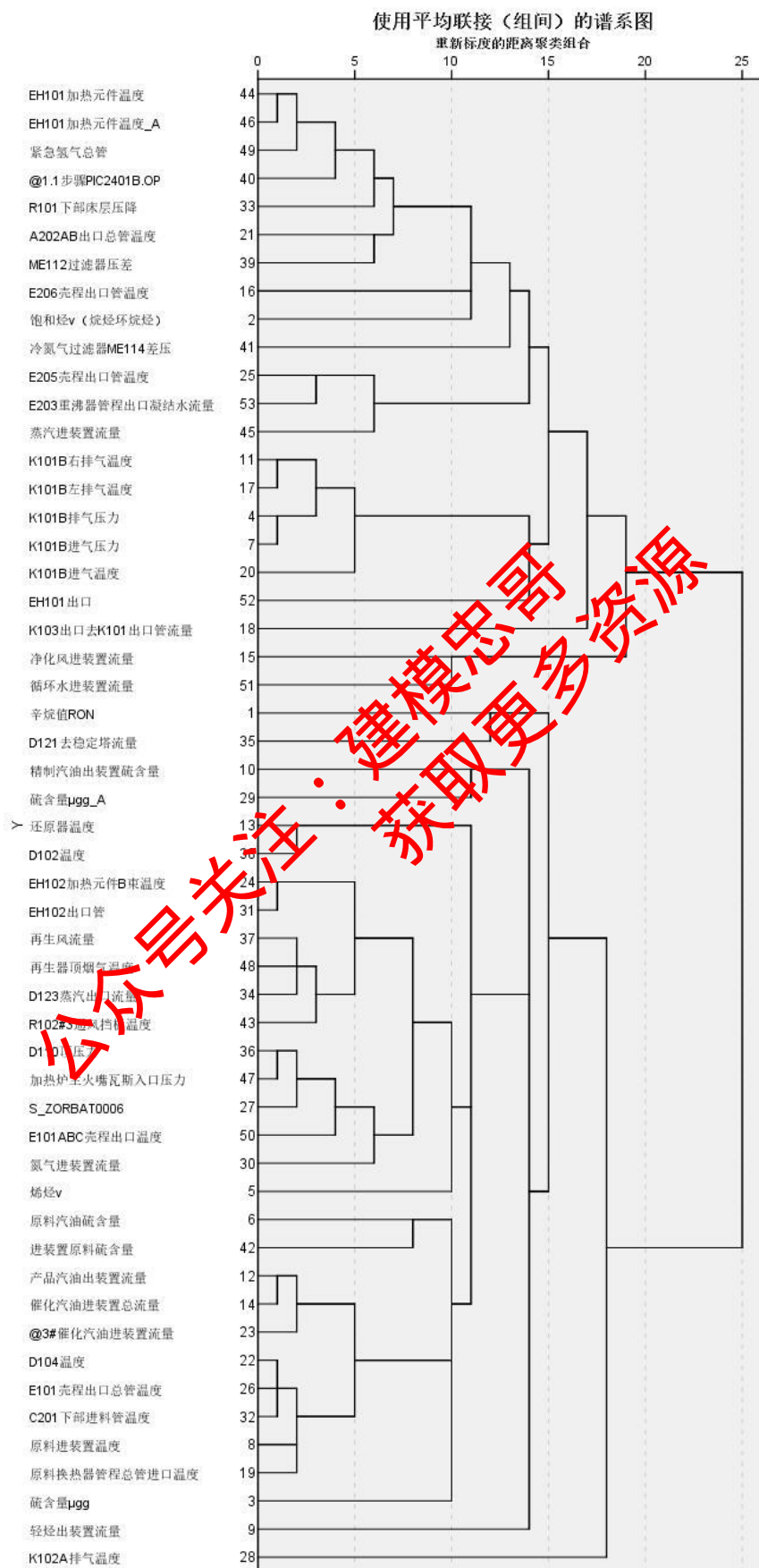


图 8 聚类谱系图

该谱系图将变量分成 23 类，我们将该谱系图作为最后选取主要变量的依据。

5.2.4 综合确定主要变量

以下是通过文献查阅对辛烷值 RON 损失影响较大的几个变量，都予以保留。

第一，原料硫含量。原料中的硫含量越高，达到相同的脱后油流含量，条件越苛刻，比如需要高吸附剂活性、高氢分压等。因此原料汽油性质的优化是降低汽油辛烷值损失的主要因素。

第二，吸附剂。再生剂上的硫越低，其活性越高，烯烃加氢反应程度越大，辛烷值损失越大。所以要在保证产品硫含量合格的前提下，提高再生剂的硫含量降低辛烷值损失。生产中，反应器中吸附剂含硫量低、活性偏高，加速了汽油烯烃饱和反应，增加了吸附脱硫中的辛烷值损失。

第三，氢油比。氢油比越大，烯烃加氢反应也越大，辛烷值损失越大。

第四，反应温度。烯烃加氢饱和反应是强放热反应，因此提高反应温度能抑制烯烃饱和反应。控制其他条件不变的情况下，随着反应温度提高，产品的辛烷值损失逐渐减小。但不能超过反应器要求的最高值。

第五，反应压力。烯烃加氢反应是体积减小的化合反应，增大反应压力或提高氢分压，会加快烯烃反应，导致烯烃含量减少，辛烷值损失增大。反应压力主要是通过氢分压影响反应进行。所以反应压力对辛烷值损失的影响要同时考虑氢气纯度和循环量影响。

结合相关性分析，聚类 and 文献参考，我们综合考虑选取以下 29 个主要变量，包括 11 个非操作变量和 18 个操作变量。

表 8 主要非操作变量表

序号	变量名	简述
1	原料硫含量,μg/g	原料硫含量越高，辛烷值损失越大。
2	原料辛烷值 RON	产品中的辛烷值和原料中的辛烷值有很大关系。
3	原料饱和烃,v%(烷烃+环烷烃)	饱和烃包括环烷烃和链状烷烃。
4	原料烯烃,v%	烯烃加氢饱和变成烷烃后会导致辛烷值下降。
5	原料芳烃,v%	指分子中含有苯环、芳香环结构的碳氢化合物。
6	原料溴值,gBr/100g	溴值是 100g 燃料所能吸收溴的克数。
7	原料密度(20℃), kg/m ³	密度等于物体的质量除以体积。
8	待生吸附剂性质焦炭,wt%	待生吸附剂的持碳率越低，则其活性越高。
9	待生吸附剂性质 S, wt%	待生吸附剂的持硫率越低，则其活性越高。
10	再生吸附剂性质焦炭,wt%	经空气烧焦，脱除炭和部分硫，由氢气还原表面的镍组分，活性得到恢复。
11	再生吸附剂性质 S, wt%	其硫含量一般为 3%~6%。

表 9 主要操作变量表

序号	变量名	简述
----	-----	----

1	氢油比	氢油比过大，烯烃加氢反应程度也越大，产品辛烷值损失会增大。
2	反应过滤器压差	过滤介质阻碍流体流动,产生了压力差。
3	反应器质量空速	简称 WHSV，指反应器进料量与吸附剂藏量的比值，增加质量空速，会减缓烯烃通过加氢饱和和成烷烃的反应速率，令汽油 RON 损失变低。
4	反应器入口温度	反应器和稳定塔是两个重要的发生装置。因为烯烃加氢饱和反应是强放热反应，因此提高反应温度能抑制烯烃饱和反应。在其他条件不变的情况下，反应温度越高，辛烷值损失减小。
5	加热炉循环氢出口温度	
6	反应器底部温度	
7	稳定塔下部温度	
8	稳定塔顶压力	烯烃加氢反应是化合反应，使得产物压力变小。
9	反应系统压力	因此增大反应压力会使烯烃加氢反应剧烈，增大辛烷值损失。
10	原料进装置流量	聚类结果包括以上部分指标，以及左侧 10-18 号指标。
11	原料汽油硫含量	
12	原料进装置温度	
13	精制汽油出装置硫含量	
14	还原器温度	
15	催化汽油进装置总流量	
16	A-202A/B 出口总管温度	
17	氮气进装置流量	
18	D121 去稳定塔流量	

六、问题三的分析与求解

6.1 问题分析

问题三要求依据样本和建模主要变量，通过数据挖掘建立辛烷值（RON）损失预测模型，并进行模型验证。

由于炼油工艺过程的复杂性和设备的多样性，它们的操作变量之间具有高度非线性的关系。并且传统的数据关联模型中变量比较少，急需优化。我们分别建立 BP 神经网络模型和多元非线性回归模型进行辛烷值损失预测。人工神经网络作为一种智能机器学习算法，具有良好的非线性特征，在函数逼近中应用很广泛。这种算法在多因素变量回归中优势较大，适用难以用常规数学模型描述的过程。

6.2 BP 神经网络预测模型的建立与求解

BP 神经网络是基于误差反向传播算法的若干层前向神经网络，代表神经网络里最精华的一部分，是广泛应用神经网络学习算法之一。其非循环多级网络训练算法使其具有实用性。它能以任意精度逼近任意连续函数，实现非线性映射。

辛烷值损失预测模型由典型三层全连接神经网络构建，神经网络由输入层，隐含层和输出层构成。针对输入信号，首先向前传播到隐含层，经过作用函数后，把隐含层的输出信号传到输出节点，最后给输出结果。激励函数选 Sigmoid 型函数。BP 算法学习过程由正向和反向传播组成。正向传播中，输入信息由输入层经过隐含层加权处理传到输出层，由作用函数运算后得到输出值，与期望值比较。如果有误差，则误差反向传播，沿原路返回，利用反向传播调整网络权值和阈值，令网络的误差平方和最小。不断循环一直到满足输出要求。神经网络模型要选择适宜的训练参数和函数，使模型泛化能力好，预测误差小最小。

在本实验中使用的激活函数为 relu 函数，relu 函数是一种功能强大、有间接的非线性函数，在 BP 神经网络中得到了广泛的应用。BP 神经网络模型内部构架如下图所示：

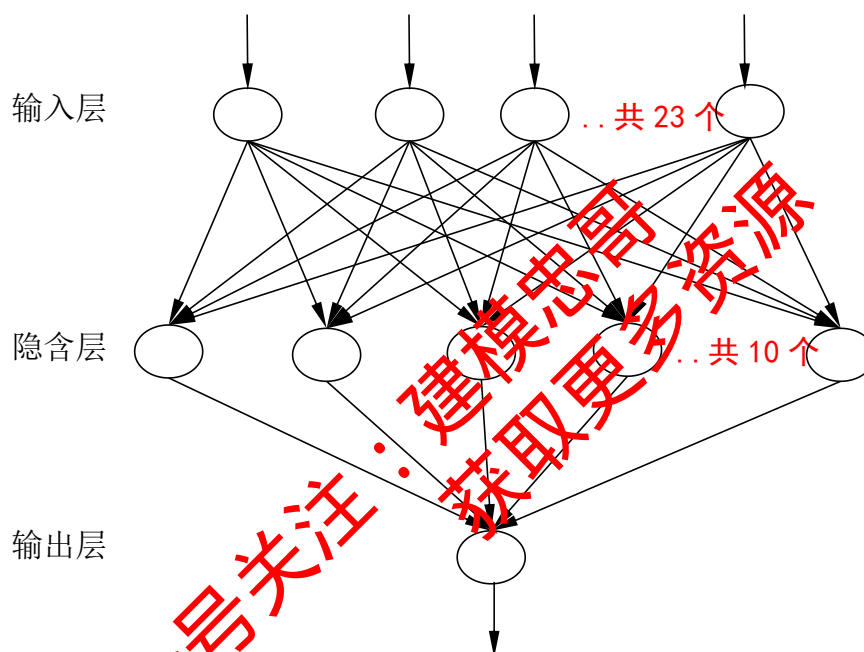


图 9 BP 神经网络模型内部图

在本神经网络中，将输入层神经元的个数设定为 23 个，这是由聚类后以及综合分析得到的指标重要性差别而特别设立的。模型输入层的 23 个神经元由 20 个重要变量和 3 个次重要变量构成。通过实验证实，相比较构建一个输入为 29（20 个重要变量以及 9 个次重要变量）个变量的神经网络模型，其训练结果以及均方误差远远差于构建三个输入为 23（将上述网络改为 3 个次重要变量）的神经网络模型。其原因在于，本实验的样本只有 325 个，其样本值不能远远大于参数总量。所以我们将次重要变量分批建模，其实也是在减少模型的参数，从而提高模型训练的效果。模型的整体架构如下图所示：

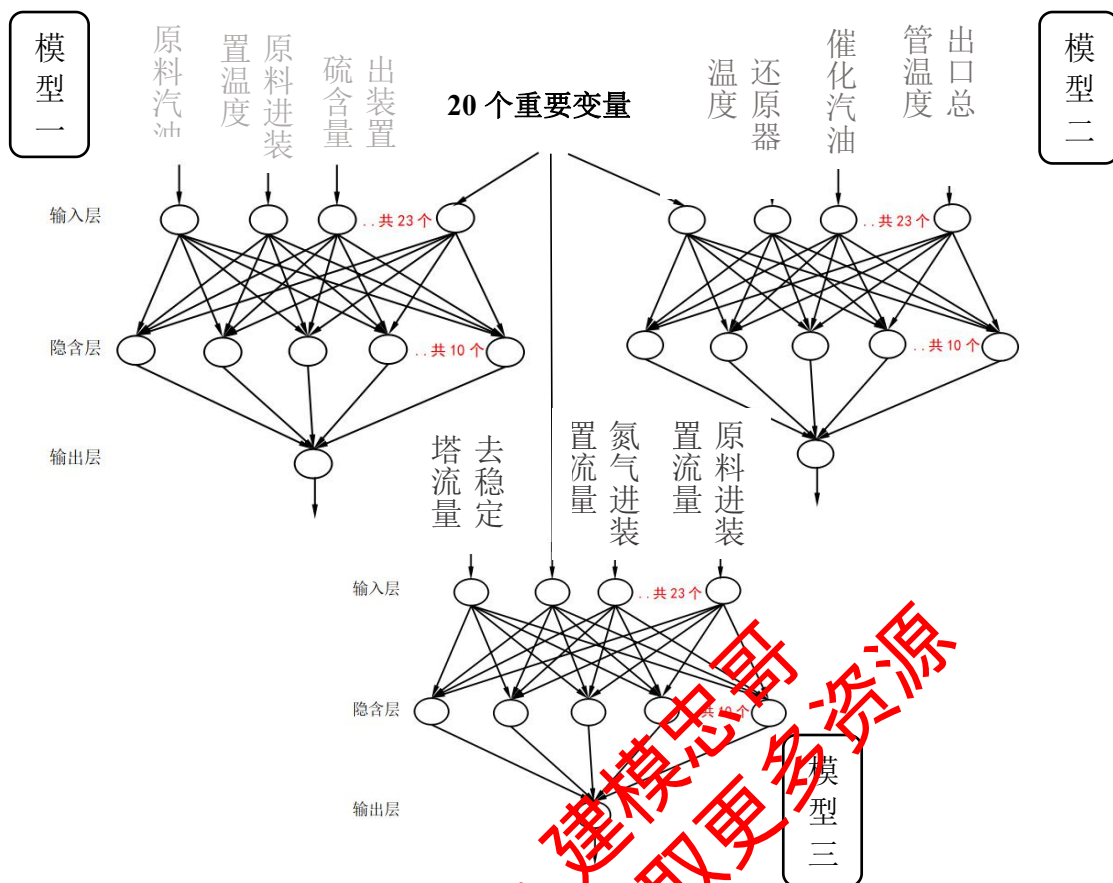


图 10 基于问题三的 BP 神经网络模型

我们将 20 个重要变量输入到模型中，这是因为重要操作变量是预测辛烷值损失必不可少的指标。除此以外，我们将剩余 9 个次重要变量随机分组，将（原料汽油、原料进装置温度、出装置硫含量）作为第一个模型的输入；将（还原器温度、催化汽油、出口总管温度）作为第二个模型的输入；将（去稳定塔流量、氮气进装置流量、原料进装置流量）作为第三个模型的输入。经过实验证明，三个模型的均方误差都在我们的接受范围内，所以我们对三个模型的 \hat{y}_{hat} 预测值进行取均值的操作作为预测的最后结果。

我们的总样本量为 325 个样本，为了使得模型的训练更加准确，我们将训练集和测试集的比例设定为 5: 1，即取得最后 50 个样本为测试集。模型预测的测试集结果如下表所示：

表 10 三种模型预测均值表

样本序号	模型一	模型二	模型三	预测均值	目标值
1	89.2249	86.9574	89.3854	88.5225	88.8
2	89.4968	89.9027	89.4619	89.6204	88.72
3	89.281	89.1831	90.7825	89.7488	88.45
4	89.0092	88.2596	90.7795	89.3494	88.12
5	89.5853	84.8716	90.8854	88.4474	88.75
6	89.2703	88.3635	89.5983	89.0773	88.76
7	88.9118	87.4037	89.4928	88.6027	88.46
...

我们将结果分为两个部分，第一部分为三个模型各自的预测值，第二部分我们计算了三个模型的均值作为模型最后的预测结果。子模型的折线图如下图所示：

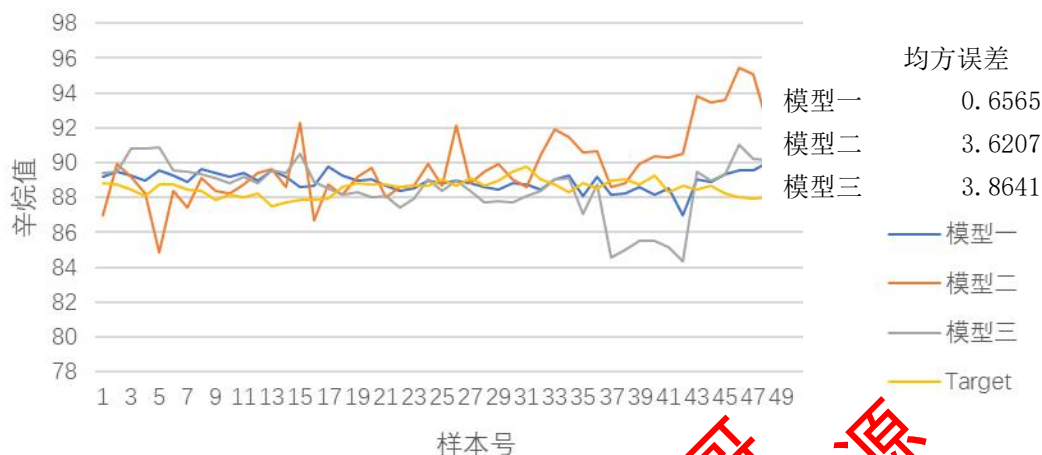


图 11 三个模型辛烷值预测图

可以看到三个模型基本上还是在目标值上下波动范围波动。从预测图的波动形势来看，模型一与目标值最为相近，波动范围小，比较稳定；模型二在最后预测时出现了一点偏差；模型二与目标值差别较大，出现误差大的预测值比较多。通过和三个模型的均方误差的验证，发现模型的趋势是正常的。模型二和模型三的均方误差接近，均为小于 4；而模型一的均方误差最少，仅有 0.65，这与预测样本的结果走向是符合的。除此以外，我们将三个预测值进行了均值操作，得到的折线图如下所示。



图 12 三个模型均值与目标值对比图

可以看到预测均值和目标值整体是符合规律的，仅有最后几个值偏差比较大。但对三个模型进行均值处理也有一定的优势，我们将提取的九个次重要变量进行了整合，可以使得辛烷值的预测更加稳定。

6.3 多元非线性回归预测模型的建立与求解

与多元非线性回归相对应的是多元线性回归，然而多元线性非回归能够对因变量的影响产生更加重要的意义。如本题中预测辛烷值损失中，共有 29 个变量对因变量产生影响，应用多元线性回归的实用意义很大，在实际应用中变量之间的非线性变化，以及交互项对

因变量的预测产生重要的影响，非线性变化的加入也能够让因变量得到更加准确的效果。多元非线性回归的公式如下所示：

$$Zy = f(Z_1) + f(Z_2) + f(Z_3) + \dots + f(Z_k)$$

这里的 f 指非线性函数，包括幂次、指数、 S 曲线、对数函数等。这里还应该注意的，由于不同的指标具有不同的单位，如本示例中筛选出来的指标单位就有：摄氏度、密度、速度等。不同的单位加入到公式运算中，即使得到了如上公式中的系数 β_i ，但这时的系数值大小不能代表该指标的重要性权重。所以本例中应用了 *SPSS* 软件进行标准化的操作，得到了标准化指标。

基于第二问主要变量的确定，记产品的辛烷值 RON 为 y ， RON 损失为 r 。以下将构建产品辛烷值 y 与 29 个主要变量的回归模型。自变量符号如下表所示。

表 11 变量符号表

变量符号	变量名	变量符号	变量名
x_1	原料汽油硫含量	x_{16}	原料进装置流量
x_2	原料进装置温度	x_{17}	加热炉循环氢出口温度
x_3	精制汽油出装置硫含量	x_{18}	反应过滤器压差
x_4	还原器温度	x_{19}	原料辛烷值 RON
x_5	催化汽油进装置总流量	x_{20}	饱和烃, v% (烷烃+环烷烃)
x_6	A-202A/B 出口总管温度	x_{21}	硫含量, $\mu\text{g/g}$
x_7	氮气进装置流量	x_{22}	烯烃, v%
x_8	D121 去稳定塔流量	x_{23}	芳烃, v%
x_9	氢油比	x_{24}	溴值, $\text{gBr}/100\text{g}$
x_{10}	反应器底部温度	x_{25}	密度(20°C), kg/m^3
x_{11}	稳定塔顶压力	x_{26}	待生吸附焦炭, wt%
x_{12}	稳定塔下部温度	x_{27}	待生吸附 S, wt%
x_{13}	反应系统压力	x_{28}	再生吸附焦炭, wt%
x_{14}	反应器质量空速	x_{29}	再生吸附 S, wt%
x_{15}	反应器入口温度		

首先，假设产品辛烷值 y 与 29 个主要变量之间是多元线性回归模型，即：

$$y = b_0 + b_1x_1 + b_2x_2 + \dots + b_{29}x_{29}$$

$$r = x_{19} - y$$

式中： b_0 、 b_1 ... b_{29} 为待估系数。

首先我们用 *SPSS* 软件的相关性分析功能，对各自变量之间进行相关性检验，发现变量之间没有很高的相关性，由此我们可以排除自变量之间共线性对回归模型带来的影响，之后我们借助 *SPSS* 软件分析-回归功能，以此来构建多个主要变量和因变量之间的回归关

系。为了更好地对模型进行验证，我们将附件一的样本分为两个部分：前 275 个样本作为模型拟合样本，后 50 个样本作为预测样本并与实际值进行比较。

通过 SPSS 软件进行计算，得出各回归系数，由此得到回归方程为：

$$y = 8.651 + 0.001x_1 - 0.008x_2 + 0.002x_3 + 0.003x_4 + 0.001x_5 - 0.002x_6 - 0.000128x_7 + 0.036x_8 - 2.931x_9 - 0.009x_{10} + 4.004x_{11} - 0.013x_{12} - 0.687x_{13} + 0.012x_{14} - 0.002x_{15} - 0.004x_{16} + 0.001x_{17} - 0.005x_{18} + 0.962x_{19} - 0.009x_{20} + 0.00001773x_{21} + 0.0001x_{22} + 0.006x_{23} - 0.000347x_{24} - 0.001x_{25} + 0.03x_{26} - 0.027x_{27} - 0.064x_{28} + 0.037x_{29}$$

$$r = x_{19} - y$$

根据 SPSS 软件运行结果，如下表所示，表中调整后 R^2 值用来衡量估计的模型对观测值的拟合程度，此时该模型的拟合优度 R^2 值为 0.958，接近于 1，表示模型的拟合优度较好。

表 12 拟合优度 R^2 值表

模型	R	R 方	调整后 R 方	标准估算的错误	德宾-沃森
1	.981 ^a	.962	.958	.90700	1.965

为了更好地提高模型的拟合优度，我们对一些数值较大的变量进行取自然对数处理，以此来降低其数值大小，之后用 SPSS 软件进行回归拟合，得到新的非线性回归方程模型：

$$y = -25.312 + 0.016 \ln x_1 - 0.005x_2 + 0.003x_3 - 0.985 \ln x_4 + 0.096 \ln x_5 - 0.001x_6 - 0.034 \ln x_7 + 0.043x_8 - 2.5x_9 - 0.009x_{10} + 2.082x_{11} - 0.013x_{12} - 0.765x_{13} + 0.021x_{14} - 0.006x_{15} - 0.002x_{16} + 0.858 \ln x_{17} - 0.008x_{18} + 0.959x_{19} + 5.396 \ln x_{20} - 0.011 \ln x_{21} + 0.113x_{22} + 0.118x_{23} - 0.000345x_{24} - 0.001x_{25} + 0.028x_{26} - 0.027x_{27} - 0.062x_{28} + 0.035x_{29}$$

$$r = x_{19} - y$$

由此可得以下辛烷值损失预测函数。

$$r = x_{19} - y = 25.312 - 0.016 \ln x_1 + 0.005x_2 - 0.003x_3 - 0.985 \ln x_4 - 0.096 \ln x_5 + 0.001x_6 + 0.034 \ln x_7 - 0.043x_8 + 2.5x_9 + 0.009x_{10} - 2.082x_{11} + 0.013x_{12} + 0.765x_{13} - 0.021x_{14} + 0.006x_{15} + 0.002x_{16} - 0.858 \ln x_{17} + 0.008x_{18} + 0.041x_{19} - 5.396 \ln x_{20} - 0.011 \ln x_{21} - 0.113x_{22} - 0.118x_{23} + 0.000345x_{24} + 0.001x_{25} - 0.028x_{26} + 0.027x_{27} + 0.062x_{28} - 0.035x_{29}$$

如下表所示，新的回归模型的拟合优度 R^2 为 0.962，相较于之前的模型有了一定的提高，而且更加接近于 1，有着很高的拟合度，因此可以用来进行产品辛烷值的预测。

表 13 优化后模型拟合优度

模型	R	R 方	调整后 R 方	标准估算的错误	德宾-沃森
1	.988 ^a	.976	.962	.20811	1.946

接下来进行模型验证。为了检验上述模型的准确度，我们将附件一后五十个样本数据代入上述模型，得到 50 组产品辛烷值预测结果（见附件），将预测结果与附件一样本中的产品实际辛烷值进行比较，结果如图 13 所示。

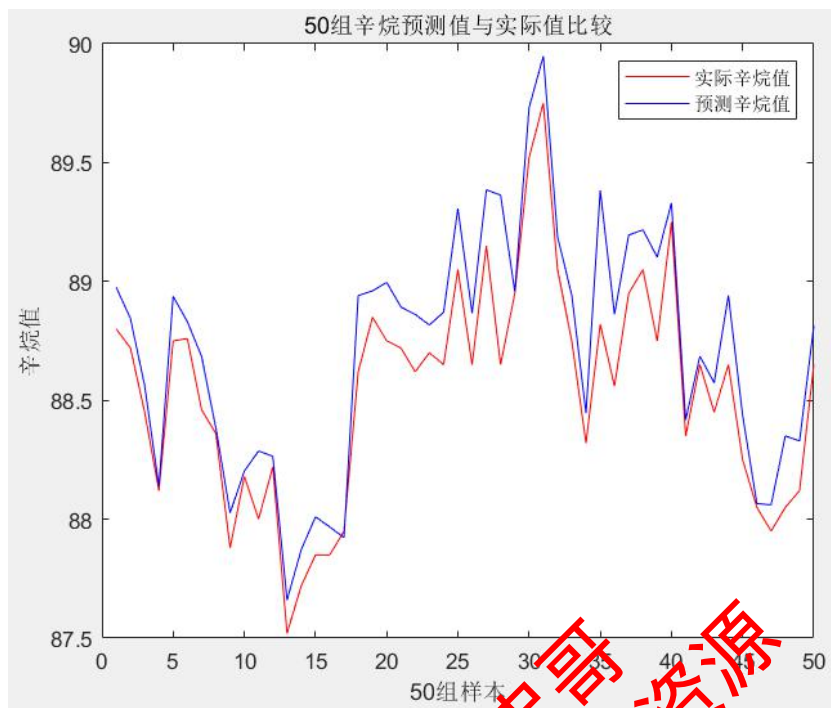


图 13 预测辛烷值与实际辛烷值对比图

从图中可以看出，预测的产品辛烷值与实际辛烷值有着极高的准确度，由于题目需求是求出辛烷值损失模型，所以需要计算出辛烷值损失的预测值，并将预测值与实际值进行比较。如图 14 所示为辛烷损失预测值与实际值的比较。

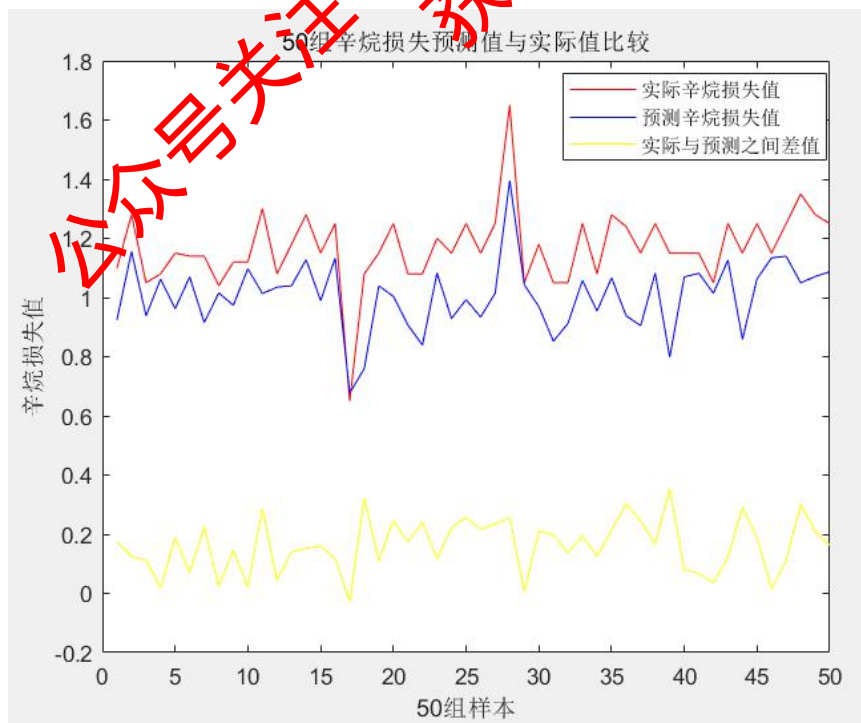


图 14 辛烷损失预测值与实际值的对比图

从图中可以看出，辛烷损失预测值与实际值之间的差值在 0.3 以内，有着较高的准确度，可以用来进行辛烷损失值的预测。

七、问题四的分析与求解

7.1 问题分析

问题四要求保证产品硫含量不大于 $5 \mu\text{g/g}$ 的前提下，利用预测模型获得对应的样本。由这些样本得到它们对应的主要变量优化后的操作条件。由于非操作变量不变，我们依据问题三得到的多元非线性公式，算出这些变量和常数的和。建立最小化辛烷值损失的非线性规划模型，决策变量是 18 个操作变量，约束条件分别有操作变量的可取值范围，硫含量不大于 $5 \mu\text{g/g}$ ，得到的目标函数小于原先辛烷值损失的 70%。

7.2 非线性规划模型的建立与求解

由于目标函数中含有非线性函数，因此这种规划问题叫非线性规划问题。

非线性规划的数学模型常表示成以下形式：

$$\begin{aligned} \min & F(X) \\ \text{s.t.} & \begin{cases} AX \leq b \\ Aeq \cdot X = beq \\ G(X) \leq 0 \\ Ceq(X) = 0 \\ VLB \leq X \leq VUB \end{cases} \end{aligned}$$

其中 X 为 n 维变元向量， $G(X)$ 与 $Ceq(X)$ 均为非线性函数组成的向量，其它变量的含义与线性规划、二次规划中相同。

基于问题四的要求，我们将最小化辛烷值损失 $c(X)$ 作为目标函数，18 个操作变量作为决策变量，约束条件有操作变量的取值范围，硫含量不大于 $5 \mu\text{g/g}$ 。再看得到的目标函数值是否小于原先样本中的辛烷值损失值。初步建立的非线性规划模型如下：

$$\begin{aligned} \min & F(X) \\ \text{s.t.} & \begin{cases} G(X) \leq 0 \\ VLB \leq X \leq VUB \end{cases} \end{aligned}$$

其中 $X = (x_1, x_2, \dots, x_{18})^T$ ，包括 18 个操作变量。

通过初步拟合得到产品硫含量函数为

$$\begin{aligned} S(X) = & 11.392 + 0.12 \ln x_1 - 0.608x_2 + 0.408x_3 + 7.186 \ln x_4 + 0.567 \ln x_5 + 0.46x_6 \\ & - 1.863 \ln x_7 + 0.128x_8 + 3.862x_9 + 0.128x_{10} - 67.518x_{11} + 0.028x_{12} \\ & - 7.201x_{13} + 0.946x_{14} - 0.041x_{15} - 0.066x_{16} - 0.64 \ln x_{17} + 0.148x_{18} \\ & - 0.06x_{19} - 8.134 \ln x_{20} - 1.438 \ln x_{21} - 0.153x_{22} - 0.298x_{23} - 0.018x_{24} \\ & + 0.056x_{25} - 0.184x_{26} + 0.2x_{27} + 0.394x_{28} - 0.059x_{29} \end{aligned}$$

由于 x_{19} 到 x_{29} 是非操作变量，问题中要求非操作变量不改变。我们计算后直接将数值代入，作为常数。 b 是产品最大硫含量 $5 \mu\text{g/g}$ 。 VLB 是 18 个操作变量的下界列向量， VUB 是 18 个操作变量的上界列向量。

辛烷值损失预测函数 $F(X)$ 中后 11 个非操作变量项和 $F(X)$ 中常数相加得到的和，作为目标函数中的常数 c 。

表 14 每个样本 $F(X)$ 常数项表

样本编号	常数项 c	样本编号	常数项 c

15	15.1339	166	15.1234
59	15.7489	173	15.1856
79	15.0797	180	14.7619
84	15.1464	182	14.8704
101	14.9248	192	15.0081
110	14.9256	208	14.8926
114	14.9701	262	14.3818
131	15.1847	264	14.8059
135	15.4004	268	14.5738
138	15.4004	303	14.4403
151	15.1265	310	14.6403
163	15.1185	314	14.4268

因此目标函数

$$\begin{aligned}
 F(X) = & -0.016 \ln x_1 + 0.005x_2 - 0.003x_3 - 0.065 \ln x_4 - 0.096 \ln x_5 \\
 & + 0.001x_6 + 0.034 \ln x_7 - 0.043x_8 + 2.5x_9 + 0.009x_{10} - 2.082x_{11} \\
 & + 0.013x_{12} + 0.765x_{13} - 0.021x_{14} + 0.006x_{15} + 0.002x_{16} \\
 & - 0.858 \ln x_{17} + 0.008x_{18} + c
 \end{aligned}$$

由于 11 个非操作变量按照原先的计算不用改变，我们将硫含量函数 $S(X)$ 中后 11 个非操作变量项和 $S(X)$ 中的常数相加得到的和作为最终常数，如下表。

表 15 每个样本 $S(X)$ 最终常数项表

样本编号	常数项 d	样本编号	常数项 d
15	19.6506	166	20.0667
59	20.2654	173	19.2995
79	19.9499	180	19.9338
84	19.5381	182	18.9952
101	19.3498	192	19.8704
110	19.2588	208	19.5591
114	19.0811	262	17.1907
131	19.5608	264	18.6607
135	19.8751	268	18.2437

138	19.9955	303	18.1921
151	19.1269	310	18.6781
163	19.0554	314	18.7954

因此 $G(X) = 0.12 \ln x_1 - 0.608x_2 + 0.408x_3 + 7.186 \ln x_4 + 0.567 \ln x_5 + 0.46x_6 - 1.863 \ln x_7 + 0.128x_8 + 3.862x_9 + 0.128x_{10} - 67.518x_{11} + 0.028x_{12} - 7.201x_{13} + 0.946x_{14} - 0.041x_{15} - 0.066x_{16} - 0.64 \ln x_{17} + 0.148x_{18} + d - 5$

我们采用 MATLAB 解决上述问题，以样本 15 为例，分为三步。

第一步，建立 M 文件 *fun.m*，定义目标函数 $F(X)$ 。

Function f = fun(X);

f = F(X);

第二步，因为约束条件中有非线性约束： $G(X) \leq 0$ ，则建立 M 文件 *nonlcon.m*，定义函数 $G(X)$ ：

Function [G,ceq] = nonlcon(X)

$G = 0.12 \ln x_1 - 0.608x_2 + 0.408x_3 + 7.186 \ln x_4 + 0.567 \ln x_5 + 0.46x_6 - 1.863 \ln x_7 + 0.128x_8 + 3.862x_9 + 0.128x_{10} - 67.518x_{11} + 0.028x_{12} - 7.201x_{13} + 0.946x_{14} - 0.041x_{15} - 0.066x_{16} - 0.64 \ln x_{17} + 0.148x_{18} + d - 5$

第三步，主程序 *you.m* 为：

load guihua.mat %里面含 VLB 和 VUB

$X0 = [112.1389; 42.2199; 0.2057; 271.7690; 49.7254; 58.6000; 519.3539; 9.8267; 0.2889; 418.4399; 0.6599; 123.8879; 2.2599; 3.2574; 409.9799; 120.2429; 439.5920; 27.1475]$

$A = []; b = [];$

$Aeq = []; beq = [];$

$[X, fval] = fmincon('fun', X0, A, b, Aeq, beq, VLB, VUB, 'nonlcon')$

接着我们对得到的最优值进行检验，查看是否小于原来损失值的 70%。

我们针对若干样本进行计算，得到如下优化操作表。

表 16 样本的主要变量优化操作表

样本编号	变量的优化操作
15	原料汽油硫含量提高 10 次，每次提高 10； 氢油比降低 2 次，每次降低 0.01； 反应器底部温度提高 2 次，每次提高 1℃。
79	精制汽油出装置硫含量提高 2 次，每次提高 1μg/g； 反应器质量空速提高 2 次，每次提高 0.5h-1； 还原器温度提高 6 次，每次提高 1℃。

101	A-202A/B 出口总管温度提高 2 次，每次提高 1℃； 还原器温度提高 8 次，每次提高 1℃； 氢油比降低 3 次，每次降低 0.01。
114	还原器温度提高 5 次，每次提高 1℃； 稳定塔顶压力降低 2 次，每次降低 0.05MPa； 氢油比降低 2 次，每次降低 0.01。
138	反应器底部温度提高 2 次，每次提高 1℃； 氢油比降低 4 次，每次降低 0.01； 还原器温度提高 10 次，每次提高 1℃。
151	氮气进装置流量提高 1 次，每次提高 50Nm ³ /h； 稳定塔下部温度提高 3 次，每次提高 1℃； 反应器质量空速提高 2 次，每次提高 0.5h ⁻¹ 。
192	D121 去稳定塔流量提高 2 次，每次提高 1t/h； 反应器质量空速提高 6 次，每次提高 0.5h ⁻¹ ； 稳定塔顶压力降低 2 次，每次降低 0.05MPa。
264	A-202A/B 出口总管温度提高 10 次，每次提高 1℃； 反应系统压力降低 1 次，每次降低 0.1MPa； 稳定塔下部温度提高 3 次，每次提高 1℃。
310	氢油比降低 1 次，每次降低 0.01； 氮气进装置流量降低 2 次，每次降低 50Nm ³ /h； 稳定塔顶压力降低 2 次，每次降低 0.05MPa。

以上是数据挖掘的结果。另外我们从文献查阅知，针对汽油辛烷值损失较大的原因，应从反应压力、反应温度、质量空速等方面采取优化措施。

第一，提高反应温度。在 399-438℃ 范围内，随着温度升高，辛烷值降低。由于反应器的温度升高会使烯烃加氢反应减少，反应温度上升的幅度范围也会随着温度升高而减小。将加热炉出口温度由 410℃ 提高至 420℃。反应器出口温度控制在 427℃-432℃ 以下，反应器床层温度上升控制在 5℃ 左右。

第二，降低反应压力。增加反应压力使氢分压增加，使烯烃加氢的速率增加形成烷烃，使汽油中烯烃占比减小，辛烷值损失增大。降低压力会让辛烷值损失降低，并带来硫含量上升。因此要保证脱硫率和辛烷值损失间寻找平衡。

第三，降低氢油比。在系统压力大小不变的情况下，增大氢油比可以抑制结焦。但是会增加辛烷值损失。降低氢油比可以减少辛烷值损失。

第四，增大质量空速可以降低烯烃加氢饱和变成烷烃的速率，降低辛烷值损失，但硫含量会提高。反应器的吸附剂藏量由 33t 减少到 25t，会减少烯烃加氢饱和反应形成烷烃的反应速率，降低辛烷值损失。

八、问题五的分析与求解

5.1 问题分析

题目五要求我们在保持非操作变量不改变的条件下，只对操作变量的值进行调整。我们根据 133 号样本数据提取了我们所筛选的 29 个指标的数据值，29 个变量中包括 18 个操作变量和 11 个非操作变量。我们应用控制变量法对 18 个操作变量在其有效区间内取值，从而计算出有效区间内每个取值对应的辛烷值和产品硫含量值。在上面建模过程中，我们对 18 个操作中的 5 个变量进行了优化，我们通过对五个优化变量进行了非线性变化的方法，从而提高辛烷值和产品硫含量预测的准确率。

5.2 图示整体说明

对于产品的硫含量指标来说，我们画出的所有轨迹图，其指标都符合题目的要求——有效区间内的最大值都小于 5 ng/g ，从另一角度也证实我模型建立的辛烷值以及产品中硫含量的关系式是相对准确的。

由于我模型的硫含量稳定低于 5 ng/g ，则辛烷值的可变动范围也比较大，通过下面各指标的轨迹分析可以看到，很多指标的调整可以使得辛烷值增大，使得辛烷值损失降低，从而可以产生一定的经济效益。如果考虑指标之间共同变化的结果，那么可以实现降低辛烷值损失的 30%，会实现较高的经济效益。

对于各个指标的轨迹图，我们都考虑了指标的变化范围以及变化单位。在下面的轨迹图中，蓝色的虚线“-----”表示该指标的非变化范围；蓝色虚线中间的白色区域为该指标的有效变化范围；中间白色区域中的红点“•”与红点“•”之间的距离表示该指标的单位变化量；而红点中的粉点“•”表示 133 号样本的指标值。具体分析如下节所示。

5.3 轨迹分析

(1) 反应系统压力

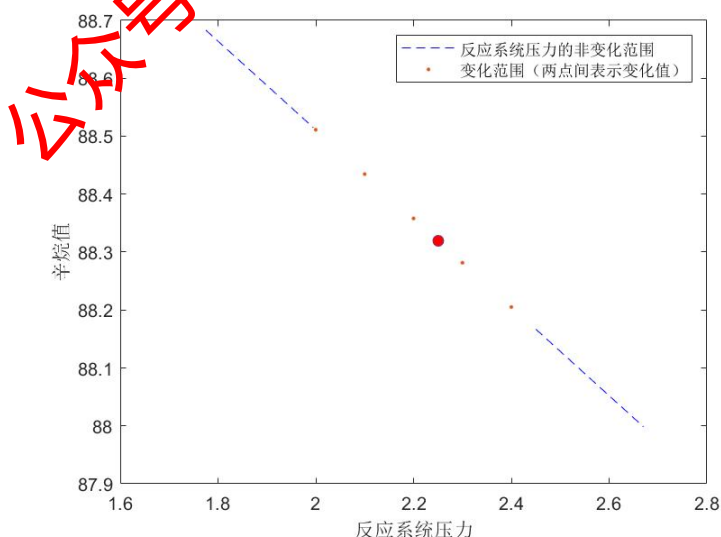


图 15 反应系统压力和产品辛烷值关系图

反应系统压力的指标范围为 2-2.45，变化单位是 0.1。从轨迹上可以看到将反应系统压力调低可以使得辛烷值增大，以 133 号样本为例，133 号样本的反应系统压力为 2.25，如果将反应系统压力变小，则最多可以调整两个单位（即 0.2），那么辛烷值将提升约 0.2（从 88.3 至 88.5）。133 号样本原料中的辛烷值为 89.4，那么辛烷值的损失值约减小了 20%，

可以看到反应系统压力对于辛烷值来说是比较重要的。

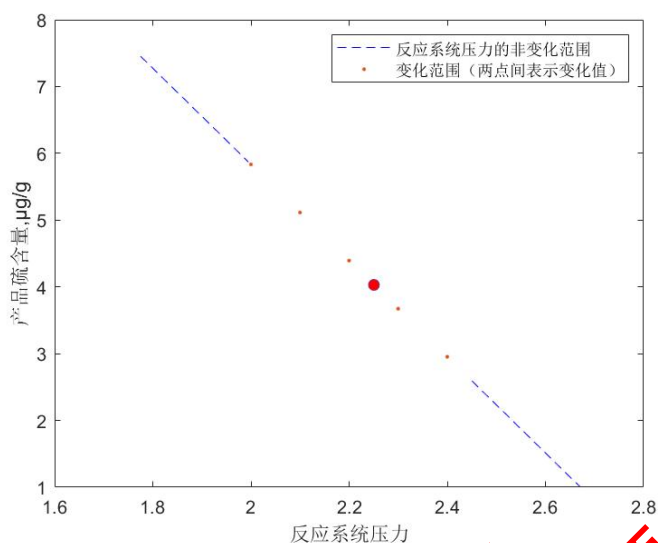


图 16 反应系统压力和产品硫含量关系图

对于产品中硫含量来说，反应系统压力也起到了比较重要的作用。反应系统压力越小，产品中硫含量越大。

(2) 氢油比

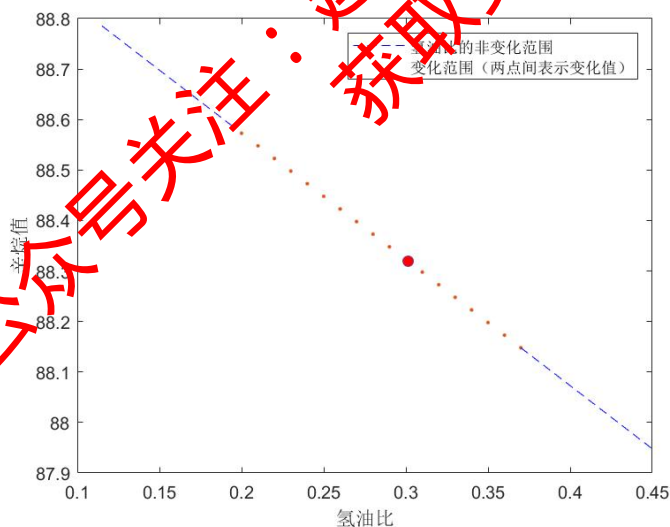


图 17 氢油比和产品辛烷值关系图

氢油比的有效区间是 0.2-0.37，变化单位为 0.01。对于氢油比，当氢油比减小，则辛烷值加大，则辛烷值的损失减少。对于样本 133 来说，指标氢油比为 0.3014，133 号样本可以控制氢油比减少十个单位，那么辛烷值会增加 0.26，辛烷值损失会减少越 20%。

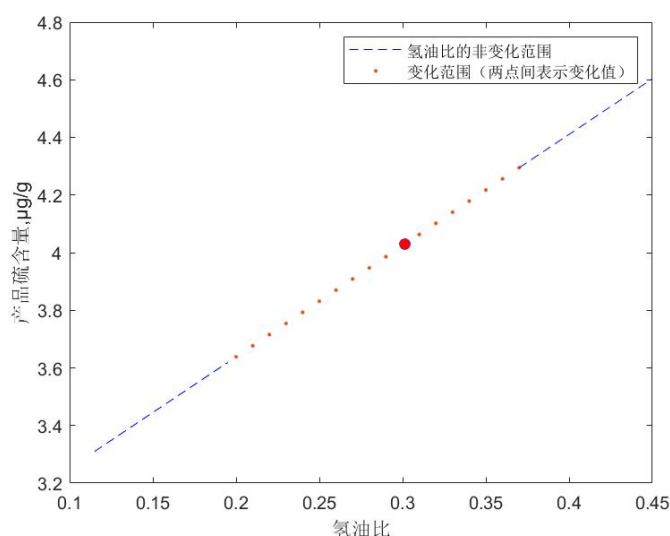


图 18 氢油比和产品硫含量关系图

对于产品中硫含量来说，氢油比增加，产品中硫含量会随之增加；反之氢油比减少，产品中硫含量也会减少。

(3) 催化汽油进装置总流量

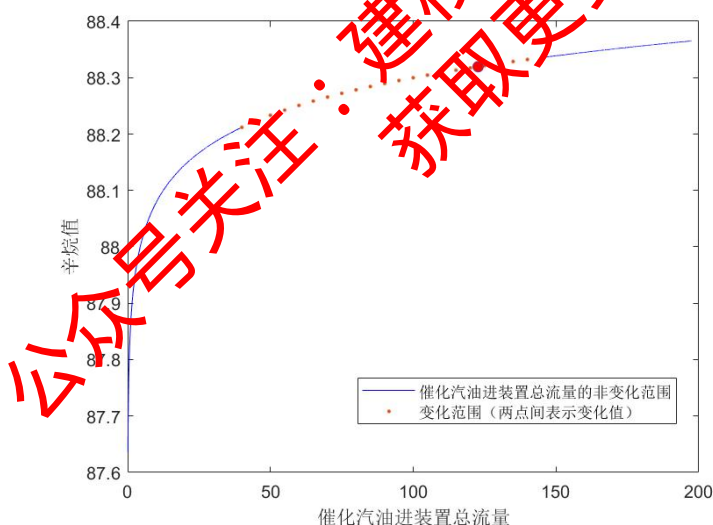


图 19 催化汽油进装置总流量和产品辛烷值关系图

催化汽油进装置总流量与辛烷值的关系呈现非线性的趋势，进行测试，我们发现二者的关系接近对数运算。催化汽油进装置总流量的变化范围是 40-145，变化单位为 5。当指标的值较低时，辛烷值上升的速度比较大，当辛烷值较高时，则辛烷值上升的速度就有所减慢。如 133 号样本的该指标值为 122.7，属于变化范围内较高的数值，经过计算仅有 4 个单位可以调整。当 133 号样本的催化汽油进装置总流量上升 4 个单位后，辛烷值上升越 0.3，损失值减少 27%。可以看到该变量对辛烷值损失的影响程度很大。

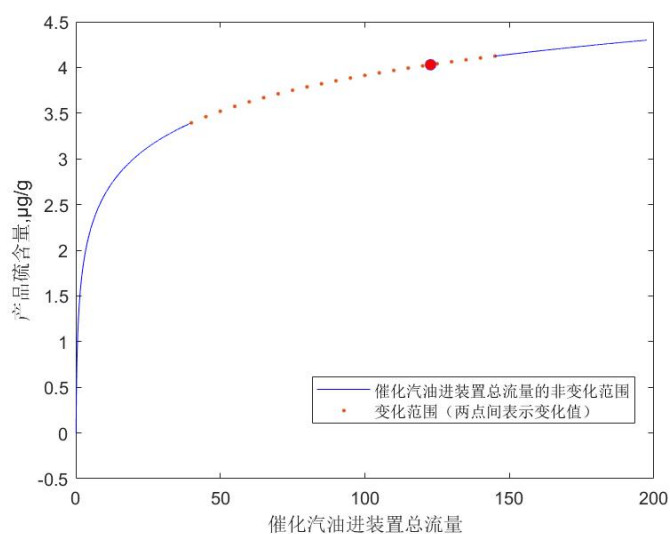


图 20 催化汽油进装置总流量和产品硫含量关系图

催化汽油进装置总流量对产品中硫含量也有正向的影响。随着催化汽油进装置总流量的上升，产品中硫含量也相应上升；反之催化汽油进装置总流量下降时，则产品中硫含量也相应下降。

(4) 反应器质量空速

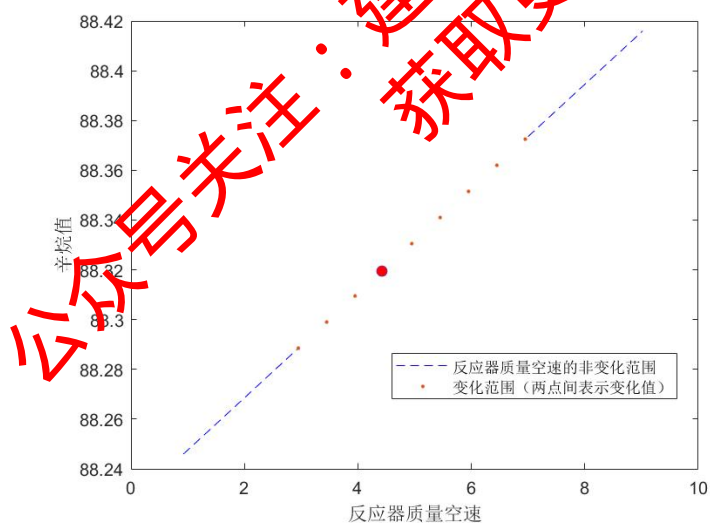


图 21 反应器质量空速和产品辛烷值关系图

反应器质量空速的变化范围在 2.95-7，变化单位为 0.5。133 号样本的反应器质量空速为 4.42，在此变化范围内还能增加 5 个变化单位，从而辛烷值加大约 0.06，辛烷值损失减少 0.06，约为 4%。可以看出，反应器质量空速虽然不能对辛烷值损失产生极大的影响，但通过调试，还是可以使得辛烷值显著减少的。

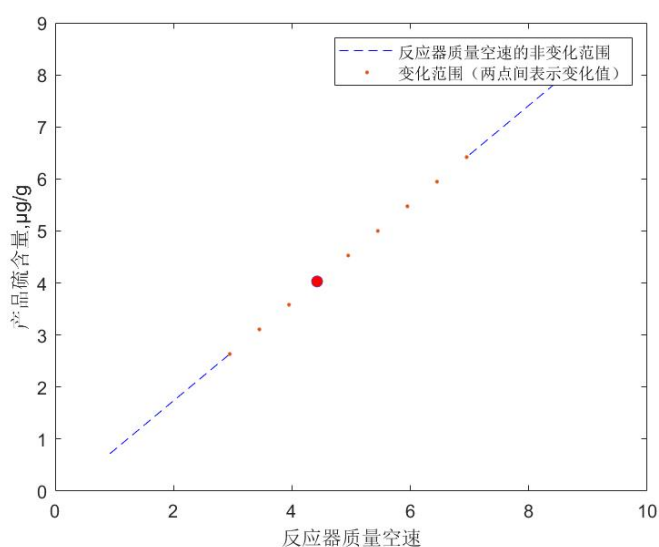


图 22 反应器质量空速和产品硫含量关系图

反应器质量空速与产品中硫含量呈现正比的趋势，当反应器质量空速越大，则产品中硫含量也会升高；反之当反应器质量空速减少，则产品中硫含量也会相应减少。

(5) 氮气进装置流量

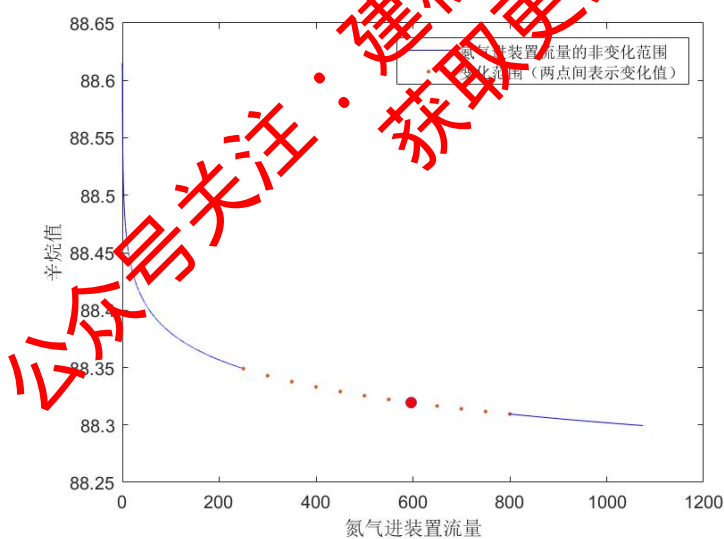


图 23 氮气进装置流量和产品辛烷值关系图

氮气进装置流量和辛烷值呈现指数的形态，当指标在变化范围内的值越小时，辛烷值变化的速率比较快，当指标在变化范围内的值越大时，辛烷值变化的速率比较慢。当指标的值增大时，辛烷值损失率下降，对于样本号 133 的数据而言，损失率最多下降百分之 4。

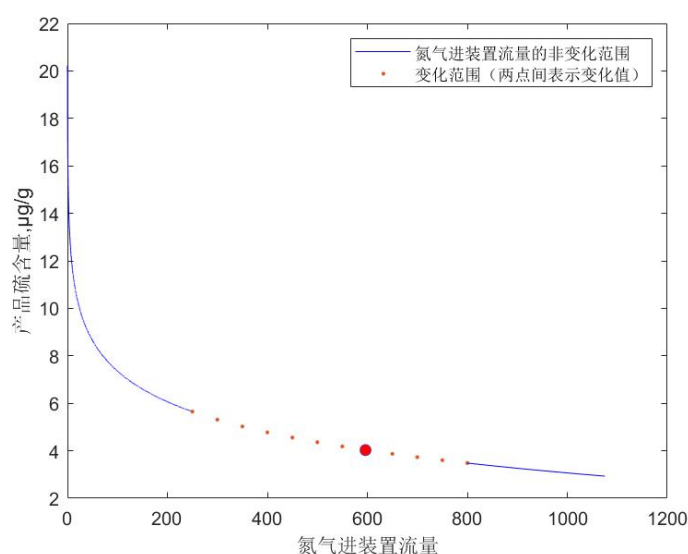


图 24 氮气进装置流量和产品硫含量关系图

氮气进装置流量和产品硫含量成非线性的关系，当氮气进装置流量减少时，产品硫含量会随之减少。

其余关系图在附录中。

九、模型的评价与改进建议

9.1 模型评价

我们分别采用了 BP 神经网络模型和多元非线性回归模型进行辛烷值损失预测。BP 神经网络作为一种智能机器学习算法，具有良好的非线性特征，在函数逼近中应用很广泛。

我们对 BP 神经网络预测的评价如下。第一，输入和输出之间高度非线性的映射，使它更适应非线性预测。第二，对权值的非严格特点，保存了各种预测方法提供的有效信息。第三，无后效性的特点，减少了预测结果对真实值的偏离。

由于主要变量有 19 个，我们又考虑建立多元非线性回归模型进行预测。通过不断尝试，最后得到一个非线性函数。

根据多元非线性函数我们建立了非线性规划模型。非线性规划是求解目标函数或约束条件中有一个或几个非线性函数的最优化问题的方法。通常用 MATLAB 或 LINDO 软件进行求解。

9.2 模型的改进建议

BP 神经网络模型应选取适量多的样本参数来进行学习训练。随着样本的数量增多，结果将更精确。

多元非线性回归模型有更好的函数形式。由于变量过多，可以再进行多次尝试得到更优的一个函数。

非线性规划模型的建立基于多元非线性函数，模型的关键是写出目标函数，决策变量和约束条件。约束条件中有产品中的硫含量的非线性函数，可进一步优化。

参考文献

- [1] 马强, 赵昌明. 降低 S-Zorb 装置汽油辛烷值损失的优化操作[J]. 当代化工研究, 2020 (15): 43-45. 2020.
- [2] 田勇震, 杨忠义, 马健波. 降低汽油加氢装置辛烷值损失的优化措施[J]. 石化技术与应用, 2019, 37 (05): 345-348. 2019.
- [3] 周欢, 齐万松, 李宏勋. S Zorb 装置辛烷值损失大原因的分析与措施[J]. 云南化工, 2019, 46 (09): 90-91+95. 2019.
- [4] 孟祥雷. 1.0 Mt/a 催化汽油加氢装置开工运行总结[J]. 齐鲁石油化工, 2020, 48 (01): 53-58. 2020.
- [5] 张大观. 80 万吨/年催化汽油加氢装置的模拟与优化[D]. 中国石油大学(华东), 2016.
- [6] 刘宝, 倪维起. S Zorb 装置汽油辛烷值损失影响因素分析[J]. 齐鲁石油化工, 2019, 47 (02): 102-104+124. 2019.
- [7] 高洁, 王莉娟, 孙丽琳. 优化操作条件降低汽油辛烷值损失[J]. 石油化工应用, 2011, 30 (11): 97-101. 2011.
- [8] 周小伟. 应用 BP 神经网络的二次反应清洁汽油辛烷值预测[J]. 西安交通大学学报, 2010, 44 (12): 82-86. 2010.
- [9] 孙忠超, 山红红. 用于 FCC 汽油辛烷值预测的非线性数学模型[J]. 炼油技术与工程, 2012, 42 (02): 60-64. 2012.

关注公众号：建模忠告
获取更多资源

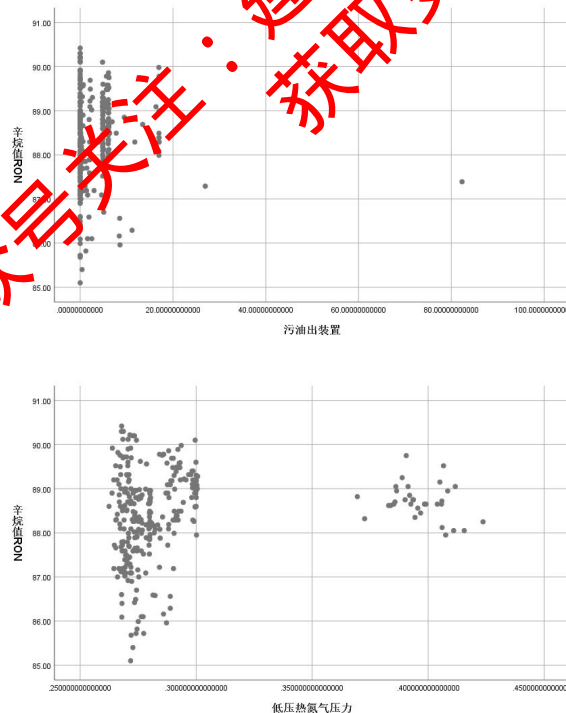
附录

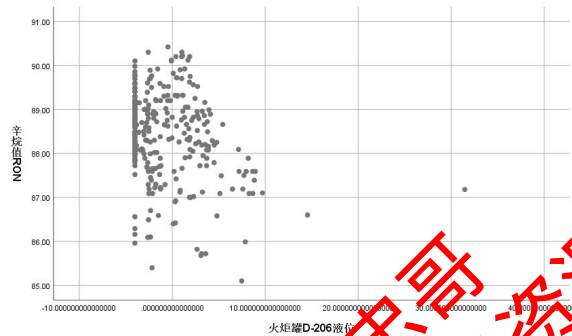
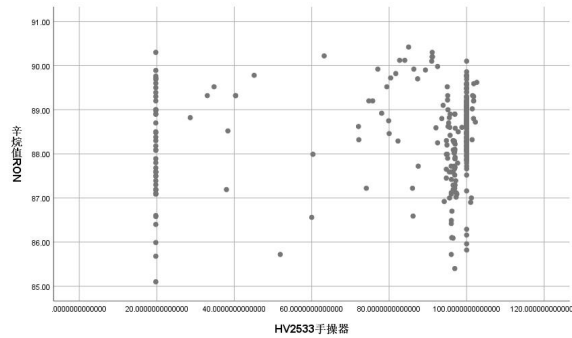
问题二

利用 matlab 计算相似度和显著性

```
1. clear;clc
2. load 'Test-data.mat' %Test-data 保存为所有指标的数据，我们对这些数据直接计算相关系数和显著性
3. %%通过 qq 检验发现我们的指标较多，不完全符合正态分布，所以这里我们使用斯皮尔曼相关系数
4. %进行 qq 检验，抽样检查各个指标是否符合正态分布，qq 图呈现一条直线则为符合正态分布。
5. qqplot(Test(:,1))
6. qqplot(Test(:,34))
7. %qq 图呈现结果不完全符合正态分布，这里我们使用斯皮尔曼相关系数进行计算
8. [R,P]=corr(Test,'type','Spearman')
9. %上一步骤中我们得到了 p 值，在下面两个步骤中，我们要对 p 值进行比较，从而确定在不同置信区间内显著的指标
10. re=P<0.01%在 99%的置信区间下显著
11. re2=(P<0.05).*(P>0.01)%在 95%的置信区间下显著
```

线性相关分析——所筛除变量的散点图





问题三

神经网络——用 python 实现模型的建立

```

1. """
2. 构建数据集,将 20 个主要变量和 1 个次要变量进行读入
3. """
4. from torch.utils.data import Dataset, DataLoader
5. import xlrd
6. import 简单的测试.config as cg
7. from tqdm import tqdm
8. import torch
9.
10.
11. class MyDataset(Dataset):
12.     def __init__(self, black_data_begin=0):
13.         red_sheet = xlrd.open_workbook(cg.redData_path).sheet_by_index(0)
14.         black_sheet = xlrd.open_workbook(cg.blackData_path).sheet_by_index(0)
15.
16.         self.red_data = []
17.         self.black_data = []
18.         for i in range(1, red_sheet.nrows):
19.             self.red_data.append(red_sheet.row_values(i))
20.             self.black_data.append(black_sheet.row_values(i))
21.         self.labels = []
22.         for i in range(len(self.red_data)):

```

```

22.         self.labels.append(self.red_data[i][0])
23.         self.labels = torch.LongTensor(self.labels)
24.         self.train_datas = []
25.         for i in range(len(self.red_data)):
26.             self.train_datas.append(self.red_data[i][1:])
27.             self.train_datas[i].extend(self.black_data[i][black_data_begin:black_data_end])
28.         self.train_datas = torch.FloatTensor(self.train_datas)
29.
30.         def __getitem__(self, index):
31.             return self.train_datas[index], self.labels[index]
32.
33.         def __len__(self):
34.             return len(self.red_data)
35.
36.
37.     def get_dataLoader():
38.         dataset = MyDataset()
39.         data_loader = DataLoader(dataset, batch_size=50, shuffle=False)
40.         return data_loader
41.
42.
43. if __name__ == '__main__':
44.     for index, (data, label) in enumerate(get_dataLoader()):
45.         print(index)
46.         print(label)
47.         print(data)
48.         break

```

```

1. """
2. 构建模型，每次模型输入 23 个指标，将 9 个非主要指标分三次输入到模型中
3. """
4. import torch.nn as nn
5. from torch.optim import Adam
6. import torch.nn.functional as F
7. from 简单的测试.dataset import get_dataLoader
8. import torch
9. class MyModel(nn.Module):
10.

```

```

11.     def __init__(self):
12.         super(MyModel, self).__init__()
13.         self.fc = nn.Linear(23, 10)
14.
15.     def forward(self, input):
16.         out = self.fc(input)
17.         return F.log_softmax(out, dim=-1)
18.
19.
20. model = MyModel()
21. optimizer = Adam(model.parameters(), 0.001)
22.
23.
24. def train(epoch):
25.     for idx, (input, target) in enumerate(get_dataLoader()):
26.         optimizer.zero_grad()
27.         output = model(input)
28.         loss = F.nll_loss(output, target)
29.         loss.backward()
30.         optimizer.step()
31.         print(loss.item())
32.
33.
34. if __name__ == '__main__':
35.     for i in range(30):
36.         train(i)
37.     print("done")

```

问题五

应用 matlab, 对所有指标画对辛烷值的线性图, 得到线性轨迹图

clear;clc

```

1. load 5.变化&画图_data.mat;
2. %down 下界 up 上界 change 变化值
3. a0 = -25.312; %常数
4. all_a = []; %记录所有操作变量的常熟项值
5. label = ["原料辛烷值 RON", "饱和烃, v% (烷烃+环烷烃)", "硫含量, µg/g", "烯烃, v%", "原料汽油
硫含量", "原料进装置温度", "精制汽油出装置硫含量", "还原器温度", "催化汽油进装置总流量", "A-202A/B 出
口总管温度", "氮气进装置流量", "D121 去稳定塔流量", "氢油比", "反应器底部温度", "稳定塔顶压力", "稳定塔
下部温度", "反应系统压力", "芳烃, v%", "溴值 gBr/100g", "密度(20℃), kg/m³", "待生吸附焦炭, wt%", "待生
吸附 S, wt%", "再生吸附焦炭, wt%", "再生吸附 S, wt%", "反应器质量空速", "反应器入口温度", "原料进装置流
量", "加热炉循环氢出口温度", "反应过滤器压差"];
6. for i = [6, 7, 10, 12:17, 25, 26, 29] i = 13

```

```

7.
8.     all = 0; %将 all 常数项归零
9.     %[down(i):change(i):up(i)]
10.     all = sum(x(1:i-1).*a(1:i-1))+sum(x(i+1:29).*a(i+1:29)); %非 i 元素的常数项和
11.     all = all +a0; %得到非 i 元素的最终常数项和（包括回归方程的系数 a0）
12.     all_a(i) = all; %把所有的 all_a 存下来，论文可能会展示
13.     y_i = all+[down(i):change(i):up(i)]*a(i);%在范围内的 y 值
14.     y_i_for = all+[(down(i)-0.5*(up(i)-down(i))):0.01:down(i)]*a(i);%前面不在范围内的辅助线
15.     y_i_bac = all+[up(i):0.01:up(i)+0.5*(up(i)-down(i))]*a(i);%后面不在范围内的辅助线
16.     %计算样本 133 的 x、y 值
17.     x133 = x(i);%133 号样本的值
18.     y133 = all+x133*a(i);%在范围内的 y 值
19.     %指标和辛烷值的关系图
20.     figure
21.     plot([(down(i)-0.5*(up(i)-down(i))):0.01:down(i)],y_i_for,'b--',[down(i):change(i):up(i)],y_i,'.',[up(i):0.01:up(i)+0.5*(up(i)-down(i))],y_i_bac,'b--')
22.     hold on
23.     scatter(x133,y133,'MarkerFaceColor','r')
24.     ylabel('辛烷值')
25.     xlabel(label(i))
26.     legend(label(i)+'的非变化范围','变化范围（两点间表示变化值）')
27.     %指标和辛烷值损失的关系图
28.     % figure
29.     % y_loss = x(1)-y_i;
30.     % y_loss_for = x(1)-y_i_for;
31.     % y_loss_bac = x(1)-y_i_bac;
32.     % plot([(down(i)-0.5*(up(i)-down(i))):0.01:down(i)],y_loss_for,'b--',[down(i):change(i):up(i)],y_loss,'.',[up(i):0.01:up(i)+0.5*(up(i)-down(i))],y_loss_bac,'b--')
33.     % ylabel('辛烷值损失')
34.     % xlabel(label(i))
35.     % legend(label(i)+'的非变化范围','变化范围（两点间表示变化值）')
36.     end

```

应用 matlab，对所有指标画对辛烷值的非线性图，得到非线性轨迹图

```

1. clear;clc
2. load 5.变化&画图_data2.mat;
3. %down 下界 up 上界 change 变化值
4. a0 = -25.312; %常数

```

```

5. label = ["原料辛烷值 RON", "饱和烃, v% (烷烃+环烷烃)", "硫含量,  $\mu\text{g/g}$ ", "烯烃, v%", "原料汽油  
硫含量", "原料进装置温度", "精制汽油出装置硫含量", "还原器温度", "催化汽油进装置总流量", "A-202A/B 出  
口总管温度", "氮气进装置流量", "D121 去稳定塔流量", "氢油比", "反应器底部温度", "稳定塔顶压力", "稳定塔  
下部温度", "反应系统压力", "芳烃, v%", "溴值 gBr/100g", "密度(20 $^{\circ}\text{C}$ ),  $\text{kg/m}^3$ ", "待生吸附焦炭, wt%", "待生  
吸附 S, wt%", "再生吸附焦炭, wt%", "再生吸附 S, wt%", "反应器质量空速", "反应器入口温度", "原料进装置流  
量", "加热炉循环氢出口温度", "反应过滤器压差"];
6. for i = [5,8,9,11,28]
7.     %非线性指标和辛烷值的关系图
8.
9.     y_i = log([down(i):change(i):up(i)])*a(i)+all_a(i);
10.    y_i_for = all_a(i)+log([0.1:0.01:down(i)])*a(i);%前面不在范围内的辅助线
11.    y_i_bac = all_a(i)+log([up(i):0.01:up(i)+0.5*(up(i)-down(i))])*a(i);%后面  
不在范围内的辅助线
12.    %计算样本 133 的 x、y 值
13.    x133 = x_original(i);%133 号样本的值
14.    y133 = all_a(i)+log(x133)*a(i);%在范围内的 y 值
15.    figure
16.    plot([0.1:0.01:down(i)], y_i_for, 'b--', [down(i):change(i):up(i)], y_i, '.', [up  
(i):0.01:up(i)+0.5*(up(i)-down(i))], y_i_bac, 'b--');
17.    hold on
18.    scatter(x133, y133, 'MarkerFaceColor', 'r')
19.    ylabel('辛烷值')
20.    xlabel(label(i))
21.    legend(label(i)+'的非变化范围', '变化范围 (两点间表示变化值)')
22.
23.    %非线性指标和辛烷值损失的关系图
24.    % figure
25.    % y_loss = x(1)-y_i;
26.    % y_loss_for = x(1)-y_i_for;
27.    % y_loss_bac = x(1)-y_i_bac;
28.    % plot([0.1:0.01:down(i)], y_loss_for, 'b--', [down(i):change(i):up(i)], y_lo  
ss, '.', [up(i):0.01:up(i)+0.5*(up(i)-down(i))], y_loss_bac, 'b--')
29.    % ylabel('辛烷值损失')
30.    % xlabel(label(i))
31.    % legend(label(i)+'的非变化范围', '变化范围 (两点间表示变化值)')
32. end

```

应用 matlab，对所有指标画对产品中硫的线性图，得到线性轨迹图

```

1. load 5.变化&画图 s_data.mat;
2. %down 下界 up 上界 change 变化值
3. a0 = 11.392; %常数
4. all_a = []; %记录所有操作变量的常熟项值
5. for i = [2,3,6,8:16,18] i= 16

```



```

6.      all = 0; %将 all 常数值归零
7.      %[down(i):change(i):up(i)]
8.      all = sum(x(1:i-1).*a(1:i-1))+sum(x(i+1:29).*a(i+1:29)); %非 i 元素的常数项和
9.      all = all +a0; %得到非 i 元素的最终常数项和（包括回归方程的系数 a0）
10.     all_a(i) = all; %把所有的 all_a 存下来，论文可能会展示
11.     y_i = all+[down(i):change(i):up(i)]*a(i);%在范围内的 y 值
12.     y_i_for = all+[(down(i)-0.5*(up(i)-down(i))):0.01:down(i)]*a(i);%前面不在范围内的
的辅助线
13.     y_i_bac = all+[up(i):0.01:up(i)+0.5*(up(i)-down(i))]*a(i);%后面不在范围内的辅助
线
14.     %计算样本 133 的 x、y 值
15.     x133 = x(i);%133 号样本的值
16.     y133 = all+x133*a(i);%在范围内的 y 值
17.     %指标和辛烷值的关系图
18.     figure
19.     plot([(down(i)-0.5*(up(i)-down(i))):0.01:down(i)],y_i_for,'b--',[down(i):chan
ge(i):up(i)],y_i,'.',[up(i):0.01:up(i)+0.5*(up(i)-down(i))],y_i_bac,'b--')
20.     hold on
21.     scatter(x133,y133,'MarkerFaceColor','r')
22.     ylabel('产品硫含量,μg/g')
23.     xlabel(label(i))
24.     legend(label(i)+'的非变化范围','变化范围（两点间表示变化值）')
25.     %指标和辛烷值损失的关系图
26.%     figure
27.%     y_loss = x(1)-y_i;
28.%     y_loss_for = x(1)-y_i_for;
29.%     y_loss_bac = x(1)-y_i_bac;
30.%     plot([(down(i)-0.5*(up(i)-down(i))):0.01:down(i)],y_loss_for,'b--',[down(i):
change(i):up(i)],y_loss,'.',[up(i):0.01:up(i)+0.5*(up(i)-down(i))],y_loss_bac,'b--')
31.%     ylabel('辛烷值损失')
32.%     xlabel(label(i))
33.%     legend(label(i)+'的非变化范围','变化范围（两点间表示变化值）')
34.end
35.% for i =[1:18]
36.%     all = 0; %将 all 常数值归零
37.%     %[down(i):change(i):up(i)]
38.%     all = sum(x(1:i-1).*a(1:i-1))+sum(x(i+1:29).*a(i+1:29)); %非 i 元素的常数项
和
39.%     all = all +a0; %得到非 i 元素的最终常数项和（包括回归方程的系数 a0）
40.%     all_a(i) = all; %把所有的 all_a 存下来，论文可能会展示
41.% end

```

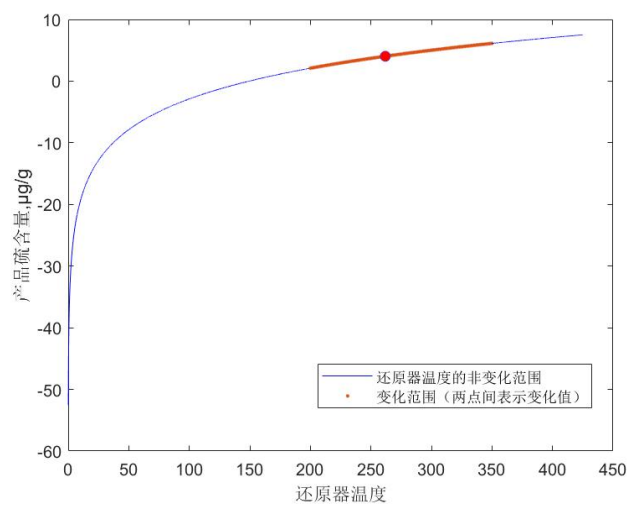
应用 matlab，对所有指标画对产品中硫的非线性图，得到非线性轨迹图

```

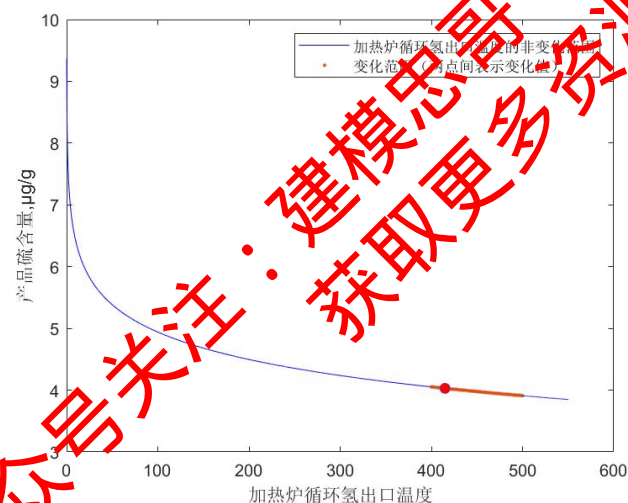
1. clc;clear
2. load 5.变化&画图 s_data2.mat;
3. %down 下界 up 上界 change 变化值
4. a0 = 11.392; %常数
5. for i = [1,4,5,7,17]
6.     %非线性指标和辛烷值的关系图
7.
8.     y_i = log([down(i):change(i):up(i)])*a(i)+all_a(i);
9.     y_i_for = all_a(i)+log([0.1:0.01:down(i)])*a(i);%前面不在范围内的辅助线
10.    y_i_bac = all_a(i)+log([up(i):0.01:up(i)+0.5*(up(i)-down(i))])*a(i);%后面不在
范围内的辅助线
11.    %计算样本 133 的 x、y 值
12.    x133 = x_original(i);%133 号样本的值
13.    y133 = all_a(i)+log(x133)*a(i);%在范围内的 y 值
14.    figure
15.    plot([0.1:0.01:down(i)],y_i_for,'b',[down(i):change(i):up(i)],y_i,'.',[up(i):
0.01:up(i)+0.5*(up(i)-down(i))],y_i_bac,'b')
16.    hold on
17.    scatter(x133,y133,'MarkerFaceColor','b')
18.    ylabel('产品硫含量,μg/g')
19.    xlabel(label(i))
20.    legend(label(i)+'的非变化范围','变化范围(两点间表示变化值)')
21.
22.    %非线性指标和辛烷值损失的关系图
23.%    figure
24.%    y_loss = x(1)-y_i;
25.%    y_loss_for = x(1)-y_i_for;
26.%    y_loss_bac = x(1)-y_i_bac;
27.%    plot([0.1:0.01:down(i)],y_loss_for,'b--',[down(i):change(i):up(i)],y_loss,'.
',[up(i):0.01:up(i)+0.5*(up(i)-down(i))],y_loss_bac,'b--')
28.%    ylabel('辛烷值损失')
29.%    xlabel(label(i))
30.%    legend(label(i)+'的非变化范围','变化范围(两点间表示变化值)')
31.end

```

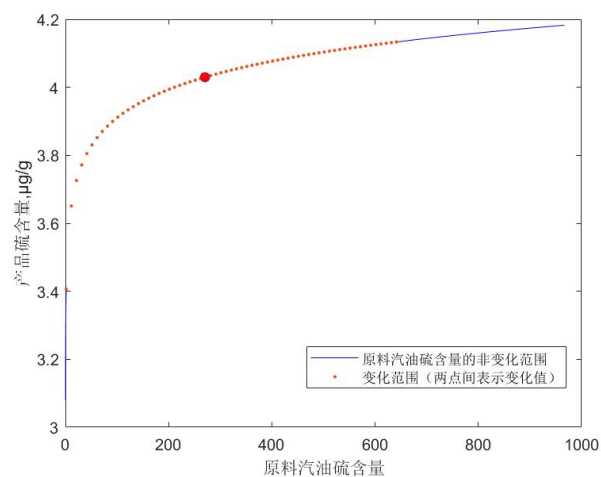
所有指标相对辛烷值和产品中硫的轨迹图



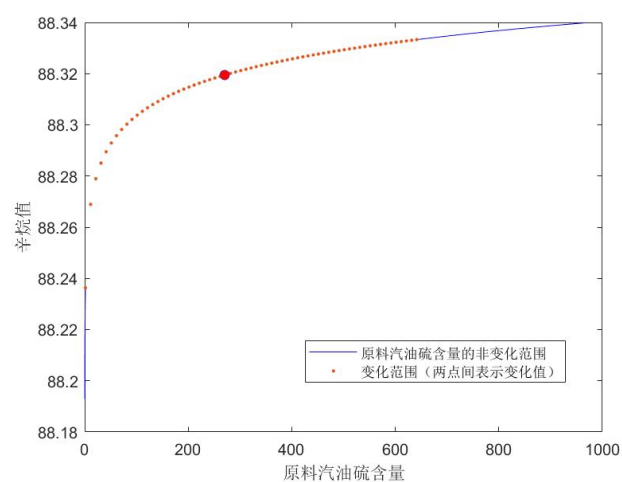
还原器温度和产品硫含量关系图



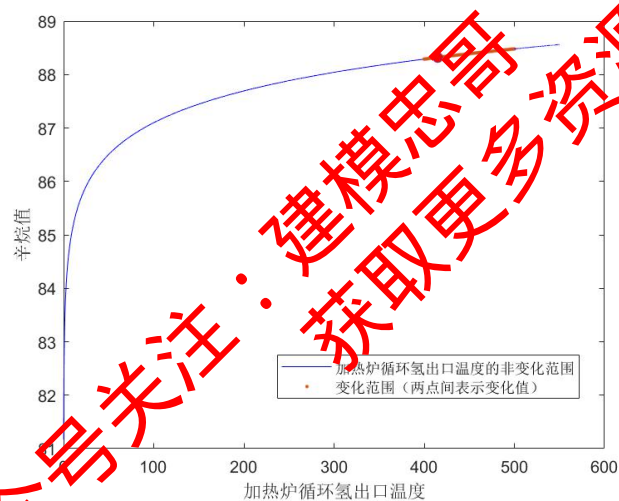
加热炉循环氢出口温度和产品硫含量关系图



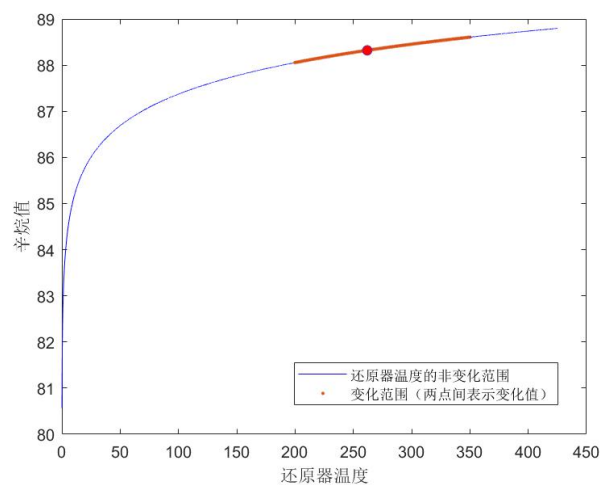
原料汽油硫含量和产品硫含量关系图



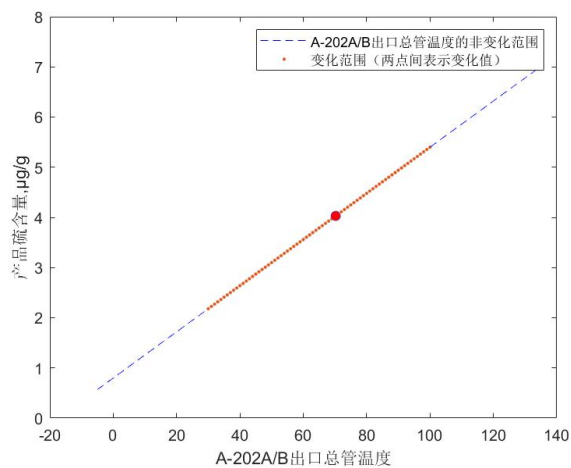
原料汽油硫含量和产品辛烷值关系图



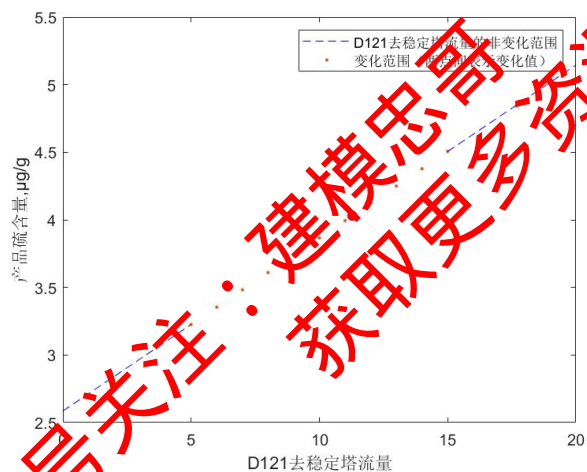
加热炉循环氢出口温度和产品辛烷值关系图



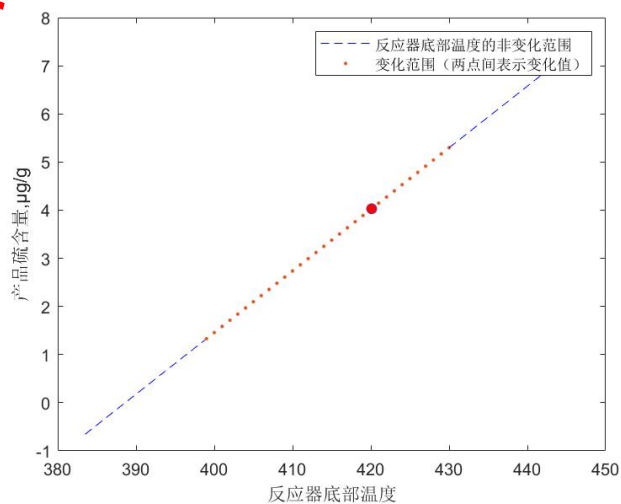
还原器温度和产品辛烷值关系图



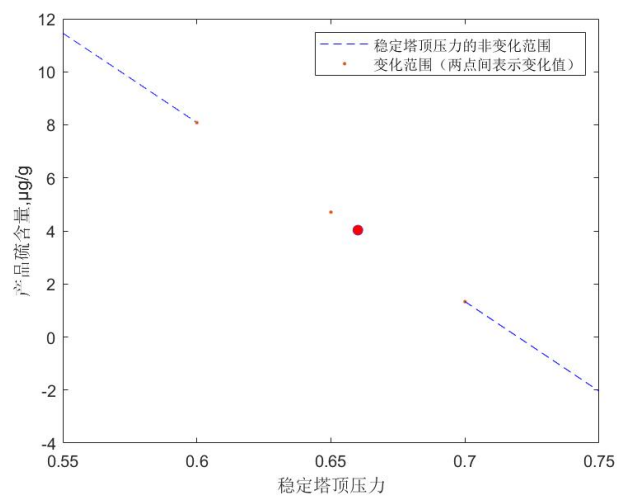
A-202A/B 出口总管温度和产品硫含量关系图



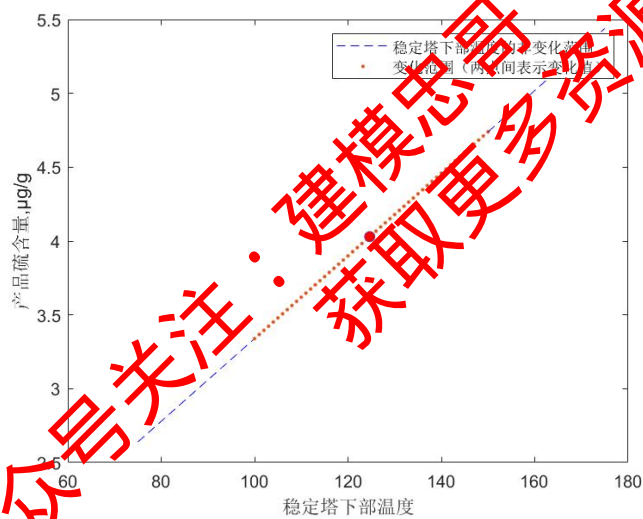
D121 去稳定塔流量和产品硫含量关系图



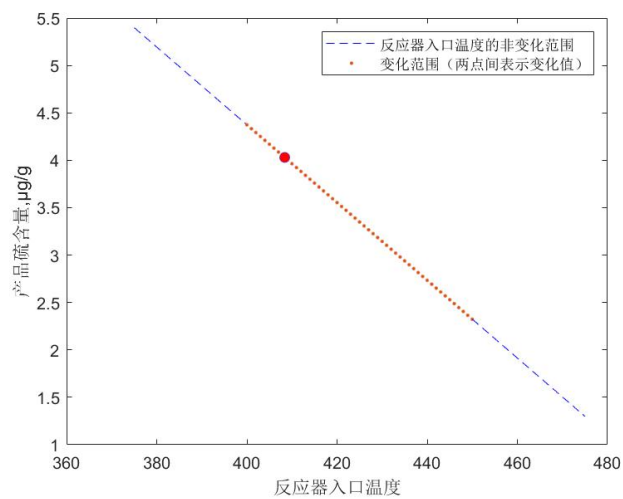
反应器底部温度和产品硫含量关系图



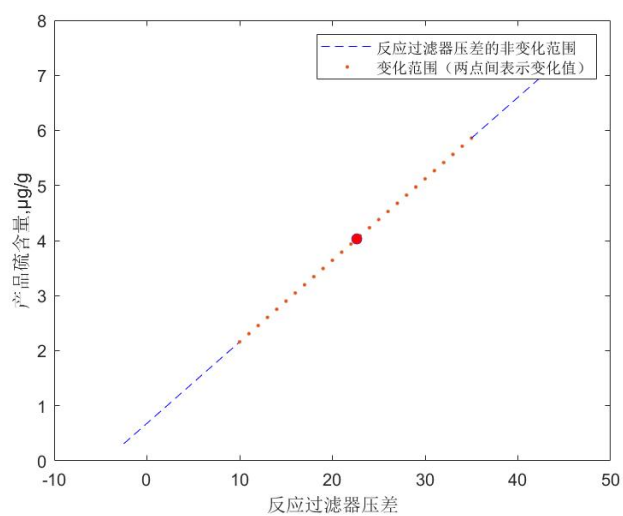
稳定塔顶压力和产品硫含量关系图



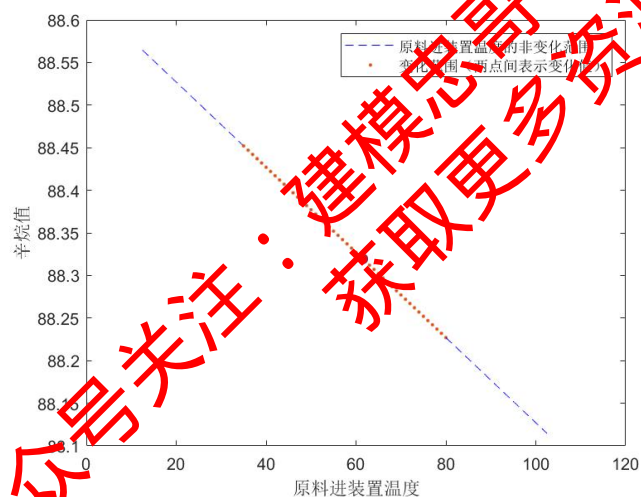
稳定塔下部温度和产品硫含量关系图



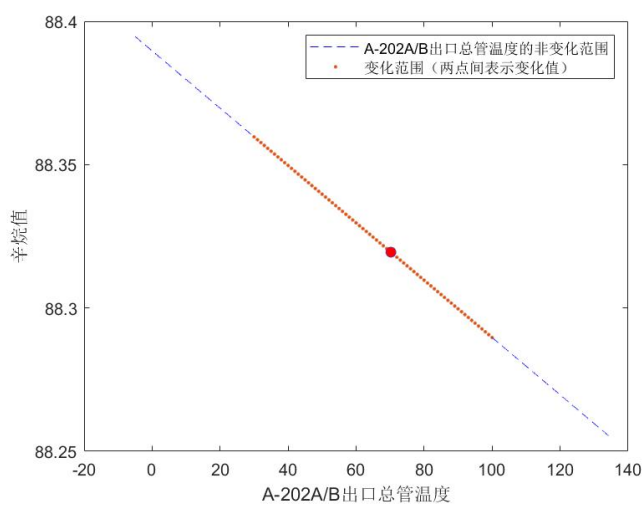
反应器入口温度和产品硫含量关系图



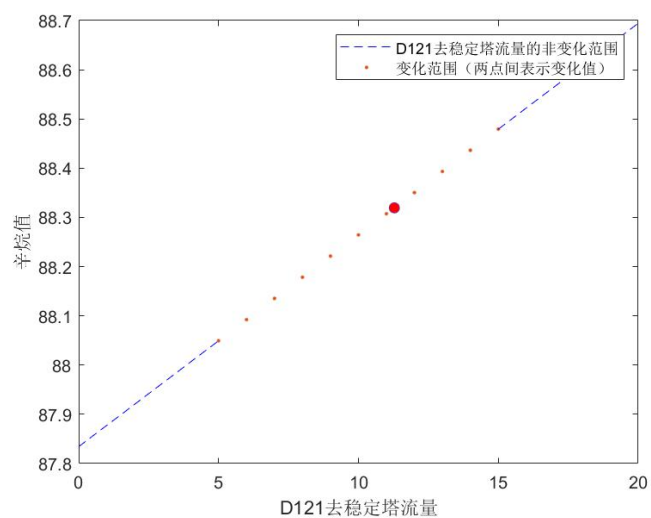
反应过滤器压差和产品硫含量关系图



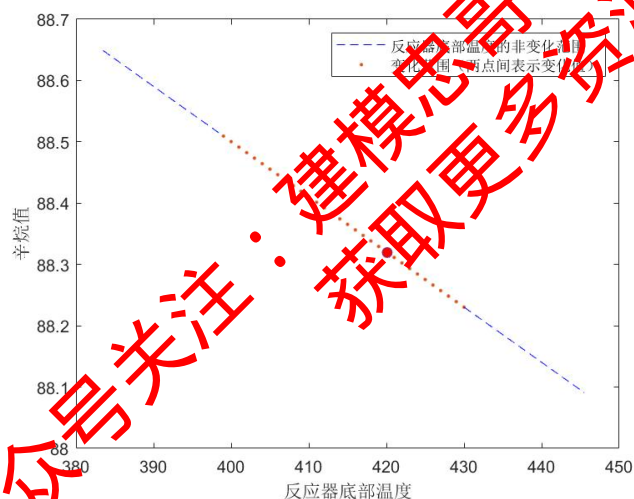
原料进装置温度和产品辛烷值关系图



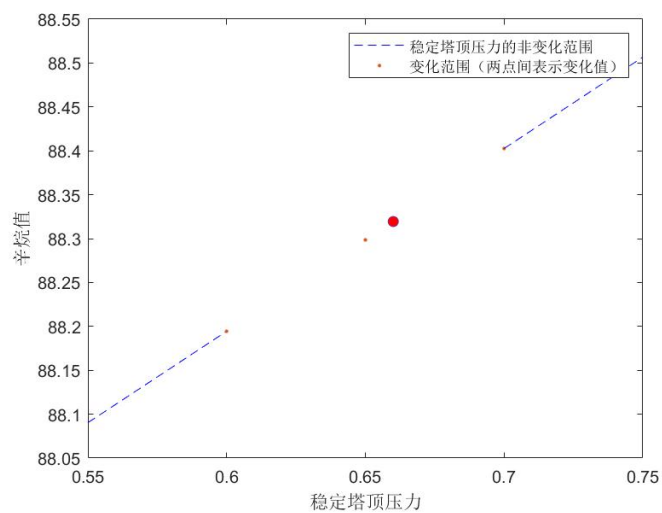
A-202A/B 出口总管温度和产品辛烷值关系图



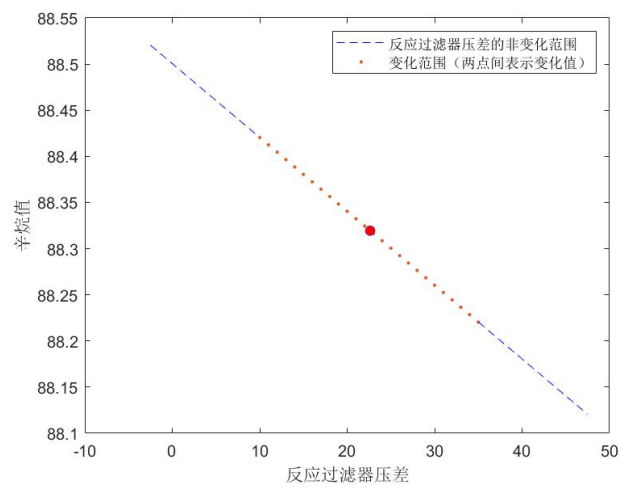
D121 去稳定塔流量和产品辛烷值关系图



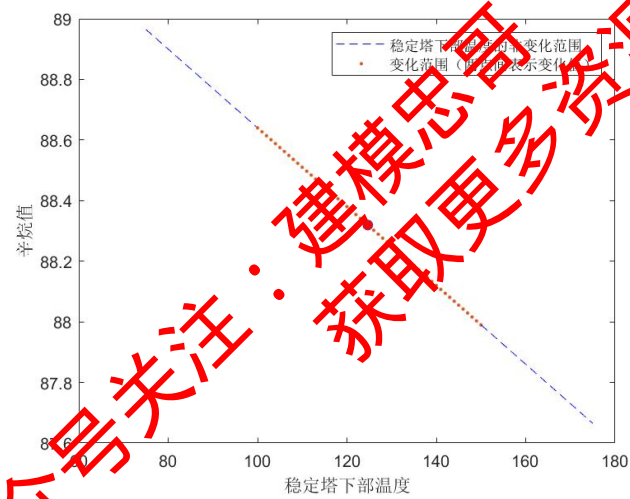
反应器底部温度和产品辛烷值关系图



稳定塔顶压力和产品辛烷值关系图



反应过滤器压差和产品辛烷值关系图



稳定塔下部温度和产品辛烷值关系图