

降低汽油精制过程中的辛烷值损失模型

一、背景

汽油是小型车辆的主要燃料，汽油燃烧产生的尾气排放对大气环境有重要影响。为此，世界各国都制定了日益严格的汽油质量标准（见下表）。汽油清洁化重点是降低汽油中的硫、烯烃含量，同时尽量保持其辛烷值。

欧盟和我国车用汽油主要规格

车用汽油标准	辛烷值	硫含量/($\mu\text{g/g}$)	苯含量/%	芳烃含量/%	烯烃含量/%
		≦	≦	≦	≦
国III（2010 年）	90-97	150	1	40	30
国IV（2014 年）	90-97	50	1	40	28
国V（2017 年）	89-95	10	1	40	24
国VI-A（2019 年）	89-95	10	0.8	35	18
国VI-B（2023 年）	89-95	10	0.8	35	15
欧 V（2009 年）	95	10	1	35	18
欧 VI（2013 年）	95	10	1	35	18
世界燃油规范（V 类汽油）	95	10	1	35	10

注: $\mu\text{g/g}$ 是一个浓度单位，也有用 mg/kg 或 ppm 表示的（以下同）

我国原油对外依存度超过 70%，且大部分是中东地区的含硫和高硫原油。原油中的重油通常占比 40-60%，这部分重油（以硫为代表的杂质含量也高）难以直接利用。为了有效利用重油资源，我国大力发展了以催化裂化为核心的重油轻质化工艺技术，将重油转化为汽油、柴油和低碳烯烃，超过 70%的汽油是由催化裂化生产得到，因此成品汽油中 95%以上的硫和烯烃来自催化裂化汽油。故必须对催化裂化汽油进行精制处理，以满足对汽油质量要求。

辛烷值（以 RON 表示）是反映汽油燃烧性能的最重要指标，并作为汽油的商品牌号（例如 89#、92#、95#）。现有技术在对催化裂化汽油进行脱硫和降烯烃过程中，普遍降低了汽油辛烷值。辛烷值每降低 1 个单位，相当于损失约 150 元/吨。以一个 100 万吨/年催化裂化汽油精制装置为例，若能降低 RON 损失 0.3 个单位，其经济效益将达到四千五百万元。

化工过程的建模一般是通过数据关联或机理建模的方法来实现的，取得了一定的成

果。但是由于炼油工艺过程的复杂性以及设备的多样性，它们的操作变量（控制变量）之间具有高度非线性和相互强耦联的关系，而且传统的数据关联模型中变量相对较少、机理建模对原料的分析要求较高，对过程优化的响应不及时，所以效果并不理想。

某石化企业的催化裂化汽油精制脱硫装置运行 4 年，积累了大量历史数据，其汽油产品辛烷值损失平均为 1.37 个单位，而同类装置的最小损失值只有 0.6 个单位。故有较大的优化空间。请参赛研究生探索利用数据挖掘技术来解决化工过程建模问题。

二、目标

依据从催化裂化汽油精制装置采集的 325 个数据样本（每个数据样本都有 354 个操作变量），通过数据挖掘技术来建立汽油辛烷值（RON）损失的预测模型，并给出每个样本的优化操作条件，在保证汽油产品脱硫效果（欧六和国六标准均为不大于 $10\text{ }\mu\text{g/g}$ ，但为了给企业装置操作留有空间，本次建模要求产品硫含量不大于 $5\text{ }\mu\text{g/g}$ ）的前提下，尽量降低汽油辛烷值损失在 30% 以上。

三、问题

1. 数据处理：请参考近 4 年的工业数据（见附件一“325 个数据样本数据.xlsx”）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（原始数据见附件三“285 号和 313 号样本原始数据.xlsx”）并将处理后的数据分别加入到附件一中相应的样本号中，供下面研究使用。

2. 寻找建模主要变量：

由于催化裂化汽油精制过程是连续的，虽然操作变量每 3 分钟就采样一次，但辛烷值（因变量）的测量比较麻烦，一周仅 2 次无法对应。但根据实际情况可以认为辛烷值的测量值是测量时刻前两小时内操作变量的综合效果，因此预处理中取操作变量两小时内的平均值与辛烷值的测量值对应。这样产生了 325 个样本（见附件一）。

建立降低辛烷值损失模型涉及包括 7 个原料性质、2 个待生吸附剂性质、2 个再生吸附剂性质、2 个产品性质等变量以及另外 354 个操作变量（共计 367 个变量），工程技术应用中经常使用先降维后建模的方法，这有利于忽略次要因素，发现并分析影响模型的主要变量与因素。因此，请你们根据提供的 325 个样本数据（见附件一），通过降维的方法从 367 个操作变量中筛选出建模主要变量，使之尽可能具有代表性、独立性（为了工程应用方便，建议降维后的主要变量在 30 个以下），并请详细说明建模主要变量的筛选过程及其合理性。（提示：请考虑将原料的辛烷值作为建模变量之一）。

3. 建立辛烷值（RON）损失预测模型：采用上述样本和建模主要变量，通过数据挖掘技术建立辛烷值（RON）损失预测模型，并进行模型验证。
4. 主要变量操作方案的优化：要求在保证产品硫含量不大于 $5\text{ }\mu\text{g/g}$ 的前提下，利用你们的模型获得 325 个数据样本（见附件四“325 个数据样本数据.xlsx”）中，辛烷值（RON）损失降幅大于 30% 的样本对应的主要变量优化后的操作条件（优化过程中原料、待生吸附剂、再生吸附剂的性质保持不变，以它们在样本中的数据为准）。
5. 模型的可视化展示：工业装置为了平稳生产，优化后的主要操作变量（即：问题 2 中的主要变量）往往只能逐步调整到位，请你们对 133 号样本（原料性质、待生吸附剂和再生吸附剂的性质数据保持不变，以样本中的数据为准），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。（各主要操作变量每次允许调整幅度值 Δ 见附件四“354 个操作变量信息.xlsx”）。

附件：

附件一：325 个样本数据.xlsx

附件二：样本确定方法.docx

附件三：285 号和 313 号样本原始数据.xlsx

附件四：354 个操作变量信息.xlsx