

DOI: 10.11949/0438-1157.

基于 TabPFN 的汽油辛烷值预测与模型可解释性

李亚洁^{1,2,3}, 王文春^{1,2,3}, 李炜^{1,2,3}

(¹ 兰州理工大学自动化与电气工程学院, 甘肃 兰州 730050; ² 甘肃省工业过程先进控制重点实验室, 甘肃 兰州 730050; ³ 兰州理工大学电气与控制工程国家级实验教学示范中心, 甘肃 兰州 730050)

摘要: 针对汽油在线调和配方优化中传统机器学习模型存在超参数调优复杂、建模效率偏低及可解释性不足等问题, 提出了一种基于表格先验数据拟合网络 (TabPFN) 的汽油辛烷值预测与模型可解释分析方法。首先, 以企业实际汽油调和组分及配比为输入、辛烷值为输出, 建立基于 TabPFN 的辛烷值预测模型。其次, 通过引入沙普利加性解释 (SHAP) 和偏依赖图 (PDP) 方法, 从全局与局部、单变量与双变量多维度解析各组分及配比对辛烷值的影响机制。结果表明, 本文的辛烷值预测模型无需复杂调参, 测试集均方误差低至 0.001, 训练时间仅 4 s, 进而 SHAP 量化了各组分的贡献与影响程度, PDP 揭示了各组分的非线性响应与交互作用, 可为汽油在线调和配方优化提供高精度预测与决策支持。

关键词: 油品调和; 汽油辛烷值; 表格先验数据拟合网络; 沙普利加性解释; 偏依赖图; 预测; 燃料

中图分类号: TP 273

文献标志码: A

文章编号: 0438-1157 (XXXX) XX-0001-15

Gasoline octane number prediction and model interpretability based on TabPFN

LI Yajie^{1,2,3}, WANG Wenchun^{1,2,3}, LI Wei^{1,2,3}

(¹ College of Electrical and Information Engineering, Lanzhou University of Technology, Lanzhou 730050, Gansu, China; ² Key Laboratory of Gansu Advanced Control for Industrial Processes, Lanzhou 730050, Gansu, China; ³ National Demonstration Center for Experimental Electrical and Control Engineering Education, Lanzhou University of Technology, Lanzhou 730050, Gansu, China)

Abstract: Gasoline octane number is a key quality index in the blending process of finished gasoline. Its accurate prediction is of great significance for improving oil quality, optimizing production efficiency and reducing production costs. Aiming at the problems of complex hyper-parameter tuning, low modeling efficiency and insufficient interpretability of existing machine learning models in octane number prediction, a gasoline octane number prediction and model interpretability analysis method based on tabular prior-data fitted network (TabPFN) is proposed. Firstly, based on the actual gasoline blending data of the enterprise, each blending component and its ratio are used as input variables, and the octane number is used as the output variable. Then, a gasoline octane number prediction model based on TabPFN is constructed and compared with several typical machine learning models in the verification process. Secondly, a variety of evaluation indicators are used to comprehensively evaluate

收稿日期: 2026-04-14 修回日期: 2026-05-19

通信作者: 李炜 (1963—), 女, 硕士, 教授, liwei@lut.edu.cn

第一作者: 李亚洁 (1981—), 女, 博士, 副教授, liyaj@lut.edu.cn

基金项目: 国家自然科学基金项目 (62163022); 甘肃省科技重大专项计划 (23ZDGA007); 甘肃省产业支撑计划项目 (2024CYZC-18); 甘肃省博士基金项目 (2021QB-044)

引用本文: 李亚洁, 王文春, 李炜. 基于 TabPFN 的汽油辛烷值预测与模型可解释性[J]. 化工学报, XXXX, XX(X): 1–15

Citation: LI Yajie, WANG Wenchun, LI Wei. Gasoline octane number prediction and model interpretability based on TabPFN[J]. CIESC Journal, XXXX, XX(X): 1–15

the performance of the model from the three dimensions prediction accuracy, training efficiency and reasoning efficiency. Finally, the interpretability of the prediction model is analyzed by introducing Shapley additive explanation (SHAP) and partial dependence plot (PDP) methods. The experimental results show that the mean square error of the octane number prediction of the TabPFN model is as low as 0.001 on the test set without a lot of hyper-parameter tuning, and the model training takes time is only 4 s, which is better than the comparison model. SHAP analysis identified and quantified the contribution of each gasoline blending component and its ratio to the octane number prediction results. PDP analysis further revealed the marginal effect, nonlinear relationship and interaction of each blending component and its combination on the octane number prediction results. This method can achieve efficient, accurate and strong explanatory power of gasoline octane number prediction, which can provide high-precision prediction and decision support for gasoline online blending formulation optimization.

Keywords: gasoline blending; gasoline octane number; tabular prior-data fitted network; Shapley additive explanation; partial dependence plot; prediction; fuel

引 言

在“双碳”战略目标引领下,中国成品油质量升级持续推进。降低汽油烯烃含量、提升辛烷值不仅是降低机动车尾气排放、助力“双碳”目标的重要途径,也是现阶段国内汽油质量升级的核心方向。自2023年全面实施国VIB汽油标准以来,对辛烷值、烯烃含量等关键指标的要求不断提高^[1]。汽油调和作为成品汽油生产的关键环节^[2],通过科学配比加氢汽油、甲基叔丁基醚(Methyl tert-Butyl Ether, MTBE)等调和组分,不仅能够精准调控辛烷值和抗爆指数等核心指标,还能为炼化企业实现提质增效、满足国VIB标准提供重要技术支撑。其中,辛烷值是衡量汽油抗爆性能的重要指标,直接影响燃料燃烧质量和发动机运行性能^[3-4]。因此,高精度的辛烷值预测对于优化汽油调和工艺、提升油品质量以及支撑绿色低碳生产具有重要意义。在此背景下,本文聚焦汽油在线调和配方优化场景,旨在构建一个高精度、低调参且可解释的辛烷值预测模型,为配方实时调整提供决策支持。

现有辛烷值预测方法大致可分为三类:第一类是基于光谱的化学计量学方法,通过近红外光谱^[5]、中红外光谱^[6]、拉曼光谱^[7-8]和核磁共振光谱^[9-10]等光谱信息实现快速检测;第二类是基于分子结构描述符的定量结构性质关系(Quantitative Structure - Property Relationship, QSPR)模型^[11-13],通过分子结构描述符建立辛烷值与分子特征之间的关联;第三类是基于宏观理化性质的经验或物性模型,利用密度、蒸汽压和蒸馏曲线等宏观指标进行预测^[14-16]。总体来看,上述方法在特定条件下均取得了一定成

效,但仍存在各种不同的局限:光谱化学计量学方法对输入数据质量和样本代表性较为敏感,QSPR模型对复杂多组分混合体系的适应性有限,而经验或物性模型对未知体系的泛化能力相对不足。

随着汽油调和工艺逐步由经验驱动向数据驱动转变,机器学习技术因能够从历史生产数据中自适应提取关键特征,并揭示复杂非线性关系^[17],在辛烷值预测及炼化过程建模中显示出明显优势。例如,Sun等^[18]基于随机森林(Random Forest, RF)构建了多特征机器学习模型,有效捕捉了汽油组分物性指标与辛烷值之间的非线性映射,其预测性能接近实验不确定度;Zhang等^[19]通过最大信息系数筛选关键变量并优化组合模型,以高维工艺数据为输入,同时预测生产过程中的辛烷值损失与硫含量,表现出较好的泛化能力。此外,将稀疏自编码器、集成学习等深度特征提取方法引入工业数据建模,还可以提升模型在高维噪声数据下的稳健性^[20]。由此可见,数据驱动方法正在成为辛烷值预测与在线质量控制的重要技术手段。

然而,高精度预测并不意味着模型可以直接应用于工业生产。如果模型缺乏可解释性,其预测结果往往难以转化为可执行的调和策略和操作方案。可解释性分析能够在一定程度上打开机器学习模型的“黑箱”,揭示模型的决策依据与关键变量之间的关系,从而为工艺参数调控与工艺优化提供更具针对性的决策支持^[21]。目前,沙普利加性解释(Shapley Additive exPlanations, SHAP)和偏依赖图(Partial Dependence Plot, PDP)等方法已被广泛应用于复杂机器学习模型的可解释性研究。SHAP基于博弈论中的Shapley值进行计算,能够衡量每个特征

对模型预测的贡献,提供全局与局部解释,提高模型的透明性和可信度^[22]。而PDP则通过边际化其他特征的影响,直观反映单变量或多变量组合对模型预测输出的全局平均边际效应,使研究人员能够更清晰地理解特征变化对模型决策的影响^[23]。

基于SHAP和PDP的模型可解释性分析已在多个领域取得了较为丰富的研究进展。李炜等[24]基于改进遗传算法和轻量级梯度提升机模型构建了汽油性能与环保指标预测模型,并利用SHAP进行全局与局部致因分析,为缺陷油品配方提供了可操作的单变量定性修正建议。信自成等[25]利用表格先验数据拟合网络(Tabular Prior-data Fitted Network, TabPFN)预测钢包炉精炼Si元素收得率,并结合SHAP方法实现了高精度且可解释的预测。郑可欣等[26]在蒸汽裂解产物成分预测中引入SHAP分析关键特征,提升了模型对主要影响因素的解释能力。陈霖等[27]将决策树与SHAP结合,用于量化炼化含油污水急性毒性关键驱动因子。进一步,Huang等[28]和Panicker等[29]通过集成模型结合SHAP与PDP,分析作物残留物及生物炭产率,揭示了输入特征与输出之间的非线性关系及阈值效应。Yun等[30]利用极端梯度提升算法(eXtreme Gradient Boosting, XGBoost)结合SHAP和PDP,分析污水处理脱氮过程关键因素及非线性交互作用,为工艺优化提供了可解释性决策支撑。由此可见,将机器学习预测模型与SHAP、PDP等可解释性方法相结合,不仅能够实现高精度预测,还能揭示复杂模型的预测依据及各变量作用规律,为复杂工艺条件下的参数优化与过程控制提供有效支持。

尽管机器学习方法在辛烷值预测中已取得一定进展,但在汽油在线调和配方优化场景下,其实际应用中仍面临两个突出问题:一是RF、XGBoost或神经网络模型等传统方法超参数调优复杂,建模效率偏低;二是上述模型多为黑箱,仅输出预测结果,难以揭示各组分如何影响辛烷值以及组分之间是否存在交互作用,无法为配方调整提供明确的决策依据。因此,亟需一种兼顾预测精度、建模效率及可解释性的方法。TabPFN是近年来提出的一种面向表格数据的预测模型,其依托大规模合成数据预训练机制,能够对新表格数据任务进行直接推理,无需复杂超参数调优,且能有效捕捉特征间的非线性关系和潜在依赖结构。在此基础上,引入SHAP从全局和局部两个层面量化各组分的贡献度

与影响方向,引入PDP从单变量和双变量维度揭示组分的非线性响应规律与协同交互作用。上述方法协同配合,可在保证预测精度的同时,为配方优化提供从预测到解释的完整链路。

基于上述分析,本文提出将TabPFN与SHAP及PDP方法相结合,构建可解释的汽油辛烷值预测模型。该模型以企业实际汽油调和组分及其配比为输入、以辛烷值为输出,通过TabPFN实现高效预测,并利用SHAP与PDP从全局与局部、单变量与双变量多维度解析各组分及其对比对辛烷值的影响机制,为汽油在线调和配方优化提供数据支撑与决策支持。

1 基于TabPFN的汽油辛烷值预测与可解释性分析方案

1.1 成品汽油在线调和工艺简介

某炼油化工企业的成品汽油在线调和工艺流程如图1所示。

由图1所示,在成品汽油在线调和过程中,四种调和组分即加氢汽油、MTBE、车用异辛烷及混合芳烃(主要由汽油重芳烃、甲苯、乙苯及二甲苯等组成)分别储存在独立组分罐中,经对应压力泵加压后进入流量调节阀,由集散控制系统(Distributed Control System, DCS)结合调和优化控制系统进行精确调控,实现各流路流量的闭环控制,并在静态混合器中完成高效、均匀的混合。在调和过程中,在线分析仪实时采集各组分及混合汽油的物性参数,并将数据反馈至DCS,用于动态调整各组分比例,从而精确控制辛烷值及其他关键质量指标,确保成品汽油满足企业标准与市场需求。

1.2 汽油辛烷值预测建模与模型可解释性分析的方案设计

辛烷值是炼油化工企业在汽油调和过程中判断产品是否合格的核心指标。在实际生产中,汽油在线调和配方优化需要根据组分配比准确预测辛烷值,现有辛烷值预测方法多依赖经验决策、线性配方模型或传统机器学习算法。然而,在多组分、相互耦合且非线性的复杂汽油调和场景下,这些方法往往存在预测精度不足、建模效率偏低以及参数调优复杂等问题。此外,模型的推理决策过程缺乏可解释性,难以满足精细化生产与产品质量的精准调控需求。为此,本文提出了一种基于TabPFN的

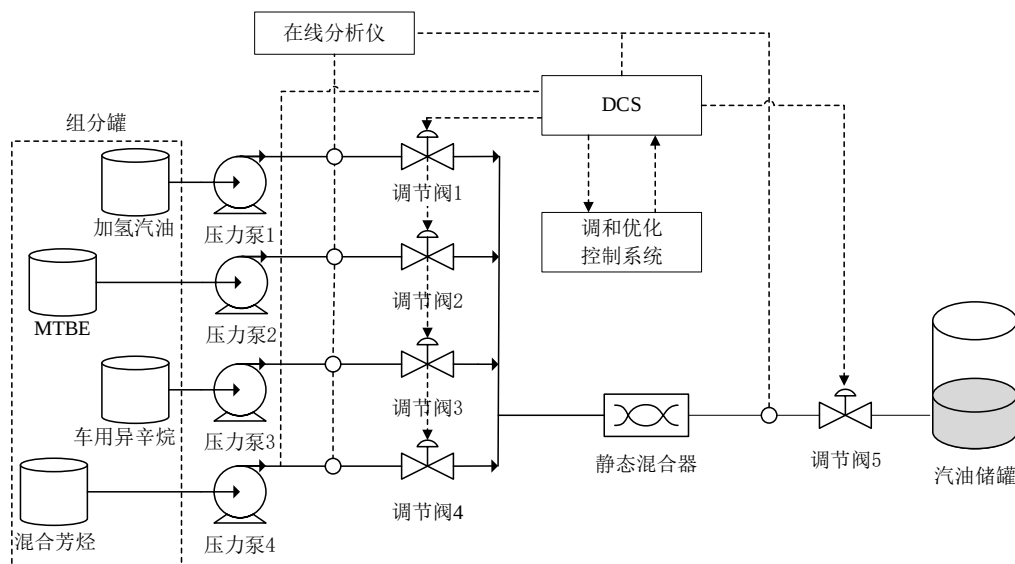


图1 成品汽油在线调和工艺流程示意图

Fig.1 Schematic diagram of the online blending process for finished gasoline

辛烷值预测与模型可解释性分析方案,旨在为汽油在线调和配方提供高精度、低调参且可解释的代理

模型,其整体框架如图2所示。

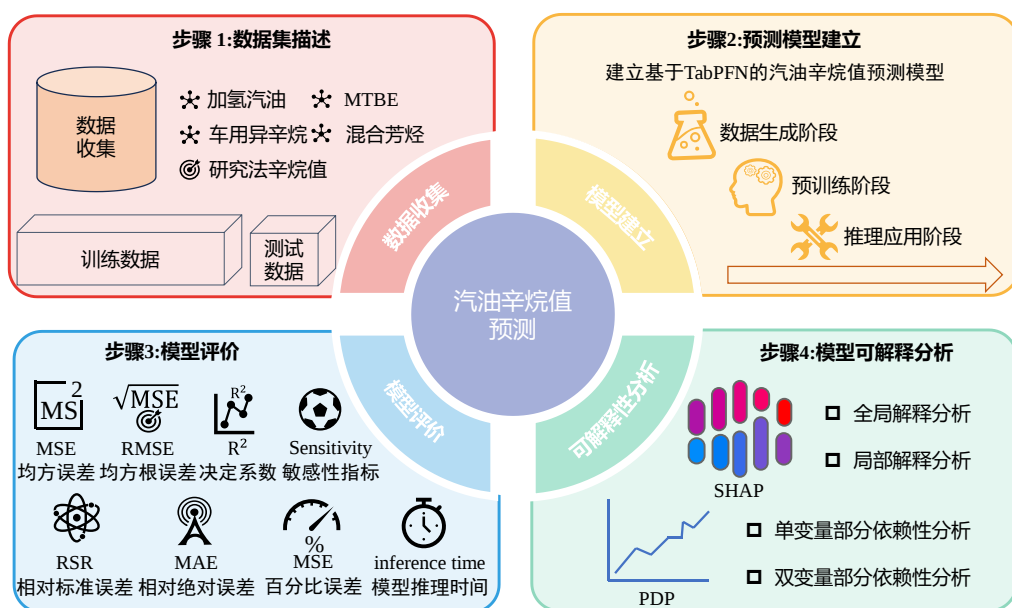


图2 基于TabPFN的成品汽油在线调和辛烷值预测建模与可解释性分析方案

Fig.2 Schematic of Finished Gasoline Online Blending Octane Number Prediction and Explainability Analysis Based on TabPFN

由图2可知,本研究在完成工艺机理解析和调和数据分析的基础上,首先结合成品汽油实际生产数据,将加氢汽油、MTBE、车用异辛烷和混合芳烃四种调和组分及其配比作为模型输入,以辛烷值作为模型输出;然后,构建基于TabPFN的汽油辛烷值预测模型,并在相同数据集条件下与XGBoost等典型机器学习模型进行对比分析;其次,采用多种模型性能评价指标,对TabPFN模型在预测精度、训练

效率、推理效率及敏感性方面的优势进行验证;最后,引入SHAP和PDP方法对TabPFN模型进行可解释性分析,揭示各调和组分及其配比对预测结果的贡献和影响规律,从而提升模型的透明度和可解释性,为配方优化提供从预测到解释的决策支持。

为保证对比实验的公正性,本研究在同一设备及相同训练集与测试集条件下,对所有模型进行训练与评估。实验设备主要配置为:处理器为AMD

Ryzen 7 6800H, 主频 3.20 GHz, 集成 AMD Radeon (TM)

Graphics; 内存 16 GB, 运行频率 6400 MT/s; 存储容量 477 GB。软件环境方面, 操作系统为 Windows 11 64 位。本文使用的所有算法均基于 Python 3.12 编程语言实现, 主要依赖的库包括 sklearn、tabpfn、catboost、xgboost、shap、numpy、pandas 及 time 等。

2 预测模型构建与分析方法

2.1 数据集描述

本研究采用某大型炼油化工企业油品储运厂

2024 年 92# 成品汽油在线调和的阶段性生产数据。数据涵盖加氢汽油、MTBE、车用异辛烷及混合芳烃四种调和组分及其配比, 以及对应的辛烷值等关键质量指标。数据采集期间为在线调和优化系统调试阶段, 各调和组分物性相对稳定, 并以 180 秒间隔实时监测, 为后续模型训练与分析提供了可靠的数据基础。

2.2 TabPFN 算法简介

TabPFN 由 HOLLMANN N 等^[31]提出, 是一种专门针对表格数据设计的神经网络模型, 其核心在于利用自注意力机制高效建模特征间的非线性关系及潜在依赖结构。该模型的训练过程可分为三个阶段, 如图 3 所示。

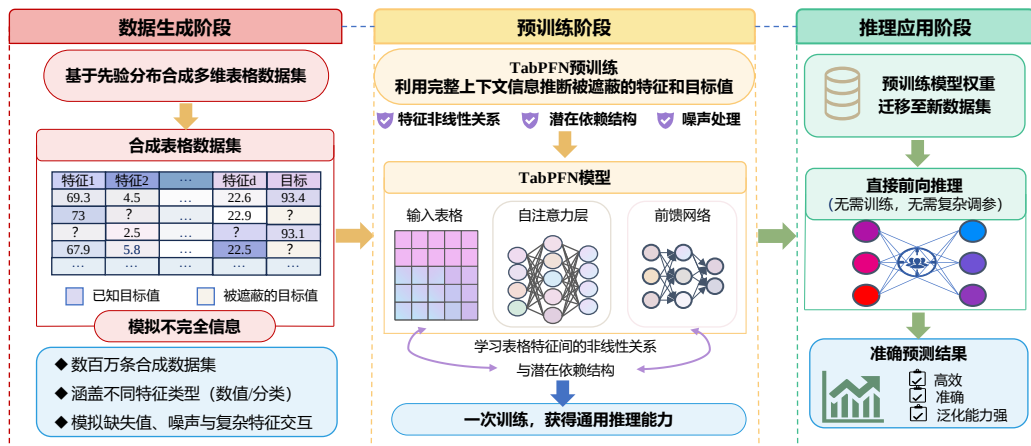


图3 TabPFN 模型训练原理示意图

Fig.3 Schematic diagram of the TabPFN model training principle

首先, 在数据生成阶段, 模型基于先验分布生成大规模合成表格数据, 覆盖数值型、类别型等多种特征类型, 并随机遮蔽部分特征与目标值, 模拟实际监督学习中信息不完整的情况, 为模型预训练提供多样化、高覆盖度的训练样本。其次, 在预训练阶段, 模型利用完整上下文信息, 通过自注意力层捕捉特征间的非线性依赖和潜在交互结构, 并借助前馈网络对被遮蔽的特征或目标值进行预测。通过大规模离线预训练, TabPFN 可有效学习表格数据的通用分布规律, 进而获得面向新任务的直接推理能力, 无需针对新数据集开展额外超参数调优。最后, 在推理应用阶段, 预训练完成后的模型权重可直接迁移至新数据集, 实现高精度预测。这种设计使 TabPFN 能够在小样本条件下, 稳健处理缺失值、噪声及复杂特征交互等问题, 在表格数据预测任务中表现出优异的泛化性能与建模效率, 尤其适用于数据规模有限、对部署时效性要求较高的工业

场景。

由于汽油调和过程中各组分及其配比对辛烷值存在显著非线性影响, 不同组分之间可能存在复杂交互作用, 传统基于经验公式或线性模型的预测方法难以充分捕捉这些关系。在此背景下, TabPFN 凭借其对表格特征非线性依赖和复杂交互的强大建模能力, 以及在小样本、多类型特征等条件下的稳健性, 成为辛烷值预测的理想选择。本研究拟基于 TabPFN 构建辛烷值预测模型, 用于快速评估汽油调和配方, 从而为汽油在线调和优化提供可靠的数据支持。

2.3 模型评价方法

为了综合评价 TabPFN 预测模型的预测精度及其在配方扰动条件下的输出稳定性, 本文从预测性能评价和配方扰动敏感性评价两个方面开展模型评价。一方面, 采用常用回归评价指标对不同模型在相同测试集上的预测性能进行对比分析; 另一方

面,针对汽油在线调和过程中可能存在的组分流量波动、计量误差及配方微调等情况,构造配方扰动样本,进一步分析模型对配方比例小幅变化的敏感性。

2.3.1 预测性能评价指标 为了评估 TabPFN 预测模型的性能优势,本文在相同测试集上采用回归分析中常用评价指标对各模型性能进行对比评估,包括均方误差(Mean Square Error, MSE)、均方根误差(Root Mean Square Error, RMSE)、平均绝对误差(Mean Absolute Error, MAE)、决定系数(R^2)、相对标准误差(Relative Standard Deviation, RSR)和百分比误差(Percent Bias, PBIAS)。其中, R^2 用于衡量模型的拟合优度,其值越接近1表示模型拟合数据的能力越强;MSE、RMSE及MAE用于评估预测值与实际值之间的误差大小,其中MSE和RMSE关注平方误差及其标准差,MAE关注平均绝对偏差,值越低表示预测精度越高;RSR反映预测误差相对于数据波动的比例,值越小表示模型拟合度越好,通常低于0.5视为良好;PBIAS则用于量化模型预测的总体偏差,负值表示低估,正值表示高估,零值则表示模型无偏。各评价指标的计算公式如式(1)式~(6)所示。

$$\text{MSE} = \frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2 \quad (1)$$

$$\text{RMSE} = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2} \quad (2)$$

$$\text{MAE} = \frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i| \quad (3)$$

$$R^2 = 1 - \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y})^2} \quad (4)$$

$$\text{RSR} = \sqrt{\frac{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2}{\frac{1}{n-1} \sum_{i=1}^n (y_i - \bar{y})^2}} \quad (5)$$

$$\text{PBIAS} = \left[\frac{\sum_{i=1}^n (y_i - \hat{y}_i)}{\sum_{i=1}^n y_i} \right] \times 100\% \quad (6)$$

式中: n 为测试集样本数, $y_i, \hat{y}_i, \bar{y}$ 分别为第*i*个样本的实际值、预测值和实际值的平均值。

此外,为全面评估模型的计算效率,本文同时记录各模型的训练耗时,并对 TabPFN 进行单样本推理耗时测试,以评估其在配方优化高频迭代调用场景下的实时性。

2.3.2 配方扰动敏感性评价指标 为进一步评估模型对配方比例变化的敏感性,本文从单组分扰动和整体配方扰动两个角度,分别分析 TabPFN 模型对汽油调和配方比例变化的响应稳定性。

单组分实验分别对四种调和组分施加相对扰

动。设某一组分的原始配比为 x ,扰动幅度为 δ ,扰动后的配比为 x' :

$$x' = x(1 + \delta)$$

其中, δ 分别取 $\pm 0.5\%$ 、 $\pm 1\%$ 、 $\pm 2\%$ 和 $\pm 3\%$ 。扰动后对四种组分配比进行归一化处理,使每个样本的组分配比总和与原始样本保持一致。由于扰动后的配方样本缺少对应的实测辛烷值,本文以扰动前后模型预测辛烷值变化量作为评价依据,计算公式如式(7)所示。

$$\Delta \hat{y}_i = |\hat{y}_i' - \hat{y}_i| \quad (7)$$

整体配方扰动实验同时对四种组分施加相对随机扰动。设第*i*个样本中第*j*个组分的原始配比为 x_{ij} ,扰动幅度为 r ,扰动后配比为 x'_{ij} :

$$x'_{ij} = x_{ij}(1 + \varepsilon_{ij}), \quad \varepsilon_{ij} \sim U(-r, r)$$

式中, r 分别设置 $\pm 0.5\%$ 、 $\pm 1\%$ 、 $\pm 2\%$ 和 $\pm 3\%$ 四个扰动等级,每个扰动等级重复扰动50次。扰动后同样进行归一化处理。在 $\Delta \hat{y}_i$ 的基础上,采用以下四个统计指标进行评价,各计算公式如式(8)~式(11)所示。

$$\overline{\Delta \hat{y}} = \frac{1}{N} \sum_{i=1}^N \Delta \hat{y}_i \quad (8)$$

$$Q_{95} = P_{95}(\Delta \hat{y}) \quad (9)$$

$$\sigma_{\Delta \hat{y}} = \sqrt{\frac{1}{N-1} \sum_{i=1}^N (\Delta \hat{y}_i - \overline{\Delta \hat{y}})^2} \quad (10)$$

$$S = \frac{\overline{\Delta \hat{y}}}{r_{\%}} \quad (11)$$

式中, N 为每个扰动等级下的重复扰动次数, $P_{95}(\cdot)$ 表示95%分位数计算函数, $r_{\%}$ 为扰动幅度的百分数形式。其中,平均预测变化量 $\overline{\Delta \hat{y}}$ 反映模型在给定扰动幅度下的整体响应水平,95%分位预测变化量 Q_{95} 反映大多数样本的变化范围,预测变化标准差 $\sigma_{\Delta \hat{y}}$ 反映不同样本之间预测变化的离散程度,单位扰动敏感性 S 反映每1%配方扰动引起的平均预测变化量。

2.4 模型可解释性分析

尽管 TabPFN 在少样本条件下能够实现高精度预测,但其基于复杂概率推理的机制使模型呈现典型的“黑箱”特性,用户难以直观理解先验信息如何影响具体预测结果,从而限制了模型的透明度和可信度^[32]。在汽油调和过程中,各组分及其配比对辛烷值预测的贡献尚不明确,因此对模型进行可解释分析对优化调和配方和保障产品质量具有重要意义。为此,本文针对训练好的 TabPFN 模型,采用 SHAP 方法进行全局与局部特征贡献分析,并结合

PDP方法探究单个特征及其交互作用对辛烷值预测的边际效应。这两种方法均不依赖于模型内部结构,能够从多个维度解释模型行为,为辛烷值预测与汽油调和决策提供可靠的解释与参考。

2.4.1 SHAP分析 SHAP方法结合了Shapley值和合作博弈论的原理,用于在考虑所有可能特征组合的情况下,量化每个特征及其组合对模型预测输出的贡献。其核心思想是基于博弈论机制,通过对所有可能特征子集的平均边际贡献进行加权求和,实现特征重要性的公平分配。同时,Shapley值满足线性可加性,即所有特征贡献之和等于模型预测输出与基准预测之间的差值,从而增强了黑箱模型的可解释性。SHAP值的计算公式如式(12)所示。

$$\phi_j = \sum_{S \subseteq F \setminus \{j\}} \frac{|S|!(|F| - |S| - 1)!}{|F|!} [f(S \cup \{j\}) - f(S)] \quad (12)$$

式中, F 表示所有特征的集合, $S \subseteq F \setminus \{j\}$ 为任意不含特征 j 的子集, $|S|$ 表示子集 S 的元素个数,函数 $f(S)$ 表示仅使用特征子集 S 的模型预测输出,而 $f(S \cup \{j\})$ 表示在 S 的基础上加入特征 j 后的预测输出。通过对所有子集进行加权求和,即可得到特征 j 的平均边际贡献 ϕ_j 。

2.4.2 PDP分析 偏依赖图PDP是机器学习模型常用的可视化工具,用于量化单个或多个特征对模型预测输出的平均边际效应。其核心思想是在固定目标特征值的情况下,对其余特征进行取样,并计算模型预测输出的平均值。PDP能够直观地揭示特征与预测目标之间的关系及潜在交互作用,从而提

高复杂模型的透明度和可信度。

对于单个特征 x_k ,其偏依赖函数定义如式(13)所示。

$$f_{\text{PDP}}(x_k) = \frac{1}{n} \sum_{l=1}^n f(x_k, x_{-k}^{(l)}) \quad (13)$$

式中,函数 $f_{\text{PDP}}(x_k)$ 表示特征 x_k 的平均边际效应。 $x_{-k}^{(l)}$ 表示第 l 个样本中除 x_k 外的其他所有特征取值, n 为测试集样本数。函数 $f(x_k, x_{-k}^{(l)})$ 表示在固定特征 x_k 并取第 l 个样本的其他特征值时的模型预测输出。通过对所有样本取平均,即可得到特征 x_k 对模型输出的全局效应。当分析两个特征的联合影响时,通过固定这两个特征的组合取值,对其余特征取样并计算模型预测平均值,从而绘制二维PDP图,直观反映两特征对预测输出的交互作用。

3 实验结果对比分析

3.1 数据预处理及数据集划分

为了便于与典型机器学习模型进行结果对比分析,本文采用相同的汽油在线调和实际生产数据集进行预测模型的验证研究。为消除异常值和量纲差异对模型的影响,采用拉依达准则(3σ 原则)对原始数据进行了筛选。最终选取了1620组实际汽油调和组分配比数据,以及对应的成品汽油辛烷值实测值。输入变量包括加氢汽油、MTBE、车用异辛烷和混合芳烃四类组分的调和比例。为清晰展示数据分布,表1给出了主要变量的描述性统计特征。

表1 主要变量描述性统计

Table 1 Descriptive statistics of the main variables

变量	统计指标						
	平均值	标准差	最小值	25%分位数	中位数	75%分位数	最大值
加氢汽油质量分数/%	71.21	3.20	63.00	69.10	71.10	73.00	82.90
MTBE质量分数/%	3.92	1.61	0.10	2.50	3.50	5.00	9.50
车用异辛烷质量分数/%	3.41	1.90	0.10	2.50	3.50	4.50	13.50
混合芳烃质量分数/%	21.45	3.39	7.80	19.70	21.90	23.50	29.20
辛烷值	93.08	0.17	92.60	93.00	93.00	93.20	93.70

由表1可知,四种调和组分的配比覆盖了较宽的范围(如加氢汽油63.0%~82.9%,MTBE 0.1%~9.5%),辛烷值集中在92.6~93.7区间,符合92#汽油的质量标准。数据具有一定的多样性,能够反映实际生产中的典型配方变化。为了构建高精度且具有强泛化能力的辛烷值预测模型,并验证其性能,按照8:2的比例将1620组数据随机划分为训练集

(1296组)和测试集(324组),确保不同模型在相同数据条件下的可比性。为检验数据划分的合理性,绘制了训练集与测试集在各输入特征及辛烷值上的左右半小提琴图,如图4所示。

由图4可见,两组数据在各特征上的分布形态、范围及分位数位置非常接近,未出现明显偏差,表明随机划分后的训练集和测试集具有良好的一致

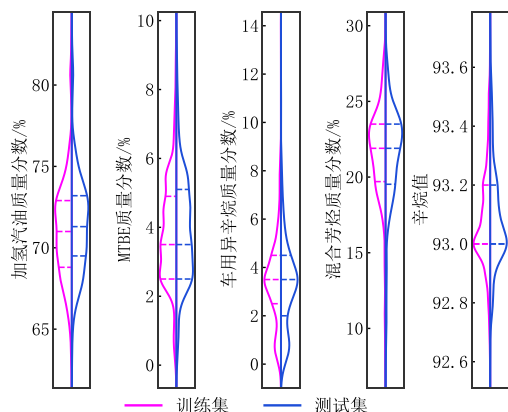


图4 训练集和测试集分布小提琴图

Fig.4 Violin plots of training and test set distributions

性,能够保证后续模型性能评估的公平性和可靠性。

3.2 基于TabPFN的汽油辛烷值预测模型构建

基于上述数据,本研究构建了基于TabPFN的汽油辛烷值预测模型。模型以加氢汽油、MTBE、车用异辛烷及混合芳烃四种调和组分及其配比作为输入,实际辛烷值作为预测目标。初始化TabPFN回归模型,模型自动加载预训练权重以快速适应新的数据模式,同时保留对表格特征非线性关系和潜在依赖结构的学习能力。在训练集上,模型学习各调和组分及其配比与辛烷值之间的复杂非线性关系,并在测试集上进行性能评估,实现高精度预测,为后续配方优化提供可信的定量基础。

3.3 各模型预测结果对比验证

3.3.1 各模型预测结果对比验证 基于六种不同机器学习模型预测汽油辛烷值的散点图如图5所示。在图中,蓝色点代表训练数据集,红色点代表测试数据集,且各数据点的实际值与预测值之间的关系可通过直线拟合表示。

如图5(a)–(f)所示,不同辛烷值预测模型的性能存在明显差异,其中,TabPFN模型整体表现最优。支持向量回归(Support Vector Regression, SVR)通过核函数将输入映射到高维特征空间,具备一定的非线性拟合能力。然而,其性能在很大程度上依赖于核函数形式及参数设置。在本研究中,SVR在训练集和测试集上的 R^2 分别仅为0.759和0.750,整体拟合效果相对较弱。相比之下,TabPFN在测试集上的MAE和RMSE分别由0.080和0.092降至0.008和0.032,测试集 R^2 由0.750提高至0.970,表明其在误差控制和预测精度方面均具有明显优势。

RF采用Bagging策略集成多个决策树学习器,

具有较好的稳定性和抗噪能力。RF在训练集和测试集的 R^2 分别为0.970和0.934,表明其拟合能力较强,但训练集与测试集之间仍存在一定性能差距,显示模型存在轻微过拟合现象。与RF相比,TabPFN的MAE由0.027降至0.008, RMSE由0.047降至0.032,测试集 R^2 由0.934提高至0.970,说明其在泛化性能和预测稳定性方面优于RF。

梯度提升决策树(Gradient Boosting Decision Tree, GBDT)、XGBoost和分类梯度提升(Categorical Gradient Boosting, CatBoost)均属于Boosting类集成模型,通过逐步拟合残差增强对复杂非线性关系的表征能力,因此整体性能优于SVR和RF。其中,GBDT在训练集和测试集上的 R^2 分别为0.995和0.944, XGBoost进一步通过引入L1和L2正则化增强泛化能力,其在训练集和测试集上的 R^2 分别为0.992和0.963; CatBoost则通过有序目标统计编码和排列机制处理类别特征,在表格数据建模中具有较好适应性,其测试集 R^2 达到0.968,已接近TabPFN。尽管如此,与GBDT相比,TabPFN的MAE和RMSE分别降低约11%和27%;与XGBoost相比,分别降低约47%和10%;与CatBoost相比,MAE和RMSE仍分别降低约47%和4%。这表明TabPFN在保持高预测精度的同时,在误差控制方面亦具有一定优势。

为进一步全面评估各模型在汽油辛烷值预测任务中的性能表现,本文在图4对比结果的基础上,引入多个量化评价指标进行综合分析,具体结果如表2所示。

由表2可知,TabPFN在各项评估指标中表现最佳。该模型测试集上的MSE仅为0.0010,在所有对比模型中最低,表明模型预测误差更小、预测精度更高。同时,TabPFN的RSR和PBIAS分别为0.174和0.0039%,均处于较低水平,说明在保持高预测精度的同时,系统残差和整体偏差也得到了有效控制,具备良好的预测稳定性和可靠性。此外,在训练效率方面,TabPFN训练仅需4秒,而SVR、RF、GBDT、XGBoost和CatBoost的训练耗时分别为69秒、60秒、30秒、14秒和90秒,各对比模型的训练耗时分别约为TabPFN的17倍、15倍、8倍、4倍和23倍。

为进一步评估推理效率,在AMD Ryzen 7 6800H笔记本CPU环境下对TabPFN进行了推理耗时测试,结果如表3所示。

由表3可知,单样本平均推理时间仅76.74毫

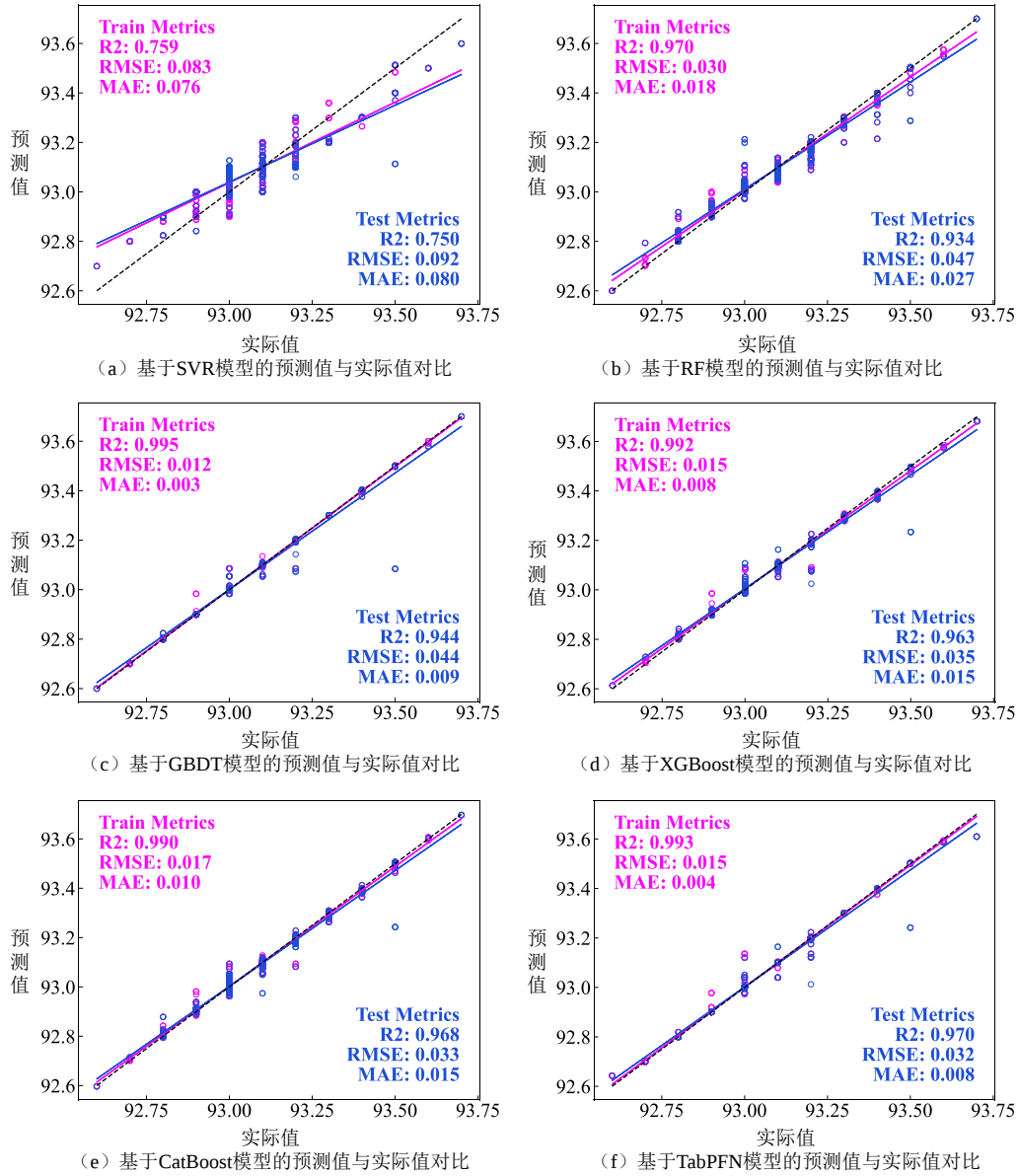


图5 基于不同模型预测辛烷值的结果对比图

Fig.5 Comparison of octane number predictions based on different models

注:Train Metrics—训练集指标;Test Metrics—测试集指标

表2 不同模型下辛烷值预测性能的比较

Table 2 Comparison of octane number prediction performance across different models

模型	MSE	RMSE	R ²	RSR	PBIAS/%	训练时间/s
SVR	0.0084	0.0918	0.750	0.500	-0.0009	69
RF	0.0022	0.0473	0.934	0.258	0.0034	60
GBDT	0.0019	0.0436	0.944	0.238	0.0055	30
XGBoost	0.0012	0.0352	0.963	0.192	0.0038	14
CatBoost	0.0011	0.0331	0.968	0.180	0.0042	90
TabPFN	0.0010	0.0318	0.970	0.174	0.0039	4

表3 TabPFN推理效率测试结果

Table 3 TabPFN inference efficiency test results

测试环境	测试集 样本数	重复 次数	平均总耗 时/s	耗时标 准差/s	单样本推理 时间/ms
AMD Ryzen 7 6800H	324	30	24.864	0.678	76.74

秒,30次重复测试耗时标准差为0.678秒,推理速度稳定。考虑到现场数据采样周期为180秒,单周期内可完成超过2000次推理,能够充分满足配方优化中高频迭代调用的实时性需求。由此可见,TabPFN不仅在预测性能上具有明显优势,同时训练与推理效率均表现优异,适合作为配方优化中高频调用的

代理模型。

综上所述,尽管SVR、RF、GBDT、XGBoost和CatBoost等模型均表现出一定的拟合能力,但在测试集上的预测精度、误差控制、训练效率和推理效率方面仍存在不同程度的局限。相比之下,TabPFN依托大规模合成数据预训练机制以及针对表格数据设计的归纳偏置,能够更有效地捕捉特征之间的复杂依赖关系,在无需复杂超参数调优的条件下实现预测精度与计算效率的协同提升,表现出较强的综合性能优势。

3.3.2 配方扰动条件下的敏感性验证 为进一步验证TabPFN模型在配方比例变化的输出稳定性,本文分别从单组分扰动和整体配方扰动两个角度进行分析。采用测试集中的324组92#汽油调和样本,分别对加氢汽油、MTBE、车用异辛烷和混合芳烃四种调和组分施加相对扰动,单组分扰动实验结果如下表4所示。

表4 单组分扰动下辛烷值预测变化量实验结果
Table 4 Experimental results of predicted octane number changes under single-component perturbations

扰动组分	扰动幅度			
	$\pm 0.5\%$	$\pm 1\%$	$\pm 2\%$	$\pm 3\%$
加氢汽油	0.0463	0.0592	0.0802	0.0941
MTBE	0.0403	0.0408	0.0456	0.0533
车用异辛烷	0.0411	0.0417	0.0447	0.0495
混合芳烃	0.0450	0.0564	0.0766	0.0879

由表4可知,随着扰动幅度由 $\pm 0.5\%$ 增大至 $\pm 3\%$,各组分扰动下的平均预测变化量整体呈增加趋势,模型能够对单组分配比变化产生连续响应。即使在 $\pm 3\%$ 扰动条件下,四种组分的平均预测变化量分别为0.0941、0.0533、0.0495和0.0879,均低于0.10,说明模型对单组分小幅相对扰动的预测输出较为稳定,未出现明显异常波动。需要说明的是,该结果主要反映模型在当前数据分布及配方约束条件下的局部敏感性,并不代表各组分真实增辛能力的强弱排序。

为模拟在线调和过程中多组分同时波动的实际情况,进一步开展了整体配方扰动实验。对四种组分配比施加相对随机扰动。整体配方扰动实验结果如下表5所示。

由表5可知,随着整体配方扰动幅度由 $\pm 0.5\%$ 增大至 $\pm 3\%$,模型预测辛烷值的平均预测变化量由0.0445增加至0.0773,整体变化幅度较小,说明模型

表5 整体配方扰动下辛烷值预测变化量实验结果

Table 5 Experimental results of predicted octane number changes under overall formulation perturbations

扰动幅度	平均预测变化量	95%分位预测变化量	预测变化标准差	单位扰动敏感性
$\pm 0.5\%$	0.0445	0.1788	0.0595	0.0890
$\pm 1\%$	0.0515	0.1919	0.0624	0.0515
$\pm 2\%$	0.0661	0.2210	0.0708	0.0330
$\pm 3\%$	0.0773	0.2375	0.0771	0.0258

对配方比例小幅扰动的平均响应较弱,未表现出明显过度敏感现象。从95%分位预测变化量来看,在 $\pm 3\%$ 扰动条件下,95%的扰动样本预测变化量不超过0.2375,表明绝大多数测试样本在整体配方扰动后预测输出仍保持相对稳定。预测变化标准差由0.0595增加至0.0771,说明随着扰动幅度增大,不同样本之间的预测变化离散程度略有增加,但整体仍处于较低水平。单位扰动敏感性未随扰动幅度增大而升高,说明模型未出现扰动放大现象。

综合单组分扰动和整体配方扰动实验结果,本文建立的TabPFN辛烷值预测模型在92#汽油调和配方范围内,对同牌号内配方比例小幅变化具有一定稳定性。

综上,TabPFN在汽油辛烷值预测任务中具有较好的应用前景,可为汽油在线调和配方优化提供高精度、高效率的代理模型支撑。后续结合SHAP和PDP分析,可进一步为配方优化提供多维度决策参考。

3.4 汽油辛烷值预测模型的可解释性分析验证

3.4.1 汽油辛烷值预测模型的SHAP分析 SHAP方法主要从全局和局部两个层面对TabPFN辛烷值预测模型进行可解释分析,基于TabPFN模型的全局SHAP特征重要性环形图如图6所示。

由图6可知,各汽油调和组分及其配比对TabPFN模型预测结果的贡献存在显著差异。其中,混合芳烃配比贡献最高,占比31.1%,其次为MTBE和加氢汽油,贡献度分别为27.4%和22.9%。车用异辛烷配比贡献相对较低,占比为18.6%。这表明,混合芳烃、MTBE和加氢汽油是影响辛烷值预测的主要因素,而车用异辛烷作用相对较弱。

辛烷值SHAP全局可解释性分析能够从整体层面揭示各调和组分及其配比在模型预测过程中的相对重要性,而局部可解释性分析则侧重于揭示单一样本中各调和组分配比对预测结果的具体贡献。

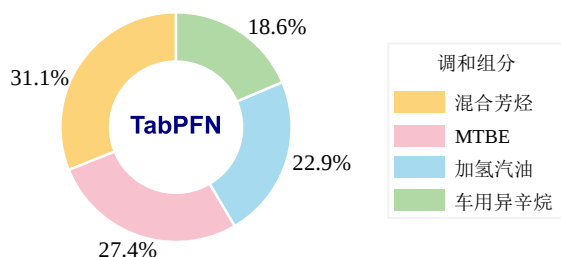


图6 基于TabPFN的SHAP特征重要性环形图

Fig.6 SHAP feature importance ring chart based on TabPFN

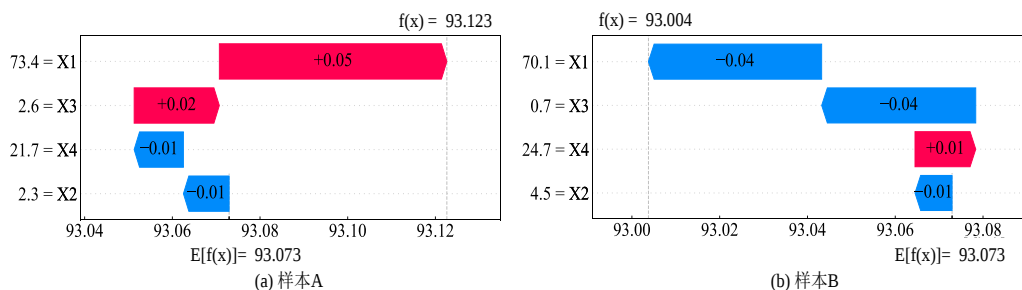


图7 不同样本的SHAP分析瀑布图

Fig.7 SHAP waterfall plots for different samples

注:X1—加氢汽油;X2—MTBE;X3—车用异辛烷;X4—混合芳烃

如图7所示,以样本A为例,其输入变量包括加氢汽油、车用异辛烷、混合芳烃及MTBE,TabPFN模型对应的基线值为93.073。在该样本中,加氢汽油的SHAP值为0.05,表明该配比组分对预测结果具有最大的正向贡献,使预测辛烷值提高了0.05。混合芳烃配比对应的SHAP值为-0.01,说明其对预测结果具有最大的负向贡献,使预测值降低了0.01。在综合基线值和各配比组分贡献后,模型最终预测值为93.123,而该样本的实际辛烷值为93.10,实际值与预测值吻合较好,说明TabPFN模型具有较好的预测能力和较高的局部解释一致性。在样本B中,加氢汽油对比对预测值的负向贡献最大,而混合芳烃配比则是该样本中唯一具有正向贡献的组分。该结果表明,不同样本中同一组分对辛烷值预测的影响方向和作用程度可能存在差异,进一步体现了SHAP局部解释在刻画样本个体差异和揭示模型决策依据方面的优势。

3.4.2 汽油辛烷值预测模型的PDP分析 为了深入揭示各汽油调和组分配比对模型输出的影响规律,各组分如何影响辛烷值以及组分之间是否存在交互作用,本节采用部分依赖图PDP开展可解释性分析,以清晰刻画各调和组分对辛烷值预测结果的边际影响及其变化趋势。

基于TabPFN模型构建的单变量部分依赖图如

不同样本的SHAP瀑布图如图7所示,其中,基线值 $E[f(x)]$ 表示模型对辛烷值的平均预测水平,各组分对应的SHAP值则反映其相对于基线值的正向或负向作用。正值表示该汽油调和组分配比推动预测值升高,负值表示其使预测值降低。通过将基线值与各组分SHAP值逐项累加,可以得到该样本的最终预测结果。

图8所示。其中,横轴为目标汽油组分的调和配比(质量分数)范围,纵轴为模型预测辛烷值随该组分变化的边际响应,直观反映了组分配比对辛烷值预测结果的非线性影响规律。

如图8所示,各汽油调和组分配比对辛烷值的影响表现出显著的非线性特征,同时存在阈值效应和最优作用区间。由图8(a)可知,加氢汽油对辛烷值的边际效应随配比增加呈先升后降趋势。在低配比阶段,辛烷值提升较为平缓;当配比约为78.5%时,预测辛烷值达到峰值,随后随配比进一步增加逐步下降。加氢汽油辛烷值相对较低,主要作为基础调和组分参与配方体系,其最优配比由多组分的协同作用决定。由图8(b)可知,MTBE在低配比范围内对辛烷值具有正向作用,辛烷值随配比快速上升。当配比超过约5%后,其边际效应逐渐减弱。MTBE作为典型高辛烷值调和组分,过量添加并不能持续提高辛烷值。实际调和中,需要在提高辛烷值的同时兼顾经济性。图8(c)表明,车用异辛烷对辛烷值的提升作用存在明显阈值。当配比低于6.9%时,辛烷值随配比增加快速上升;配比进一步增加仅能维持辛烷值在较高水平。这表明车用异辛烷作为高品质高辛烷值调和组分,在适宜调和配比范围内能够有效改善成品汽油的抗爆性能。由图8(d)可知,混合芳烃对辛烷值的影响呈复杂非线性

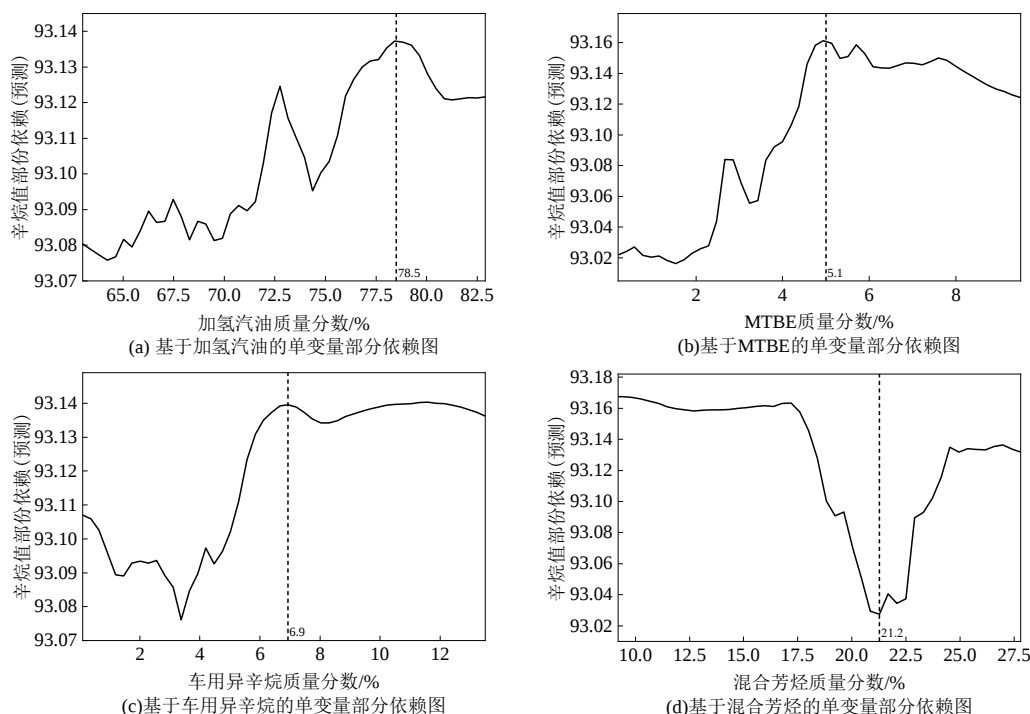


图8 基于TabPFN的各汽油调和组分单变量部分依赖图

Fig.8 Univariate partial dependence plots of gasoline blending components based on TabPFN

性波动。在配比约为10%~17.5%范围内,辛烷值保持较高水平,随后随配比增加持续下降,在约21.2%处达到最低值,超过该阈值后,辛烷值又逐渐回升。混合芳烃属于高辛烷值调和组分,但在本研究体系下,并非配比越高越好。在实际调和中,需结合整体配方进行精准调控。整体来看,各组分均存在最佳调和阈值,通过合理控制各组分配比,可在不同作用区间实现辛烷值的优化调控。

基于单变量分析结果,本研究进一步构建了二维部分依赖图,以系统揭示汽油组分间的交互作用,并探讨加氢汽油、MTBE、异辛烷与混合芳烃对辛烷值的协同效应,如图9所示。

由图9(a)可知,当MTBE配比约 $>4.5\%$ 时,加氢汽油在约67%~76%范围内能够将辛烷值稳定在93.11以上,最优组合可达93.16;而当MTBE配比约 $<4\%$ 时,辛烷值整体偏低,表明在MTBE充足的条件下,加氢汽油表现出明显的协同增效作用。由图9(b)可知,当车用异辛烷配比约 $>5\%$ 时,可与加氢汽油协同提升辛烷值,最优组合同样可达93.16;若车用异辛烷配比不足,其增效作用明显受限;由图9(c)可知,混合芳烃存在约14%~18%的最佳配比区间,当其与加氢汽油约72%~76%配比协同时,辛烷值最高达93.22;而当混合芳烃配比约 $>20\%$ 时,辛烷值则明显下降。由图9(d)可知,当MTBE配比约

$>4.5\%$ 时,可增强车用异辛烷的提辛效果,最优组合辛烷值为93.15。当MTBE约 $<3\%$ 时,辛烷值整体偏低。由图9(e)可知,当MTBE配比约 $>4.5\%$ 且混合芳烃约17%~19%时,辛烷值最高达93.20;若MTBE配比不足,则混合芳烃的提效作用受限,且并非配比越高效果越好。由图9(f)可知,当混合芳烃约 $<19\%$ 时,辛烷值随车用异辛烷配比增加而升高,两者呈现明显的协同提辛效应。

综合双变量PDP分析结果表明,MTBE与车用异辛烷是提升汽油辛烷值的核心调控组分,同时也是其他组分发挥协同作用的重要前提。当MTBE配比约 $\geq 4.5\%$ 、车用异辛烷配比约 $>5\%$ 时,加氢汽油与混合芳烃能够有效协同提升辛烷值,为汽油调合工艺优化及原料配比提供了重要参考和理论依据。

4 结 论

本文建立了基于TabPFN的汽油辛烷值预测模型,并结合SHAP与PDP方法对模型进行了可解释性分析,得出的主要结论如下:

(1)TabPFN免调参建模的优势。现有辛烷值预测研究多采用SVR、RF、XGBoost等传统机器学习方法,需反复调整核函数类型、惩罚参数、学习率、树深等多个超参数,建模过程较为繁琐。本文引入

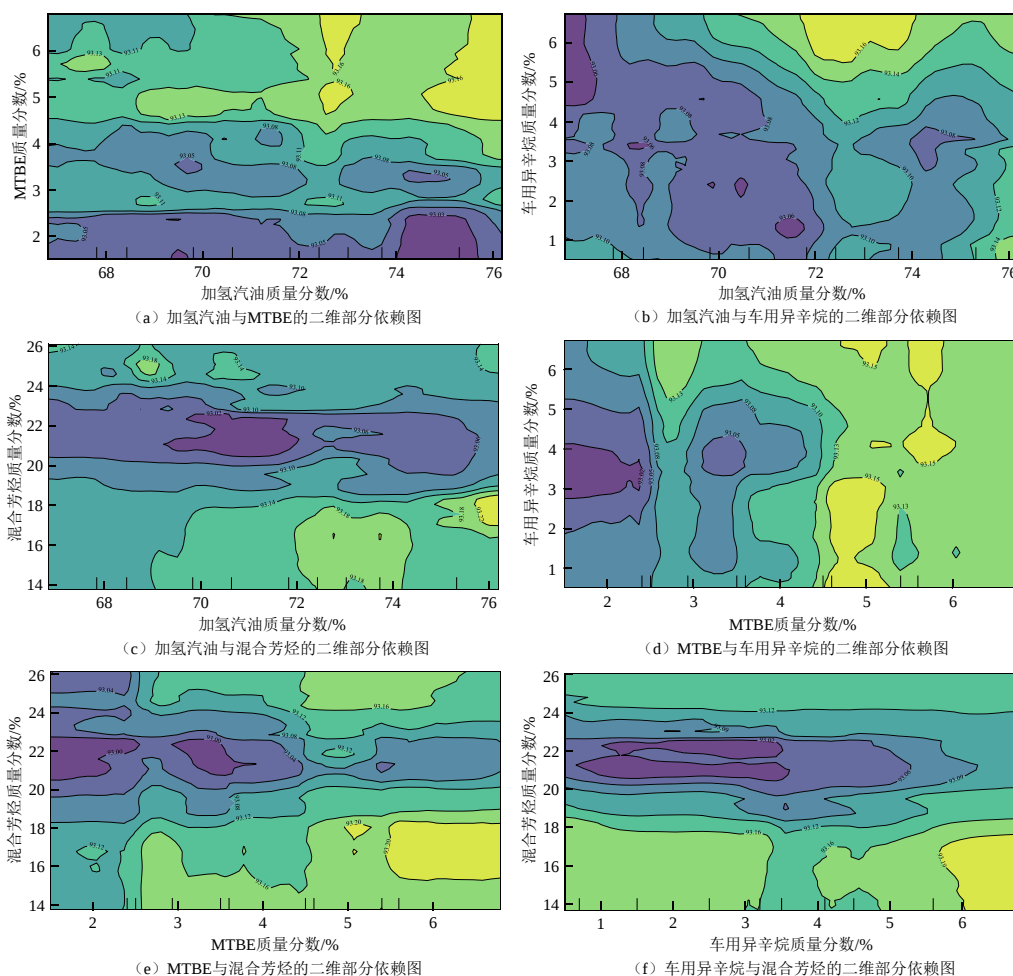


图9 基于TabPFN的任意两个汽油调和组分的二维偏依赖图

Fig.9 Two-dimensional partial dependence plots of any two gasoline blending components based on TabPFN

TabPFN 构建辛烷值预测模型,依托其预训练机制,无需复杂超参数调优即可获得优异的预测性能,测试集的 R^2 达到 0.970,表现出良好的预测性能。

(2) 训练与推理效率及稳定性优势。在训练效率方面,TabPFN 训练仅需 4 秒,各对比模型的训练耗时约为其 4 至 23 倍,在推理效率方面,TabPFN 单样本推理时间达到毫秒级。同时,配方扰动实验表明,模型对同牌号内配方比例小幅变化具有一定稳定性,在 $\pm 3\%$ 扰动条件下四种组分的平均预测变化量均低于 0.10,未出现扰动放大现象。

(3) TabPFN-SHAP-PDP 预测与解释一体化框架。SHAP 从全局和局部两个层面量化各组分对辛烷值预测的贡献度与影响方向,PDP 从单变量和双变量两个维度揭示组分的非线性响应规律与协同交互作用。该框架增强了模型的透明度和可信度,上述先验知识也可后续配方优化提供帮助,在一定程度上降低强非线性带来的求解难度。

本研究通过将 TabPFN 与 SHAP、PDP 方法结

合,构建了面向汽油在线调和配方优化的预测与解释一体化方案。本文模型基于 92# 成品汽油在线调和数据构建,主要适用于组分油物性相对稳定条件下的辛烷值预测。未来可进一步采集多批次组分油物性数据与多牌号调和数据,验证模型在更宽工况范围内的适用性,并引入因果推断等方法增强决策支持能力。

参考文献

- [1] 何仁初,张朝晖,杨明磊,等.考虑碳排放因素的汽油调合在线优化[J].化工学报,2023,74(2): 818-829.
He R C, Zhang Z H, Yang M L, et al. Online optimization of gasoline blending considering carbon emissions[J]. CIESC Journal, 2023, 74(2): 818-829.
- [2] He R C, Yan X Y, Hua J J, et al. Closed-loop gasoline blending scheduling based on real-time optimized slack feedback[J]. Chemical Engineering Science, 2025, 309: 121426.
- [3] Lyu D L, Cui X J, Qu Y Q, et al. Fuel octane number measurement method based on high-frequency combustion pressure and its application to fuel octane number over 100[J].

- Energy, 2025, **335**: 138234.
- [4] Fernanda M A, Irawan B. The effect of different octane number on power and specific fuel consumption in gasoline compression ignition engine[J]. *Evrinata: Journal of Mechanical Engineering*, 2025, **2**(2): 32–36.
- [5] Wu Y S, Liu Y S, Li X L, et al. Gasoline octane number prediction from near-infrared spectroscopy with an ANN-based model[J]. *Fuel*, 2022, **318**: 123543.
- [6] 高永红. 中红外光谱法在汽油生产控制中的应用[J]. *石化技术*, 2018, **25**(10): 135–137.
- Gao Y H. Application of mid-infrared spectroscopy in gasoline production control[J]. *Petrochemical Industry Technology*, 2018, **25**(10): 135–137.
- [7] 王拓, 戴连奎, 马万武. 拉曼光谱结合后向间隔偏最小二乘法用于调和汽油辛烷值定量分析[J]. *分析化学*, 2018, **46**(4): 623–629.
- Wang T, Dai L K, Ma W W. Quantitative analysis of blended gasoline octane number using Raman spectroscopy with backward interval partial least squares method[J]. *Chinese Journal of Analytical Chemistry*, 2018, **46**(4): 623–629.
- [8] 蒋书波, 林锦国, 程明霄, 等. 基于拉曼技术的汽油辛烷值测定系统设计[J]. *化工学报*, 2011, **62**(8): 2188–2194.
- Jiang S B, Lin J G, Cheng M X, et al. System design of gasoline octane number detection based on Raman technology[J]. *CIESC Journal*, 2011, **62**(8): 2188–2194.
- [9] Wang B Y, Cao R Y, Pi Z Y, et al. Displacement-imbibition coupling mechanisms between matrix and complex fracture during injecting-shut in-production process using pore-scale simulation and NMR experiment[J]. *Geoenergy Science and Engineering*, 2024, **242**: 213254.
- [10] Cheng H R, Wei J G, Zhou X F, et al. Experimental study on the effect of microscale wettability of shale to chemical agents by using nuclear magnetic resonance method[J]. *Powder Technology*, 2024, **445**: 120084.
- [11] Jiao L, Liu H H, Qu L, et al. QSPR studies on the octane number of toluene primary reference fuel based on the electrotopological state index[J]. *ACS Omega*, 2020, **5**(8): 3878–3888.
- [12] Wu Y S, Guan C, Huang Z, et al. Fuel octane number prediction based on topological indices and active subspace method[J]. *Fuel*, 2021, **293**: 120494.
- [13] Al-Fahemi J H, Albis N A, Gad E A M. QSPR models for octane number prediction[J]. *Journal of Theoretical Chemistry*, 2014, **2014**: 520652.
- [14] Issa H M. Prediction of octane numbers for commercial gasoline using distillation curves: a comparative regression analysis between principal component and partial least squares methods[J]. *Petroleum Science and Technology*, 2024, **42**(10): 1233–1249.
- [15] 仇爱波, 周如金, 邱松山, 等. 基于原子类型电拓扑状态指数的QSPR模型预测链烷烃辛烷值[J]. *天然气化工—C1化学与化工*, 2014, **39**(5): 52–55.
- Qiu A B, Zhou R J, Qiu S S, et al. QSPR modeling with electrotopological state index for atom type for predicting octane number of paraffins[J]. *Natural Gas Chemical Industry*, 2014, **39**(5): 52–55.
- [16] Mohan B, Alsaleh F, AlRamadan A S, et al. In silico estimation of density, vapor pressure, distillation curve, and octane numbers of gasoline fuels using machine learning[J]. *Fuel*, 2025, **399**: 135640.
- [17] Wang C C, Chien C H. Machine learning for industrial optimization and predictive control: a patent-based perspective with a focus on Taiwan's high-tech manufacturing[J]. *Processes*, 2025, **13**(7): 2256.
- [18] Sun X Y, Zhang F, Liu J P, et al. Prediction of gasoline research octane number using multiple feature machine learning models[J]. *Fuel*, 2023, **333**: 126510.
- [19] Zhang F Y, Su X C, Tan A L, et al. Prediction of research octane number loss and sulfur content in gasoline refining using machine learning[J]. *Energy*, 2022, **261**: 124823.
- [20] Lin X C, Yin S T, Li X. A trustworthy framework for predicting gasoline octane loss using sparse autoencoder and stacking ensemble learning in petrochemical processes[J]. *Results in Engineering*, 2025, **28**: 107519.
- [21] Hettikankanamage N, Shafiabady N, Chatteur F, et al. eXplainable artificial intelligence (XAI): a systematic review for unveiling the black box models and their relevance to biomedical imaging and sensing[J]. *Sensors*, 2025, **25**(21): 6649.
- [22] Zhao C, Liu J, Parilina E. The Shapley value contribution to explainable artificial intelligence: a comprehensive survey[J]. *Dynamic Games and Applications*, 2025: 1–38.
- [23] Pelosi D, Cacciagrano D, Piangerelli M. Explainability and interpretability in concept and data drift: a systematic literature review[J]. *Algorithms*, 2025, **18**(7): 443.
- [24] 李炜, 郑明杰, 李亚洁, 等. 融合可解释机器学习的成品汽油调和配方质量预测评价与致因分析[J]. *石油学报(石油加工)*, 2024, **40**(1): 126–136.
- Li W, Zheng M J, Li Y J, et al. Predictive evaluation and cause analysis of finished gasoline blending formulation quality by interpretable machine learning[J]. *Acta Petrolei Sinica (Petroleum Processing Section)*, 2024, **40**(1): 126–136.
- [25] 信自成, 张江山, 张军国, 等. TabPFN与SHAP融合的LF精炼Si元素收得率预测模型[J]. *中国冶金*, 2025, **35**(11): 178–186.
- Xin Z C, Zhang J S, Zhang J G, et al. Prediction model of Si element yield in LF refining based on TabPFN and SHAP fusion[J]. *China Metallurgy*, 2025, **35**(11): 178–186.
- [26] 郑可欣, 江雨欣, 毕可鑫, 等. 用于蒸汽裂解产物成分预测的集成迁移学习框架[J]. *化工进展*, 2024, **43**(5): 2880–2889.
- Zheng K X, Jiang Y X, Bi K X, et al. Ensemble transfer learning framework for outflow compositions prediction in steam cracking process[J]. *Chemical Industry and Engineering Progress*, 2024, **43**(5): 2880–2889.
- [27] 陈霖, 祖思雅, 胡方杰, 等. 炼化含油污水全流程急性生物毒性演变与关键影响因子识别[J/OL]. *化工进展*. <https://doi.org/10.16085/j.issn.1000-6613.2025-1280>.
- Chen L, Zu S Y, Hu F J, et al. Evolution of acute ecotoxicity and identification of key influencing factors in full-process refinery oily wastewater[J/OL]. *Chemical Industry and Engineering Progress*. <https://doi.org/10.16085/j.issn.1000-6613.2025-1280>.
- [28] Huang W Y, Li K Q, Miao S S. Real-time process design enabled by an interpretable GBDT model for high-fidelity prediction of pyrolysis products from crop residues[J]. *Bioresource Technology*, 2026, **442**: 133742.
- [29] Panicker T F, Sarkar D, Krishna R, et al. A machine learning-assisted prediction of potential biochar yield subjected to the physicochemical properties of biomass[J]. *Fuel*, 2026, **409**: 137836.

- [30] Yun J H, Yu Y, Tao C L, et al. Machine learning-based optimization of enhanced nitrogen removal in a full-scale urban wastewater treatment plant with ecological combination ponds[J]. *Water Research*, 2025, **285**: 123976.
- [31] Hollmann N, Müller S, Purucker L, et al. Accurate predictions on small data with a tabular foundation model[J]. *Nature*, 2025, **637** (8045): 319–326.
- [32] Wang H Y, Pei J, Yan F Q, et al. Application of TabPFN model on the energy performance improvement of high-power multistage centrifugal pump[J]. *Energy*, 2025, **336**: 138399.