

文章编号: 1671-6906(2026)02-0000-00

基于全信息竞争自适应高斯建模策略的煤炭发热量建模方法

李赛赛¹, 佟鹏^{2,3}, 王圣毫¹, 张纪峰¹, 朱永胜¹

(1. 中原工学院 自动化与电气工程学院, 河南 郑州 450007; 2. 保定电力职业技术学院, 河北 保定 071051; 3. 国网冀北电力有限公司技能培训中心, 河北 保定 071051)

摘要: 针对煤炭发热量传统检测方法周期长、效率低的问题,提出了一种基于近红外光谱技术并结合全信息竞争自适应高斯回归建模策略的建模方法 FICAGM。该方法首先采用自扩展全信息优化策略 SFIOS,自动寻找最佳光谱预处理方法组合,然后利用竞争性自适应重加权采样 CARS 算法筛选与发热量最相关的特征波长,最后基于高斯过程回归 GPR 建立定量分析模型。实验结果表明,其预测决定系数 R^2 可达 0.920,预测均方根误差 RMSEP 为 0.909,性能显著优于主成分分析、连续投影算法等传统特征提取方法及支持向量机模型。本研究提出的面向煤炭品质快速无损检测的近红外建模方法,在所研究样本与实验条件下表现出较好的预测精度与稳健性,建模流程完整,可用于火电、冶金等工业场景的煤质快速监测与质量控制。

关键词: 近红外光谱;煤炭;发热量;FICAGM;CARS

中图分类号: TM464 文献标志码: A DOI:10.3969/j.issn.1671-6906.2026.02.000

A prediction model for coal calorific value based on full-information competitive adaptive reweighted sampling and gaussian process regression

LI Saisai¹, TONG Peng^{2,3}, WANG Shenghao¹, ZHANG Jifeng¹, ZHU Yongsheng¹

(1. School of Automation and Electrical Engineering, Zhongyuan University of Technology, Zhengzhou 450007, China; 2. Baoding Electric Power Technical College, Baoding 071051, China; 3. Skills Training Center of State Grid Jibei Electric Power Company, Baoding 071051, China)

Abstract: To address the long analytical cycle and low efficiency of conventional calorific value measurements, this study develops a near-infrared spectroscopy modeling framework called Full-Information Competitive Adaptive Gaussian Modeling, or FICAGM. The method first employs a self-extending full-information optimization strategy to automatically determine the most suitable preprocessing sequence for the spectra. It then applies competitive adaptive reweighted sampling to extract wavelength variables most relevant to calorific value, followed by a Gaussian process regression model for quantitative calibration. Experiments show that the proposed framework achieves a prediction coefficient of determination of 0.920 and a root-mean-square error of prediction of 0.909, outperforming traditional chemometric approaches such as principal component analysis, successive projection algorithms and support vector machines. Under the tested conditions, the framework exhibits strong predictive accuracy and robustness, and its complete mod-

收稿日期: 2025-11-11

基金项目: 河南省高等学校校重点科研项目(25A520001); 河南省科技攻关项目(242102210177); 河南省自然科学基金项目(252300421881); 国家自然科学基金项目(62473391)

引文格式: 李赛赛, 佟鹏, 王圣毫, 等. 基于全信息竞争自适应高斯建模策略的煤炭发热量建模方法[J]. 中原工学院学报, 2026, 37(2): 00-00.

LI Saisai, TONG Peng, WANG Shenghao, et al. A prediction model for coal calorific value based on full-information competitive adaptive reweighted sampling and gaussian process regression[J]. Journal of Zhongyuan University of Technology, 2026, 37(2): 00-00.

eling pipeline is well suited for rapid, non-destructive coal quality monitoring in coal-fired power generation, metallurgical processes and other industrial applications.

Key words: near-infrared spectroscopy; coal; calorific value; FICAGM; CARS

煤炭发电作为工业文明的重要支撑,因其经济性好、运行可靠,在全球能源结构中具有不可替代的作用^[1]。然而煤炭的化学组成高度复杂,固定碳、挥发分、灰分及发热量等多种指标相互耦合,共同决定其燃烧特性与能量转化效率^[2-3]。其中,发热量作为评价煤炭能源品质的核心指标,其高低直接影响发电系统的热效率和经济性。当前国内外普遍采用氧弹量热法测定煤炭发热量,该方法虽具有较高精度且符合国家标准,但存在检测周期长、操作繁琐、样品制备要求严苛以及资源消耗大等局限。这些问题使煤炭质量评价面临显著挑战,尤其在智能电厂要求对煤质参数进行快速、连续和大规模监测时,现有方法已难以满足实际需求^[4-6]。

在近红外光谱(Near-Infrared Spectroscopy, NIR)技术获得广泛应用前,煤炭发热量的预测主要依赖于工业分析和元素分析数据,通常采用经验公式法或回归模型进行估算。Chen等^[7]利用挥发分、固定碳、灰分等指标构建 Cubist 回归模型,实现了发热量预测,并获得了较好的拟合结果。Munshi等^[8]通过对比传统经验模型与集成学习模型指出,传统线性/经验类方法虽然实现过程简单、成本低,但对煤样来源与成分变化较为敏感,稳定性与适用范围受限。在此背景下,近红外光谱分析技术因其基于分子振动倍频与合频吸收效应,能够实现煤炭中发热量的快速、无损检测^[9-11]。与传统化学分析方法相比,NIR 不仅大幅缩短了检测周期,且无需使用化学试剂,兼具在线监测和绿色检测的优势,已成为智能电厂建设的重要支撑工具。要实现煤炭发热量的高精度近红外光谱建模,关键在于需对光谱预处理、数据降维与模型选择这三个环节进行合理的协同优化^[11-13]。当前,相关研究虽已在煤炭分析的个别环节取得进展:冯晓涵等^[7]系统梳理了煤质快速检测技术的演进脉络;张海军^[12]构建了基于发热量的定量预测模型;徐鼎等^[11]通过融合连续小波变换与卷积神经网络,实现了光谱特征的高效提取与分类;雷利维等^[14]则深入探讨了数据挖掘技术在煤质近红外光谱分析中的应用潜力。然而,现有研究普遍聚焦于单一环节的优化,尚未形成预处理、特征降维与建模环节的有机整合体系,这种碎片化研究模式制约了煤炭发热量预测模型在精度提升与抗干扰能力方面的突破性进展。

为有效应对上述挑战,本研究提出了一种系统化的 NIR 建模方案——全信息竞争自适应高斯建模策略(Full-Information Competitive Adaptive Gaussian Modeling, FICAGM)。首先基于自扩展全信息优化策略构建预处理方案空间,并结合蒙特卡洛交叉验证对预处理组合进行自动寻优与优选;随后引入竞争性自适应重加权采样策略,并通过指数函数对波长变量进行逐轮淘汰;最后利用高斯过程回归的贝叶斯非参数回归与核函数学习完成定量预测,并通过 R^2 与预测均方根误差 RMSEP 等指标对模型的预测性能与稳健性进行综合验证。该策略在建模过程中能够实现各环节的协同优化,从而系统挖掘煤质参数的最优预测路径。

1 算法原理

FICAGM 是一种面向煤炭近红外光谱的系统化建模策略,旨在实现对煤质参数的可靠预测。该策略以“适者生存”为设计理念,通过协同作用的智能化机制高效利用光谱信息:全信息优化机制自适应调节光谱以突出有效信息;动态竞争与衰减机制引导特征变量优胜劣汰,形成与预测目标高度相关的子集;在此基础上通过概率回归推演机制实现精确预测并量化不确定性。整个闭环流程不仅有效降低了数据维度,还提升了模型在当前数据条件下的解释能力与预测稳定性。

1.1 FICAGM 工作原理

FICAGM 由自动光谱预处理、特征变量筛选和回归建模三部分构成,其流程如图 1 所示。具体而言,首先,对原始光谱数据进行多次随机放回抽样,并结合预处理方法全信息生成器(Full-Information Generator, FIG)与蒙特卡罗交叉验证(Monte Carlo Cross-Validation, MCCV),实现光谱预处理方案的优选。其次,将数据再次抽样划分为校正集和验证集,并在校正集中应用所选预处理方案。经过预处理的数据再通过竞争性自适应重加权采样(Competitive adaptive reweighted sampling, CARS)进行特征变量筛选,并基于高斯过程回归(Gaussian Process Regression, GPR)建立预测模型。最后,利用验证集数据,通过均方根误差(RMSE)和决定系数(R^2)等指标对模型性能进行评估,定量地检验模型的预测精度和稳健性,从而形

成一套闭环、系统化的建模策略。所使用的评估指标定义如式(1)、式(2)所示:

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

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

式中: y_i, \hat{y}_i 分别为验证集中第*i*个样本的真实观测值与模型预测值; \bar{y} 为真实均值; n 为测试集样本量。

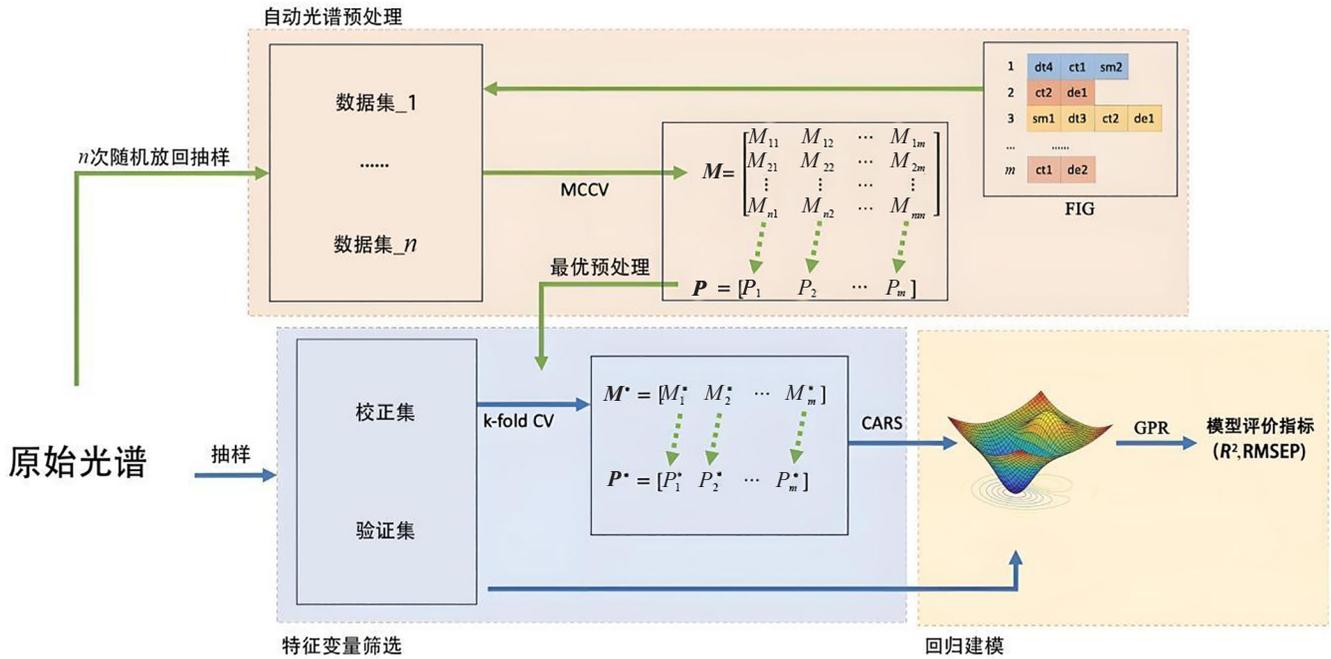


图 1 FICAGM 工作原理图

Fig. 1 FICAGM working principle diagram

1.2 FIG 原理

FIG 是自扩展全信息优化策略 (Self-Extending Full-Information Optimization Strategy, SFIOS)^[15] 的核心模块之一, FIG 旨在生成光谱数据预处理方案的全信息, 从而避免重复和无效的数据预处理^[1]。FIG 的实现包括生成初始方案、生成预处理序列、扩展预处理方法、计算模型性能和获取最优预处理组合 5 个步骤, 其原理如图 2 所示。具体流程为: 首先生成光谱数据预处理初始方案, 其中“1”表示该类别被激活并纳入后续计算, “0”表示未选中, SM、DF、NM、CT 和 DE 分别表示平滑、微分、归一化、校正和去趋势 5 种光谱数据预处理类型, 这些数据预处理类型对应的具体预处理算法如表 1 所示。随后根据全因子实验设计原理, 构建计算序列附加阶段, 即产生不同执行顺序的预处理方案。然后再按照全排列原理, 加入每类预处理方案的具体算法 (实际参数设置如表 1 所示)。最后采用 MCCV 进行预处理算法有效性评估, 并根据性能统计指标对所有方案进行综合排序, 以确定最优预处理策略。

1.3 CARS 算法

CARS 是一种基于回归系数权重的迭代变量筛选策略, 通过多轮抽样与加权选择, 从高维数据中动态

提取与预测目标最相关的特征子集^[16-18]。该方法借鉴自然选择中的“优胜劣汰”机制, 将各波长视为一个候选个体, 通过多轮筛选逐步剔除冗余变量。其核心是依据回归系数的大小来评估波长重要性, 第*k*个波长在第*j*次采样中的重要性权重 w_k^j 为:

$$w_k^j = \frac{\beta_k^j}{\sum_{i=1}^m |\beta_i^j|} \quad (3)$$

式中: β_k^j 为第*j*次采样所建偏最小二乘 (PLSR) 模型中第*k*个波长的回归系数; m 为变量总数。同时, 该方法通过引入指数衰减函数 (EDF) 来动态调节每一轮的波长保留比例。第*j*次采样时保留的变量比例 r_j 为:

$$r_j = \mu e^{-kj} \quad (4)$$

式中: μ 和*k*为衰减参数, 用于在采样初期保留较多变量, 在采样后期保留较少变量。

在建模过程中, CARS 通过上述自适应加权和采样操作, 优先保留那些权重 w_k^j 较大的波长。每次采样后, 基于新的变量子集建立 PLSR 模型, 并通过选择最小的五折交叉验证均方根误差 (RMSECV) 值对应的变量组合作为最优特征子集, 从而提升模型的解释能力与预测性能。

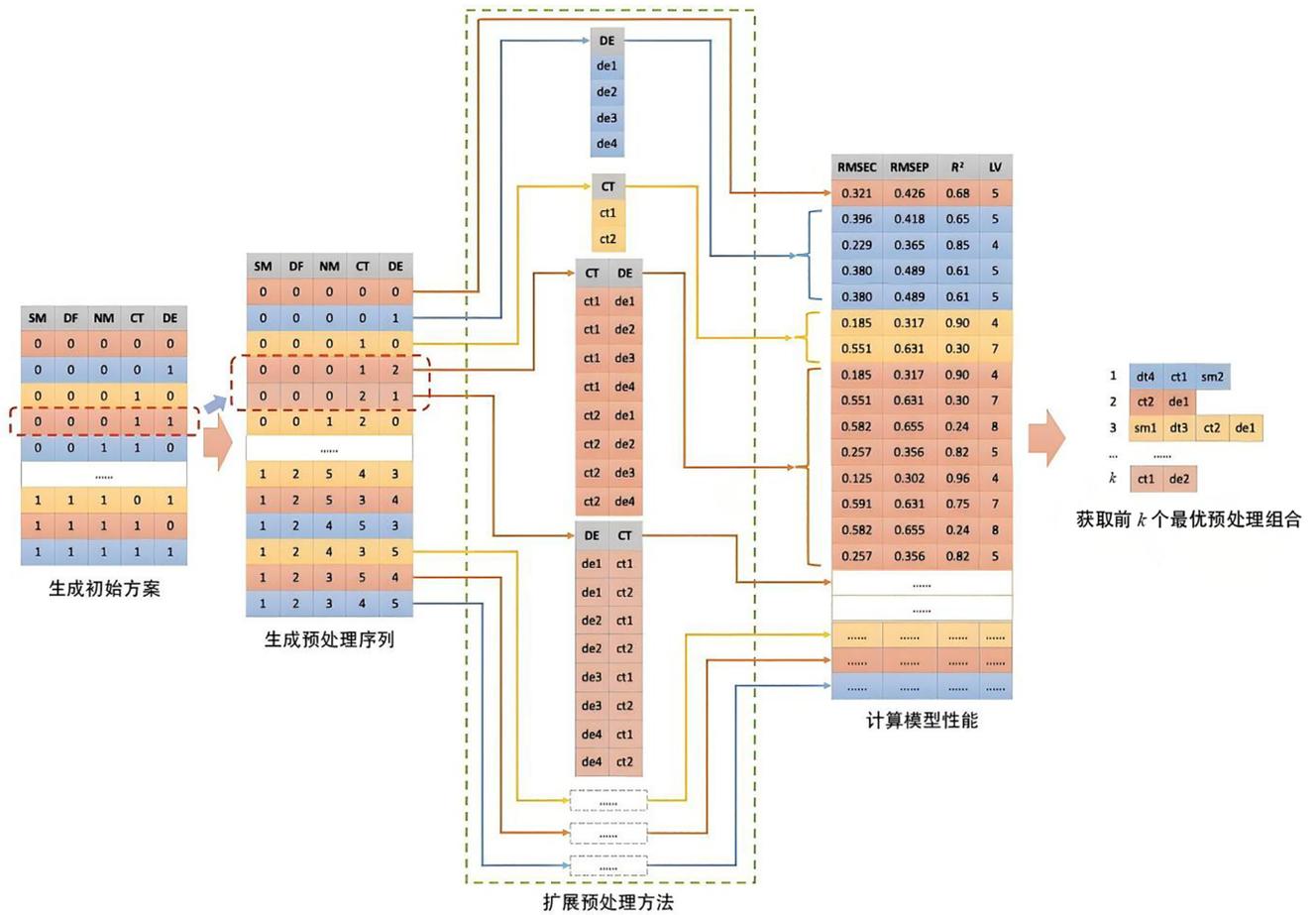


图2 FIG工作原理图

Fig. 2 Working principle diagram of FIG

表1 SFIOS在默认配置下对应的光谱预处理方法参数设置

Tab.1 The parameter settings of the spectral preprocessing method corresponding to SFIOS under the default configuration

预处理类别	预处理方法	参数设置			缩写编号
		K	F	M	
SM	Savitzky Golay 平滑滤波器, 多项式阶数 K, 窗口宽度 F	2	7		sm1
		3	7		sm2
		2	15		sm3
		3	15		sm4
		2	21		sm5
		3	21		sm6
DF	Savitzky Golay 微分器, 多项式阶数 K, 窗口宽度 F, 微分次数 M	2	7	1	df1
		2	15	1	df2
		2	21	1	df3
		2	7	2	df4
		2	15	2	df5
		2	21	2	df6
		3	7	1	df7
		3	15	1	df8
		3	21	1	df9
		3	7	2	df10
		3	15	2	df11
		3	21	2	df12

续表

预处理类别	预处理方法	参数设置			缩写编号
		K	F	M	
CT	多元散射校正(MSC)				ct1
	变量标准化(SNV)				ct2
NM	单位向量归一化				nm1
	范围归一化				nm2
DE	去趋势,多项式阶数K	1			de1
		2			de2
		3			de3
		4			de4

1.4 GPR算法

高斯过程回归(Gaussian Process Regression, GPR)是一种基于贝叶斯理论的非参数回归算法,通过将目标函数建模为高斯过程并利用核函数刻画样本在输入空间的相似性,从而实现预测与不确定性量化^[19]。相比传统方法,GPR在小样本、高维非线性问题中具备灵活性和可解释性,并可提供置信区间,为模型可靠性评估与工程应用提供重要支持。

GPR计算时,首先假设待定的输入变量与输出变量的函数关系 $y=f(x)$ 服从高斯过程先验 $GP(m(x),k(x,x'))$,其中 $m(x)$ 为均值函数,通常取值为0; $k(x,x')$ 为核函数,用于描述点间相关性。接着构造训练核矩阵式(5),并加入噪声项 $K_y = K + \sigma_n^2 I$ 。

$$K = \begin{bmatrix} k(x_1, x_1) & \cdots & k(x_1, x_n) \\ \vdots & \ddots & \vdots \\ k(x_n, x_1) & \cdots & k(x_n, x_n) \end{bmatrix} \quad (5)$$

计算新样本预测分布时,对于一个新输入 x_* ,按照下面的式子计算训练点与新点的协方差向量 $k_* = [k(x_*, x_1), k(x_*, x_2), \dots, k(x_*, x_n)]^T$,并计算预测输出的均值和方差分别为:

$$\bar{y}_* = k_*^T K_y^{-1} y \quad (6)$$

$$Var[y_*] = k(x_*, x_*) - k_*^T K_y^{-1} k_* \quad (7)$$

对多个新样本 $X_* = \{x_*^1, x_*^2, \dots, x_*^m\}$,只需构造对应的核矩阵 K_{**} 和交叉协方差矩阵 K_* 即可得到联合预测分布:

$$y_* \sim N(K_*^T K_y^{-1} y, K_{**} - K_*^T K_y^{-1} K_*) \quad (8)$$

2 实验材料与方法

2.1 样品来源及制备

实验煤样来自东北地区若干电厂的入炉燃料管道,使用前先将煤样破碎,保持尺寸在0.2 mm以下,每个样品质量约500 g,共计133个原始煤样。按照GB/T213—2008执行标准,手工分析得到其干燥基高

位发热量为10.93~24.80 MJ/kg。在提取煤粉的近红外光谱前,将样品陈放在实验室中2~3 h,使其温度均衡;装样时,采用压片使样品表面基本保持平整。所有样本均采用反射率模式采集其近红外光谱,光谱范围为1000~2500 nm。本研究的分析对象为煤炭关键质量指标——发热量。

2.2 NIR光谱采集方法

通过美国Thermo公司生产的漫反射型傅立叶变换近红外光谱仪Antaris进行光谱采集,检测器和光源分别为铟镓砷(In Ga As)检测器与卤钨灯,测试中采用积分球扫描附件及旋转封闭式样品杯。采集在温度(20±2)℃、相对湿度(50±5)%条件下进行。每个样品扫描32次,分辨率为0.4 cm⁻¹,扫描范围为1000~2500 nm。每个样品分别进行3次重复采集,共获得399组光谱数据;对同一样品的3次光谱取平均后,最终得到133组煤粉的NIR光谱数据。

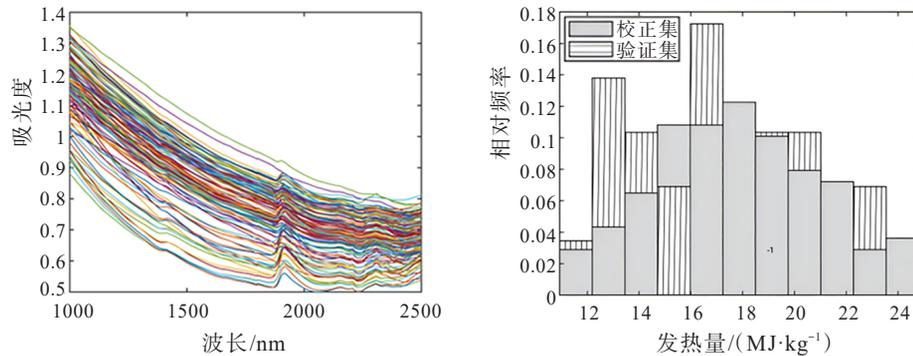
3 实验结果分析

3.1 参数设置

本研究采用SPXY(Sample set Partitioning based on joint X-Y distances)方法对数据集进行样本划分,其中校正集包含110个样本,验证集包含23个样本。图3(a)给出了133个样本的原始近红外反射光谱曲线,可用于直观表征样本间光谱形态差异及整体谱形特征。图3(b)展示了校正集与验证集在发热量取值区间上的分布直方图,用于说明两子集在目标变量覆盖范围与分布形态上的差异,为后续模型训练与结果对比分析提供数据分布背景。从图3(a)可见,原始光谱整体呈非线性衰减趋势,并伴随散射效应引起的轻微尺度变化及明显的基线漂移。基于上述光谱特性,FIG在运行过程中选用CT、DE和DF这三类预处理方法;CARS的参数设置为:蒙特卡洛采样次数为50、交叉验证折数为5、最大主成分数为

15,其余变量选择标准与样本排序方式采用默认值; GPR的参数设置为:核函数采用RBF(平方指数核),

基函数为常数项,近似方法与优化选项均采用默认设置。



(a) 133个样本的原始近红外反射光谱图 (b) 校正集与验证集在发热量取值区间上的分布直方图

图3 原始近红外反射光谱与煤炭发热量分布特征

Fig. 3 Raw near-infrared reflectance spectra and distribution of coal calorific value

3.2 预处理方法优选结果

CT、DE和DF三类光谱预处理方法通过组合与排列共形成755种预处理方案。基于RMSECV中位数排序,图4给出了性能最优的前50种方案与性能最差的后50种方案在RMSECV、校正均方根误差(RM-

SEC)及RMSEP指标上的统计结果。结果表明,前50种方案的模型性能差异较小,尤其是排名靠前的若干方案,其各项误差指标高度接近;并且,随着方案排名后移,模型误差整体呈上升趋势,导致后50种方案的预测性能明显下降。

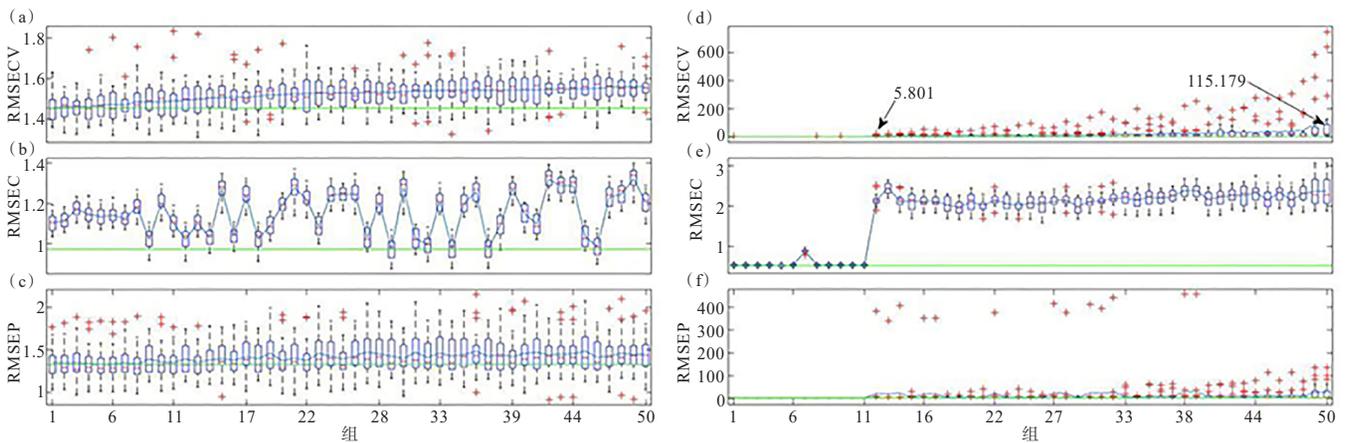


图4 前50种(a-c)与后50种(d-f)光谱预处理方案性能统计

Fig. 4 Performance statistics of the top 50 (a-c) and the bottom 50 (d-f) pretreatment schemes

为分析预处理顺序对模型性能的影响,图5统计了前10种与后10种方案中各预处理方法在不同计算步骤中的出现频次。结果显示,在高性能方案中,DF与DE更倾向于出现在前两步,而CT主要位于最后一步;相反,在低性能方案中,一旦将CT或DE操作前置、DF步骤后置,模型的误差便会显著增大。上述结果表明,预处理顺序对模型性能具有重要影响,在本研究条件下,“先DE/DF、后CT”的预处理顺序更有利于提高建模稳定性与预测精度。

图6给出了FICAGM筛选得到的部分代表性预处

理组合及其对应的光谱结果。由图6(a)可见,经de4→df2→ct2处理后的谱线在特征吸收区(如约1897 nm附近)峰谷结构清晰,同时整体噪声水平较低。尽管图6(b)与图6(a)的整体谱形相近,但其在短波端与长波端均出现更明显的高频波动,并且在1000~1750 nm区间仍存在轻微的基线漂移。df11对应的是二次微分算法,该算法虽能有效消除光谱中的线性基线,但也会放大整个谱区的噪声,如图6(c)所示。经df9→de3→ct1处理后,谱线的关键吸收特征未被有效突显,反而引入了额外的高频扰动,具体见图6(d)。经ct1

→de3→df7 与 de3→ct1→df10 处理后, 谱线的高频波动显著增强, 部分吸收特征被削弱(见图 6(e, f)), 主要表现为谱线波动幅度增大、有效特征弱化, 并局部出现异常尖峰或剧烈振荡, 这严重影响后续建模的稳定性。

综合比较不同预处理组合的光谱平滑性、噪声水平及有效特征保留情况, de4→df2→ct2 的预处理组合在抑制噪声与增强关键信息之间取得了较优平衡, 因此被选为 FICAGM 的最优光谱预处理方案。

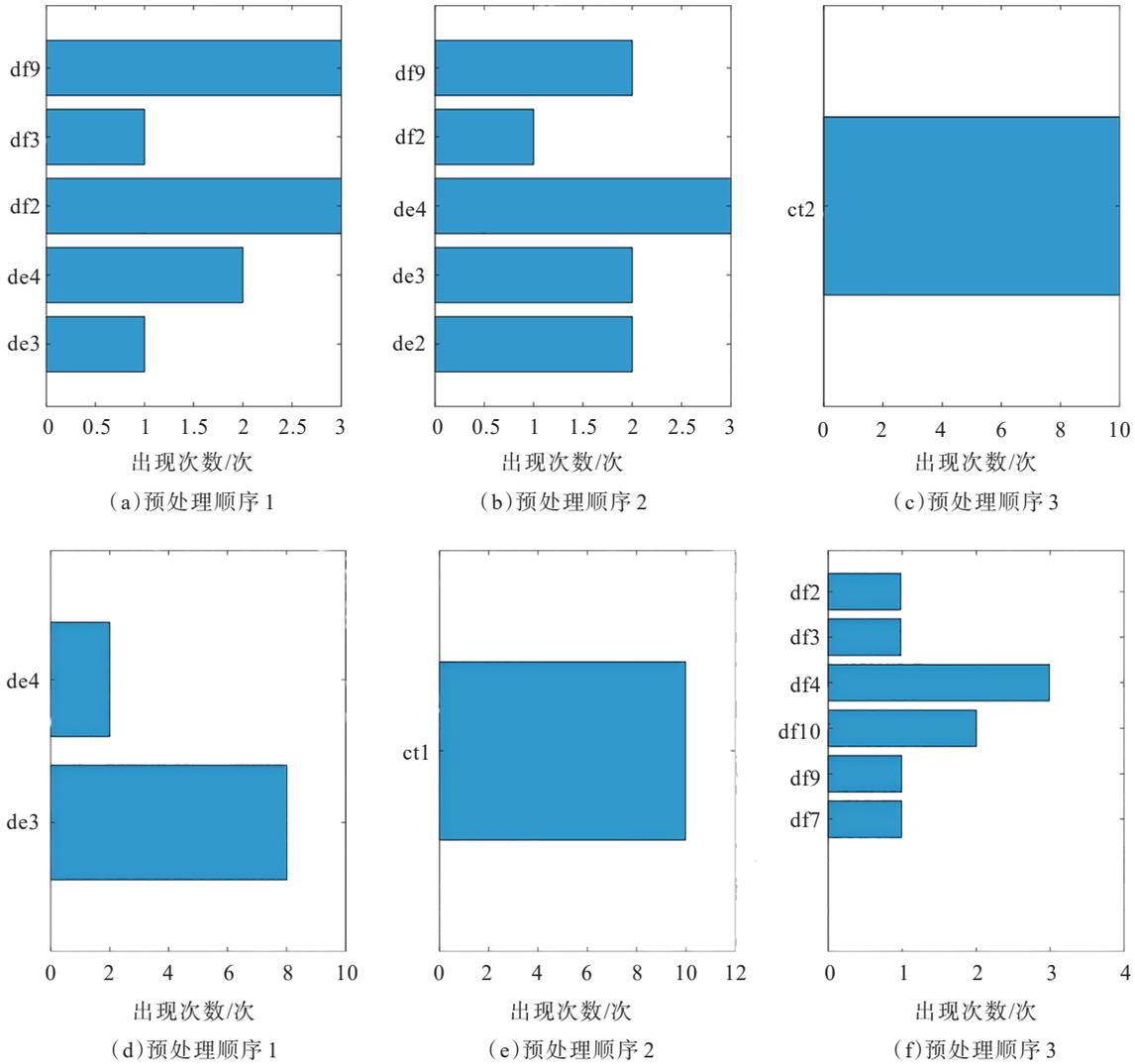
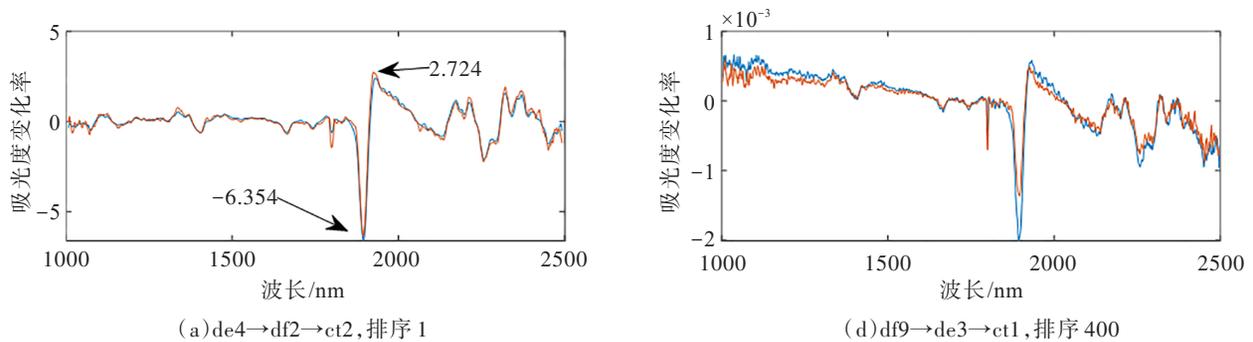


图 5 前 10 种 (a-c) 与后 10 种 (d-f) 光谱预处理方案中采用的具体预处理方法计算顺序统计结果

Fig. 5 The statistical results of the calculation sequence of the pretreatment methods adopted in the first 10 (a-c) and the last 10 (d-f) spectral pretreatment schemes



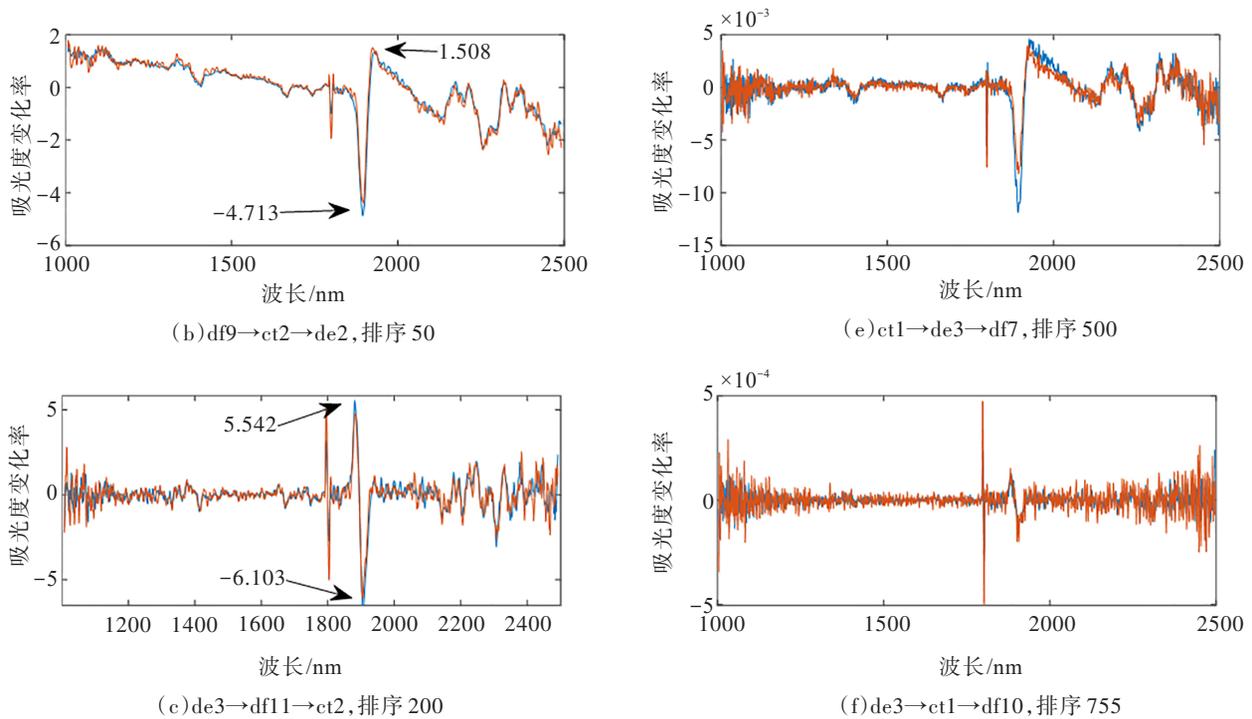


图6 FICAGM选择的部分预处理方法

Fig. 6 Some of the pretreatment methods selected by FICAGM

3.3 特征选择结果

图7展示了基于CARS算法的变量筛选结果。由图7(a)可知,在EDF的作用下,变量数量在前10次采样中迅速下降,表明算法在初期的“粗选”阶段会快速剔除大量冗余变量。这些冗余变量主要是和发热量无关的波长变量,其参与建模会导致RMSECV值增大。随后变量下降速度明显减缓,表明算法进入“精选”阶段,该阶段更侧重于保留具有较高预测能

力的核心变量。图7(c)通过回归系数的变化进一步揭示了变量筛选过程中各波长点重要性的动态调整。最终,根据图7(b)中RMSECV在第22次采样达到最小值所对应的最优采样轮次,并结合图7(a)在该轮次下的入选变量数,确定最优变量子集由67个关键变量组成。因此,该算法能够在最大限度降低模型复杂度的同时,确保对煤炭发热量预测的准确性与稳健性。

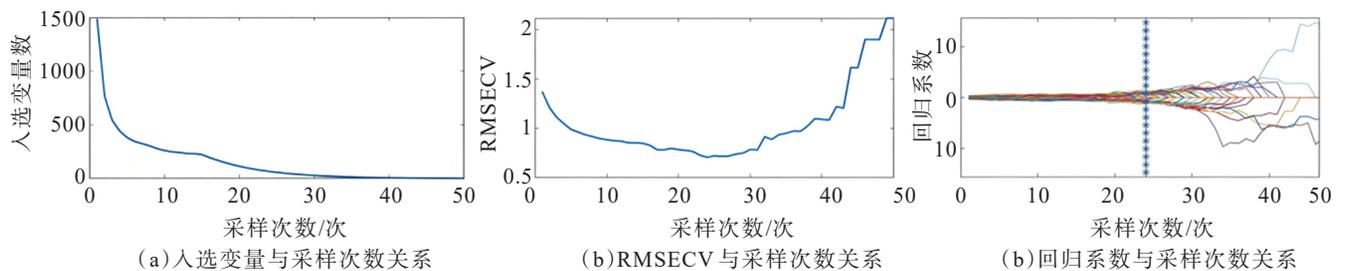


图7 基于CARS算法的关键变量筛选图

Fig. 7 Key variable selection using the CARS algorithm

图8展示了经SFIOS预处理后,采用CARS算法对光谱变量进行筛选的结果。曲线为校正集中第1个样本的光谱,星号(☆)标记的波长为CARS算法最终保留的特征波段。所选波长主要分布在1450 nm、1900 nm及2100~2300 nm等典型吸收区,这些谱区与水分子及含氢官能团的吸收特性密切相关,因而对煤炭的发热量具有较强的表征能力。图8表明,基于

CARS算法的特征提取能将光谱变量有效聚焦于一组“信息密集波段”。这些波段与煤样组成变化密切相关,吸收机理明确,能够为后续模型精度的提升提供具有可解释性的波段依据,从而避免了在全谱范围内进行随机或均匀筛选的不足。被剔除的波长多位于吸收较弱或信噪比较低的区域,其光谱变化主要受背景噪声、散射效应及无关组分的影响,难以提供与发

热量相关的有效信息。剔除这些冗余变量后,模型最终保留下来的均为与样品化学组成及能量特性高度相关的有效波段。

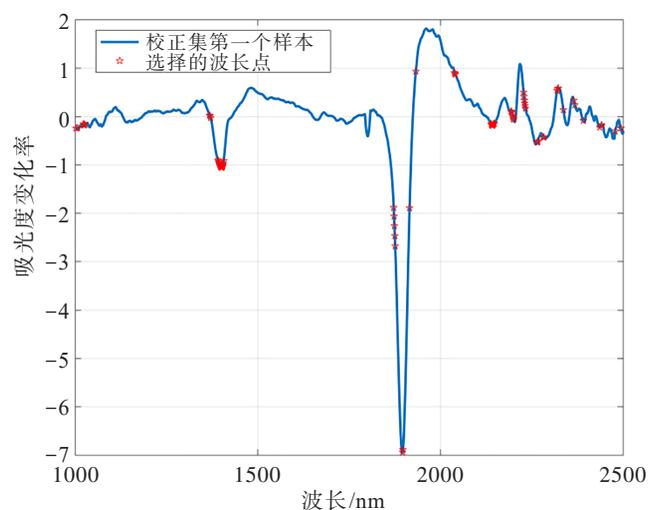


图8 CARS算法选择的波长变量

Fig. 8 Wavelength variables selected by the CARS algorithm

3.4 回归模型结果

本研究系统比较了不同回归算法与特征选择策略在煤炭近红外发热量建模中的性能表现。采用的回归方法包括偏最小二乘回归(PLS)、支持向量回归(SVR)和高斯过程回归(GPR),分别代表线性建模与两类典型非线性建模范式;特征选择方法涵盖主成分分析(PCA)、逐步投影算法(SPA)及竞争性自适应重加权采样(CARS)。通过构建覆盖线性与非线性建模范式的多种算法组合,系统评估了不同特征选择策略与回归算法对模型预测性能的影响。

所有实验均在MATLAB环境下完成。PLS在校正集上采用5折交叉验证来确定最优潜在变量数,并限制最大潜在变量数为30;SVR采用径向基核函数(RBF),惩罚因子设为1,核尺度自动估计;GPR采用平方指数核函数,相关超参数由系统自动优化,其余参数均采用默认设置,以保证实验的稳定性与可复现性。

在特征选择方面,PCA作为线性降维基准方法,通过削弱噪声与共线性确定主成分数量;SPA通过逐步正交投影降低波长间相关性,从而保留少量具有代表性的特征变量;CARS则结合蒙特卡洛采样与回归权重的竞争性筛选机制,自适应提取与发热量最相关的特征区间,并在变量压缩与预测性能之间实现较优平衡。基于上述配置,本文构建了多组模型组合进行对比分析,其预测结果汇总于表2。

为定量评估各建模方法的拟合与预测性能,本研

究采用RMSEC、RMSEP和决定系数 R^2 作为主要评价指标。其中,RMSEC与RMSEP分别用于衡量模型在校正集、验证集上的平均预测误差,误差值越小,表明模型预测精度越高; R_c^2 与 R_v^2 分别表示模型在校正集、验证集上的决定系数,数值越接近1,表明模型对目标变量的解释能力越强。表2表明,原始光谱条件下不同算法的性能差异显著;而引入适当的特征选择方法后,多数模型的预测精度与稳健性均得到提升。其中,本文提出的FICAGM方法在校正集和验证集上均表现出最优或接近最优的综合性能。

表2 FICAGM与其他方法的模型性能比较

Tab. 2 Performance comparison of FICAGM with other modeling method

建模方法	特征选择	校正集		验证集	
		RMSEC	R_c^2	RMSEP	R_v^2
原始光谱 + PLS		0.890	0.931	1.153	0.873
原始光谱 + SVR		1.836	0.700	1.275	0.772
SVR + PCA	PCA	1.470	0.790	1.386	0.868
SVR + SPA	SPA	1.512	0.781	1.185	0.897
SVR + CARS	CARS	0.887	0.930	0.937	0.907
原始光谱 + GPR		1.732	0.733	0.866	0.895
GPR + PCA	PCA	1.293	0.840	1.085	0.890
GPR + SPA	SPA	1.526	0.777	0.946	0.916
FICAGM	CARS	0.856	0.929	0.909	0.920

注:表中CARS及FICAGM的性能指标均来自10次独立运行中RMSEP最小(或 R^2 最高)的结果。由于CARS算法具有随机采样特性,通过T检验分析了各次运行结果之间的差异。

以原始光谱结合PLS的模型作为基线,其在校正集上具有较好的拟合能力(RMSEC=0.890, $R_c^2=0.931$),但在验证集上的预测误差明显增大(RMSEP=1.153, $R_v^2=0.873$),表明模型存在一定的过拟合风险。原始光谱结合SVR的模型在校正集与验证集上均表现较差,表明在未进行特征选择的情况下,SVR对高维冗余光谱特征的适应性不足,整体呈现欠拟合特征。

在引入PCA或SPA进行特征选择后,SVR模型的预测性能有所改善,但其验证集指标仍普遍优于校正集,表明模型对训练数据的拟合能力有限,欠拟合问题尚未完全解决。相比之下,结合CARS特征选择后,SVR模型性能显著提升(RMSEP=0.937, $R_v^2=0.907$),校正集与验证集性能差距明显缩小,说明有效变量筛选可以显著增强模型的稳定性与泛化能力。对于GPR模型,直接基于原始光谱建模时,其在校正集上的拟合能力有限(RMSEC=1.732, $R_c^2=0.733$),但在验

证集上却取得了相对较好的预测结果(RMSEP=0.866, $R_v^2=0.895$)。需要指出的是,在小样本条件下,即使采用SPXY方法进行样本划分,校正集与验证集在目标变量分布上仍可能存在一定差异,这种差异可能导致个别情况下验证集误差低于校正集误差。GPR结合PCA后,模型性能有所下降,说明线性降维可能削弱了GPR所需的关键信息;而GPR+SPA虽在验证集上取得较高的预测精度 $R_v^2(0.916)$,但其校正集拟合能力明显不足($R_c^2=0.777$),模型稳定性依然较差。相比之下,FICAGM在校正集与验证集上均表现出较优且均衡的预测性能(RMSEC=0.856, $R_c^2=0.929$; RMSEP=0.909, $R_v^2=0.920$)。以原始光谱+PLS为基线模型,FICAGM使验证集的RMSEP降低了约21.2%, R_v^2 提升了约5.4%,表明该方法在降低预测误差的同时,有效增强了模型的泛化能力与预测稳健性。

综上所述,FICAGM在各类建模方法中的表现最为优异。该模型不仅克服了传统PLS模型预测稳健性不足的问题,也避免了GPR与SPA结合时模型稳定性欠佳的缺陷,从而进一步证实了CARS特征选择与GPR建模相结合的有效性。基于本文的实验条件与数据范围,FICAGM可作为预测煤炭发热量的一种最优建模方案。

4 结论

本文提出的FICAGM方法,通过光谱预处理、竞争性变量筛选与高斯过程回归3个步骤的协同,实现了对煤炭发热量的快速、无损及稳健预测。实验结果表明,该方法在显著降低维度的同时提升了预测精度,性能优于传统回归方法,且所筛选的关键波段具备良好的物理可解释性,可适用于燃料质量在线监控与燃烧调度辅助决策。但需要指出的是,该方法的泛化能力受到样本来源与测量设备的区域局限性制约,后续工作需通过跨厂区样本验证、多模态数据融合以及模型轻量化等途径推动其实用化。尽管FICAGM在煤炭发热量近红外建模中表现突出,但其在更广泛场景下的适用性仍需进一步系统验证。由于本研究模型基于有限样本构建,当煤样产地或煤阶结构发生显著变化时,有必要引入更具代表性的样本,对模型进行持续验证与适应性更新,以进一步评估和提升其在跨样本条件下的泛化潜力。

参考文献

[1] 卜昌森. 煤炭市场三次大幅下行情况的差异[J]. 中国煤炭工业, 2025(9): 32-34.

- [2] 张冬练, 王庆松, 荏方, 等. 煤炭成分对煤炭发热量的影响研究[J]. 能源与环保, 2025, 47(5): 203-207.
- [3] 侯娟萍. 煤炭成分对其发热量的影响研究[J]. 山西化工, 2025, 45(1): 148-150.
- [4] 李金金. 煤化验室发热量测定质量控制措施[J]. 内蒙古煤炭经济, 2025(13): 103-105.
- [5] 赵积坤. 煤炭中灰分及挥发分对发热量的影响研究[J]. 化工管理, 2025(1): 63-65.
- [6] 贾鹏. 分析煤质化验对提高煤炭质量的作用[J]. 内蒙古煤炭经济, 2024(16): 106-108.
- [7] CHEN J L, HE Y L, LIANG Y X, et al. Estimation of gross calorific value of coal based on the cubist regression model[J]. Scientific Reports, 2024, 14: 23176.
- [8] MUNSHI T A, JAHAN L N, HOWLADAR M F, et al. Prediction of gross calorific value from coal analysis using decision tree-based bagging and boosting techniques [J]. Heliyon, 2023, 10(1): e23395.
- [9] 冯晓涵, 杜明生, 崔柳, 等. 煤质快速检测分析技术发展与应用[J]. 煤炭加工与综合利用, 2025(8): 116-122.
- [10] 丁友智, 刘峰, 罗玉琳, 等. 红外光谱技术在煤质检测分析中的应用[J]. 中国石油和化工标准与质量, 2025, 45(6): 38-40.
- [11] 梁南山, 梁瑜祯. 某厂部分入厂煤近红外光谱特性的应用研究[J]. 燃料与化工, 2025, 56(2): 4-8.
- [12] 张海军. 煤炭发热量的近红外光谱定量检测研究[J]. 煤炭加工与综合利用, 2025(2): 88-91.
- [13] 徐鼎, 闵红, 郭升阳, 等. 近红外光谱结合连续小波变换-卷积神经网络判定煤炭类别[J]. 冶金分析, 2024, 44(10): 45-54.
- [14] 雷利维, 闫海军, 屈晓渊, 等. 煤质近红外光谱分析中数据挖掘技术的应用[J]. 数字技术与应用, 2024, 42(2): 154-156.
- [15] WANG S H, LIN M M, MENG Y H, et al. Self-expansion full information optimization strategy: Convenient and efficient method for near infrared spectrum auto-analysis [J]. Spectrochimica Acta Part A, Molecular and Biomolecular Spectroscopy, 2023, 303: 123224.
- [16] 王正新, 何凌阳. 全信息变权缓冲算子的拓展、优化及其应用[J]. 控制与决策, 2019, 34(10): 2213-2220.
- [17] 母雯竹, 张贵宇, 张维, 等. 基于CARS-SPA特征提取的黄水淀粉近红外光谱定量模型优化[J]. 食品科学, 2024, 45(19): 8-14.
- [18] 付博, 杨永锋, 刘向真, 等. 数据集划分及预处理方法对烟叶化学成分近红外定量模型的影响[J]. 河南农业大学学报, 2025, 59(3): 516-527.
- [19] 马鑫鹏, 陈宇, 崔志成, 等. 高斯过程回归误差标定辅助的室内惯性测量单元与超宽带融合定位算法研究[J]. 电子与信息学报, 2025, 47(10): 3772-3782.

作者简介:李赛赛,男,硕士研究生,主要研究方向为机器学习

在光谱信息中的处理与应用。E-mail:1300667287@qq.com
通信作者:佟鹏,男,副教授,博士,主要研究方向为机器学习在
光谱信息中的处理与应用。E-mail:tptong@163.com
通信作者:王圣毫,男,高级工程师/讲师,博士,主要研究方向

为机器学习在光谱信息中的处理与应用。E-mail:wangsheng-
hao@zut.edu.cn

(编辑 JJL, 校对 ZDC)