

污水处理与回用

崔杰, 张波, 王荣建, 等. 基于可解释性机器学习的餐厨垃圾浆液厌氧产气性能与关键影响因素[J]. 净水技术, 2026, 45(2): 89-100, 119.
Cui J, Zhang B, Wang R J, et al. Performance and key influencing factors of anaerobic biogas production for kitchen wastes leachate based on interpretable machine learning [J]. Water Purification Technology, 2026, 45(2): 89-100, 119.

基于可解释性机器学习的餐厨垃圾浆液厌氧产气性能与关键影响因素

崔杰^{1,2}, 张波^{2,*}, 王荣建³

(1. 上海交通大学中英国际低碳学院, 上海 200240; 2. 上海交通大学环境科学与工程学院, 上海 200240; 3. 上海东石塘再生能源有限公司, 上海 201400)

摘要 【目的】餐厨垃圾浆液厌氧消化系统因进料特征复杂、多因素动态耦合, 使得其预测与调控具有较大挑战。【方法】为实现工业规模厌氧反应过程的精准管理与优化, 本文收集 2019 年—2023 年上海某工业规模厌氧反应器运行数据, 分别利用随机森林、梯度提升、K 邻近回归与人工神经网络(ANN)建立产气预测模型, 并通过沙普利加和解释(SHAP)算法量化各输入变量对产甲烷效率的贡献与抑制作用。【结果】梯度提升模型在测试集中 R^2 最高, 均方误差(MSE)与平均绝对误差(MAE)最低, 预测精度与稳健性优于 K 邻近回归、随机森林以及 ANN 模型。结合最优模型的 SHAP 分析发现: 进水化学需氧量(COD)较高时易引发酸化失稳; 氨氮质量浓度约 1 500 mg/L 时有利于微生物活性, 但过高则抑制产气; pH 值低于 6.2 时, 产气效率显著下降。【结论】可解释性机器学习方法可有效提升沼气产量预测的准确性, 并帮助快速定位与调控核心影响因素, 实现对餐厨垃圾厌氧消化系统的精准管理与持续优化, 对其规模化推广具有重要意义。

关键词 厌氧消化 餐厨垃圾 机器学习 性能预测 沙普利加和解释(SHAP)

中图分类号: X703 文献标志码: A 文章编号: 1009-0177(2026)02-0089-13

DOI: 10.15890/j.cnki.jsjs.2026.02.009

Performance and Key Influencing Factors of Anaerobic Biogas Production for Kitchen Wastes Leachate Based on Interpretable Machine Learning

Cui Jie^{1,2}, Zhang Bo^{2,*}, Wang Rongjian³

(1. China-UK Low Carbon College, Shanghai Jiao Tong University, Shanghai 200240, China;

2. School of Environmental Science and Engineering, Shanghai Jiao Tong University, Shanghai 200240, China;

3. Shanghai Dongshitang Renewable Energy Co., Ltd., Shanghai 201400, China)

Abstract [Objective] The anaerobic digestion system for kitchen waste leachate faces great challenges in prediction and regulation due to the complex feeding characteristics and dynamic coupling of multiple influencing factors. [Methods] To realize the precise management and optimization of the industrial-scale anaerobic reaction process, this paper collected the operational data of an industrial-scale anaerobic reactor in Shanghai from 2019 to 2023. Methane production prediction models were established using random forest, gradient boosting, K-nearest neighbor regression and neural network, respectively. Furthermore, the Shapley additive explanations (SHAP) algorithm was adopted to quantify the contributory and inhibitory effects of each input variable on methanogenic efficiency. [Results] The gradient boosting model achieved the highest R^2 value, along with the lowest mean square error (MSE) and mean absolute error (MAE) in the test set, exhibiting superior prediction accuracy and robustness over the K-nearest neighbor

[收稿日期] 2025-01-06

[作者简介] 崔杰(1999—), 女, 硕士研究生, 研究方向为餐厨垃圾厌氧消化机制及优化, E-mail: cui_jie@sjtu.edu.cn。

[通信作者] 张波, 女, 研究员, E-mail: zhangbo214@sjtu.edu.cn。

regression, random forest and neural network models. SHAP analysis based on the optimal model revealed that a high chemical oxygen demand (COD) of influent water tended to induce acidification instability; an ammonia nitrogen mass concentration of approximately 1 500 mg/L was conducive to microbial activity, while an excessively high concentration inhibited methane production; and the methane production efficiency decreases significantly when the pH value was lower than 6.2. [**Conclusion**] The interpretable machine learning method can effectively improve the accuracy of biogas production prediction and help quickly identify and regulate the core influencing factors, thus enabling the precise management and continuous optimization of the anaerobic digestion system for kitchen waste. This study holds important significance for the large-scale popularization of the system.

Keywords anaerobic digestion kitchen waste machine learning performance prediction Shapley additive explanation (SHAP)

随着社会经济的发展和人们生活水平的提高,餐厨垃圾的产生量日益增加。餐厨垃圾浆液主要来源于家庭厨房及其他食品加工行业。餐厨垃圾浆液中的有机物含量和含水率高,资源化与能源化利用潜力大。目前,厌氧消化是处理餐厨垃圾的主要技术途径之一^[1]。通过微生物的协同代谢过程,餐厨垃圾中的有机物能够被转化为甲烷等清洁能源,这对于能源转型和实现碳中和具有重要意义。然而,厌氧消化过程本身高度复杂,涉及多种微生物群落与物理化学过程,且受温度、pH、负荷率等多种环境因素的动态变化影响^[2-3]。传统的人工调控依赖于经验,往往耗时且难以准确把握关键节点。另外,经典的厌氧消化模型(如 ADM1)在大规模应用中由于模型参数众多、对基质特征要求严苛而难以实现准确预测^[4]。因此,在对厌氧消化系统进行过程控制与优化时,亟需一种能够有效处理多维数据、具备良好泛化性能的方法。

近年来,机器学习由于在处理高维度、非线性及动态数据方面的优势,逐渐成为厌氧消化过程监测与预测的重要技术工具^[5]。石悦等^[6]用人工神经网络(ANN)耦合遗传算法对规模化的牛粪厌氧消化工厂进行甲烷产量模拟及优化。Cai 等^[7]使用各种机器学习算法的分类模型,通过微生物群落组成预测底物类型。Clereq 等^[8]开发了一个在线网络工具,主要使用回归模型来估计沼气产量和相应的收入。Wang 等^[9]对比了随机森林、GLMNET、支持向量机以及 K-邻近(KNN)算法 4 种算法,最后认为在回归模型中,KNN 表现出最佳预测精度。但是,已有研究多集中于实验室规模或中试规模数据,对工业大规模厌氧消化系统的泛化能力仍有待提高。此外,除少数最新研究外,大多数应用于厌氧消化的机器学习模型都缺乏可解释性分析,难以量化各输

入特征(如进料特征、操作参数、微生物信息)对模型输出(产甲烷量、沼气产量等)的影响。沙普利加和解释(SHAP)、置换特征重要性及部分依赖图等可解释性方法的缺失,导致对模型性能提升背后的原因难以追溯,也不利于工艺操作人员在实际生产中做出针对性调整和优化。为克服这类“黑箱”局限,未来亟需将模型可解释性手段与机器学习建模深度融合,提升对厌氧消化机理的理解。

本文结合工业实际厌氧消化反应器的运行数据,综合考虑多种运行参数的动态变化,利用机器学习算法建立反应预测模型。与传统“黑箱”模型不同的是,本文在模型构建后进一步融合 SHAP 可解释性分析方法,从而量化各关键因素对于厌氧消化运行状态和产气性能的影响权重。通过这种可解释性机器学习框架,不仅能显著提升对厌氧消化系统的预测准确度,也有助于指导操作人员进行精准调控,提高餐厨垃圾厌氧消化的资源化与能源化利用效率,为工业规模的推广应用提供数据支撑和技术参考。

1 模型和方法

1.1 数据收集与预处理

本文聚焦于某废水处理厂厌氧反应器,收集了该反应器 2019 年—2023 年的进出水运行数据,共包含 1 796 个数据点。经过数据质量检查发现,数据缺失情况主要分为 2 类:一类为监控过程中偶发的个别数据缺失,即在连续监控中偶尔出现的少量间断性缺失;另一类则是设备定期维护而导致的长时间数据缺失。鉴于设备维护期间进水和产气均处于停滞状态,其数据对于建立厌氧产气预测模型贡献有限,因此,予以剔除。剩余空缺数据由线性插值法填补,线性插值法适用于数据缺失较少且缺失位置相对简单的情况^[10]。由于本数据集呈现出较好

的连续性,且同一时期内数据波动较小,采用线性插值法进行补全较为适宜。线性插补的计算如式(1)。

$$y = y_0 + \frac{(y_1 - y_0)}{(x_1 - x_0)} \cdot (x - x_0) \quad (1)$$

其中: x_0, y_0, x_1, y_1 ——已知数据点,且 $x_0 \neq x_1$;
 x ——待插值的点,位于 $x_0 \sim x_1$;
 y ——线性插值计算所得值。

对完成线性插补的数据集进行极端值去除,在机器学习建模过程中,剔除极端值有助于降低由测量误差或异常情况引起的噪声干扰,从而使模型更专注于数据的主要分布,提高模型的稳定性和泛化能力;同时,去除这些异常数据能够改善评估指标的准确性,使得模型性能更真实地反映在大部分数据上的表现^[11]。本文采取四分位距法(IQR)剔除极端值,IQR是一种常用的异常值检测方法,它基于数据的四分位数来识别和剔除极端值,该方法不依赖数据分布的假设,因此,适用于非正态分布的数据集^[12]。其判断标准如式(2)~式(4)。

1) 计算 IQR。

$$I = Q_3 - Q_1 \quad (2)$$

其中: I ——IQR;

Q_1 ——第一四分位数;

Q_3 ——第三四分位数。

2) 确定判断界限。

$$S_1 = Q_1 - 1.5I \quad (3)$$

$$S_2 = Q_3 + 1.5I \quad (4)$$

其中: S_1 ——下界;

S_2 ——上界。

3) 判断异常值:如果数据点小于下界或大于上界,则该数据点被视为异常值。

产气量为 0 及其他变量缺失的数据剔除后,共剩余 306 个数据点。对原始数据集经过缺失值插补和极端值去除后,剩余 1 570 个有效数据点。每个数据点包括 11 个变量:进水量、进水 pH、出水 pH、进水化学需氧量(COD)、出水 COD、进水氨氮、出水氨氮、出水挥发性脂肪酸(VFA)、进水悬浮物(SS)、出水 SS 和每日生物产气量,其中每日生物产气量作为评价厌氧反应效率的指标作为模型因变量,其余变量为自变量。

1.2 数据特征分析

皮尔逊(Pearson)相关性分析主要用于衡量 2 个连续型变量间的线性关联程度^[13]。其核心思想是通过计算相关系数(-1~1)来评估,在一个变量变动时,另一变量是否随之呈同向或反向的线性变化。若相关系数为正且数值较大,表示两者呈显著正相关,即其中一个变量上升或下降时,另一个变量也相应地同向变化;若相关系数为负且绝对值较高,则意味着存在稳定的负相关关系。当相关系数接近 0 时,说明两者间未能体现明显的线性趋势,通过 Pearson 相关性分析,可以为变量间潜在的交互效应提供定量依据,并为后续建模或机制探讨奠定基础。

本厌氧数据集影响因素多、数据维度高,存在许多冗余或相关性较高的变量,这会影响模型的性能和解释性。主成分分析(PCA)是一种线性降维方法,通过将高维数据投影到低维空间,在降低数据维度的同时尽可能保留其主要特征^[14]。PCA 通过消除变量间的冗余信息和相关性,使得少量的主成分能够代替原有的多变量,实现数据的简化^[15]。

1.3 机器学习模型训练与优化

将经过预处理的数据分为训练集(70%)和测试集(30%),训练集用于模型训练,而测试集则用于评估模型的准确性。本文根据过往文献中厌氧消化机器学习建模所用的算法,并根据数据集样本较小、数据维度高的特点选择了 4 种算法:随机森林、梯度提升(XGBoost)、ANN 和 KNN 来进行数据拟合。本文使用 Python 语言,并利用第三方库如 Numpy、Pandas 等处理数据,使用 Sklearn 库、Scikit-optimize 库来搭建模型和超参数调整。

随机森林属于 Bagging 类算法,通过构建多棵相互独立的二叉决策树,利用随机抽样和随机选取特征子集来进行回归预测,具有高预测准确率、抗噪声和异常值能力,同时不易过拟合^[16]。XGBoost 是一种集成学习方法,其核心原理是采用加法模型的方式迭代优化损失函数,每一轮迭代中,新加入的决策树用于拟合当前模型的残差,并利用梯度下降优化损失函数,从而不断降低预测误差,适用于处理非线性问题,但计算复杂度较高且需调节超参数^[17]。KNN 方法基于距离度量,通过选择最近的 K 个邻居来进行预测,回归任务中取邻居数值的平均值作为预测输出,KNN 无需训练过程,依赖于整个数据集,适合小规模数据,但对噪声和特征尺度敏感^[18]。神

神经网络(NN)模拟生物神经系统,由输入层、隐藏层和输出层构成,输入层接收原始数据,隐藏层进行数据处理和特征提取,输出层产生最终结果。每个连接都有一个权重,表示该连接的重要性。每个神经元可以有一个偏差项,用于调整输出,激活函数决定了神经元的输出。常用激活函数包括 Sigmoid、ReLU 和 Tanh 等。激活函数的选择对网络的性能有显著影响。NN 通过调整神经元之间的权重进行学习。这通常通过一种称为反向传播的算法实现,该算法结合了梯度下降方法,以最小化预测输出和实际输出之间的差异^[19]。

1.3.1 超参数优化

本文使用了网格搜索法和贝叶斯优化法,网格

搜索法是一种基于穷举策略的超参数优化方法,其核心思想是在预设的参数空间内对所有可能的参数组合进行全面搜索,通过交叉验证等评估手段比较每组参数下模型的表现,并最终选出在特定评估指标上达到最佳性能的参数组合^[20]。贝叶斯优化法则通过构建超参数与模型性能之间的概率映射,利用后验分布和采集函数指导每一步的参数选择,从而在较少的迭代次数内高效地探索参数空间并定位全局最优解^[20]。2 种参数优化方法的参数范围如表 1 所示。

在使用随机森林进行建模时,需要对诸如决策树数量、树的最大深度、节点最小分裂样本数以及每棵树选取的特征数量等参数进行调整,以获得最佳

表 1 各模型超参数范围

Tab. 1 Hyperparameter Ranges for Each Model

模型	超参数	参数范围	
		网格搜索法	贝叶斯优化法
随机森林	n_estimators	50, 100, 150, 200, 250, 300	50~300
	max_depth	10, 15, 20, 25, 30	10~30
	min_samples_split	2, 4, 6, 8, 10	2~10
	min_samples_leaf	1, 2, 3, 4, 5	1~5
XGBoost	max_depth	4, 5, 6, 7, 8	4~8
	min_samples_split	2, 5, 10	2~10
	subsample	0.6, 0.7, 0.8, 0.9	0.6~0.9
	max_features	0.6, 0.7, 0.8, 0.9	0.6~0.9
	learning_rate	0.01, 0.03, 0.05, 0.1, 0.3, 0.5	0.01~0.5
	n_estimators	100, 200, 300	100~300
KNN	n_neighbors	—	1~30
	weights	—	统一权重(uniform), 距离加权(distance)

模型^[16]。使用 Grid Search CV 进行网格搜索和交叉验证时,由于同时调优多个超参数且每个参数具有多个候选值,若一次性搜索所有组合将显著增加计算成本和搜索复杂度。因此,本文将超参数分组,逐轮调优,以既降低计算量又精确评估各参数对模型性能的影响。第一轮主要针对森林中树的数量(n_estimators)和树的最大深度(max_depth)进行搜索;随后,在固定第一轮最优参数的基础上,第二轮进一步调优节点分裂前的最小样本数(min_samples_split)和叶节点的最小样本数(min_samples_leaf)。

XGBoost 超参数的调整同样使用了 Grid Search CV 进行网格搜索和交叉验证。将这 6 个参数分成 3 组,依次固定部分参数并逐轮进行调优。第一轮主要针对最大树深(max_depth)和子节点分裂前的最小样本数(min_samples_split)进行搜索;第二轮进

一步调整子采样率(subsample)和每棵树的特征采样比例(max_features),并固定第一轮的最优值;第三轮则在前两轮的基础上,调优学习率(learning_rate)和基学习器数量(n_estimators)。这样做既能显著降低计算量,又能更准确地评估各个参数对模型性能的影响,从而得到更合理的最终参数组合。

在试验中,采用了多层感知器回归模型(MLPRegressor),并通过设置特定超参数来优化模型性能。模型配置中,隐藏层结构(hidden_layer_sizes)参数设为 100,表示仅含 1 层隐藏层,该层拥有 100 个神经元,此设置既控制了模型复杂度,又确保了足够的表达能力;激活函数采用线性整流函数(ReLU),有效缓解了梯度消失问题并具备较高计算效率;优化算法选择自适应矩估计优化器(Adam),它结合了动量和 RMSProp 的优势,常在实际应用中

表现出色。为防止过拟合, L2 正则化系数 (Alpha) 参数设为 0.05, 利用 L2 正则化提高模型泛化性能。此外, 采用 Constant 学习率策略, 训练过程中保持学习率 (0.001) 恒定, 从而确保权重稳定更新并逐步逼近最优解。

1.3.2 模型评价方法

R^2 是常用于评估回归模型性能的指标, 反映了模型对数据的拟合优度, 即模型解释因变量波动的程度^[21]。其值为 0~1, 越接近 1 表示模型对数据的拟合效果越好。 R^2 的计算如式(5)。

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

其中: n ——样本个数;

y_i ——实际值;

\hat{y}_i ——预测值;

\bar{y}_i ——实际值的均值。

式(5)的分子表示模型的残差平方和, 分母表示实际数据与均值的总平方和。 R^2 越大, 说明模型对数据的拟合能力越强, R^2 越小, 意味着模型解释的变异有限, 然而 R^2 并非真正的误差度量, 仍需结合绝对误差指标以全面评估模型。

均方误差 (MSE) 是预测值与真实值之差的平方均值^[22], 计算如式(6)。

$$S_{MSE} = \frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2 \quad (6)$$

其中: S_{MSE} ——MSE。

平均绝对误差 (MAE) 是预测误差绝对值的平均值^[22], 计算如式(7)。

$$S_{MAE} = \frac{1}{n} \sum_{i=1}^n |(y_i - \hat{y}_i)| \quad (7)$$

其中: S_{MAE} ——MAE。

MSE 通过对预测误差取平方并取均值的方式突出极端误差, 在优化算法中也因可微特性而广为采用, 平方运算会放大异常值影响, 并使误差指标的物理解释相对不直观。MAE 以绝对误差度量模型偏差大小, 既与原数据单位保持一致, 便于理解, 也能在一定程度上减弱异常值对评价的冲击, 惟其对极端误差的敏感度不足, 且绝对值函数在优化时难

以直接求解导数^[22]。因此, 综合使用 R^2 、MSE 和 MAE 不仅能衡量模型对整体趋势的解释能力, 也能兼顾预测精度与稳健性, 有助于更加全面地评估回归模型的性能。

1.4 基于 SHAP 的模型解释

SHAP 算法是一种基于合作博弈论的模型解释方法, 主要用于解释机器学习模型的预测结果。其核心思想是通过计算每个特征对模型输出的边际贡献, 来量化各个特征的重要性。SHAP 结合了 Shapley 值的概念与机器学习模型, 能够提供公平且一致的特征贡献分配^[23]。

对于特定的特征 x_j 在给定模型预测值 $f(x)$ 中的贡献, 其 Shapley 值定义如式(8)。

$$\phi_j = \sum_{S \subseteq N \setminus \{j\}} \frac{|S|! \cdot (|N| - |S| - 1)!}{|N|!} [f(S \cup \{j\}) - f(S)] \quad (8)$$

其中: ϕ_j ——第 j 个变量的 Shapley 值;

N ——特征的全集;

S ——特征子集 (不包括特征 j);

$|S|$ ——子集 S 中特征的数量;

$f(S)$ ——模型的预测值;

$|N|$ ——总特征量;

$S \subseteq N \setminus \{j\}$ ——不包含特征 j 的子集;

$f(S \cup \{j\}) - f(S)$ ——特征 x_j 的边际贡献, 即加入 j 后对预测值的增量。

2 结果与讨论

2.1 数据集统计分析

本文所用数据集涵盖了餐厨垃圾浆液厌氧处理过程中关键影响因素及产气量等指标, 其变量包括进水量、沼气产量、进出水 pH、COD、氨氮、SS 及 VFA。由表 2 可知, 各指标均表现出一定程度的离散性。首先以 COD 为例, 如图 1 所示, 进水 COD 平均质量浓度约 43 680 mg/L, 标准偏差近 17 712 mg/L, 说明进料 COD 波动较大; 而出水 COD 平均质量浓度显著下降至 3 575 mg/L, 表明该厌氧反应器的处理效率高, 能够有效应对餐厨垃圾浆液有机质浓度高的问题。从进出水氨氮来看, 氨氮则在处理后略有升高, 平均质量浓度增至 1 807 mg/L, 推测与厌氧分解过程中蛋白质与有机氮的转化有关。出水 pH 均值为 7.85 而进水 pH 均值为 6.22, 这一差异可归因于反应器内酸碱平衡的动态调节以及外部碱度的补充。

厌氧消化过程中,水解、酸化、乙酸化及甲烷生成等阶段依次进行,使初期生成的酸性物质在后续转化过程中逐步生成碳酸盐和重碳酸盐,从而增强系统缓冲能力^[1]。此外,为维持微生物活性,工厂会适量补充碱度,进一步提升反应器内整体 pH 水平。同时,SS 在进水端平均质量浓度约 4 198.655 mg/L,出水端平均质量浓度为 4 126.577 mg/L,但其离散度在出水侧明显增大(标准偏差为 7 074.568 mg/L),反映污泥浓度或泥水分离的波动性。值得注意的是,出水 VFA 平均质量浓度达 2 756.36 mg/L,标准偏差为 2 458.281 mg/L,提示系统酸化与产甲烷平衡容易受到外界干扰,需在工艺控制上加以关注,沼气产量均值为 2 991 m³,标准偏差约 2 052 m³,亦说明产气情况在不同工况下差异明显。总体而言,该数据集较全面地反映了厌氧处理进出水水质特征及

表 2 数据集各变量平均值及标准差
Tab. 2 Mean and Standard Deviation of Each Variable in the Dataset

指标	平均值	标准差
进水量/m ³	123.476 1	61.085 73
生物气/m ³	2 991.411 0	2 051.585
进水 pH 值	6.621 105	0.482 999
进水 COD/(mg · L ⁻¹)	43 680.24	17 712.31
进水氨氮/(mg · L ⁻¹)	1 185.991	345.504 4
进水 SS/(mg · L ⁻¹)	4 198.655	2 977.173
出水 pH 值	7.850 559	0.176 024
出水 COD/(mg · L ⁻¹)	3 575.471	1 735.94
出水氨氮/(mg · L ⁻¹)	1 806.978	1 127.264
出水 SS/(mg · L ⁻¹)	4 126.577	7 074.568
出水 VFA/(mg · L ⁻¹)	2 756.360	2 458.281

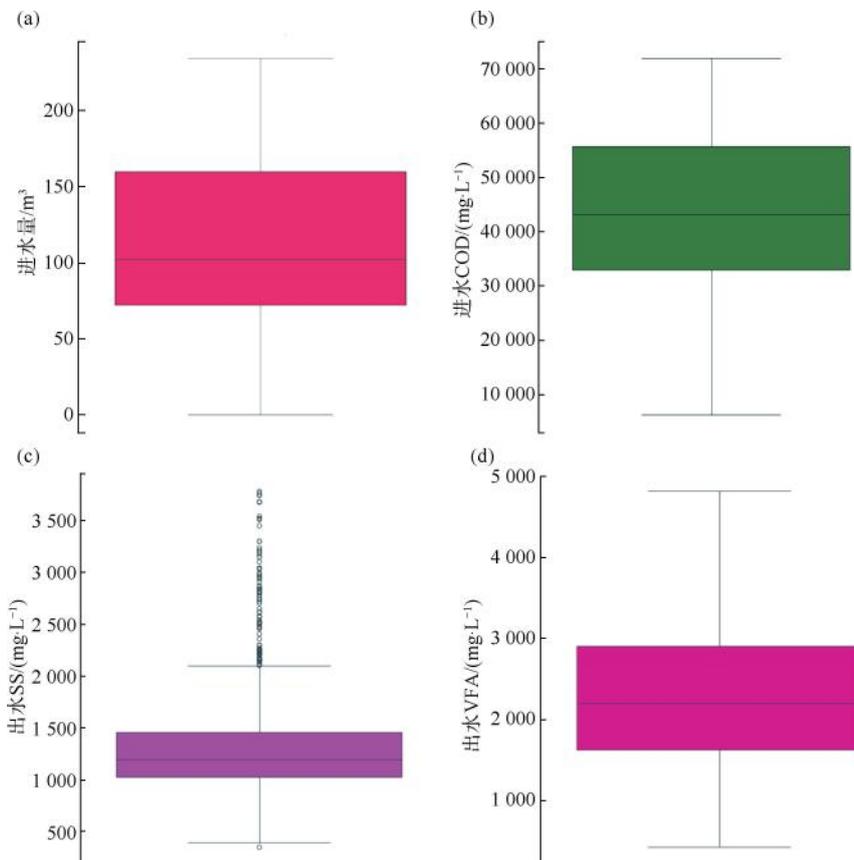


图 1 关键指标的分布

Fig. 1 Distribution of Key Indices

产气水平,并为后续深入分析各变量之间的关联性 及处理效率评价奠定了基础。

2.2 数据相关性分析

本文基于 Pearson 相关系数对厌氧消化过程中

若干关键指标进行热图分析,旨在识别各变量间的 依赖关系并探讨其对沼气产量的线性影响程度,分 析结果如图 2、图 3 所示。

从进水端来看,进水量与沼气产量呈显著正相

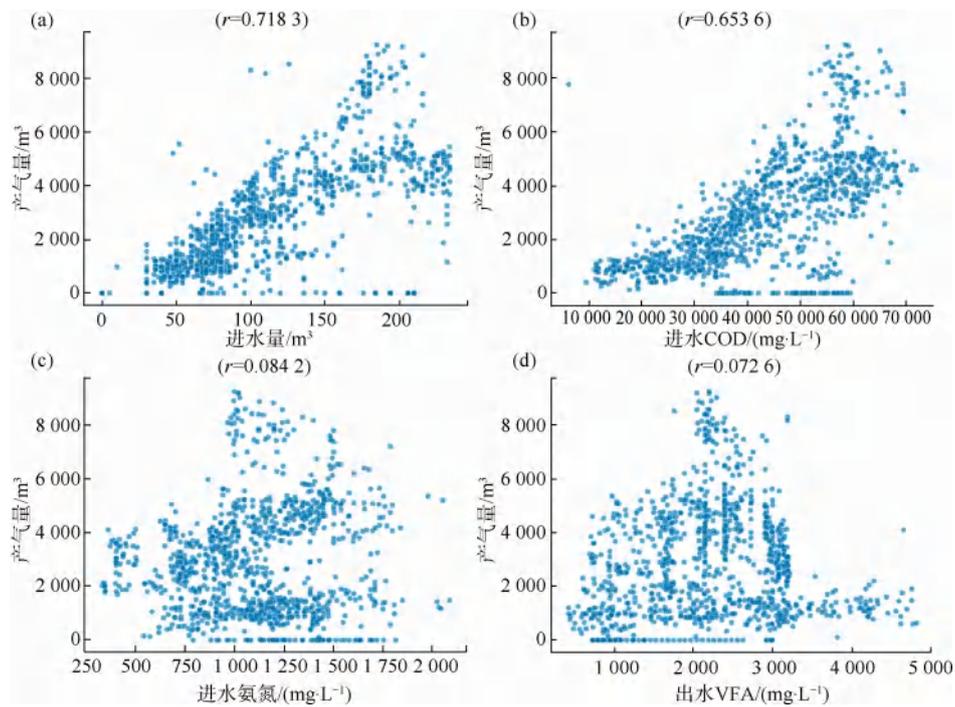


图2 (a) 进水量; (b) 进水 COD; (c) 进水氨氮; (d) 出水 VFA 与产气量的相关关系

Fig. 2 Correlation among (a) Influent Volume; (b) Influent COD; (c) Influent Ammonia Nitrogen; (d) Effluent VFA and Biogas Production

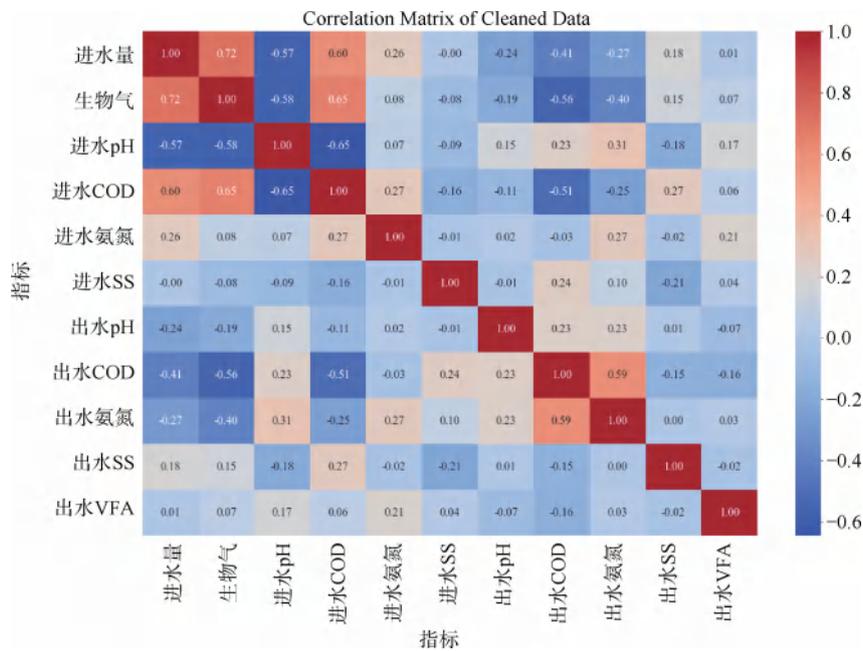


图3 数据集指标相关性分析

Fig. 3 Correlation Analysis of Dataset Indices

关(r 为0.72),指示进水量增加可提升微生物底物供应,从而促进产甲烷过程;但若进水负荷过大,也易引发系统不稳定^[24],此外,数据还显示,进水量与

进水 COD 之间亦有正相关(r 约0.60),这意味着,当进水量提高时,水中所携带的有机物总量也随之增加,从而导致 COD 测定值上升。进水 COD 与沼

气产量的 r 达 0.65,表明 COD 浓度越高,潜在产气能力越强,但过量有机负荷可能导致酸化抑制,影响甲烷菌活性。在 pH 方面,进水 pH 与沼气产量呈负相关(r 为 -0.58),或意味底物在偏离适宜 pH 区间时难以被甲烷菌有效利用^[1],需要结合其他运行参数(如负荷率、VFA)综合评估。进水 pH 与进水 COD 之间的负相关(r 为 -0.65),表明 COD 积累常伴随有机酸生成,进而致使 pH 降低。

出水端方面,出水 COD 与出水氨氮呈正相关(r 约 0.59),可能与氨氮升高抑制微生物降解效率有关;同时,沼气产量与出水 COD 呈一定负相关(r 为 -0.56),说明高产气通常伴随更彻底的 COD 去除。出水 VFA 与产气量则表现出负向分布趋势,高 VFA 常与酸化环境相关,易阻碍产甲烷进程。

厌氧消化过程的反应环境由氨氮、VFA 等酸碱物质调节,从反应环境来看,针对氨氮,其与沼气产量的线性相关并不突出,但从数据分布可见氨氮质量浓度在 500~1 000 mg/L,有助于提升产气,超过 1 000 mg/L 后,则呈抑制特征,反映出氨氮在提供营养与导致毒害间的平衡作用。出水氨氮与出水 VFA 略呈负相关,提示二者或存在此消彼长的动态关系。

此外,出水 SS 与 VFA 的正相关(r 约 0.32)可能暗示较高 SS 含量与 VFA 累积相伴,操作人员可从固体停留时间(SRT)、水力停留时间(HRT)等方面进行调控。出水 pH 与多数参数的相关系数相对较低,表明该系统在微生物代谢后 pH 趋于稳定。

综上,相关性分析显示负荷与抑制效应并存,COD、氨氮、pH、VFA 等指标均需适宜区间内平衡。但仅凭线性分析尚不足以全面解释复杂耦合作用,后续可结合机器学习和 SHAP 等可解释性工具进行多维度深度挖掘,以实现厌氧消化过程更精准调控与高效产气。

2.3 数据特征选择

该厌氧数据集影响因素多、数据维度高,存在许多冗余或相关性较高的变量,增加了模型复杂度并降低预测性能。为提炼关键影响因素并简化后续建模工作,采用 PCA 对数据进行降维,从而在减少过拟合风险的同时提高模型泛化能力^[15]。

图 4 展示了通过 PCA 提取的各主成分所解释的累计方差比例。横轴代表主成分的数量,纵轴代表累计解释方差比例,表 3 是 PCA 得到的各主成分

贡献的方差解释,随着主成分数量的增加,累计解释方差比例逐渐上升,最终趋近于 1。由表 3 可知,PCA 提取的第 1 个主成分解释了数据中 29.65% 的方差,表明该成分捕捉了数据集中较大比例的信息。第 2 个主成分进一步解释了 14.54% 的方差,使累计解释方差比例达到了 44.19%,说明前 2 个主成分能够覆盖数据中接近一半的变异信息。第 3~5 主成分,方差解释比例分别为 12.72%、12.31% 和 8.46%,累计解释方差比例逐步上升至 77.68%,表明前 3~5 个主成分对数据的主要信息进行了较好的表征。第 6~8 个主成分的方差贡献逐渐减小,分别解释了 8.10%、5.48% 和 3.86% 的方差,到第 8 个主成分时,累计解释方差比例已达到 95.12%,表明通过前 8 个主成分可以有效保留数据中 95% 以上的变异信息,剩余成分对数据方差的贡献较小,大部分数据的方差可以通过这 8 个主成分来解释。

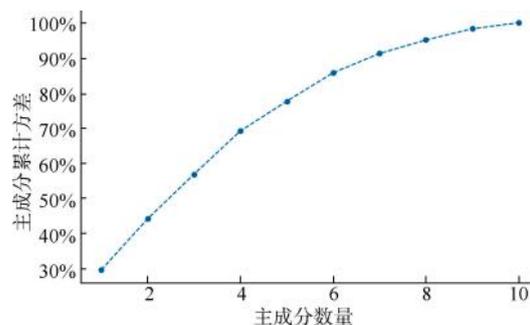


图 4 数据集 PCA 分析结果

Fig. 4 PCA Analysis Results of the Dataset

表 3 PCA 总方差解释

Tab. 3 Total Variance Explanation by PCA

主成分	解释的方差比例	累计解释的方差比例
1	29.05%	29.05%
2	14.54%	44.19%
3	12.72%	56.91%
4	12.31%	69.22%
5	8.46%	77.68%
6	8.10%	85.78%
7	5.48%	91.26%
8	3.86%	95.12%
9	3.18%	98.30%
10	1.70%	100.00%

2.4 基于机器学习的厌氧甲烷产量预测模型建立

在确认关键变量后,本文采用了随机森林、XGBoost、KNN 和 ANN 4 种机器学习算法对生物气

产量进行拟合预测,并对拟合后的模型进行超参调优。表4展示了不同超参调优方法下的各模型的评价指标值,包括 R^2 、MSE和MAE,用于评估模型的拟合能力和误差控制情况。结果表明:XGBoost模型在所有评估指标上均表现最佳,其网格搜索和贝叶斯优化均取得了相近的 R^2 值,分别为0.7342和0.7339,表明该模型具有较强的泛化能力和稳定的预测精度。在网格搜索得到的最优超参数下,XGBoost的MSE和MAE均达到最低水平,分别为0.1583和0.2645,表明其误差控制能力优于其他模型。相比之下,随机森林模型的2种优化方法表现

稳定,其 R^2 值约0.7245,略低于XGBoost,其MSE为0.1666和MAE为0.2658,均高于XGBoost,表明尽管该模型仍具备一定的预测能力,但在高精度需求的场景中,其表现不及XGBoost。KNN模型贝叶斯优化的 R^2 与随机森林相近,表明其在中等范围内的预测较为准确,但其MSE值为0.1736比随机森林更高,说明在极端值上的预测误差较大。然而,该方法的MAE值为0.2503低于随机森林,表明其整体误差分布较稳定。ANN的 R^2 相对较低(0.7176),其MSE和MAE最高,表明其在本文中的预测稳定性和准确性均不及树模型随机森林。

表4 各模型在不同超参调优方法下的评价指标值及最优超参数

Tab. 4 Evaluation Index and Optimal Hyperparameters for Each Model under Different Hyperparameter Tuning Methods

基于不同优化方法的各模型	最优超参数	R^2	MSE	MAE
随机森林 网格搜索	n_estimators = 150, max_depth = 15 min_samples_split = 2, min_samples_leaf = 1	0.7242	0.1671	0.2673
随机森林 贝叶斯优化	n_estimators = 296, max_depth = 14 min_samples_split = 3, min_samples_leaf = 1	0.7245	0.1666	0.2658
XGBoost 网格搜索	max_depth = 5, min_samples_split = 2 subsample = 0.8, max_features = 0.9 learning_rate = 0.03, n_estimators = 300	0.7342	0.1583	0.2645
XGBoost 贝叶斯优化	max_depth = 8, min_samples_split = 2 subsample = 0.6, max_features = 0.6 learning_rate = 0.031, n_estimators = 268	0.7339	0.1588	0.2651
KNN 贝叶斯优化	n_neighbors = 5, weights = distance	0.7247	0.1736	0.2503
ANN 手动调参	hidden_layer_sizes = 100, activation = relu solver = adam, alpha = 0.05 learning_rate = constant learning_rate_init = 0.001	0.7176	0.2025	0.3053

图5展示了不同模型的预测值与实际值之间的关系,X轴表示实际值,Y轴表示预测值,蓝色数据点代表每个样本的实际值和预测值,虚线表示理想情况下预测值与实际值完全一致的参考线。随机森林、XGBoost、KNN和ANN4个模型均呈现出一定的线性趋势,说明它们能够在一定程度上捕捉生物气产量的变化规律。其中:随机森林与XGBoost模型的预测值较为集中,在2500~6500 m³/d的表现较为理想,大部分数据点分布在虚线附近,表明预测值与实际值接近。整体拟合效果较优,尽管在高数值区间略有波动,但仍保持较强的泛化能力。相比之下,KNN模型在极端值区域的预测偏差较大,表明其对复杂关系的学习能力可能相对有限。ANN模型的预测值分散程度较高,特别是在高数值区域,其稳定性和准确性均不及树模型。综合来看,树模型(随机森林与XGBoost)在本任务中的表现更为稳健和精

准,这一结论亦与表4中的评价指标结果相一致。

综合来看,XGBoost模型在预测准确性和误差控制方面表现最佳,具有较高的 R^2 值和最低的MSE及MAE,能够较好地捕捉生物气产量与输入变量之间的非线性关系。KNN模型和随机森林次之,其在中等范围内的预测较为准确,但在边缘区间表现出一定的误差。ANN模型预测结果分布较为分散,误差较大。因此,在实际应用中,XGboost模型可以作为预测生物气产量的首选工具,具有较好的稳定性和解释能力。

2.5 基于SHAP值的模型可解释性分析

基于网格搜索法调参后的XGboost模型,本文利用SHAP方法为模型预测结果提供可解释性。SHAP值作为一种基于博弈论的模型解释工具,通过计算各特征对预测输出贡献的分布,从而帮助研究者更直观地识别模型决策背后的驱动因素^[15]。

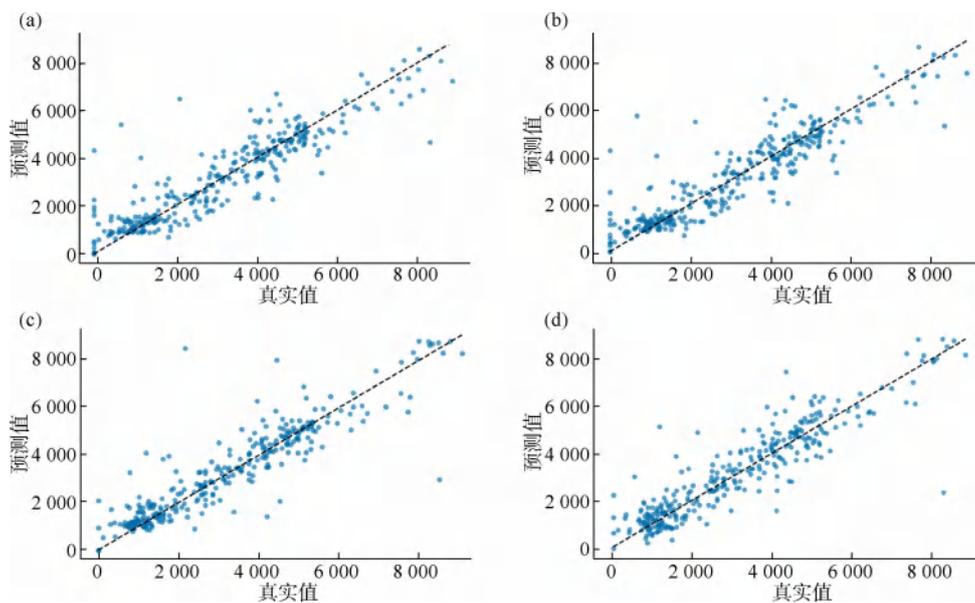


图 5 机器学习模型预测情况: (a) 随机森林; (b) XGboost; (c) KNN; (d) ANN

Fig. 5 Machine Learning Model Prediction: (a) Random Forest; (b) XGboost; (c) KNN; (d) ANN

图 6 显示了不同特征对模型输出的影响分布,其中每一行代表一个输入特征,横坐标为 SHAP 值,即对预测结果的贡献值。特征按照 SHAP 的平均绝对值

进行排序,由上至下对模型的影响程度依次降低。一个散点代表一个样本,颜色越红说明特征本身数值越大,颜色越蓝说明特征本身数值越小。

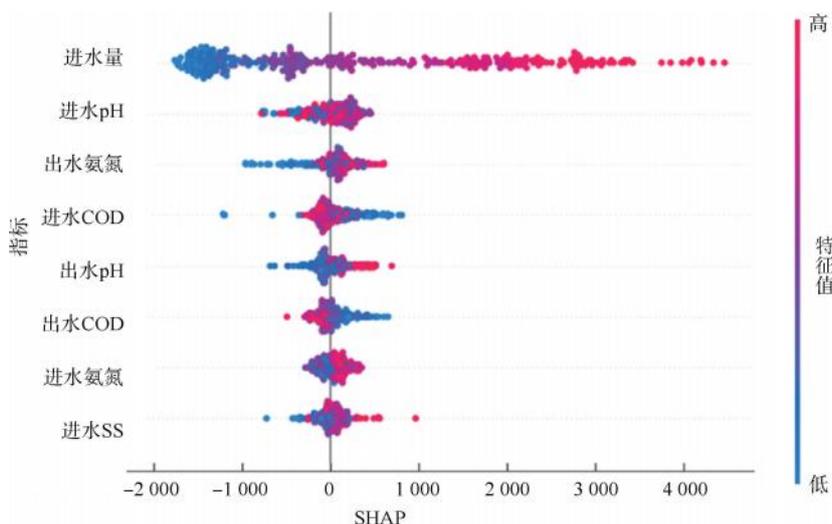


图 6 SHAP 分析结果

Fig. 6 Analysis Results of SHAP

在特征影响排序中,整体上可观察到进水量、进水 pH、出水氨氮和进水 COD 是对产气量影响较为显著的前 4 项因素。其中,进水量对产气量的影响尤为突出,由图 6 可知,进水量较大时,对产气量预测的贡献值为正,进水量较小时,对产气量预测的贡献值为负,这一结论说明,当进水量较大时,系统能够提供足够的底物和营养,使微生物在厌氧消化过

程中表现出较高的活性,从而显著促进产气;而当进水量较小时,由于底物供应不足,微生物活性受限,产气量则相对降低^[25]。

进水 pH 是排名第 2 的重要特征,pH 越低时,对甲烷产量的负贡献越大,pH 越高时,对甲烷产量的正贡献越大,但极高的 pH 也会对产气造成负影响,pH 对厌氧消化过程的性能和稳定性起着重要作

用。厌氧系统中的微生物包括产甲烷古菌、发酵菌和水解菌都对 pH 非常敏感。研究^[1]表明,不同微生物的最佳 pH 范围有所不同。产甲烷古菌的最佳 pH 值为 6.5~8.0,发酵细菌能够在 pH 值为 4.0~8.5 生长,并在 pH 值为 5.0~6.0 表现最佳,该体系内 pH 值偏低为 6.0 附近,不利于甲烷菌生存。

出水氨氮浓度则从氮源与抑制效应 2 个层面影响生物反应过程。由图 6 可知,出水氨氮浓度偏低时,对产气量的贡献为负,浓度偏高时对产气量的贡献为正,但部分浓度过高的样本点对产气量产生了负影响。氨氮是厌氧微生物成长的必要元素,同时在反应体系中也有缓冲稳定 pH 的作用^[26],当反应体系中的氨氮浓度过低时,必需的氮源不足会限制厌氧微生物的生长和代谢活性,导致蛋白质和其他细胞组分合成受阻,从而降低了对有机底物的降解效率且减少产气量^[2]。中间浓度数据点分布幅度最宽,此时氨氮质量浓度约 1 500 mg/L,对生物气预测的贡献度较低,进水氨氮浓度过高时,模型显示对生物气产生了负贡献,此时的氨氮质量浓度为 3 000 mg/L 左右,高氨氮往往导致体系中游离氨的积累,而游离氨对微生物细胞代谢过程有较强的抑制作用,使微生物群落的活性大大降低。这种生物抑制效应在高氨氮条件下表现尤为明显,从而显著影响甲烷菌的正常代谢与产气能力^[27]。随着出水氨氮浓度升高,相应散点的分布幅度明显更宽,这说明该大型反应器中高氨氮条件出现频率和波动性较大。

在这一系列特征中,进水 COD 对产气量也有着极为显著的影响。由图 6 可知,进水 COD 在高浓度中分布点幅度更宽,其中极高值对生物气产量预测的贡献为负。作为有机负荷的表征指标,COD 浓度过高的进水会导致系统中快速富集的有机物在水解与产酸菌的作用下迅速转化为 VFA。VFA(如乙酸、丙酸、丁酸等)的堆积主要是由于酸化和产甲烷微生物的世代时间差异,以及 VFA 降解的热力学壁垒引起^[28],VFA 积累过多、体系 pH 骤降,将削弱甲烷菌的活性,进而降低产甲烷效率。如果在高负荷运行条件下未能及时稳定系统 pH,则微生物群落对 COD 的利用效率会下降,最终抑制产气量的提升,这与理论研究^[27,29]结果一致。

综上所述,SHAP 方法在本文中不仅提供了对预测模型的可解释性,也从定量角度凸显了各运行

参数的重要性,找到了影响该厌氧反应器的核心的运行条件,分别是进水量、进水 pH、进水 COD 以及出水氨氮,从大规模实际运行数据的角度揭示了高水力负荷、低 pH、高 COD 及高氨氮等因素对产气效率的抑制效应,不仅为厌氧消化系统的精准调控提供了可操作的指标范围,也为深入探究厌氧微生物生态和工艺优化策略奠定了坚实的理论基础。

3 结论

本文针对厌氧消化系统在工业规模应用中面临的变量多重耦合及模型预测不稳定等问题,综合运用相关性分析、PCA 降维 4 种机器学习算法(随机森林、XGBoost、KNN、ANN)和 SHAP 可解释性方法,系统探讨了沼气产量预测及影响沼气产量的关键因素,主要结论如下。

1) 多维数据相关性凸显负荷与抑制效应并存,产水量与产气量、进水 COD 呈强正相关,而进水 pH、出水 VFA 和高氨氮等因素在特定范围内会对产气产生抑制。显著的正负相关关系意味着该系统中既有正向促进的高负荷潜力,也存在酸化与氨抑制风险,需要在运行中谨慎平衡。

2) 机器学习预测对比凸显梯度提升优势,随机森林、XGBoost、KNN、ANN 4 种模型均具备一定预测能力,但 XGBoost 在 R^2 评价中表现最佳,KNN 和随机森林次之。ANN 受数据规模和模型复杂度影响,泛化能力相对不足。可见在当前数据与工况下,XGBoost 更具鲁棒性与准确性。

3) SHAP 定量剖析产气关键影响因素,SHAP 分析表明,进水量、进水 pH、进水 COD、出水氨氮是影响产气量的核心因素。高 COD 虽能带来有机底物增量,但若超出微生物利用能力,则会导致抑制。氨氮在中等浓度时有积极意义,过高则触发氨毒害。pH 越偏离适宜区间,微生物活性越不稳定,影响产气效率。

综上所述,XGBoost 模型在复杂厌氧系统预测中兼具高精度与鲁棒性,配合 SHAP 方法可有效识别和量化关键影响因素,为工业规模的厌氧消化调控提供科学依据。未来可在进水负荷与 pH、氨氮的联合控制上进一步优化,并结合实时在线监测与动态调度策略,持续提升沼气产量与系统稳定性,为餐厨垃圾及多源有机废水的资源化利用提供更全面的技术支撑。

参考文献

- [1] Wang Z Z, Hu Y S, Wang S, et al. A critical review on dry anaerobic digestion of organic waste: Characteristics, operational conditions, and improvement strategies [J]. *Renewable and Sustainable Energy Reviews*, 2023, 176: 113208. DOI: 10.1016/j.rser.2023.113208.
- [2] Rocamora I, Wagland S T, Hassard F, et al. Inhibitory mechanisms on dry anaerobic digestion: Ammonia, hydrogen and propionic acid relationship[J]. *Waste Management*, 2023, 161: 29–42. DOI: 10.1016/j.wasman.2023.02.009.
- [3] Steiniger B, Hupfauf S, Insam H, et al. Exploring anaerobic digestion from mesophilic to thermophilic temperatures—Operational and microbial aspects [J]. *Fermentation*, 2023, 9 (9): 798.
- [4] Mo R R, Guo W J, Batstone D, et al. Modifications to the anaerobic digestion model no. 1 (ADMI) for enhanced understanding and application of the anaerobic treatment processes—A comprehensive review [J]. *Water Research*, 2023, 244: 120504. DOI: 10.1016/j.watres.2023.120504.
- [5] Wu D, Li L, Peng Y, et al. State indicators of anaerobic digestion: A critical review on process monitoring and diagnosis [J]. *Renewable and Sustainable Energy Reviews*, 2021, 148: 111260. DOI: 10.1016/j.rser.2021.111260.
- [6] 石悦,肖本益,张宁,等. 基于神经网络耦合遗传算法对规模化牛粪厌氧消化产沼气的模拟与优化 [C]. 北京:中国环境科学学会 2022 年科学技术年会—环境工程技术创新与应用分会场论文集(四), 2022.
- Shi Y, Xiao B Y, Zhang N, et al. Simulation and optimization of large-scale anaerobic digestion of cowmanure for biogas production based on artificial neural network coupled with genetic algorithm [C]. Beijing: 2022 Science and Technology Annual Meeting of the Chinese Society of Environmental Sciences-Field Theory Collection of the Environmental Engineering Technology Innovation and Application Branch (IV), 2022.
- [7] Cai W F, Lesnik K L, Wade M J, et al. Incorporating microbial community data with machine learning techniques to predict feed substrates in microbial fuel cells [J]. *Biosensors and Bioelectronics*, 2019, 133: 64–71. DOI: 10.1016/j.bios.2019.03.021.
- [8] Clercq D, Jalota D, Shang R, et al. Machine learning powered software for accurate prediction of biogas production: A case study on industrial-scale Chinese production data [J]. *Journal of Cleaner Production*, 2019, 218: 390–399. DOI: 10.1016/j.jclepro.2019.01.031.
- [9] Wang Z X, Peng X G, Xia A, et al. The role of machine learning to boost the bioenergy and biofuels conversion [J]. *Bioresource Technology*, 2022, 343: 126099. DOI: 10.1016/j.biortech.2021.126099.
- [10] Noor N M, Abdullah M M A B, Yahaya A S, et al. Comparison of linear interpolation method and mean method to replace the missing values in environmental data set [J]. *Materials Science Forum*, 2014, 803: 278–281. DOI: 10.4028/www.scientific.net/MSF.803.278.
- [11] Aguinis H, Gottfredson R K, Joo H. Best-practice recommendations for defining, identifying, and handling outliers [J]. *Organizational Research Methods*, 2013, 16(2): 270–301.
- [12] Wan X, Wang W Q, Liu J M, et al. Estimating the sample mean and standard deviation from the sample size, median, range and/or interquartile range [J]. *BMC Medical Research Methodology*, 2014, 14: 1–13. DOI: 10.1186/1471-2288-14-135.
- [13] Yetilmezsoy K, Karakaya K, Bahramian M, et al. Black-, gray-, and white-box modeling of biogas production rate from a real-scale anaerobic sludge digestion system in a biological and advanced biological treatment plant [J]. *Neural Computing & Applications*, 2021, 33(17): 11043–11066.
- [14] Kurita T. Principal component analysis (PCA) [M]. *Computer Vision: A Reference Guide*. London: Springer. 2021: 1013–1016.
- [15] 李笑梅. 多指标综合评价方法综述 [J]. *统计与管理*, 2022, 37(2): 45–48.
- Li X M. Overview of multi index comprehensive evaluation methods [J]. *Statistics and Management*, 2022, 37(2): 45–48.
- [16] 方匡南,吴见彬,朱建平,等. 随机森林方法研究综述 [J]. *统计与信息论坛*, 2011, 26(3): 32–38.
- Fang K N, Wu J B, Zhu J P, et al. A review of technologies on random forests [J]. *Journal of Statistics and Information*, 2011, 26(3): 32–38.
- [17] Chen T Q, Guestrin C. XGBoost: A scalable tree boosting system [C]. San Francisco: Proceedings of the Proceedings of the 22nd Acm Sigkdd International Conference on Knowledge Discovery and Data Mining, 2016: 785–794.
- [18] Zhang S, Cheng D, Deng Z, et al. A novel kNN algorithm with data-driven k parameter computation [J]. *Pattern Recognition Letters*, 2018, 109: 44–54. DOI: 10.1016/j.patrec.2017.09.036.
- [19] Lecun Y, Bengio Y, Hinton G. Deep learning [J]. *Nature*, 2015, 521: 436–444. DOI: 10.1038/nature14539.
- [20] 李海霞,宋丹蕾,孔佳宁,等. 传统机器学习模型的超参数优化技术评估 [J]. *计算机科学*, 2024, 51(8): 242–255.
- Li H X, Song D L, Kong J N, et al. Evaluation of hyperparameter optimization techniques for traditional machine learning models [J]. *Computer Science*, 2024, 51(8): 242–255.
- [21] Cabuk K S, Cengiz S K, Guler M G, et al. Chasing the objective upper eyelid symmetry formula; R^2 , RMSE, POC, MAE, and MSE [J]. *International Ophthalmology*, 2024, 44(1): 303.

(下转第 119 页)

- Dai X H, Zhang C, Zhang L W, et al. Thoughts on the development direction of sludge treatment and resource recovery under the background of carbon neutrality [J]. *Water & Wastewater Engineering*, 2021, 47(3): 1-5.
- [10] 成珊, 罗睿, 田红, 等. 水热碳化温度对污泥有机氮固液相迁移转化路径影响研究[J]. *化工学报*, 2022, 73(11): 5220-5229.
- Cheng S, Luo R, Tian H, et al. Effect of hydrothermal carbonization temperature on transformation path of organic nitrogen in sludge[J]. *CIESC Journal*, 2022, 73(11): 5220-5229.
- [11] 程寒飞, 高卫民, 张镭. 污泥水热碳化技术的应用[J]. *中国环保产业*, 2023(1): 15-20.
- Cheng H F, Gao W M, Zhang L. Application of the hydrothermal carbonization technology for sludge [J]. *China Environmental Protection Industry*, 2023(1): 15-20.
- [12] Wang L P, Chang Y Z, Li A M. Hydrothermal carbonization for energy-efficient processing of sewage sludge: A review [J]. *Renewable and Sustainable Energy Reviews*, 2019, 108: 423-440. DOI: 10.1016/j.rser.2019.04.011.
- [13] Zheng X Y, Chen W, Ying Z, et al. Thermodynamic investigation on gasification performance of sewage sludge-derived hydrochar: Effect of hydrothermal carbonization [J]. *International Journal of Hydrogen Energy* 2019, 44 (21): 10374-10383.
- [14] 范准. 反应条件对污泥水热碳化产物的影响分析[D]. 重庆: 重庆大学, 2020.
- Fan Z. Effect of reaction conditions on the products of sludge hydrothermal carbonization [D]. Chongqing: Chongqing University, 2020.
- [15] 许劲, 徐军, 吕秋颖, 等. 水热碳化技术用于污泥处理处置前景分析[J]. *中国给水排水*, 2020, 36(16): 54-59.
- Xu J, Xu J, Lü Q Y, et al. Perspectives on hydrothermal carbonization technology for municipal sludge treatment and disposal [J]. *China Water & Wastewater*, 2020, 36(16): 54-59.
- [16] 孙飞凡, 袁世震, 卢如飞, 等. 固体回收燃料掺烧对污泥燃烧特性的影响[J]. *洁净煤技术*, 2023, 29(9): 78-86.
- Sun F F, Yuan S X, Lu R F, et al. Influence of the added solid recovered fuel on the combustion characteristics of sewage sludge [J]. *Clean Coal Technology*, 2023, 29(9): 78-86.
- [17] 雷浩洋, 乔加飞, 王睿坤, 等. 污泥秸秆共混水热碳化固体炭产物的燃烧及 NO_x 生成特性[J]. *动力工程学报*, 2024, 44(2): 181-187, 327.
- Lei H Y, Qiao J F, Wang R K, et al. Combustion characteristics and NO_x generation of hydrochar from co-hydrothermal carbonization of sewage sludge and corn straw [J]. *Journal of Chinese Society of Power Engineering*, 2024, 44(2): 181-187, 327.
- [18] 况怡婷, 黄芳, 米铁, 等. 中药渣水解炭燃料特性及热解燃烧动力学分析[J]. *煤炭转化*, 2024, 47(3): 45-48.
- Kuang Y T, Huang F, Mi T, et al. Fuel characteristics and kinetics analysis of pyrolysis and combustion of Chinese herbal residue hydrochars [J]. *Coal Conversion*, 2024, 47(3): 45-48.

(上接第 100 页)

- [22] 文洁. MSE 与 MAE 对机器学习性能优化的作用比较 [J]. *信息与电脑*, 2018, (15): 42-43.
- Wen J. Comparison of the effects of MSE and MAE on machine learning performance optimization [J]. *Information and Computers*, 2018, (15): 42-43.
- [23] Clercq D D, Wen Z G, Fei Fan, et al. Interpretable machine learning for predicting biomethane production in industrial-scale anaerobic co-digestion [J]. *Science of the Total Environment*, 2020, 712: 134574. DOI: 10.1016/j.scitotenv.2019.134574.
- [24] Yang Z Q, Yang D H, Hua Y, et al. Dual optimization in anaerobic digestion of rice straw: Effects HRT and OLR coupling on methane production in one-stage and two-stage systems [J]. *Journal of Environmental Management*, 2024, 370: 123041. DOI: 10.1016/j.jenvman.2024.123041.
- [25] Appels L, Baeyens J, Degève J, et al. Principles and potential of the anaerobic digestion of waste-activated sludge [J]. *Progress in Energy and Combustion Science*, 2008, 34(6): 755-781.
- [26] Fuchs W, Wang Xuemei, Gabauer W, et al. Tackling ammonia inhibition for efficient biogas production from chicken manure: Status and technical trends in Europe and China [J]. *Renewable and Sustainable Energy Reviews*, 2018, 97: 186-199. DOI: 10.1016/j.rser.2018.08.038.
- [27] Peng X Y, Zhang S Y, Li L, et al. Long-term high-solids anaerobic digestion of food waste: effects of ammonia on process performance and microbial community [J]. *Bioresource Technology*, 2018, 262: 148-158. DOI: 10.1016/j.biortech.2018.04.076.
- [28] Li Q, Li H, Wang G J, et al. Effects of loading rate and temperature on anaerobic co-digestion of food waste and waste activated sludge in a high frequency feeding system, looking in particular at stability and efficiency [J]. *Bioresource technology*, 2017, 237: 231-239. DOI: 10.1016/j.biortech.2017.02.045.
- [29] Qin S Y, Wainaina S, Liu H M, et al. Microbial dynamics during anaerobic digestion of sewage sludge combined with food waste at high organic loading rates in immersed membrane bioreactors [J]. *Fuel*, 2021, 303: 121276. DOI: 10.1016/j.fuel.2021.121276.