DOI:10.11918/202512009

全球城市生活垃圾产生量智能算法与非线性演化特征

杨泽国[1,2],邹秋霞[1,2],段华波[2],杨家宽[2]

(1.华中科技大学 人工智能与自动化学院,武汉 430074;2.华中科技大学 环境科学与工程学院,武汉 430074)

摘 要:准确估算城市生活垃圾(MSW)产生量并识别其关键关联特征,是实现全球城市环境精细化管控及差异化治理的重要前提。然而,全球特别是发展中国家长时序数据的匮乏,以及传统模型对非线性关联特征解析能力的不足,导致现有研究难以精准还原全球MSW的历史产生规模与演化规律。为此,构建了基于XGBoost和K-Means++算法的全球MSW产生量智能估算与聚类分析框架。结果表明:XGBoost表现出优异的泛化能力,对未知国家MSW产生量的决定系数(R^2)达0.91±0.07;交叉验证策略对比证实,模型成功捕捉了社会经济特征对MSW产生量预测的内在关联,而非单纯依赖时间趋势,具备良好的空间泛化可靠性。SHAP解释分析证明,人口规模与经济体量是影响MSW产生量的核心关联因素,并揭示了人均GDP(按购买力平价,PPP)与产生量之间呈现独特的"N型"非线性关联轨迹,突破了传统线性外推模型的局限。测算结果表明,全球MSW产生总量由1991年的(12.9±0.4)亿t攀升至2021年的(21.4±1.6)亿t,年均增长1.7%。聚类分析识别出全球各国的农业生存型、新兴工业型、成熟稳定型及高耗富裕型四类典型演化模式,其中,新兴工业型国家是全球MSW增量的主要来源。构建的预测框架具备良好的普适性,可为全球固废精细化管理、区域差异化政策制定及基础设施规划提供科学依据。

关键词:生活垃圾;机器学习;SHAP解释;聚类分析;固废管理

中图分类号:X7 **文献标志码**:A **文章编号**:0367-6234(2026)06-0080-10

Intelligent algorithm and non-linear evolutionary characteristics of global municipal solid waste generation

YANG Zeguo[1,2], ZOU Qiuxia[1,2], DUAN Huabo[2], YANG Jiakuan[2]

(1. School of Artificial Intelligence and Automation, Huazhong University of Science and Technology, Wuhan 430074, China;
2. School of Environmental Science & Engineering, Huazhong University of Science and Technology, Wuhan 430074, China)

Abstract: Accurate estimation of municipal solid waste (MSW) generation and identification of its key associative characteristics are important prerequisites for achieving refined global urban environmental management and differentiated governance. However, the scarcity of long-time-series data globally, especially in developing countries, and the insufficient capability of traditional models in resolving non-linear associative characteristics make it difficult for existing studies to accurately reconstruct the historical generation scale and evolutionary laws of global MSW. To this end, this paper established an intelligent estimation and cluster analysis framework for global MSW generation based on XGBoost and K-Means++ algorithms. The results indicate that XGBoost exhibits excellent generalization capability, with a coefficient of determination (R^2) of 0.91±0.07 for MSW generation prediction in unknown countries; the comparison of cross-validation strategies confirms that the model successfully captures the intrinsic associations of socio-economic characteristics for MSW generation prediction, rather than merely relying on time trends, showing good spatial generalization reliability. SHAP interpretation analysis proves that population scale and economic volume are the core associative factors affecting MSW generation and reveals a unique "N-shaped" non-linear associative trajectory between GDP (in terms of purchasing power parity, PPP) per capita and generation, breaking the limitations of traditional linear extrapolation models. The estimation results indicate that the total global MSW generation has climbed from (1.29±0.04) billion t in 1991 to (2.14±0.16) billion t in 2021, with an average annual growth of 1.7%. Cluster analysis identifies four typical evolutionary patterns among global countries: agricultural subsistence, emerging industrial, mature stable, and high-consumption affluent patterns, among which emerging industrial countries are the main source of global MSW increment. The constructed prediction framework possesses good universal applicability and can provide a scientific basis for global solid waste's refined management, regionally differentiated policy formulation, and infrastructure planning.

收稿日期:2025-12-03;**录用日期**:2026-02-03;**网络首发日期**:2026-04-01
网络首发地址:https://link.cnki.net/urlid/23.1235.T.20260331.0947.004
基金项目:国家重点研发计划重点专项(2023YFC3902803)
作者简介:杨泽国(1996—),男,博士研究生;段华波(1981—),男,副教授,博士生导师
通信作者:段华波,huabo@hust.edu.cn

Keywords：municipal solid waste；machine learning；SHAP interpretation；cluster analysis；solid waste management

随着全球城市化进程的加速与新兴经济体消费水平的提升，城市生活垃圾（municipal solid waste，MSW）产生量迅速增长，给生态环境与公共健康带来了巨大压力，已成为制约全球可持续发展的关键问题之一[1-2]。准确量化全球特别是数据稀缺的发展中国家 MSW 产生量数据[3]，并探究其背后的社会经济关联特征，是制定差异化国际环境治理策略与实现城市减污降碳目标的迫切需求。

针对 MSW 产生量的估算和预测研究，学术界已经历了从传统计量经济学模型到人工智能技术的演进。早期研究多采用多元线性回归、系统动力学或灰色模型等，解析人口、GDP 等宏观指标与垃圾产生量的线性关联[4-6]。然而，MSW 产生是一个涉及社会、经济、人口及环境等多重因素耦合的复杂过程，传统模型难以捕捉变量间复杂的非线性关系。近年来，机器学习技术凭借其在处理高维非线性数据及挖掘潜在关系方面的显著优势，被广泛应用于 MSW 产生量预测[7-10]，如人工神经网络、支持向量机及长短期记忆网络等模型在特定城市或国家的短期预测中展现出优异性能[6,11-13]。

尽管机器学习在局部案例中已表现优异，但应用于全球尺度的预测任务时仍面临严峻挑战。现有研究多聚焦于数据完备的单一城市（如北京、上海）或特定国家（如美国、马来西亚）的案例分析[14-17]，模型严重依赖高频、高精度的本地化数据，难以在跨国尺度上处理全球各国间高度的区域异质性，导致模型缺乏空间泛化能力。同时，在涉及跨国面板数据的预测研究中，普遍存在验证策略的不足，大多沿用标准的随机 K 折交叉验证策略[18-21]，忽略了时序数据的自相关特征，高估了其在未知时间节点的真实预测性能。因此，如何构建一套兼顾时空泛化验证与全球尺度重构的综合方案，并利用机器学习系统解析全球格局下的关键影响因子关联特征，是当前该领域研究亟待探索的方向。

为解决上述问题，通过整合全球多源面板数据，构建了基于 XGBoost 和 K-Means + + 算法的 MSW 产生量智能估算与聚类分析框架。采用严格的时空互斥交叉验证策略评估模型捕捉社会经济结构横截面规律的能力；利用最优模型重构全球 159 个主要国家长时序 MSW 缺失数据，并结合 SHAP 可解释性算法与 K-Means + + 聚类分析解析关键关联因子的非线性演化特征及全球典型固废产生模式。旨在突破数据稀缺与传统线性假设的局限，明确适用于不同发展阶段国家的固废演化特征，为全球固废精细化管理、区域差异化政策制定及基础设施规划提供理论依据与数据支撑。

1　研究方法

1.1　数据来源与特征选择

基于 Hoy 等[6] 的研究成果，获取涵盖全球 43 个国家 1990—2020 年的 MSW 产生量数据。系统查询各国统计年鉴（如中国国家统计局、日本环境省）及区域性权威数据库（如欧盟统计局），将数据时间跨度更新至 2022 年，并将国家样本量扩充至 69 个。

在影响因子的选取上，参考 Smailbegovic 等[20] 关于区域 MSW 预测的指标体系，结合全球数据的可获取性，从世界银行世界发展指标（world development indicator，WDI）数据库中筛选了 170 个国家 1990—2022 年涵盖经济水平、人口特征、社会发展及环境指标 4 个维度的 14 个潜在因子（表 1）。

1.2　训练集与预测集构建

本研究剔除了特征严重缺失的年份及异常样本，并采用基于国家分组的线性插值法对部分非连续缺失数据进行补全，详细的处理流程及缺失值统计结果见附录及附表 S1、S2。经过预处理，最终构建了两个核心数据集。一是训练集，由上述扩充后的国家经异常样本筛选后得出，包含 65 个国家 1991—2021 年的 1 839 条影响因子以及 MSW 产生量，用于预测模型的训练、性能评估及可解释性分析。二是预测集，涵盖全球 159 个国家（包含训练集国家）1991—2021 年的 4 790 条影响因子，用于全球范围内 MSW 历史缺失数据的重构与聚类分析（各数据集国家名单见附表 S3）。

根据世界银行数据统计，上述 159 个样本国家涵盖了全球约 95% 的人口与 96% 的经济总量，在统计学意义上具备极强的全球代表性，能够有效反映全球尺度的 MSW 演化规律。

1.3　建模预处理

在模型训练前，对影响因子进行进一步的建模预处理：通过评估影响因子间的共线性及方差特性，剔除可能存在的共线性冗余因子以及低方差因子；针对连续型影响因子执行 Z-score 标准化以消除量纲差异；将地理区域等分类因子转化为独热编码；对目标变量（MSW 产生量）进行对数转换以改善数据分布的偏态特性，从而提升模型的拟合精度与泛化能力。

表 1 生活垃圾产生量影响因子

Tab. 1 Influencing factors of municipal solid waste generation

维度分类	指标名称	指标代码	单位	数据来源
数据属性	年份	Year	年	
数据属性	国家名称	Country Name	国家	
数据属性	地区	Region	地区分类	
数据属性	收入分组	Income Group	收入分组	
经济水平	GDP (PPP)	NY. GDP. MKTP. PP. CD	2017 年国际美元	世界银行
经济水平	人均 GDP (PPP)	NY. GDP. PCAP. PP. CD	2017 年国际美元	世界银行
经济水平	外商直接投资	BX. KLT. DINV. WD. GD. ZS	美元	世界银行
人口特征	总人口	SP. POP. TOTL	人	世界银行
人口特征	城市人口比例	SP. URB. TOTL. IN. ZS	%	世界银行
人口特征	总生育率	SP. DYN. TFRT. IN	每位女性生育数/人	世界银行
人口特征	劳动年龄人口占比	SP. POP. 1564. TO. ZS	占总劳动力比例/%	世界银行
人口特征	预期寿命	SP. DYN. LE00. IN	a	世界银行
社会发展	农业用地占比	AG. LND. AGRI. ZS	占土地面积比例/%	世界银行
社会发展	教育支出占比	NY. ADJ. AEDU. GN. ZS	占 GDP 比例/%	世界银行
社会发展	移动电话用户数	IT. CEL. SETS. P2	每百人订阅数	世界银行
社会发展	失业率	SL. UEM. TOTL. ZS	占总劳动力比例/%	世界银行
环境指标	二氧化碳排放	EN. GHG. CO2. MT. CE. AR5	百万吨 CO_2 当量/(Mt CO_2e)	世界银行
环境指标	可再生能源占比	EG. FEC. RNEW. ZS	占最终能源消费比例/%	世界银行
环境指标	人均二氧化碳排放	—	单位吨 CO_2 当量/ (t CO_2e·人$^{-1}$·a^{-1})	计算因子

注:国家名称不参与建模,人均二氧化碳排放为计算因子,仅用于聚类分析。

1.4 机器学习模型选择与性能评价

为评估不同算法在面板数据预测中的适用性,构建了基于 4 类典型算法的机器学习模型:1)基线线性模型,包括线性回归(Linear Regression,LR)与岭回归(Ridge Regression,RIDGE)算法;2)非参数与距离模型,包括支持向量机(Support Vector Machine,SVM)与 K 最近邻(K-Nearest Neighbors,KNN)算法;3)神经网络模型,采用多层感知机(Multilayer Perceptron,MLP)算法,探索数据深度非线性映射;4)集成树模型,包括随机森林(Random Forest,RF)、极端随机树(Extra Trees,ET)、极端梯度提升(eXtreme Gradient Boosting,XGBoost)以及轻量级梯度提升机(Light Gradient Boosting Machine,LightGBM)算法。

为量化评估不同机器学习模型在面板数据预测任务中的拟合精度与泛化性能,选取均方根误差(root mean square error, RMSE)与决定系数(coefficient of determination,R^2)作为核心评价指标(计算公式详见附录)。

1.5 数据集划分与交叉验证策略

为评估模型在未知区域与未来时间维度的泛化性能,设计了包含内部验证与外部测试的严格评估框架。首先,将建模训练集划分为 3 个互斥集合:Set A(训练验证集)包含 90% 的国家样本,用于模型拟合与参数调优;Set B(地理泛化测试集)包含 10% 未参与训练的独立国家样本,用于评估模型对未知区域的空间迁移能力;Set C(时间泛化测试集)包含训练集国家中后 15% 年份的数据,用于评估模型对未来趋势的时间外推能力。为消除单次随机划分带来的偶然性偏差,上述数据集划分及后续模型训练过程均基于 10 个不同的随机种子重复进行,最终结果取 10 次实验的评价指标均值与标准差来评估最优模型。

本研究采用基于扩展窗口模式的 5 折时间序列交叉验证策略[22](具体介绍见附录),并与随机交叉验证策略进行对比,定量揭示了不同验证机制对模型评估客观性的具体影响。

1.6 可解释性分析与模式识别

采用 SHAP(SHapley Additive exPlanations)算法对模型的预测结果进行可解释性分析[23],通过计算每个影响因子对预测结果的边际贡献值,量化关键驱动因子的影响力,并解析人均 GDP (按购买力平

价，PPP）、城市人口比例等关键因子与 MSW 产生量之间的关联特征与阈值效应。

采用 K-Means + + 算法对全球国家进行 MSW 产生模式识别[24]。参考世界银行报告以 2016 年基准年为起点[2]，选取 2016—2020 年近 5 年各国的人均 GDP（PPP）、城市人口比例、人均 CO_2 排放及人均 MSW 产生量的年均值以平滑年度随机波动并捕捉结构性特征，经标准化处理后输入算法进行聚类分析，随后利用聚类后不同簇国家的横截面梯度特征映射全球 MSW 产生量的动态演变规律。

1.7　不确定性分析

为了量化模型预测的不确定性，采用基于集成模型的评估策略。通过利用 10 次独立重复实验的结果，计算并呈现了各预测数值的 95% 置信区间，以反映因特征估算、参数初始化及地理外推带来的预测波动。

2　结果与讨论

2.1　特征共线性与有效性检验

各影响因子间的相关系数矩阵如图 1 所示。整体而言，绝大多数因子之间的相关系数处于较低水平，表明本研究选取的影响因子在整体上具有良好的独立性，有效避免了广泛的信息冗余。部分具有内在逻辑联系的特定因子间仍表现出较高的相关性，如 GDP 总量（PPP）与二氧化碳排放量[25]、劳动年龄人口占比与总生育率[26]，但其均未超过设定的 0.95 剔除阈值，因此，全部初选因子均保留。此外，方差筛选结果显示，归一化处理后的各影响因子方差均高于 0.05 的设定阈值，不存在无效的冗余特征，为后续机器学习模型深入挖掘变量间的复杂非线性关系提供了良好的数据基础。

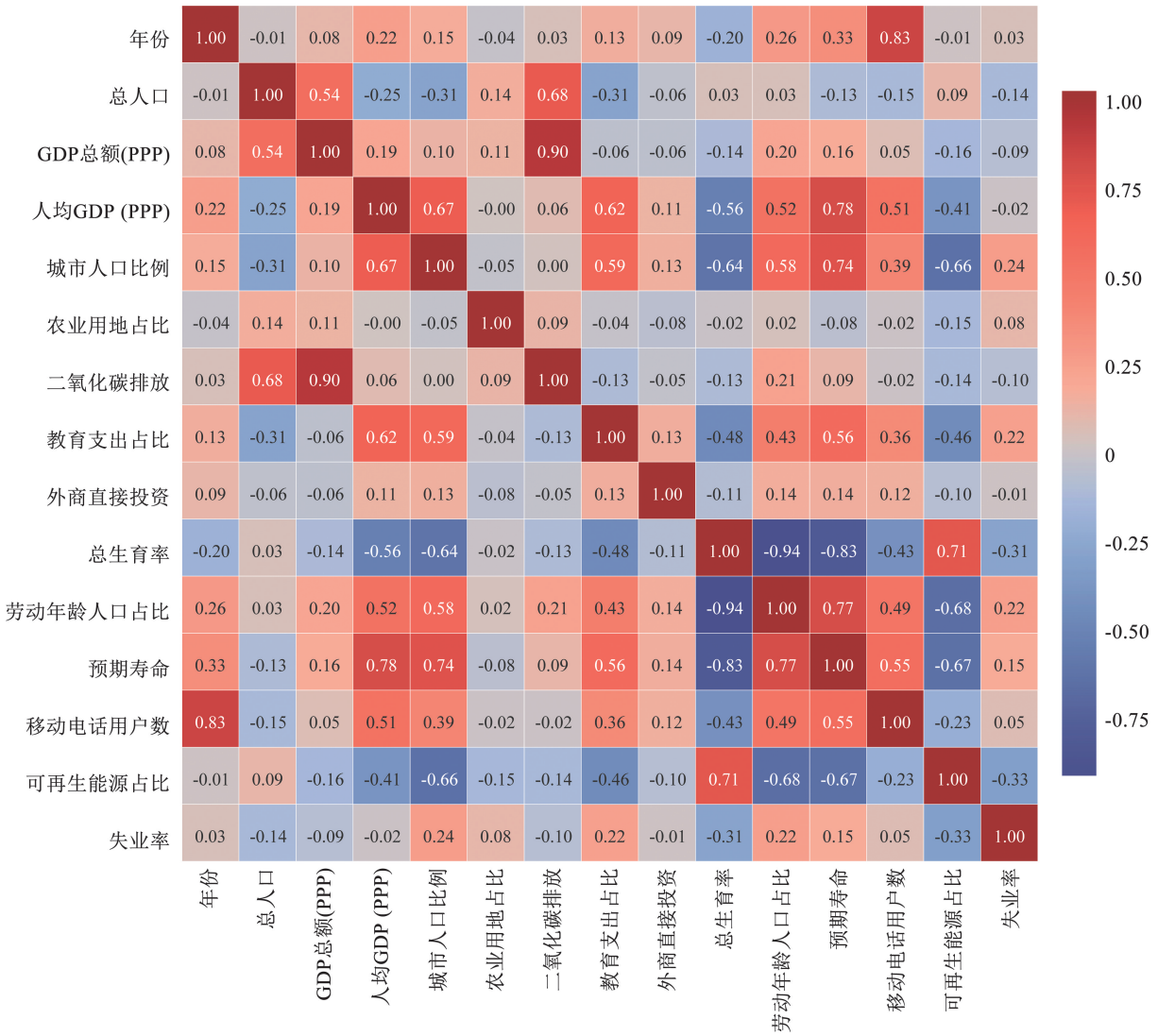
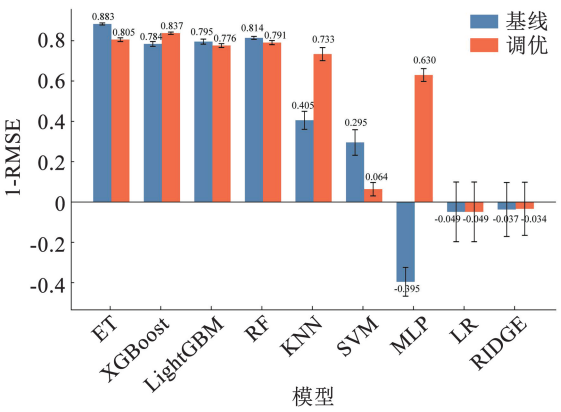


图 1　驱动因子的相关性热力图

Fig. 1　Correlation heatmap of driving factors

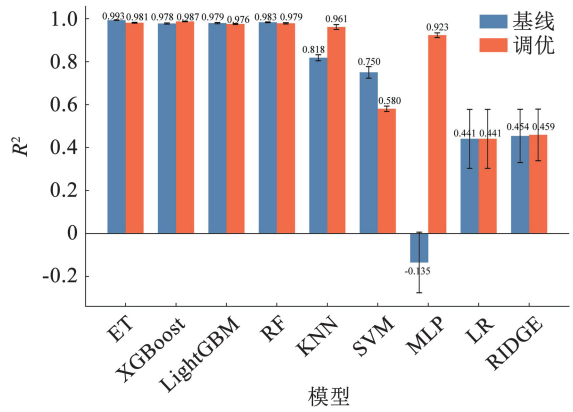
2.2 模型初步筛选与性能评价

基于时间序列交叉验证策略,对 9 种机器学习模型在默认及超参调优情况下进行了 10 次重复实验(具体的调优逻辑见附录)。结果如图 2 所示,集成树模型(ET、XGBoost、LightGBM、RF)性能显著优于其他算法。其中,ET 表现出最佳的默认参数拟合精度,其 1-RMSE 和 R^2 分别达 0.883 ± 0.005 与 0.993 ± 0.001;XGBoost 则在超参数调优后表现出最佳的综合预测性能,其 1-RMSE 和 R^2 分别提升至 0.837 ± 0.006 与 0.987 ± 0.002。相比之下,非参数与距离模型(SVM、KNN)与神经网络模型(MLP)无论是在默认参数还是调优后的性能均明显低于集成树模型。而作为基线的线性模型(LR、RIDGE)预测能力最弱,其 R^2 仅维持在 0.45 左右,证实线性假设难以捕捉复杂的非线性特征。



注:数值越高代表误差越小。

(a) 1-RMSE



注:数值越高代表拟合越好。

(b) R^2

图 2 9 种机器学习模型在时间序列交叉验证下的性能对比(10 次随机种子)

Fig. 2 Performance comparison of nine machine learning models under time-series cross-validation (10 random seeds)

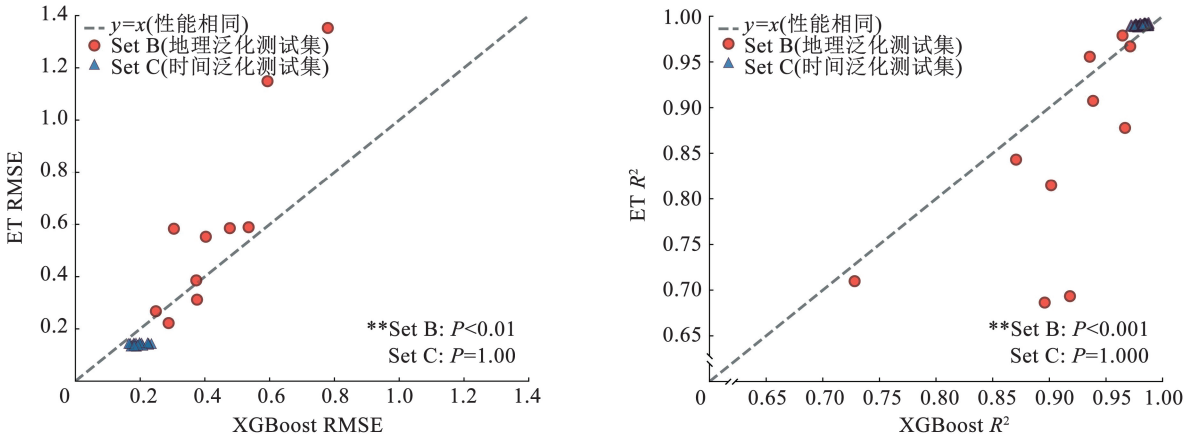
超参数调优对不同模型性能的影响存在显著差异,KNN、MLP 等模型调优后性能提升显著,ET 及 SVM 等模型则出现性能回落。由于 ET 模型在默认状态下接近 1.0 的拟合分数以及调优后性能下降的特征暗示其存在过拟合风险,选取默认状态最优的 ET 模型与调优性能最优的 XGBoost 模型进行外部泛化性测试,以甄别并确定最终用于全球数据重构的最优模型。

2.3 模型泛化能力评估与最优模型确立

基于 10 次随机划分实验,将 ET 与 XGBoost 模型置于外部测试集(Set B 与 Set C)中进行泛化性能力评估,结果如图 3 所示。在时间泛化测试集(Set C)中,两种模型的 RMSE 和 R^2 性能差异均不显著($P = 1.00$),散点紧密围绕对角线分布,表明二者在捕捉时间演变规律方面能力相当。然而,在地理泛化测试集(Set B)中,两者表现出截然不同的泛化性能,XGBoost 的平均 RMSE 为 0.44 ± 0.16,R^2 为 0.91 ± 0.07,而 ET 的平均 RMSE 为 0.60 ± 0.37,R^2 为 0.84 ± 0.11。XGBoost 的预测精度与拟合优度均优于 ET($P < 0.01$)。此外,ET 模型在部分实验批次中表现出明显的不稳定性(RMSE > 1,R^2 < 0.7),证实了其在训练阶段确实存在对局部特征的过拟合风险。相比之下,XGBoost 展现出更优的鲁棒性与抗噪能力,在保持时间维度高预测精度的同时,有效克服了空间维度的过拟合风险,更适合在数据稀缺情境下进行全球尺度数据重构[18]。因此,最终确立 XGBoost 为构建全球 MSW 预测框架的最优模型。

2.4 交叉验证策略对比与预测机制解析

基于 10 次重复实验,对比了随机与时序交叉验证策略对 XGBoost 性能的影响(图 4)。在内部验证阶段(Set A),随机划分的 RMSE 低于时间序列划分,且 R^2 接近 1.0,表现出极高的拟合优度。然而,当模型应用于未知国家(Set B)与未来年份(Set C)的外部独立测试时,随机划分策略的模型预测精度衰减至与时序划分策略相当的水平。这表明在具有显著时序特征的面板数据建模中,传统的随机划分策略可能因忽略时间维度的独立性,导致评估结果存在虚高风险[22]。相比之下,时序划分方式下模型在内部验证与外部测试中的性能表现更为一致,其评估指标更能客观反映模型的真实泛化能力。



注:点位于对角线上方表示XGBoost误差更小。

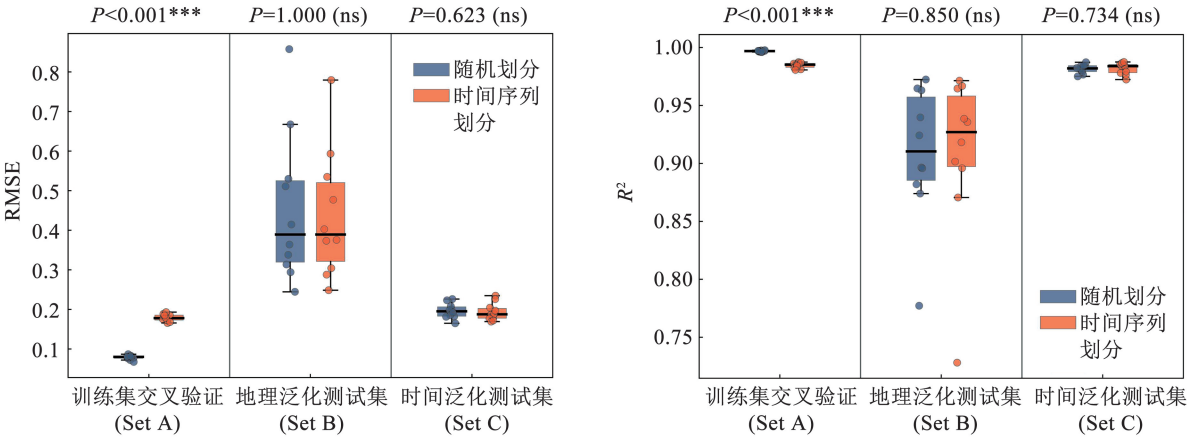
(a) RMSE对比

注:点位于对角线下方表示XGBoost拟合更好。

(b) R^2对比

图 3　ET 与 XGBoost 模型在地理与时间泛化测试集上的性能对比

Fig. 3　Performance comparison of ET and XGBoost models on geographic and temporal generalization test sets



注: ns表示无显著差异，***表示$P<0.001$。

(a) RMSE对比

(b) R^2对比

图 4　随机划分与时间序列划分策略下的 XGBoost 模型性能对比

Fig. 4　Performance comparison of XGBoost model under random split and time-series split strategies

同时,两种交叉验证策略在时空外推的预测性能一致性($P>0.5$)表明,时间序列并非模型预测的核心驱动力。初步反映人口、经济、社会等特征[2,27]与 MSW 产生量间的非线性关联规律,更有助于模型在跨时空维度上保持可靠的泛化能力。

此外,为系统评估模型在不同情境下的泛化能力,在附录中补充了 XGBoost 模型在不同地区/收入组别地理与时间泛化测试集上的性能统计(详见附表 S4 和 S5)。结果表明,模型对不同发展阶段国家的预测精度存在一定异质性(具体讨论见附录)。

2.5　模型解释性分析

基于 SHAP 算法的模型解释性分析结果如图 5 所示。全局重要性排序(图 5(a))显示,总人口与 GDP 总额(PPP)对模型输出的绝对平均 SHAP 值最高,证实了人口规模与经济体量是决定一个国家 MSW 产生总量的首要驱动因素[28]。二氧化碳排放与城市人口比例也位居前列,表明资源消耗强度与城乡结构转型也是影响 MSW 增量的关键结构性因子[29]。

SHAP 依赖图进一步显示了经济发展与产生量之间的非线性驱动关系。GDP 总额(PPP)依赖图(图 5(b))显示,随着经济体量的扩张,MSW 产生量呈现边际效应递减的对数增长趋势,表明当经济体量达到一定阈值后,单纯的经济总量扩张对 MSW 增量的驱动作用逐渐趋缓。不同于以往[2]收入与 MSW 产生量呈正相关的结论,人均 GDP(PPP)依赖图(图 5(c))结果揭示了 MSW 产生量独特的"N 型"三阶段演变规律[30]。在低收入的生存型增长阶段,SHAP 值随人均 GDP(PPP)的增加陡峭上升。进入中等收入阶段后,曲线出现震荡平台期,表明此时经济增长不再单一主导 MSW 增量。而在高收入

阶段,SHAP 值的演变呈现双轨分化特征:少部分样本点随经济增长而回落的趋势验证了传统环境库兹涅茨曲线的“倒 U 型”脱钩假设[31],表明通过政策干预与技术进步实现经济增长与 MSW 产生的相对脱钩是可行的[32];大部分样本点在极高收入阶段继续攀升的现象则对传统理论进行了重要补充,揭示了在高消费经济体中可能存在的脱钩失效风险,仅靠经济增长并不足以自动实现垃圾减量[30]。

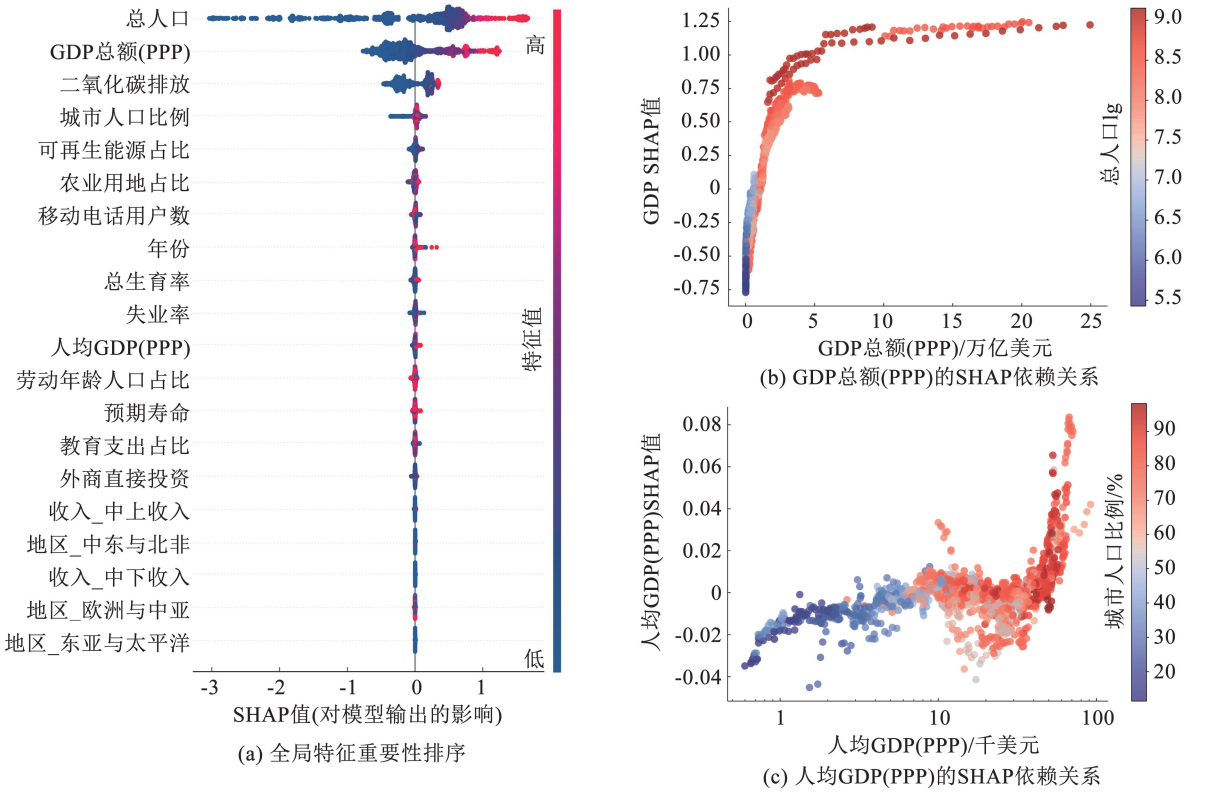


图 5 驱动因子的 SHAP 解释性分析

Fig. 5 SHAP interpretability analysis of driving factors

2.6 全球 MSW 产生量测算与现状分析

利用 XGBoost 模型填补了全球 159 个国家在 1991—2021 年的数据空白,实现了全球长时序 MSW 产生量缺失数据的全景重构。重构结果(表 2)显示,全球 MSW 产生量由 1991 年的(12.9 ± 0.4)亿 t 攀升至 2021 年的(21.4 ± 1.6)亿 t(年均增长 1.7%)。分收入组别看,低收入国家受人口激增驱动,MSW 产生量年均增速最高(3.0%);中等收入国家凭借庞大人口及经济基数贡献了全球绝大部分 MSW 增量[2];而高收入国家受源头减量政策及人口停滞影响,MSW 产生量年均增速仅 0.9%,增长趋于平缓。

表 2 不同收入组别国家 MSW 产生量及人均水平演变(1991—2021 年)

Tab. 2 Evolution of MSW generation and per capita levels in countries of different income groups (1991—2021)

收入组别	MSW 产生量						总量年均增长率/%
	1991 年总量/亿 t	2010 年总量/亿 t	2021 年总量/亿 t	1991 年人均/(t·人$^{-1}$·a^{-1})	2010 年人均/(t·人$^{-1}$·a^{-1})	2021 年人均/(t·人$^{-1}$·a^{-1})	
高收入	5.9 ± 0.1	6.8 ± 0.1	7.8 ± 1.1	0.52	0.53	0.58	0.9
中等偏上收入	3.8 ± 0.1	5.5 ± 0.1	7.1 ± 0.5	0.18	0.22	0.26	2.1
中等偏下收入	2.7 ± 0.2	4.3 ± 0.1	5.4 ± 0.9	0.17	0.18	0.19	2.3
低收入	0.5 ± 0.1	0.8 ± 0.2	1.1 ± 0.2	0.24	0.22	0.21	3.0
全球总计	12.9 ± 0.4	17.4 ± 0.4	21.4 ± 1.6	0.25	0.26	0.29	1.7

注:“总量年均增长率”为 1991—2021 年的复合年均增长率。

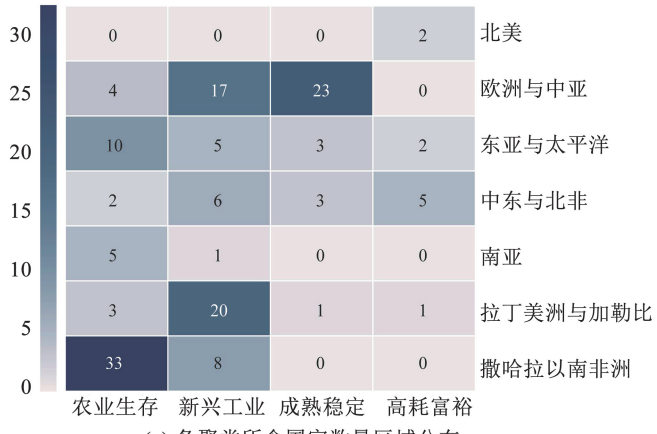
此外,本研究估算结果与世界银行 2016 年(20.1 亿 t)[2]及联合国环境规划署 2020 年估算的全球 MSW 产生量(约 21 亿 t)均高度吻合[1]。这证明了机器学习不仅能识别多维特征之间的非线性关

联关系，同样具备可靠的历史产生量估算能力，有效克服了传统模型预设 MSW 产生量单调增长的局限。

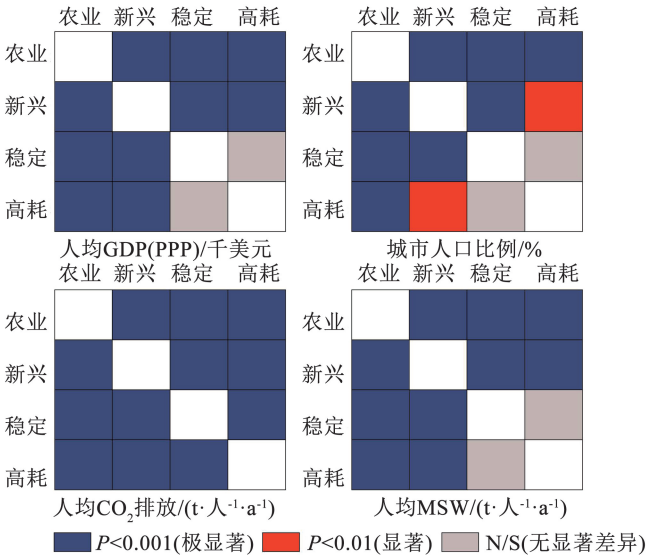
2.7　全球 MSW 产生模式聚类与时空演变

基于重构后的各国数据，选取 SHAP 分析识别出的核心影响因子作为聚类基础。为了剥离国家体量带来的规模效应，将总量指标转化为人均强度指标。具体选取 2016—2020 年各国的人均 GDP（PPP）、城市人口比例、人均 CO_2 排放量及人均 MSW 产生量的年均值构建特征向量，采用 K-Means + + 算法将全球 159 个国家划分为 4 个聚类簇（详细方法参考附录）。图 6（a）展示了 4 个聚类簇的全球国家数量地理分布，呈现出显著的地缘经济集聚特征。第 1 类聚类簇主要集中于撒哈拉以南非洲与南亚地区，构成了全球最不发达国家板块；第 2 类聚类簇广泛覆盖欧洲与中亚、东亚及拉丁美洲与加勒比的新兴经济体，构成了全球新兴工业化经济体板块；第 3 类聚类簇高度集聚于欧洲大陆及东亚发达国家，对应传统工业化核心区；第 4 类聚类簇则主要分布于北美、大洋洲及中东产油国，对应资源丰富区。

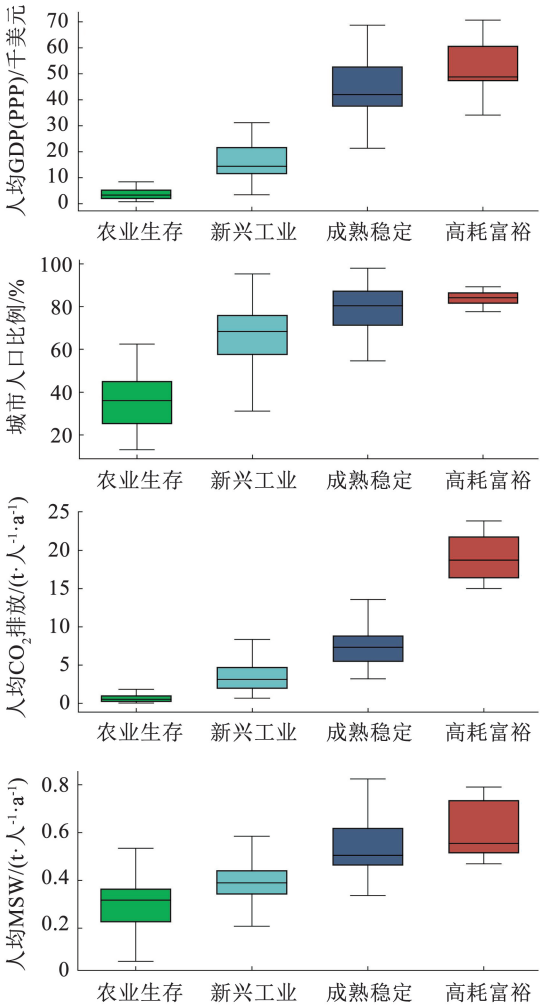
依据图 6（b）显著性热力图与图 6（c）箱线图揭示的结构性差异，对上述 4 个聚类簇进行了特征界定与命名[33-34]。第 1 类聚类簇特征为低城市人口比例与低人均 GDP（PPP），以印度及撒哈拉以南非洲等发展中国家为代表，定义为农业生存型；第 2 类聚类簇表现为城市化率的快速攀升与人均 GDP（PPP）的中速增长，以中国、东南亚及部分新兴经济体为代表，定义为新兴工业型[35]；第 3 类与第 4 类聚类簇体现了高收入国家的路径分化，以欧盟、日韩为代表的第 3 类聚类簇国家在保持高经济水平的同时，受政策约束实现了低碳排放[31,36]，定义为成熟稳定型；以美国、澳大利亚及资源型富国为代表的第 4 类聚类簇国家，人均二氧化碳排放极显著高于第 3 类（$P<0.001$），反映了以高能耗、高废弃为特征的资源密集型消费模式[37]，故定义为高耗富裕型。



(a) 各聚类所含国家数量区域分布



(b) 各聚类模式特征差异显著性检验

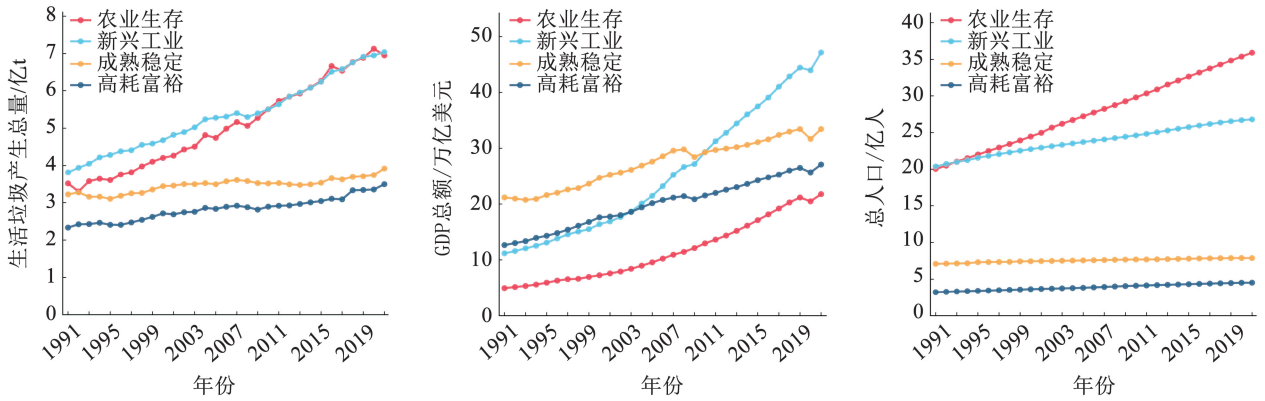


(c) 各聚类模式特征分布箱线图

图 6　全球 MSW 产生模式的空间分布、特征差异及统计显著性检验

Fig. 6　Spatial distribution, characteristic differences, and statistical significance tests of global MSW generation patterns

图 7 清晰揭示了上述 4 类模式下 MSW 增长的异质性演化特征。如图 7(a)、(b)所示,新兴工业型的 MSW 增长轨迹与 GDP 增长曲线高度耦合,呈现出同步的上升趋势,表明经济总量扩张与消费能力提升是其 MSW 激增的主导力量。农业生存型虽然 GDP(PPP)总额处于低位,但其 MSW 产生总量持续攀升,甚至逼近新兴工业型。结合图 7(c)可知,这主要因为该类国家庞大的人口基数及其快速增长的趋势,即便人均消费水平较低,巨大的人口增量仍将对全球 MSW 管理系统构成严峻挑战。成熟稳定型国家表现出独特的演变特征,尽管其 GDP 仍保持增长态势,但 MSW 产生量已趋于平稳。这种经济增长与 MSW 产生的背离,反映了欧盟及日韩等国家在源头减量与循环经济政策干预下的脱钩成效[31,36],但也可能源于产业转移带来的潜在统计假象,部分国家通过制造业外迁实现了本土产生量的结构性减量,而非完全意义上的源头减量[38]。



(a) MSW产生总量趋势 (b) GDP总额趋势 (c) 总人口趋势

图 7 全球 4 类模式国家的 MSW 产生总量、GDP 总额(PPP)及人口规模历史演变趋势(1991—2021 年)

Fig. 7 Historical evolution trends of total MSW generation, total GDP (PPP), and population scale of countries with four global patterns (1991—2021)

综上,全球 MSW 管理重心正向全球南方国家[39](农业生存型与新兴工业型)转移。高收入国家内部排放模式的分化,为正处于发展窗口期的南方国家提供了路径参照,其固废治理关键在于通过早期政策干预避免高消耗路径锁定。鉴于 MSW 与碳排放、资源消耗强度的高度耦合性,这一路径抉择将直接决定全球南方国家在碳中和背景下的低碳转型进程与固废治理成效。

3 结 论

1)构建了具备良好泛化能力的全球 159 个主要国家 MSW 产生量预测框架。基于 XGBoost 算法的机器学习模型在地理外部测试集上表现出综合最优性能(R^2 为 0.91 ±0.07);交叉验证策略对比证实,模型成功捕捉了驱动 MSW 产生的横截面社会经济结构规律,而非依赖时间惯性,有效解决了数据稀缺情境下的空间外推难题。

2)揭示了关键影响因子的非线性演化特征与全球增长态势。SHAP 分析证明人口与经济体量是核心相关因素,并发现人均 GDP(PPP)与 MSW 产生量之间呈现独特的“N 型”非线性关联轨迹,突破了传统线性假设;数据重构结果表明,1991—2021 年全球 MSW 产生总量由(12.9 ±0.4)亿 t 攀升至(21.4 ±1.6)亿 t,年均增长 1.7%。

3)识别了全球 MSW 演化模式并指出了行业治理方向。聚类分析识别出全球各国的农业生存型、新兴工业型、成熟稳定型及高耗富裕型 4 类典型演化模式,全球 MSW 管理重心正向全球南方国家转移。本研究构建的预测框架及发现的演化规律,可为各国制定匹配其发展阶段的固废管理政策及基础设施建设规划提供科学参考。

附 录

附录内容请访问 https://gitee.com/ipfas/global-msw-intelligent-projection/tree/master。

参考文献

[1] ISWA (International Solid Waste Association). Global waste management outlook 2024: beyond an age of waste: turning rubbish into a resource[R]. Environmental Science, 2024. DOI: 10.59117/20.500.11822/44939

[2] KAZA S, YAO L, BHADA-TATA P, et al. What a waste 2.0: a global snapshot of solid waste management to 2050[M]. Washington D.C.: World Bank Publications, 2018. DOI: 10.1596/978-1-4648-1329-0

[3] SUBEDI A, SHRESTHA S, GHIMIRE A, et al. Leveraging machine learning for sustainable solid waste management: a global perspective[J]. Sustainable Futures, 2025, 10: 101098. DOI: 10.1016/j.sftr.2025.101098

[4] BILLAL M M, KUMAR A. Forecasting residential and nonresidential solid waste generation, disposal, and diversion using three machine learning approaches[J]. Biofuels, Bioproducts and Biorefining, 2025, 19(6): 2179. DOI:10.1002/bbb.70010

[5] FATIH F. Hybrid modeling techniques for municipal solid waste forecasting: an application to OECD countries[J]. Statistics, Politics

and Policy, 2024, 15(3): 287. DOI:10.1515/spp-2024-0008
[6] HOY Z X, WOON K S, CHIN W C, et al. Curbing global solid waste emissions toward net-zero warming futures[J]. Science, 2023, 382(6672): 797. DOI: 10.1126/science.adg3177
[7] EL JAOUHARI A, SAMADHIYA A, KUMAR A, et al. Turning trash into treasure: exploring the potential of AI in municipal waste management-an in-depth review and future prospects[J]. Journal of Environmental Management, 2025, 373: 123658. DOI: 10.1016/j.jenvman.2024.123658
[8] HOY Z X, PHUANG Z X, FAROOQUE A A, et al. Municipal solid waste management for low-carbon transition: a systematic review of artificial neural network applications for trend prediction [J]. Environmental Pollution, 2024, 344: 123386. DOI: 10.1016/j.envpol.2024.123386
[9] ZOU Qiuxia, DUAN Huabo, YANG Zeguo, et al. Machine learning-assisted solid waste life-cycle management: applications, constrains, and future opportunities [J]. Resources, Conservation and Recycling, 2025, 219: 108320. DOI: 10.1016/j.resconrec.2025.108320
[10] VARIAN H R. Big data: new tricks for econometrics[J]. Journal of Economic Perspectives, 2014, 28(2): 3. DOI: 10.1257/jep.28.2.3
[11] ZHAO Ying, TAO Zhe, LI Ying, et al. Prediction of municipal solid waste generation and analysis of dominant variables in rapidly developing cities based on machine learning: a case study of China [J]. Waste Management and Research, 2024, 42(6): 476. DOI: 10.1177/0734242X231192766
[12] BATOOL T, GHAZALI R, NUREIZE N B, et al. An enhanced GRU model with optimized update gate: a novel approach in municipal solid waste prediction [J]. International Journal of Intelligent Engineering and Systems, 2024, 17(5): 33. DOI:10.22266/ijies2024.1031.04
[13] JASSIM M S, COŞKUNER G, SULTANA N, et al. Forecasting domestic waste generation during successive COVID-19 lockdowns by bidirectional LSTM super learner neural network[J]. Applied Soft Computing, 2023, 133: 109908. DOI: 10.1016/j.asoc.2022.109908
[14] LATIF S D, HAZRIN N A B, YOUNES M K, et al. Evaluating different machine learning models for predicting municipal solid waste generation: a case study of Malaysia[J]. Environment, Development and Sustainability, 2024, 26(5): 12489. DOI:10.1007/s10668-023-03882-x
[15] CUI Yaxin, CHIN M Y, LOH H S, et al. Forecasting municipal solid waste removal volume based on socioeconomic indicators for carbon reduction strategy in Beijing's waste management [J]. Chemical Engineering Transactions, 2024, 114: 451. DOI:10.3303/CET24114076
[16] WANG Deyun, YUAN Yingan, BEN Yawen, et al. Long short-term memory neural network and improved particle swarm optimization-based modeling and scenario analysis for municipal solid waste generation in Shanghai, China[J]. Environmental Science and Pollution Research, 2022, 29(46): 69472. DOI:10.1007/s11356-022-20438-0
[17] JOHNSON N E, IANIUK O, CAZAP D, et al. Patterns of waste generation: a gradient boosting model for short-term waste prediction in New York City[J]. Waste Management, 2017, 62: 3. DOI: 10.1016/j.wasman.2017.01.037
[18] ZHANG Chenyi, DONG Huijuan, GENG Yong, et al. Machine learning-based prediction for China's municipal solid waste under the shared socioeconomic pathways[J]. Journal of Environmental Management, 2022, 312: 114918. DOI: 10.1016/j.jenvman.2022.114918
[19] DUNKEL J, CARRETA D D, BORZDYNSKI Ó G, et al. Solid waste analysis using open-access socio-economic data [J]. Sustainability (Switzerland), 2022, 14(3): 1233. DOI: 10.3390/su14031233
[20] SMAILBEGOVIC U, KADRIC E, GLÖSER-CHAHOUD S. Assessing the determinants of municipal solid waste generation in Europe: a machine learning approach[J]. Waste Management, 2025, 205: 115033. DOI: 10.1016/j.wasman.2025.115033
[21] CAO Qilin, SONG Junnian, LIU Chaoshuo, et al. Expediting co-benefits of tailored municipal solid waste management strategies globally[J]. Nature Sustainability, 2025, 8(10): 1164. DOI:10.1038/s41893-025-01613-w
[22] BERGMEIR C, BENÍTEZ J M. On the use of cross-validation for time series predictor evaluation[J]. Information Sciences, 2012, 191: 192. DOI: 10.1016/j.ins.2011.12.028
[23] 马珍，叶程，魏卿，等. 基于机器学习的加载絮凝效能优化与机理解析[J/OL]. 哈尔滨工业大学学报，(2025-11-17). https://link.cnki.net/urlid/23.1235.T.20251117.1421.005
MA Zhen, YE Cheng, WEI Qing, et al. Machine learning-based optimization and mechanistic analysis of ballasted flocculation[J/OL]. Journal of Harbin Institute of Technology, (2025-11-17). https://link.cnki.net/urlid/23.1235.T.20251117.1421.005
[24] 黄永基，谢丹平，李世平，等. 基于 K-means 算法的生活垃圾焚烧设施烟气二恶英防控[J]. 环境工程学报，2022，16(12)：3993
HUANG Yongji, XIE Danping, LI Shiping, et al. Prevention and control of dioxins emission in domestic waste incineration facilities based on K-means algorithm[J]. Chinese Journal of Environmental Engineering, 2022, 16(12): 3993. DOI: 10.12030/j.cjee.202207056
[25] KAYA Y. Impact of carbon dioxide emission control on GNP growth: interpretation of proposed scenarios[R]. Intergovernmental panel on climate change/response strategies working group, 1989
[26] BLOOM D E, WILLIAMSON J G. Demographic transitions and economic miracles in emerging Asia[J]. The World Bank Economic Review, 1998, 12(3): 419. DOI:10.1093/wber/12.3.419
[27] BEIGL P, LEBERSORGER S, SALHOFER S. Modelling municipal solid waste generation: a review[J]. Waste Management, 2008, 28(1): 200. DOI: 10.1016/j.wasman.2006.12.011
[28] EHRLICH P R, HOLDREN J P. Impact of population growth[J]. Science, 1971, 171(3977): 1212
[29] HOORNWEG D, BHADA-TATA P, KENNEDY C. Environment: waste production must peak this century[J]. Nature, 2013, 502(7473): 615. DOI: 10.1038/502615a
[30] WANG Qiang, LI Yuanfan, LI Rongrong. Rethinking the Environmental Kuznets Curve hypothesis across 214 countries: the impacts of 12 economic, institutional, technological, resource, and social factors[J]. Humanities and Social Sciences Communications, 2024, 11(1): 1. DOI: 10.1057/s41599-024-02736-9
[31] MAZZANTI M, ZOBOLI R. Waste generation, waste disposal and policy effectiveness: evidence on decoupling from the european union[J]. Resources, Conservation and Recycling, 2008, 52(10): 1221. DOI: 10.1016/j.resconrec.2008.07.003
[32] GHISELLINI P, CIALANI C, ULGIATI S. A review on circular economy: the expected transition to a balanced interplay of environmental and economic systems [J]. Journal of Cleaner Production, 2016, 114: 11. DOI: 10.1016/j.jclepro.2015.09.007
[33] ROSTOW W W. The stages of economic growth: a non-communist manifesto (1960)[M]. The Globalization and Development Reader: Perspectives on Development and Global Change, 2014: 52
[34] KUZNETS S. Economic growth and income inequality [M]. London: Routledge, 2019: 25
[35] KURNIAWAN T A, LIANG X, O'CALLAGHAN E, et al. Transformation of solid waste management in China: moving towards sustainability through digitalization-based circular economy [J]. Sustainability, 2022, 14(4): 2374. DOI: 10.3390/su14042374
[36] 孙文君. 日本生活垃圾源头减量的政策经验及启示[J]. 海南大学学报(人文社会科学版)，2019，37(5)：85
SUN Wenjun. Policy experience of source-reduced municipal solid waste in Japan and its inspiration [J]. Humanities & Social Sciences Journal of Hainan University, 2019, 37(5): 85. DOI: 10.15886/j.cnki.hnus.2019.05.012
[37] GALBRAITH J K. The affluent society[M]. Boston: Houghton Mifflin Harcourt, 1998: 1
[38] WIEDMANN T O, SCHANDL H, LENZEN M, et al. The material footprint of nations[J]. Proceedings of the National Academy of Sciences, 2015, 112(20): 6271. DOI: 10.1073/pnas.122036211
[39] DADOS N, CONNELL R. The global south[J]. Contexts, 2012, 11(1): 12. DOI: 10.1177/1536504212436479

（编辑　刘　彤）