

工业技术

基于机器学习的有机胺吸收 CO₂ 性能预测

艾孜买提江·艾尔肯,代 琴,王亚因,张天琦,焦燕麟,张稚杰,汪黎东,于广飞*
(华北电力大学环境科学与工程学院,教育部资源环境系统优化重点实验室,北京 102206)

摘要:针对传统开发方式周期长、成本高、效率低的问题,从22种有机胺 CO₂ 吸收体系参数中筛选并采用了10种关键特征变量,通过7种不同的机器学习方法构建了常见有机胺溶剂吸收 CO₂ 性能的预测模型。研究表明,GBDT模型具有最佳的预测表现。基于SHAP、PI及MDI方法的特征重要性分析结果显示,二氧化碳分压、温度、胺基百分比是影响 CO₂ 吸收效果的关键特征,并以此揭示了其优势区间。

关键词:CO₂ 捕集;机器学习;预测模型;有机胺;液相吸收法

中图分类号:X701;TP181

文献标志码:A

文章编号:0253-4320(2025)S2-0439-05

DOI:10.16606/j.cnki.issn0253-4320.2025.S2.075

Machine learning based prediction on performance of organic amine in absorbing CO₂

AIERKEN Ai-zi-mai-ti-jiang, DAI Qin, WANG Ya-nan, ZHANG Tian-qi, JIAO Yan-lin,
ZHANG Zhi-jie, WANG Li-dong, YU Guang-fei*

(Resources and Environment System Optimization Key Laboratory of Ministry of Education,
College of Environmental Science and Engineering, North China Electric Power University, Beijing 102206, China)

Abstract: In order to solve the long cycle, high cost and low efficiency problems of traditional development methods, 10 key characteristic variables are selected from 22 kinds of parameters in organic amine-route CO₂ absorption system, and a prediction model for CO₂ absorption performance of common organic amine solvent is constructed through 7 kinds of machine learning methods. It is found that GBDT model presents the best prediction performance. The characteristic importance analysis results based on SHAP, PI and MDI methods show that the partial pressure of carbon dioxide, temperature, and amino group percentage are the key characteristics that affect the CO₂ absorption effect, and thus their advantage ranges are revealed.

Key words: CO₂ capture; machine learning; prediction model; organic amine; liquid phase absorption method

随着全球经济的快速发展,人类对能源的需求急剧增加,特别是以石油、煤炭、天然气为主的化石能源。然而,化石能源的燃烧导致了大量温室气体的产生。二氧化碳(CO₂)作为主要温室气体加剧了全球的气候变化,造成海平面上升、极地冰川融化等问题。为有效减少碳排放,CO₂的捕集、利用和封存技术(CCUS)的实施与发展至关重要^[1]。其中,CO₂捕集是CCUS技术中的重要部分,而使用有机胺溶剂的化学吸收法则是目前工业应用上最成熟的CO₂捕集技术之一^[2]。

高性能有机胺CO₂吸收体系的开发是化学吸收法的关键,例如单乙醇胺(MEA)、甲基二乙醇胺(MDEA)和哌嗪(PZ)等有机胺溶剂都已被用于开

发高性能有机胺CO₂吸收体系^[3-4]。然而,通过实验试错法设计和建立高性能有机胺CO₂吸收体系仍然存在周期长、成本高、效率低的问题。近年来,随着机器学习技术的迅速发展,已有许多研究通过构建吸收溶剂物理化学性质与其吸收能力间的构效关系模型,实现高效吸收体系的性能预测与开发^[3]。这也为高性能有机胺CO₂吸收体系的筛选与发展提供了新的思路。

考虑到目前在基于有机胺溶剂的CO₂吸收研究中,机器学习的应用在算法类型、描述符选择范围及重要特征分析方法等方面仍较为局限,采用了7种代表性机器学习算法和10种特征描述符,构建有机胺溶剂吸收CO₂的预测模型。在此基础上,结合

收稿日期:2025-03-10;修回日期:2025-07-25

基金项目:北京市自然科学基金青年基金项目(8254059);国家自然科学基金青年科学基金项目(52200126);国家自然科学基金国家杰出青年科学基金项目(52325004)

作者简介:艾孜买提江·艾尔肯(1998-),男,硕士生;于广飞(1993-),男,博士,讲师,研究方向为碳捕集、利用与封存技术,通讯联系人, gfyu@ncepu.edu.cn。

多种特征重要性分析方法与三维空间散点图、二维散点平面映射等手段,进一步揭示影响 CO₂ 吸收效果的关键因素并以此提出优势区间。该研究可为有机胺 CO₂ 吸收体系的设计与开发提供理论支撑与方法支持,助力我国碳中和、碳达峰战略目标的实现。

1 研究方法原理

1.1 机器学习算法

机器学习算法的选择是影响模型预测效果的关键。本研究选择树模型、神经网络及经典回归模型 3 类(共 7 种)具有代表性的机器学习算法进行建模并系统对比性能差异,图 1 展示了具体研究流程。

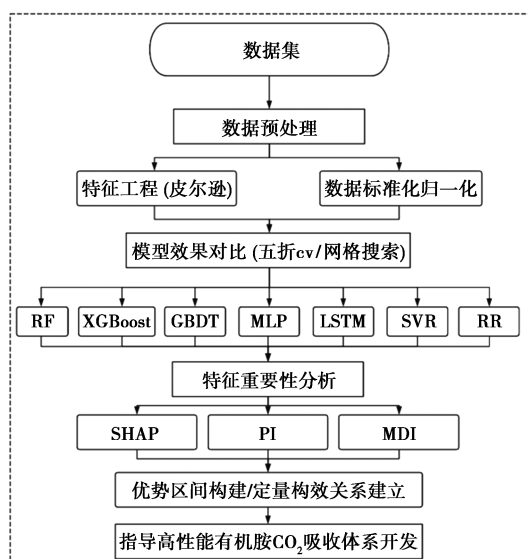


图 1 基于机器学习预测有机胺 CO₂ 吸收体系性能的流程

树模型作为一类机器学习领域中的经典算法,算法逻辑决定其在中小型数据集和模型可解释性上具有天然优势。选取了 3 种主流树模型进行对比:随机森林(RF)、极限梯度提升(XGBoost)和梯度提升决策树(GBDT)。值得注意的是,GBDT 方法可通过梯度提升集成决策树,具有高预测精度、强大的非线性关系建模能力。这些优点使得 GBDT 在环境行业污染物预测相关研究中被普遍应用^[5-6]。

神经网络(NN)是一种模仿生物神经网络的结构和功能的计算模型。在本研究中,选用多层感知机(MLP)和长短期记忆网络(LSTM)2 类 NN 模型。其中,MLP 作为基础的神经网络模型,是在过往研究中最广泛使用的模型之一^[7]。而 LSTM 则是一类较为高阶的神经网络模型,尽管参数规模较大、训练

成本较高,但凭借对复杂大量数据的建模能力,常能在数据中实现更高的预测精度。

岭回归(RR)和支持向量回归(SVR)是 2 类简单、高效的典型回归方法,由于建模快、参数少,无需复杂调优,使得这 2 类经典算法在环境领域被使用^[8-9]。

1.2 特征重要性分析方法

考虑到机器学习模型通常被认为是黑盒模型,因此需要特征重要性分析为模型提供可解释性,揭示各个输入特征在有机胺 CO₂ 吸收体系中的重要性程度。本研究使用 MDI(平均不纯度减少)、PI(排列重要性)和 SHAP(Shapley 加法解释)方法分别对机器学习模型进行特征重要性分析,通过不同的分析视角研究每个特征对整体模型预测的重要性。

其中,MDI 方法通过每个特征在决策树节点分裂时带来的不纯度减少计算每个特征在所有决策树中的 MDI 值。MDI 值越大,说明该特征对模型预测结果的影响越大;PI 方法通过随机打乱某特征值,观察模型预测准确性变化,量化其对模型决策的关键影响;SHAP 则是一种基于博弈论的分析方法,可通过计算每个特征对模型预测的贡献,从而允许对黑盒模型进行全局和局部解释。

1.3 模型评价标准

决定系数(R^2)、均方根误差(RMSE)和平均绝对误差(MAE)用于评估机器学习模型的预测准确性。 R^2 、RMSE 和 MAE 值的计算公式如式(1)~(3)。其中 R^2 取值范围为[0,1],值越接近 1 说明拟合程度越高, RMSE 与 MAE 的值则与预测变量值相关,越低的 RMSE 和 MAE 代表了越小的模型误差。

$$R^2 = 1 - \left[\sum_{n=1}^N (\hat{y} - y)^2 \right] / \left[\sum_{n=1}^N (\hat{y} - \bar{y})^2 \right] \quad (1)$$

$$RMSE = \sqrt{\left[\sum_{n=1}^N (\hat{y} - y)^2 \right] / N} \quad (2)$$

$$MAE = (1/N) \sum_{n=1}^N |\hat{y} - y| \quad (3)$$

式中, \hat{y} 、 y 和 \bar{y} 分别是目标特征的预测值、真实值和平均值; N 是数据点的总数。

2 结果与讨论

2.1 数据搜集与预处理

首先,使用多种主题关键词(包括 CO₂ 吸收、CO₂ 捕集、有机胺吸收剂等)和主流数据库(例如 Web of Science 和 Google Scholar)对使用有机胺溶

剂吸收CO₂的相关研究进行了全面的文献搜集与整理。从中收集到共500个数据点构成原始数据集^[4,10-12],包括9种常见有机胺溶剂类型:单乙醇胺(MEA)、甲基二乙醇胺(MDEA)、氨基乙基丙二醇(AEPD)、二乙醇胺(DEA)、2-氨基-2-甲基-1-丙醇(AMP)、三乙醇胺(TEA)、2-氨基-2-羟甲基-1,3-丙二醇(AHPD)、2-氨基-2-甲基-1,3-丙二醇(AMPD)、哌嗪(PZ)。

随后,对已收集数据集中的各特征进行了皮尔逊相关性分析,将皮尔逊相关系数(PCC)大于0.8的输入特征剔除,以避免高相关性变量为模型带入噪声从而影响模型预测能力。此外,特征数量的减少还可以有效降低模型复杂性,提升泛化能力。最终输入特征的数目由22个精简为10个,可分为反应条件与有机胺溶剂性质2类(表1)。精简与挑选后的模型输入特征和目标变量的皮尔逊相关性如图2所示。从图中可以发现,在输入特征和目标变量之间(PCC值在-0.42~0.44)及各输入特征变量之间(PCC值大多在-0.7~0.7)没有观察到显著的相关性。

表1 输入特征信息

特征类别	名称	缩写	备注
反应条件	温度/K	<i>T</i>	
	二氧化碳分压/kPa	<i>P</i>	
	胺摩尔分数	—	胺摩尔数/溶剂总摩尔数×100%
有机胺性质	疏水性	Log <i>P</i>	通过rdkit库获得
	极性表面积/Å ²	PSA	通过rdkit库获得
	碱度	pKa	
	N原子部分电荷/e	<i>q</i> (N)	
	N原子数/个	—	
	O原子数/个	—	
	胺基百分比	—	有机胺分子胺基数×胺浓度

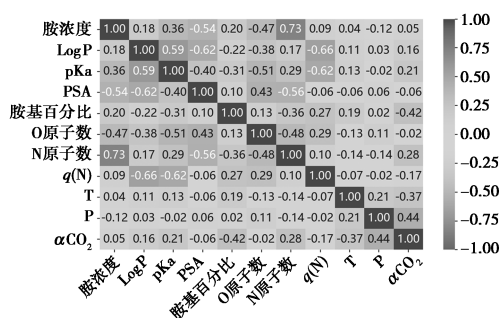


图2 皮尔逊相关性

2.2 模型构建和性能评估

经过数据预处理后,80%的数据集被划分为训练集,其余20%被划分为测试集。建模途中使用网格搜索调参优化提升模型效果并使用五折交叉验证提升模型泛化能力。如表2所示,3类模型的效果对比显示树模型整体表现优于神经网络模型与经典回归模型。具体来说,在树模型中,GBDT展现出了最低的MAE_{测试集}(0.08)、RMSE_{测试集}(0.08)以及最高的R²_{测试集}(0.915)。同时,GBDT模型中的R²_{训练集}(0.959)、R²_{测试集}(0.915)、与R²_{交叉验证}(0.879)的数值也较为一致,体现了GBDT模型出色的普适性和泛化能力。对于2种神经网络模型,MLP的RMSE_{测试集}为0.14,而R²_{测试集}仅为0.79,LSTM的RMSE_{测试集}和R²_{测试集}则为0.13与0.78。2种模型的R²_{测试集}均低于GBDT,且RMSE_{测试集}偏高。此外,2类经典的回归模型,SVR和RR的模型效果同样有限,测试集R²仅有0.78和0.64,与树模型效果具有较大差距。

表2 输入特征信息

模型	训练集	训练集	测试集	测试集	CV R ²	CV RMSE
	R ²	RMSE	R ²	RMSE		
XGBoost	0.9222	0.0656	0.8745	0.0889	0.8512	0.1060
RF	0.9195	0.0752	0.8536	0.0994	0.8398	0.1261
GBDT	0.9591	0.0568	0.9153	0.0836	0.8786	0.0967
MLP	0.7609	0.1374	0.7671	0.1386	0.7041	0.1512
LSTM	0.7760	0.1329	0.7850	0.1332	0.6593	0.1634
SVR	0.8118	0.1219	0.7851	0.1331	0.7415	0.1410
RR	0.6034	0.1769	0.6441	0.1713	0.5714	0.1823

基于上述结果,可以发现GBDT模型总体上具有最优秀的预测能力和出色的泛化能力。因此,选择GBDT模型作为最佳效果模型进行后续的研究分析。

2.3 特征重要性分析

采用SHAP、PI与MDI3种方法,基于上述最优GBDT模型分别进行特征重要性分析,通过不同的分析视角研究每个特征对整体模型预测的影响程度,揭示特征在有机胺CO₂吸收体系中的重要性程度。如表3所示,分析结果显示3种方法获得的特征重要性排序结果高度一致。前7种最重要的输入特征均依次为二氧化碳分压、温度、胺基百分比、Log*P*、胺浓度、pKa和*q*(N)。这种一致性说明了模型重要性排序结果的可靠性。值得注意的是,这些

结果也与过往有机胺溶剂吸收 CO_2 实验研究中的结果一致^[13-14]。例如对前 3 个最重要的特征而言,发现二氧化碳分压通过改变分子扩散速度,可以显著影响 CO_2 吸收效果,是气体溶解的核心驱动力,因此被认为是有机胺溶剂吸收 CO_2 最重要的影响因素^[14-16]。其次,由于温度可同时影响 CO_2 吸收反应的反应动力学与热力学 2 方面,从而影响 CO_2 吸收效率,因此在过往研究中也认为是有机胺 CO_2 吸收体系中的重要影响因素之一^[16]。胺基百分比的重要性则可归因其综合体现了胺基的数量和有机胺浓度,可通过碱性位点数量和传质阻力 2 方面影响 CO_2 吸收效果^[17]。

表 3 3 种特征重要性方法 (MDI、PI、SHAP) 的特征重要性排序

特征名称	MDI 重要性	PI 重要性	SHAP 重要性
P	0.447423	0.0955	0.160821
T	0.237863	0.0470	0.109532
胺基百分比	0.104723	0.0069	0.036264
$\text{Log}P$	0.057202	0.0044	0.025296
胺浓度	0.052742	0.0026	0.017956
pKa	0.035650	0.0023	0.017561
$q(\text{N})$	0.019950	0.0016	0.012280
N 原子数	0.017859	0.0004	0.006103
O 原子数	0.016430	0.0008	0.006474
PSA	0.010159	0.0004	0.004308

2.4 CO_2 吸收优势区间的确立与定量构效关系

前述特征重要性方法的分析结果与过往实验结果均表明,二氧化碳分压、温度和胺基百分比是影响 CO_2 吸收效果的最关键的 3 种特征。在此基础上,进一步绘制这些特征与 CO_2 吸收效果的三维空间散点图和二维散点平面映射图,以揭示 CO_2 吸收效果的优势区间。

结果如图 3 所示,点的大小代表了 CO_2 吸收量的大小。根据三维空间散点图,较大的点普遍出现在中低温区与中低胺基百分比的区间内。随后通过二维散点平面映射图,分别观察二氧化碳分压与温度、二氧化碳分压与胺基百分比和温度与胺基百分比对 CO_2 吸收效果的影响情况。从图中可以发现,对 CO_2 吸收有利的优势区间集中在 $T=300\sim 350\text{ K}$ 、胺基百分比 $=0\sim 40$ 的区域中。相对而言,压强则未表现出显著的优势范围。这些结果也可以得到过往实验研究的支持。例如 Balsora 等^[18] 提出二氧化碳

溶解度可在一定范围内随温度升高而降低,并且在 313.14 K 时达到了 0.869 mol/mol 的最大二氧化碳吸收量。同时,相对较低的胺浓度可以避免分子黏度提升带来的传质阻力,从而有利于 CO_2 的吸收。此外,使用较低的温度、分压、胺浓度的反应条件也有助于降低技术应用成本。

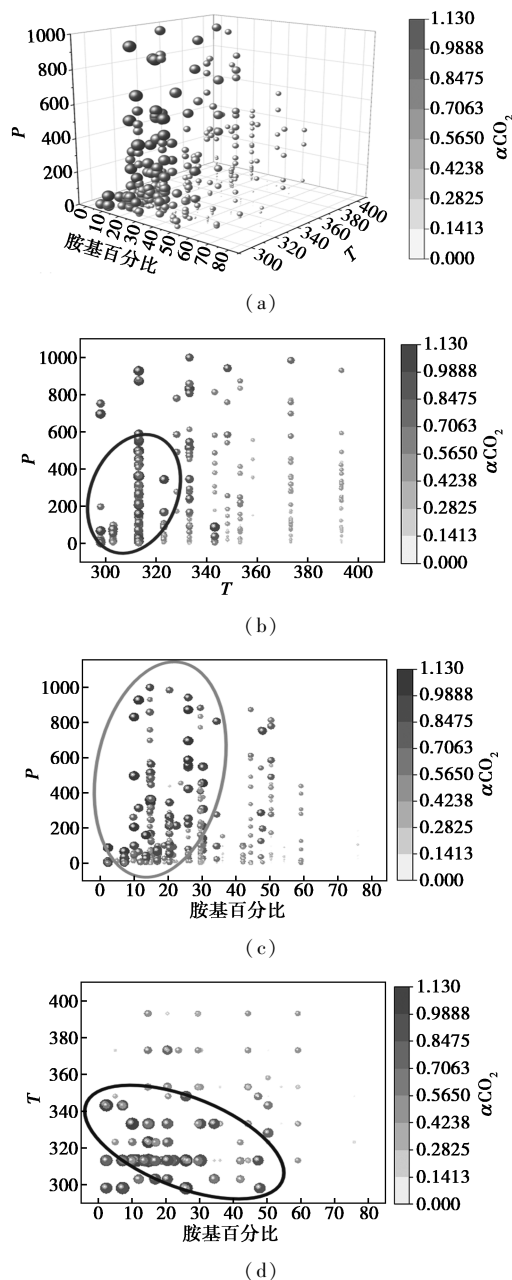


图 3 三维空间散点图与二维散点平面映射图

最后,还通过多元线性回归方法进一步建立了前述重要特征与 CO_2 吸收效果间的定量构效关系。如式(4)所示,虽然其拟合度相比机器学习模型仍较为有限,但方程中系数的正负性也可以在一定程度上说明各重要特征对 CO_2 吸收性能的影响。具体而言,方程中温度和胺基百分比的系数是负数,这

表明当2种因素处于相对较小区间时可能会对CO₂的吸收更有利,这也与前述优势区间的分析结果相一致。

$$\alpha\text{CO}_2 = 2.022 + 0.01P - 0.004T - 0.006 \times \text{胺基百分比} \\ R^2 = 0.542 \quad (4)$$

3 总结与展望

本研究针对传统有机胺CO₂吸收体系开发周期长、成本高、效率低等问题,通过机器学习建模,实现了高性能有机胺CO₂吸收体系的性能预测。研究表明,GBDT模型展现出最优预测精度和出色的泛化能力。多种特征重要性分析方法共同揭示了二氧化碳分压、温度和胺基百分比为影响CO₂吸收效果的核心因素。基于二维散点平面映射图进一步阐明了中低温度与较低胺基百分比的CO₂吸收优势区间。本工作可为高性能有机胺CO₂吸收体系的设计与开发提供新的思路,该研究策略也可延伸应用于环境治理等相关领域的未来研究。

参考文献

- [1] 赵志强,张贺,焦畅,等.全球CCUS技术和应用现状分析[J].现代化工,2021,41(4):5-10.
- [2] 盛依依.二氧化碳捕集技术研究及工业化进展[J].现代化工,2025,45(3):1-6.
- [3] Shen Y, Gong Y, Sun L, *et al.* Machine learning-driven assessment of relationship between activator properties in phase change solvent and its absorption performance for CO₂ capture[J]. Separation and Purification Technology, 2023, 309: 123092.
- [4] Qian J, Sun R, Sun S, *et al.* Computer-free group-addition method for pKa prediction of 73 amines for CO₂ capture[J]. Journal of Chemical & Engineering Data, 2017, 62(1): 111-122.
- [5] Namdeo S, Srivastava V C, Mohanty P. Machine learning implemented exploration of the adsorption mechanism of carbon dioxide onto porous carbons[J]. Journal of Colloid and Interface Science, 2023, 647: 174-187.
- [6] Yang A, Sun S, Su Y, *et al.* Insight to the prediction of CO₂ solubility in ionic liquids based on the interpretable machine learning model[J]. Chemical Engineering Science, 2024, 297: 120266.
- [7] 侯克鹏,包广拓,孙华芬.SSA-MLP模型在岩质边坡稳定性预测中的应用[J].安全与环境学报,2024,24(5):1795-1803.
- [8] Wen L, Shao H. Analysis of influencing factors of the carbon dioxide emissions in China's commercial department based on the STIRPAT model and ridge regression[J]. Environmental Science and Pollution Research, 2019, 26: 27138-27147.
- [9] 唐菲菲,胡佳赢,马英,等.基于递归特征消除选择的粒子群优化-支持向量机回归模型的滑坡位移预测方法[J].工业建筑,2024,54(11):50-60.
- [10] Fujii T, Sako M, Ishihama K, *et al.* Prediction of CO₂ absorbing performance of amine aqueous solution using random forest models[J]. Gas Science and Engineering, 2024, 129: 205417.
- [11] Yang X, Rees R J, Conway W, *et al.* Computational modeling and simulation of CO₂ capture by aqueous amines[J]. Chemical Reviews, 2017, 117(14): 9524-9593.
- [12] Li H, Yan D, Zhang Z, *et al.* Prediction of CO₂ absorption by physical solvents using a chemoinformatics-based machine learning model[J]. Environmental Chemistry Letters, 2019, 17(3): 1397-1404.
- [13] Khan A A, Halder G N, Saha A K. Experimental investigation on efficient carbon dioxide capture using piperazine (PZ) activated aqueous methyldiethanolamine (MDEA) solution in a packed column[J]. International Journal of Greenhouse Gas Control, 2017, 64: 163-173.
- [14] Kumar S, Mondal M K. Equilibrium solubility of CO₂ in aqueous blend of 2-(Diethylamine) ethanol and 2-(2-Aminoethylamine) ethanol[J]. Journal of Chemical & Engineering Data, 2018, 63(5): 1163-1169.
- [15] Ali M, Hamdi Z, Elochukwu H, *et al.* Investigation on CO₂ solubility for safer CO₂ storage; Experimental study on the effect of pH; proceedings of the offshore technology conference brasil [C]. 2023, D031S037R006.
- [16] Aghel B, Janati S, Wongwises S, *et al.* Review on CO₂ capture by blended amine solutions[J]. International Journal of Greenhouse Gas Control, 2022, 119: 103715.
- [17] Puxty G, Rowland R, Allport A, *et al.* Carbon dioxide postcombustion capture: A novel screening study of the carbon dioxide absorption performance of 76 amines[J]. Environmental Science & Technology, 2009, 43(16): 6427-6433.
- [18] Balsora H K, Mondal M K. Solubility of CO₂ in an aqueous blend of diethanolamine and trisodium phosphate[J]. Journal of Chemical & Engineering Data, 2011, 56(12): 4691-4695. ■

《现代化工》欢迎广大作者踊跃投稿,投稿系统:<http://www.xdhg.com.cn>