



人工智能机器学习助力离子液体CO₂捕集性能分析

张瑞娜^{1†}, 田媛^{2†}, 葛春亮³, 张威³, 卢晗锋¹, 崔国凯^{1*}

1. 浙江工业大学, 绿色化学合成技术国家重点实验室培育基地, 化学工程学院, 催化反应工程研究所, 杭州 310014;

2. 中国科学院过程工程研究所, 多相复杂系统国家重点实验室, 中国科学院绿色过程与工程重点实验室, 离子液体清洁过程北京市重点实验室, 北京 100190;

3. 浙江浙能科技环保集团股份有限公司, 杭州 310012

† 同等贡献

* 联系人, E-mail: chemcgk@163.com

2024-02-08 收稿, 2024-03-28 修回, 2024-04-17 接受, 2024-05-09 网络版发表

浙江省重点研发项目(2024C03108, 2023C03127)、国家自然科学基金(22378353, 22078294)、浙江省自然科学基金(LTGS24E080008)和浙江浙能科技环保集团股份有限公司项目(ZNKJ-2024-077)资助

摘要 随着我国碳达峰、碳中和战略的实施, 煤电、钢铁、水泥等产业CO₂捕集迫在眉睫. 为了替代高能耗、易挥发的有机胺CO₂吸收剂, 不挥发、高稳定、可设计的离子液体CO₂吸收剂受到广泛关注. 相对于传统实验试错方法, 人工智能机器学习成为实现快速预测离子液体CO₂捕集性能、选择最佳离子液体吸收剂的重要途径. 本文在简要介绍机器学习算法和模型的基础上, 从离子液体的描述符出发, 综述了近年来基于人工智能和机器学习的定量结构-性能关系(quantitative structure-property relationship, QSPR)方法用于离子液体CO₂捕集性能研究的进展, 重点介绍离子液体的分子结构描述符的种类及相应QSPR模型建立方法, 对机器学习应用于离子液体CO₂捕集性能研究中存在的问题和进一步的研究工作提出了建议. 本文有助于数据与化学化工交叉学科研究人员开发高效离子液体CO₂吸收剂.

关键词 人工智能, 机器学习, 离子液体, 二氧化碳捕集, 定量结构-性能关系

在应对全球气候变化和削减温室气体排放的紧迫任务中, 二氧化碳(CO₂)捕集技术的实施显得尤为关键且刻不容缓. 近年来, 离子液体(ionic liquids, ILs)因不挥发无二次污染、不可燃无安全隐患、溶解能力强、稳定性高、结构可设计性等特点和优点, 在CO₂^[1~3]、SO₂^[4~6]、H₂S^[7,8]、CO^[9,10]、NH₃^[11,12]等气体捕集分离方面具有独特的优势. 例如, 在CO₂捕集方面, 相较于通过物理溶解方式捕集CO₂的常规离子液体, 通过化学作用方式捕集CO₂的氨基功能化离子液体^[13~16]、酚阴离子功能化离子液体^[17~19]、醇阴离子功能化离子液体^[20]、唑阴离子功能化离子液体^[21~23]、酰亚胺阴离子

功能化离子液体^[24~26]、碳负功能化离子液体^[27]等得到迅速发展. 一般地, 离子液体由阴阳离子配对的结构组成方式有10¹⁸之多, 通过实验室“结构合成-性能测试”方法进行人工筛选最佳离子液体CO₂吸收剂, 程序繁琐、效率低. 如何快速、简便、智能化设计离子液体吸收剂, 是一个重大的挑战.

机器学习(machine learning, ML)是人工智能(artificial intelligence, AI)的核心, 是一类能够在高维数据中寻找规律的强大工具. 近年来, 机器学习方法的提出为快速筛选新材料注入了新活力, 已在能源材料^[28]、化学化工^[29~34]、生命医药^[35]等领域有广泛应用. 通过将

引用格式: 张瑞娜, 田媛, 葛春亮, 等. 人工智能机器学习助力离子液体CO₂捕集性能分析. 科学通报, 2025, 70: 4462–4472

Zhang R N, Tian Y, Ge C L, et al. Artificial intelligence and machine learning for analysis of CO₂ capture by ionic liquids (in Chinese). Chin Sci Bull, 2025, 70: 4462–4472, doi: 10.1360/TB-2024-0163

人工智能机器学习应用于离子液体 CO_2 捕集性能研究, 计算机可以模拟离子液体结构与 CO_2 捕集性能之间的线性或非线性关系, 学习离子液体捕集 CO_2 的经验数据, 从而预测未知离子液体或者未知条件下的 CO_2 捕集性能。此外, 与热力学状态方程(equation of state, EoS)、真实溶剂似导体屏蔽模型(conductor-like screening model for real solvents, COSMO-RS)等经典数据分析预测方法相比, 机器学习方法在数据分类和预测方面表现更优, 具有显著节约成本和提高效率的潜力。机器学习在离子液体 CO_2 捕集性能研究中的应用, 拓展了人工智能在能源环保中的应用, 提供了新思路新方法。有关“人工智能机器学习”“离子液体”和“ CO_2 捕集性能”的综述目前已有报道^[36-38]。基于描述符选择的重要性, 对物理性质描述符到分子结构描述符、单一算法到集成算法进行深入总结, 目前尚没有报道, 然而这对于构建以机器学习模型为核心的定量结构-性能关系(quantitative structure-property relationship, QSPR)或定量结构-活性关系(quantitative structure-activity relationship, QSAR)模型并从分子设计层面开发应用于 CO_2 捕集的离子液体吸收剂是极其重要的。

本文基于机器学习在离子液体 CO_2 捕集性能研究中的应用, 通过对描述符的全面总结, 系统综述了基于人工智能和机器学习的QSPR方法用于离子液体 CO_2 捕集性能研究的进展, 并提出该领域存在的问题及未来趋势(图1)。本文将有助于研究人员, 特别是数据科学家(对机器学习有更深入的理解)和化学工程师(系统级问题解决者)深入了解、应用离子液体 CO_2 捕集性能研究的机器学习基本模型, 并开发对离子液体 CO_2 捕集性能预测效果更好的机器学习模型。

1 机器学习算法与模型

离子液体捕集 CO_2 性能预测中常用的机器学习预测模型, 主要有: 人工神经网络(artificial neural network, ANN)、支持向量机(support vector machine, SVM)与最小二乘支持向量机(least square support vector machine, LSSVM)、基于决策树(decision tree, DT)的树模型、多元线性回归(multivariate linear regression, MLR)、高斯过程回归(Gaussian processes regression, GPR)等(图2)。两种或者两种以上的模型可以形成集成模型。本文侧重介绍单一机器学习模型。此外, 将各类EoS热力学模型、基团贡献(group contribution, GC)、分子动力学(molecular dynamics, MD)模拟等其他预测方法融入机

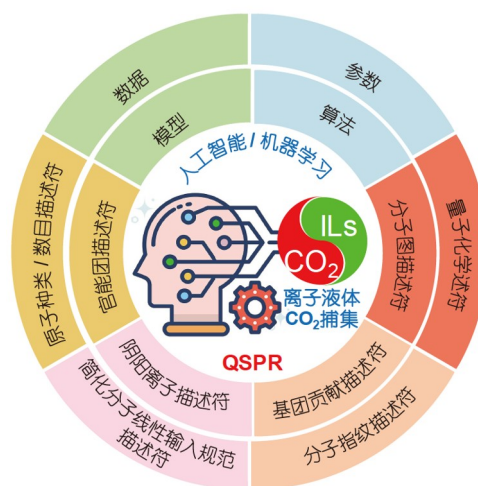


图1 机器学习的QSPR方法助力离子液体 CO_2 捕集性能分析
Figure 1 Machine learning-based QSPR methods for analysis of CO_2 capture by ionic liquids

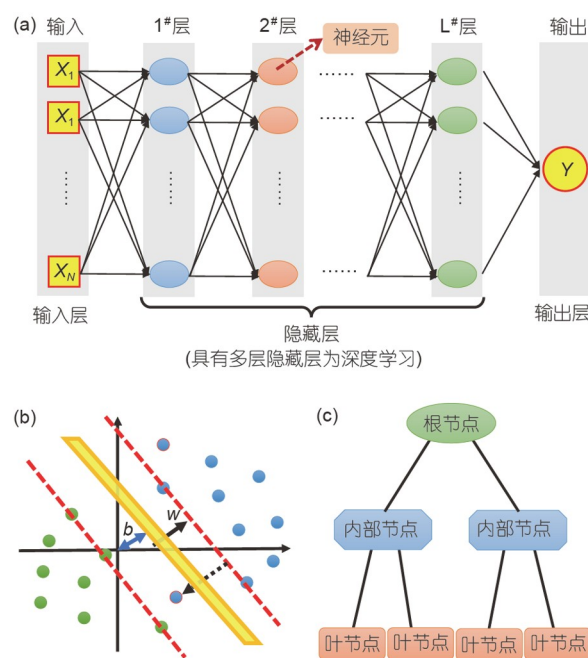


图2 三种典型的机器学习模型图。(a) ANN; (b) SVM; (c) DT
Figure 2 Schematic diagrams of three typical machine learning models. (a) ANN; (b) SVM; (c) DT

器学习, 可以提高预测的准确性。

ANN是一种非线性数学模型, 架构包括一层输入层、一层或多层隐藏层、一层输出层, 通过神经元联系在一起。ANN模型有: 多层感知器神经网络(multi-layer perceptron neural network, MLPNN)、径向基函数神经网络(radial basis function neural network,

RBFNN)、多层前馈神经网络(multi-layered feed-forward neural network, MLFNN)、前馈神经网络(feed-forward neural network, FFNN)、级联前向神经网络(cascade-forward neural network, CFNN)、广义回归神经网络(general regression neural network, GRNN)、反向传播神经网络等(back propagation neural network, BPNN). 其中, 最常用的是MLPNN和RBFNN. MLPNN中的隐藏层数可以是一层或多层, 而RBFNN中只有一层隐藏层. 此外, 基于ANN的数据处理群组方法(group method of data handling, GMDH)、极限学习机(extreme learning machines, ELM)都属于FFNN; 基于ANN的卷积神经网络(convolutional neural network, CNN)、递归神经网络(recursive neural networks, RNN)、图神经网络(graph neural network, GNN)都属于深度神经网络(deep neural network, DNN). 将ANN和模糊逻辑(fuzzy logic, FL)结合可以构筑与自适应模糊神经推理系统(adaptive neuro-fuzzy inference system, ANFIS)^[39]. 将DNN与因子分解机(factorization machines, FM)结合, 可以构筑深度因子分解机(deep factorization machine, DeepFM).

SVM是基于统计学习理论引入的算法, 通过在高维度特征空间中构造一个或一组超平面, 使之与训练数据点的距离最大化, 用于分类、回归等任务. 用来确定超平面间隔的数据点就是支持向量, n 维特征空间需要 $n+1$ 个支持向量. SVM一般采用径向基函数(radial basis function, RBF)作为核函数(kernel function)将输入数据转换为所需的形式. LSSVM是SVM的一种改进变体, 能够以较低的复杂性和较高的收敛速度处理大规模数据集. 常用的树模型算法有条件推断树(conditional inference trees, CTREE)与分类回归树(classification and regression tree, CART)、随机森林(random forest, RF)、梯度提升决策树(gradient boosting decision tree, GBDT)与极端梯度提升(extreme gradient boosting, XGBoost)等. 树模型用于拟合人们的决策行为. MLR是一种线性回归形式, 当有2个或2个以上输入变量(特征值)的时候使用, 用来表达多个输入变量如何共同影响输出变量. GPR是使用高斯过程先验对数据进行回归分析的非参数模型, 不仅能够灵活处理非线性问题, 还能够提供对预测结果的不确定性估计. 根据所使用的核函数不同, GPR分为有理二次GPR、平方指数GPR, Matérn 5/2 GPR.

典型优化算法包括: 遗传算法(genetic algorithm,

GA)、粒子群优化算法(particle swarm optimization, PSO)、模拟退火算法(simulated annealing, SA)、蚁群优化算法(ant colony optimization, ACO)、反向传播算法(back propagation, BP)等. GA、PSO、SA、ACO都属于基于种群的进化算法, 通过模拟自然过程, 搜索最优解的方法. BP算法是目前用来训练ANN的最常用、最有效的算法. 在ANN正向传播过程中, 输入信息通过逐层处理并传向输出层. 如果在输出层得不到期望的输出值, 则构造损失函数作为目标函数, 转入反向传播, 逐层求出目标函数对权值向量的梯度, 修改权值. 当输出值与真实值的误差达到所期望值时, 学习结束. 用于建模预测CO₂捕集性能的基于反向传播的优化算法有: 莱文贝格-马夸特算法(Levenberg-Marquardt, LM)、贝叶斯正则化算法(Bayesian regularization, BR)、自适应学习速率动量梯度下降反向传播算法(gradient descent with momentum and adaptive learning rate backpropagation, GDX).

机器学习模型中的参数有: 输入参数(inputs)、输出参数(outputs)、模型参数、超参数(hyperparameters)等. 机器学习的输入参数是能描述材料的分子特征并直接影响材料目标性质和功能的物理量或函数, 即描述符(descriptor). 描述符能让人类和机器学习都理解材料. 因而, 描述符的选择非常重要, 并会影响模型的性能和准确性. 目前还没有关于如何选择最佳描述符的规则, 而且也很难评估研究中描述符的使用情况, 但是普遍认为应当选择与目标属性最密切相关的描述符^[40]. 本文正是从描述符出发对应用于离子液体CO₂捕集性能研究的机器学习模型进行总结. 模型的输出参数即CO₂在离子液体中的溶解度、摩尔分数或者亨利常数.

机器学习需要将一定量的CO₂溶解度、吸收容量或者亨利常数数据形成数据集(database). 数据既可以来自实验, 也可以来自各类状态方程等热力学模型、MD模拟等计算得到的预测数据. 机器学习通常会要求将一部分大量数据作为训练集(training set)而另一部分少量数据作为测试集(test set). 此外, 还可以通过验证集(validation set)评估以及调整少量模型超参数.

机器学习模型的常见统计分析度量有: 相关系数(R^2)、平均绝对误差(mean absolute error, MAE)、均方误差(mean squared error, MSE)、均方根误差(root mean squared error, RMSE). R^2 越大、估计值的方差越小, 说明模型更好、波动更小. 此外, 还有平均绝对百分比误

差(mean absolute percentage error, MAPE)、平均相对偏差(average relative deviation, ARD)、平均绝对偏差(average absolute deviation, AAD)、平均绝对相对偏差(average absolute relative deviation, AARD)、标准差(standard deviation, STD)等。一般地,本文采用测试集的上述度量参数进行模型优劣对比。

2 物理性质描述符的机器学习模型

早期用于描述“离子液体溶解CO₂性能”的描述符有:分子量(molecular weight, M_w)、偏心因子(acentric factor, ω)、临界温度(critical temperature, T_c)、临界压力(critical pressure, T_p)、临界压缩因子(critical compressibility factor, Z_c)、操作温度(pressure, P)、操作压力(temperature, T)。如果是多组分的液体混合物,则还需加上组分的质量分数或者摩尔分数等特征作为描述符;如果是从多组分气体中吸收CO₂,则还需要加上CO₂的摩尔分数或者浓度。然而,离子液体种类繁多,其临界性质往往难以测定。

2008年, Torrecilla等人^[41]报道了关联5种离子液体CO₂溶解度的线性MLR和非线性MLPNN等机器学习模型,输入参数为 P 、 T ,输出参数为CO₂的摩尔分数,并表明非线性模型较优。Eslamimanesh等人^[42]采用 ω 、 T_c 、 T_p 、 P 、 T 为输入参数,构筑了基于24种离子液体的1140个CO₂溶解度数据的ANN模型,预测超临界CO₂在离子液体中的溶解能力。Safamirzaei等人^[43,44]采用 M_w 、 ω 、 P 、 T 为输入参数,分别构筑了预测1-丁基-3-甲基咪唑鎓六氟磷酸盐([Bmim][PF₆])和1-丁基-3-甲基咪唑鎓四氟硼酸盐([Bmim][BF₄])的CO₂溶解度的ANN机器学习模型。2015年,Hamzehie等人^[45]基于9种离子液体的38个CO₂溶解度数据,以 M_w 、 ω 、 T_c 、 T_p 、 P 、 T 为输入参数,构筑了具有两层隐藏层的MLFNN模型,架构为:6-8-4-1。随后,Fierro等人^[46]基于33种不同离子液体的2099个CO₂溶解度数据,以 T 、 P 、 T_c 、 P_c 为输入参数,形成了LM算法优化的MLPNN机器学习模型(架构为4-6-8-1),具有最佳预测效果。Baghban等人^[47]基于14种离子液体的728个CO₂溶解度数据,以 M_w 、 ω 、 T_c 、 T_p 、 P 、 T 为输入参数,构筑MLPNN和ANFIS模型来预测CO₂溶解度,结果表明,MLPNN模型在统计测量方面优秀。Tatar等人^[48]基于上述相同数据库,以相同输入变量,开发RBFNN、LSSVM、MLPNN三种方法,然后将其合并到集成学习模型委员会机器学习智能系统(Committee Machine Intelligent System, CMIS)中,结果

表明集成学习模型可以提供更准确的预测。以 M_w 、 ω 、 T_c 、 T_p 、 P 、 T 六个变量为输入参数, Baghban等人^[49]、Deng等人^[50]、Mesbah等人^[51]、Ouaer等人^[52]先后对不同数据集构建了多种机器学习模型。Ghazani等人^[53]开发了预测混合气体中CO₂溶解度的LSSVM机器学习模型。以 M_w 、 T_c 、 T_p 、 P 、 T 为输入参数, Mesbah等人^[54]、Feng等人^[55]构筑了单一或集成学习模型,来估计CO₂在不同离子液体中的溶解度,同样表明集成模型预测效果好。Mazari等人^[56]以 P 、 T 、离子液体中水含量 $Wt\%$ 作为输入参数,提出了有理二次GPR、平方指数GPR、Matérn 5/2 GPR三个GPR机器学习模型用于预测[Bmim][PF₆][PF₆]的CO₂溶解度。结果显示, Matérn 5/2 GPR预测精度最好。Daryayehsalameh等人^[57]以 P 、 T 作为输入参数,采用MLPNN、CFNN、GRNN、RBFNN、LSSVM、ANFIS六种机器学习模型,预测了CO₂在[Bmim][BF₄]中的溶解度。模型预测精度顺序为:CFNN > MLPNN > ANFIS > LSSVM > RBFNN > GRNN。物理性质描述符及相应机器学习模型如表1所示,可见六元物理性质描述符结合了分子本身的结构(M_w 、 ω)、离子液体本身的性质(T_c 、 T_p)和吸收条件(P 、 T),是常用的物理性质描述符组合;此外,MLPNN是以物理性质为描述符的机器学习模型中应用较广的模型。

3 分子结构描述符的机器学习模型

基于“结构决定性质”理念,离子液体的阴阳离子结构对其CO₂捕集性质有很大影响。以离子液体的结构特性为分子描述符,用作输入变量,形成以机器学习为核心的新型QSPR或QSAR模型。QSPR模型在数学上将物理或化学性质与分子结构联系起来,其主要步骤包括:绘制离子液体结构(一维、二维或三维)、形成离子液体的分子结构描述符(本征结构、几何结构、拓扑结构、分三维矩阵、分子性质、量子性质、电性等)、数据的选择与处理、模型建立与验证、统计分析与可视化^[58]。

(1) 原子种类和数目描述符。包括离子液体在内的所有分子都由原子构成,原子的种类和数目自然成为分子的描述符。2017年,Barati-Harooni等人^[59]基于39种离子液体的2930个CO₂溶解度数据,首次以 T 、 P 、 M_w 、阳离子中碳原子数目(C_c)、阴离子中碳原子数目(C_a)、阴离子中氢原子数目(nH_a)、离子液体中氮原子数目(N)、离子液体中硫原子数目(S)、离子液体中氧

表 1 物理性质描述符、数据量及相应机器学习模型对比

Table 1 Comparison of physical property descriptors, data volume, and corresponding machine learning models

描述符		离子液体种类	数据集	机器学习模型	参考文献
种类	具体参数				
二元	P 、 T	5		MLPNN、MLR	[41]
	P 、 T	1		MLPNN、CFNN、GRNN、RBFNN、LSSVM、ANFIS	[57]
四元	Mw 、 ω 、 P 、 T	1		ANN	[43]
	Mw 、 ω 、 P 、 T	1		ANN	[44]
	T_c 、 P_c 、 P 、 T	33	2099	MLPN	[46]
五元	ω 、 T_c 、 T_p 、 P 、 T	24	1140	ANN	[42]
	Mw 、 T_c 、 T_p 、 P 、 T	20	1386	MLPNN	[54]
	Mw 、 T_c 、 T_p 、 P 、 T	17	4107	MLPNN、SVM、RF、DT	[55]
六元	Mw 、 ω 、 T_c 、 T_p 、 P 、 T	9	38	MLFNN	[45]
	Mw 、 ω 、 T_c 、 T_p 、 P 、 T	14	728	MLPNN、ANFIS	[47]
	Mw 、 ω 、 T_c 、 T_p 、 P 、 T	14	728	RBFNN、LSSVM、MLPNN、CMIS	[48]
	Mw 、 ω 、 T_c 、 T_p 、 P 、 T	67	5368	LSSVM、MLPNN、RBFNN、ANFIS	[49]
	Mw 、 ω 、 T_c 、 T_p 、 P 、 T	13	728	DNN、CNN、RNN	[50]
	Mw 、 ω 、 T_c 、 T_p 、 P 、 T	24	1155	GMDH、RBFNN	[51]
	Mw 、 ω 、 T_c 、 T_p 、 P 、 T	13	744	MLPNN	[52]

原子数目(O)、离子液体中氟原子数目(F)、离子液体中氯原子数目(Cl)、离子液体中硼原子数目(B)等12个变量为输入参数, 构筑了PSO算法优化的RBFNN(PSO-RBF)、MLPNN、CSA算法优化的LSSVM(CSA-LSSVM)、混合方法训练的自适应神经模糊推理系统(hybrid ANFIS)四种机器学习方法, 预测 CO_2 在离子液体中的溶解度, 并与热力学预测模型Peng-Robinson状态方程(PR-EoS)进行了对比. 结果表明, 所开发的模型的预测比热力学模型的预测更好、更有效. 此外, 研究结果显示, PSO-BRF、MLPNN、CSA-LSSVM、hybrid ANFIS四种模型的测试集的AARD%分别为4.98%、4.45%、1.98%、4.19%, 表明CSA-LSSVM模型为 CO_2 -IL提供了更好的预测. 2021年, Zhang等人^[60]提出一种将多玩家蒙特卡罗树搜索(multi-player Monte-Carlo tree search, MP-MCTS)与RNN相结合的机器学习技术, 对烟气(CO_2/N_2)和合成气(CO_2/H_2)两种不同场景中的 CO_2 捕获进行了评估, 设计创建用于 CO_2 捕获的高性能离子液体. 2022年, Valeh-e-Sheyda等人^[61]基于43种不同咪唑鎓离子液体及的4397个 CO_2 溶解度数据(训练集3078个数据点+测试集660个数据点+验证集659个数据点), 以26个描述符作为输入参数, 构建了FFNN、RBFNN、SVM三种机器学习模型, 预测 CO_2 溶解度.

这26个描述符包括 P 、 T 、 Mw 、密度、水分含量、碳氢氧氮各自原子数、原子间的各类化学键信息等. 结果显示, FFNN、RBFNN、SVM三种机器学习模型对测试集的 R^2 分别为0.9458、0.8949、0.8607, 表明采用LM算法优化的FFNN模型是预测 CO_2 在咪唑基离子液体中平衡溶解度的最佳模型, 从而建立了 CO_2 溶解度、分子结构、吸收压力和温度之间的QSPR关系.

(2) 官能团描述符. 2021年, Valeh-e-Sheyda等人^[62]基于11种不同吡啶鎓离子液体的430个 CO_2 溶解度数据(训练集302个数据点+测试集64个数据点+验证集64个数据点), 分别以三属性和五属性作为输入参数, 采用LM和GDX优化算法, 建立多种MLFNN机器学习模型, 用于估计 CO_2 在吡啶鎓离子液体中的溶解度. 三属性分别为 Mw 、 P 、 T ; 五属性分别为 Mw 、 P 、 T 、是否存在醚基(分类因子)、烃链和/或醚基中的碳数. 结果表明, 以五属性分子描述符为输入变量, 用LM算法优化的MLFNN模型, 其测试集的MSE、STD、 R^2 分别为 8.23×10^{-5} 、 9.00×10^{-3} 、0.995, 可用于详细评估基于吡啶鎓的离子液体的 CO_2 捕集性能.

(3) 阴阳离子描述符. 2021年, Wu等人^[63]基于32种不同的咪唑鎓阳离子离子液体的155个 CO_2 溶解度数据(训练集132个数据点+测试集23个数据点), 以源自Dra-

gon软件的11个阴阳离子结构描述符和 T 共计12个描述符作为输入参数,分别通过MLPNN、RF、MLR三种机器学习方法来构建相应QSPR模型,预测 CO_2 在离子液体中的亨利常数。11个分子描述符分别为:阴离子 PF_6 或 BF_4 或 PF_3R_3 为描述符 a_1 、阴离子 Tf_2N 为描述符 a_2 、阴离子-CN为描述符 a_3 、阴离子 SO_4R 为 a_4 、阴离子 PO_4R_2 为描述符 a_5 、阴离子氟原子数目为描述符 a_6 、阴离子烷基链长度为描述符 a_7 、阳离子氟原子数目为描述符 c_1 、阳离子烷基链长度为描述符 c_2 、阳离子环取代基数目为描述符 c_3 、阳离子-CN为描述符 c_4 。结果表明,基于MLPNN的QSPR模型的预测性能最佳,其 R^2 、MAE、MSE分别为0.9801、0.5185、0.5643,说明该模型的准确性远远不够,原因是实验数据点太少。

(4) 基团贡献描述符。基团贡献法(group contribution, GC)基于假设不同分子中的同一基团对性质的贡献相同。GC易于计算且具有很强的通用性。2020年, Song等人^[64]基于124种不同离子液体的10116个 CO_2 溶解度数据(训练集8093个数据点+测试集2023个数据点),以51个构筑基团(包括13种阳离子、28种阴离子、10种取代基)和 P 、 T 共计53个描述符作为输入参数,提出了基于基团贡献的人工神经网络(ANN-GC)和基于基团贡献的支持向量机(SVM-GC)两种机器学习模型,用于预测不同温度和压力条件下 CO_2 在各种离子液体中的溶解度。结果显示,ANN-GC和SVM-GC的测试集的 R^2 分别为:0.9836和0.9783,表明这两个模型都可以对离子液体中的 CO_2 溶解度进行可靠的预测,并且ANN-GC模型的性能略好于基于SVM-GC模型。2022年, Jian等人^[65]采用上述相同数据集,以51个构筑基团和 P 、 T 共计53个描述符作为输入参数,构建了SVM、RF、XGBoost三种浅层机器学习模型和MLPNN深度学习模型,预测 CO_2 溶解度。四种机器学习对测试集的 R^2 分别为:0.8240、0.9774、0.9865、0.9873,表明基于MLPNN的QSPR模型的预测效果好。最近, Tian等^[66]基于164种离子液体13055个 CO_2 溶解度数据(训练集10444个数据点+测试集2611个数据点),采用基于GC的离子片段贡献(ionic fragments contribution, IFC)方法将这些离子液体结构分成38个离子片段和 P 、 T 共计40个描述符作为输入参数,提出基于离子片段贡献的人工神经网络(IFC-ANN)和基于离子片段贡献的支持向量机(IFC-SVM)两种机器学习模型,预测 CO_2 溶解度,以建立气体溶解度与离子液体结构之间的QSPR关系。结果显示,IFC-ANN、IFC-SVM对测试集的 R^2 分别为

0.9711和0.9763,表明两种模型都能准确可靠地预测离子液体中 CO_2 溶解度,从而指导离子液体的筛选。

(5) 分子指纹描述符。分子指纹(fingerprint, FP)通常用于在稀疏向量中表示分子内子结构的存在(或不存在)。摩根指纹是化学信息学中用于描述分子的一种特征表示方法,通过考虑分子的局部环境来生成位向量(bit vector)表示。摩根指纹计算速度快,并且能很好地代表分子的显著特征。Jian等人^[65]采用不同温度不同压力下10117个离子液体 CO_2 溶解度数据(训练集80%数据点+测试集20%数据点),使用RDKit软件^[67]产生4096个FP描述符,再加上 P 、 T 共计4098个描述符作为输入参数,构建SVM、RF、XGBoost、MLPNN四种机器学习模型,预测 CO_2 溶解度。四种机器学习对测试集的 R^2 分别为:0.8633、0.9802、0.9847、0.9883,表明基于MLPNN的QSPR模型的预测效果好。

(6) 简化分子线性输入规范描述符。简化分子线性输入规范(simplified molecular input line entry system, SMILES)是一种用ASCII字符串明确描述三维分子结构的规范。SMILES串可以被大多数分子编辑软件导入并转换成二维图形或分子的三维模型。2023年,北京化工大学Zhang等人^[68]结合NIST离子液体数据库和MD模拟,构建了由17种阳离子和18种阴离子构成的40种离子液体的216个 CO_2 亨利常数数据(随机分为5份,训练集80%数据点+测试集20%数据点),并在此基础上利用RDKit软件将离子液体的SMILES分子结构转换为208个分子描述符作为输入参数,建立了SVM、RF、MLPNN三个 CO_2 亨利常数预测模型,用于预测17种阳离子 $\times$ 18种阴离子构成的306个离子液体的 CO_2 亨利常数。其中,MLPNN模型的 R^2 、MAE、MSE分别为0.9817、0.3023和0.2104。随后, Liu等人^[69]针对稀疏的小数据集和复杂的离子结构,提出利用基于语法定向变分自动编码器(syntax-directed variational autoencoder, SDVAE)通过独热编码将离子液体的SMILES字符串转换为矩阵坐标,再经过基于梯度的PSO算法优化,构建基于MLPNN和FM的DeepFM机器学习模型,预测每个坐标的 CO_2 溶解度。自动编码器内部还具有一个解码器,将矩阵坐标解码回分子,从而得到最佳离子液体。结果显示,1-乙基-3-甲基咪唑鎓甲基苯磺酸盐([Emim][TOS])为306个离子液体中的最佳离子液体。

(7) 分子图描述符。分子图(molecular graph)是一种表示分子的强大方式,其中节点表示原子,边表示化学键。GNN是处理图数据的神经网络,将图数据信息表示

为三个维度的向量(节点、边、整体)并在处理过程中使用邻接矩阵来表示图形结构. GNN的输入参数是一个图, 输出参数也是个图. GNN对属性向量优化的方法叫作消息传递机制, 根据机制的差别可以将GNN划分为: 图卷积网络(graph convolutional networks, GCN)、图注意力网络(graph attention network, GAT)、图同构网络(graph isomorphism network, GIN). Jian等人^[65]将阳离子和阴离子视为整体, 将整个图输入到GNN网络中, 开发了GCN、GAT、GIN模型, 其中GIN模型对测试集具有最优的MAE(0.0137)和 R^2 (0.9884). 以上表明, GNN对图形数据学习的有效性. 此外, 研究表明, 与CO₂有物理相互作用的碎片往往不如具有化学相互作用的片段重要, 且与氮连接的氢较少的胺基碎片可能更有利于与CO₂形成稳定的化学相互作用.

(8) 量子化学描述符. 量子化学描述符(quantum chemical descriptors)是从分子的量子力学计算中获得的、用于描述和表征分子特性的数值参数, 可以提供关于分子结构、能量、电荷分布以及反应性等方面的信息. 2017年, Mehraein和Riahi^[70]基于21种不同离子液体的CO₂溶解度数据, 以在PM6理论水平上采用Gaussian优化计算得到的单键数量、平均键信息含量、溶剂化能图形三个结构描述符作为输入参数, 建立了遗传算法优化的多元线性回归(GA-MLR)和非线性LSSVM两种机器学习模型为核心的QSPR模型. 结果显示, LSSVM模型大大提高了QSPR研究中的预测能力. 随后, Ghaslani等人^[71]通过Gaussian优化和Dragon 6软件^[72]计算, 得到156个有关阴离子的描述符和346个有关阳离子的描述符, 从而提出MLR和LSSVM两种机器学习QSPR模型, 用于预测离子液体中CO₂的亨利常数. 研究结果显示, 非线性的LSSVM方法具有更好的性能. Venkatraman和Alsberg^[73]以 T 、 P 、 M_w 和基于量子化学和分子轨道的变量共计104个描述符为输入参数, 构筑线性的偏最小二乘回归(partial least squares regression, PLSR)以及非线性的CTREE和RF三种机器学习模型, 预测CO₂在离子液体中的溶解度, 结果表明RF模型进行预测更合适. Kang等人^[74]通过对阴阳离子结构的量子化学计算优化, 得到五种离子液体静电势表面积和两种电荷分布面积, 连同 T , 八个描述符一并作为输入参数, 构建了MLR、SVM、ELM三个机器学习模型, 用于预测离子液体中CO₂的亨利定律常数. 结果表明ELM模型预测CO₂亨利常数最有效. 此外, 2020年, Aghaie和Zendehboudi^[58]分别采用不同输入变量构建

LSSVM、DT、RF、GA-MLR四种机器学习模型, 预测CO₂溶解度. 一个包含46种离子液体的数据集用 M_w 、 ω 、 T_c 、 T_p 作为输入变量; 另一个包含40种离子液体的数据集用结构描述符作为输入变量构建QSPR模型. 研究表明, 基于结构描述符的模型更准确、更稳健. 2023年, 日本中央大学Kuroki等人^[75]通过密度泛函理论(density functional theory, DFT)计算得到402114种离子液体的阴离子与阳离子的几何结构特征和电子结构特征作为输入变量, 随机从402114种离子液体中选择20000种离子液体用COSMO-RS理论得到其CO₂的亨利常数作为输出变量, 其中10000种离子液体数据作为测试集, 建立Matérn 5/2 GPR模型, 预测剩余离子液体的CO₂亨利常数, 从而搜索确定最佳阳离子/阴离子组合. 研究表明, 402114种离子液体中十四烷基三己基磷全氟辛烷磺酸盐([P₆₆₆₁₄][PFOS])具有更高的CO₂溶解度.

以上研究表明, 基于结构描述符所形成的QSPR机器学习模型, 比物理性质描述符的机器学习模型, 具有更准确、更稳健的特点, 并可以以结果(CO₂吸收量、溶解度、亨利常数)为导向, 搜索确定最佳离子液体结构.

4 总结与展望

如何根据离子液体的结构及性质特点, 设计开发基于离子液体的新型CO₂吸收剂, 实现对烟气中CO₂的高效捕集, 是目前离子液体碳捕集技术所面临的关键问题. 本文综述了人工智能机器学习应用于离子液体CO₂捕集性能研究的进展, 重点关注了离子液体描述符的选择及相应机器学习模型的建立. 在对上述描述符及机器学习模型分析比较后可以得出, 机器学习模型由“浅”入“深”, 从最初的基于离子液体物理性质作为输入参数, 迅速进化到以离子液体结构参数作为输入参数, 且基于后者所建立的QSPR模型比基于前者所建立的机器学习模型更准确、更稳健. 此外, 机器学习模型从最初的以实验的CO₂吸收量(溶解度)作为输出参数, 迅速进化到以COSMO-RS、MD模拟的CO₂吸收量(溶解度)作为输出参数, 克服了溶解度数据无法直接实验得到的困难.

虽然目前对于人工智能机器学习应用于离子液体CO₂捕集性能研究有了较大的进展, 我们认为, 要真正将人工智能机器学习应用到适用于工业烟气的低能耗离子液体CO₂吸收剂的设计, 仍面临如下挑战: (1) 一般

地, 低浓度CO₂的化学吸收导致新化学键和新物种(如氨基甲酸盐、碳酸盐等)生成, 然而当前离子液体碳捕集的机器学习模型未包含化学反应情况, 主要是吸收数据少、描述符建立困难; (2) 离子液体结构中的原子组合方式及所处方位对CO₂捕集性能影响较大, 例如研究表明氨基通过化学作用捕集CO₂且基团位于阴离子上比位于阳离子上更有利于CO₂捕集^[13,14]; (3) 烟气中除了CO₂外, 还有其他杂气(SO₂、NO_x)、水分等, 这些

组分对离子液体CO₂捕集性能具有一定影响. 因此, 未来以系统认识离子液体CO₂捕集性能“构效关系”为基础, 面向工业应用, 建立功能离子液体捕集CO₂的数据库, 针对性地设计有关化学反应和化学平衡的反应描述符, 创新机器学习模型架构, 助力功能离子液体吸收剂的开发是一个具有广阔发展前景的课题. 总之, 机器学习在离子液体CO₂捕集性能研究中的应用, 拓展了人工智能在能源环保中的应用, 提供了新思路新方法.

参考文献

- 1 Zeng S, Zhang X, Bai L, et al. Ionic-liquid-based CO₂ capture systems: Structure, interaction and process. *Chem Rev*, 2017, 117: 9625–9673
- 2 Zhang R, Ke Q, Zhang Z, et al. Tuning functionalized ionic liquids for CO₂ capture. *Int J Mol Sci*, 2022, 23: 11401
- 3 Xu Y, Zhang R, Zhou Y, et al. Tuning ionic liquid-based functional deep eutectic solvents and other functional mixtures for CO₂ capture. *Chem Eng J*, 2023, 463: 142298
- 4 Cui G, Lyu S, Wang H, et al. Tuning the structure of pyridinolates-based functional ionic liquids for highly efficient SO₂ absorption. *Fuel*, 2021, 303: 121311
- 5 Cui G, Zheng J, Luo X, et al. Tuning anion-functionalized ionic liquids for improved SO₂ capture. *Angew Chem Int Ed*, 2013, 52: 10620–10624
- 6 Cui G K, Zhao N, Zhang F T, et al. Progress in SO₂ capture by ionic liquids (in Chinese). *Chin Sci Bull*, 2016, 61: 3115–3126 [崔国凯, 赵宁, 张峰涛, 等. 离子液体捕集二氧化硫气体的研究进展. 科学通报, 2016, 61: 3115–3126]
- 7 Liu P, Cai K, Liu M, et al. Exceptionally effective H₂S absorption and conversion into thiols in novel superbase protic ionic liquids. *AIChE J*, 2023, 69: e17944
- 8 Li F, Laaksonen A, Zhang X, et al. Rotten eggs reevaluated: Ionic liquids and deep eutectic solvents for removal and utilization of hydrogen sulfide. *Ind Eng Chem Res*, 2022, 61: 2643–2671
- 9 Cui G, Jiang K, Liu H, et al. Highly efficient CO removal by active cuprous-based ternary deep eutectic solvents [HDEEA][Cl] + CuCl + EG. *Sep Purif Technol*, 2021, 274: 118985
- 10 Peng L, Shi M, Pan Y, et al. Ultrahigh carbon monoxide capture by novel protic cuprous-functionalized dicationic ionic liquids through complexation interactions. *Chem Eng J*, 2023, 451: 138519
- 11 Li K, Zong K, Wang X, et al. Ionic liquids and deep eutectic solvents for NH₃ absorption and separation: A review. *New J Chem*, 2023, 47: 21426–21445
- 12 Qiu R, Luo X, Yang L, et al. Regulated threshold pressure of reversibly sigmoidal NH₃ absorption isotherm with ionic liquids. *ACS Sustain Chem Eng*, 2020, 8: 1637–1643
- 13 Bates E D, Mayton R D, Ntai I, et al. CO₂ capture by a task-specific ionic liquid. *J Am Chem Soc*, 2002, 124: 926–927
- 14 Gurkan B E, de la Fuente J C, Mindrup E M, et al. Equimolar CO₂ absorption by anion-functionalized ionic liquids. *J Am Chem Soc*, 2010, 132: 2116–2117
- 15 Wen S, Wang T, Zhang X, et al. Novel amino acid ionic liquids prepared via one-step lactam hydrolysis for the highly efficient capture of CO₂. *AIChE J*, 2023, 69: e18206
- 16 Yoon B, Chen S, Voth G A. On the key influence of amino acid ionic liquid anions on CO₂ capture. *J Am Chem Soc*, 2024, 146: 1612–1618
- 17 Luo X, Guo Y, Ding F, et al. Significant improvements in CO₂ capture by pyridine-containing anion-functionalized ionic liquids through multiple-site cooperative interactions. *Angew Chem Int Ed*, 2014, 53: 7053–7057
- 18 Zhao T, Zhang X, Tu Z, et al. Low-viscous diamino protic ionic liquids with fluorine-substituted phenolic anions for improving CO₂ reversible capture. *J Mol Liquids*, 2018, 268: 617–624
- 19 Qiu L, Fu Y, Yang Z, et al. Surpassing the performance of phenolate-derived ionic liquids in CO₂ chemisorption by harnessing the robust nature of pyrazolonates. *ChemSusChem*, 2024, 17: e202301329
- 20 Wang C, Luo H, Jiang D, et al. Carbon dioxide capture by superbase-derived protic ionic liquids. *Angew Chem Int Ed*, 2010, 49: 5978–5981
- 21 Wang C, Luo X, Luo H, et al. Tuning the basicity of ionic liquids for equimolar CO₂ capture. *Angew Chem Int Ed*, 2011, 50: 4918–4922
- 22 Wang Z, Wu C, Wang Z, et al. CO₂ capture by 1,2,3-triazole-based deep eutectic solvents: The unexpected role of hydrogen bonds. *Chem Commun*, 2022, 58: 7376–7379
- 23 Sun X, Zeng S, Li G, et al. Selective CO₂ separation through physicochemical absorption by triazole-functionalized ionic liquid binary absorbents. *AIChE J*, 2024, 70: e18376

- 24 Huang Y, Cui G, Zhao Y, et al. Preorganization and cooperation for highly efficient and reversible capture of low-concentration CO₂ by ionic liquids. *Angew Chem Int Ed*, 2017, 56: 13293–13297
- 25 Huang Y, Cui G, Wang H, et al. Tuning ionic liquids with imide-based anions for highly efficient CO₂ capture through enhanced cooperations. *J CO₂ Util*, 2018, 28: 299–305
- 26 Deepan Kumar M, Jaccob M. Revealing the non-covalent interactions existing between CO₂ and succinimide-based ionic liquid: A DFT exploration. *J Mol Liquids*, 2024, 399: 124325
- 27 Suo X, Fu Y, Do-Thanh C L, et al. CO₂ chemisorption behavior in conjugated carbanion-derived ionic liquids via carboxylic acid formation. *J Am Chem Soc*, 2022, 144: 21658–21663
- 28 Wen Y R, Fu J, Liu D H. Advances in machine learning-based materials research for MOFs: Energy gas adsorption separation (in Chinese). *CIESC J*, 2024, doi: 10.11949/0438-1157.20231381 [文一如, 付佳, 刘大欢. 基于机器学习的MOFs材料研究进展: 能源气体吸附分离. 化工学报, 2024, doi: 10.11949/0438-1157.20231381]
- 29 Chen Y X, Liu C L, Gong Z J, et al. Machine learning-assisted solvent molecule design for efficient absorption of ethanethiol (in Chinese). *CIESC J*, 2024, doi: 10.11949/0438-1157.20231370 [陈宇翔, 刘传磊, 龚子君, 等. 机器学习辅助乙硫醇高效吸收溶剂分子设计. 化工学报, 2024, doi: 10.11949/0438-1157.20231370]
- 30 Wu Z H, Zhou T H, Lan X Y, et al. AI-driven innovative design of chemicals in practice and perspective (in Chinese). *Chem Ind Eng Prog*, 2023, 42: 3910–3916 [吴正浩, 周天航, 蓝兴英, 等. 人工智能驱动化学品创新设计的实践与展望. 化工进展, 2023, 42: 3910–3916]
- 31 Wang L, Zhang L, Du J. High-throughput screening of zeolite materials for CO₂/N₂ selective adsorption separation by machine learning (in Chinese). *Chem Ind Eng Prog*, 2023, 42: 148–158 [王璐, 张磊, 都健. 机器学习高效筛选用于CO₂/N₂选择性吸附分离的沸石材料. 化工进展, 2023, 42: 148–158]
- 32 Yang W H, Sun W H. Application and prospect of machine learning in polyolefin catalysts (in Chinese). *Chin Sci Bull*, 2022, 67: 1870–1880 [杨文泓, 孙文华. 机器学习在聚烯烃催化剂领域的应用与展望. 科学通报, 2022, 67: 1870–1880]
- 33 Zhu L T, Ouyang B, Zhang X B, et al. Progress on application of machine learning to multiphase reactors (in Chinese). *Chem Ind Eng Prog*, 2021, 40: 1699–1714 [朱礼涛, 欧阳博, 张希宝, 等. 机器学习在多相反应器中的应用进展. 化工进展, 2021, 40: 1699–1714]
- 34 Li Y F, Liu Z P. Machine learning leads to the discovery of Cu-Al alloy for efficient CO₂ reduction (in Chinese). *Chin Sci Bull*, 2020, 65: 2074–2075 [李晔飞, 刘智攀. 机器学习助力发现高效CO₂电催化还原Cu-Al合金催化剂. 科学通报, 2020, 65: 2074–2075]
- 35 Wang W J, Zheng L M, Cheng H Y, et al. Applications and progress of machine learning in wearable intelligent sensing systems (in Chinese). *Chin Sci Bull*, 2023, 68: 4630–4641 [王文君, 郑丽敏, 程泓宇, 等. 机器学习在可穿戴智能传感系统中的应用与进展. 科学通报, 2023, 68: 4630–4641]
- 36 Sun J, Sato Y, Sakai Y, et al. A review of ionic liquids and deep eutectic solvents design for CO₂ capture with machine learning. *J Clean Prod*, 2023, 414: 137695
- 37 Koutsoukos S, Philippi F, Malaret F, et al. A review on machine learning algorithms for the ionic liquid chemical space. *Chem Sci*, 2021, 12: 6820–6843
- 38 Cui G K, Lü S Z, Wang J J. Functional ionic liquids for carbon dioxide capture and separation (in Chinese). *CIESC J*, 2020, 71: 16–25 [崔国凯, 吕书贞, 王键吉. 功能化离子液体在二氧化碳吸收分离中的应用. 化工学报, 2020, 71: 16–25]
- 39 Jang J S R. ANFIS: Adaptive-network-based fuzzy inference system. *IEEE Trans Syst Man Cybern*, 1993, 23: 665–685
- 40 Zhao Y, Mulder R J, Houshyar S, et al. A review on the application of molecular descriptors and machine learning in polymer design. *Polym Chem*, 2023, 14: 3325–3346
- 41 Torrecilla J S, Palomar J, García J, et al. Modelling of carbon dioxide solubility in ionic liquids at sub and supercritical conditions by neural networks and mathematical regressions. *Chemom Intell Lab Syst*, 2008, 93: 149–159
- 42 Eslamimanesh A, Gharagheizi F, Mohammadi A H, et al. Artificial neural network modeling of solubility of supercritical carbon dioxide in 24 commonly used ionic liquids. *Chem Eng Sci*, 2011, 66: 3039–3044
- 43 Safamirzaei M, Modarress H. Application of neural network molecular modeling for correlating and predicting Henry's law constants of gases in [bmim][PF₆] at low pressures. *Fluid Phase Equilib*, 2012, 332: 165–172
- 44 Safamirzaei M, Modarress H. Correlating and predicting low pressure solubility of gases in [bmim][BF₄] by neural network molecular modeling. *Thermochim Acta*, 2012, 545: 125–130
- 45 Hamzehie M E, Fattahi M, Najibi H, et al. Application of artificial neural networks for estimation of solubility of acid gases (H₂S and CO₂) in 32 commonly ionic liquid and amine solutions. *J Nat Gas Sci Eng*, 2015, 24: 106–114
- 46 Fierro E N, Faúndez C A, Muñoz A S, et al. Application of a single multilayer perceptron model to predict the solubility of CO₂ in different ionic liquids for gas removal processes. *Processes*, 2022, 10: 1686
- 47 Baghban A, Ahmadi M A, Hashemi Shahraki B. Prediction carbon dioxide solubility in presence of various ionic liquids using computational intelligence approaches. *J Supercrit Fluids*, 2015, 98: 50–64
- 48 Tatar A, Naseri S, Bahadori M, et al. Prediction of carbon dioxide solubility in ionic liquids using MLP and radial basis function (RBF) neural

- networks. *J Taiwan Inst Chem Eng*, 2016, 60: 151–164
- 49 Baghban A, Mohammadi A H, Taleghani M S. Rigorous modeling of CO₂ equilibrium absorption in ionic liquids. *Int J Greenh Gas Control*, 2017, 58: 19–41
 - 50 Deng T, Liu F, Jia G. Prediction carbon dioxide solubility in ionic liquids based on deep learning. *Mol Phys*, 2020, 118: e1652367
 - 51 Mesbah M, Soltanali S, Momeni M, et al. Effective modeling methods to accurately predict the miscibility of CO₂ in ionic liquids. *Chem Eng Res Des*, 2020, 154: 262–272
 - 52 Ouair H, Hosseini A H, Nait Amar M, et al. Rigorous connectionist models to predict carbon dioxide solubility in various ionic liquids. *Appl Sci*, 2020, 10: 304
 - 53 Ghazani S H H N, Baghban A, Mohammadi A H, et al. Absorption of CO₂-rich gaseous mixtures in ionic liquids: A computational study. *J Supercrit Fluids*, 2018, 133: 455–465
 - 54 Mesbah M, Shahsavari S, Soroush E, et al. Accurate prediction of miscibility of CO₂ and supercritical CO₂ in ionic liquids using machine learning. *J CO₂ Util*, 2018, 25: 99–107
 - 55 Feng H, Zhang P, Qin W, et al. Estimation of solubility of acid gases in ionic liquids using different machine learning methods. *J Mol Liquids*, 2022, 349: 118413
 - 56 Mazari S A, Siyal A R, Solangi N H, et al. Prediction of thermo-physical properties of 1-Butyl-3-methylimidazolium hexafluorophosphate for CO₂ capture using machine learning models. *J Mol Liquids*, 2021, 327: 114785
 - 57 Daryayehsalameh B, Nabavi M, Vaferi B. Modeling of CO₂ capture ability of [Bmim][BF₄] ionic liquid using connectionist smart paradigms. *Environ Technol Innov*, 2021, 22: 101484
 - 58 Aghaie M, Zendejboudi S. Estimation of CO₂ solubility in ionic liquids using connectionist tools based on thermodynamic and structural characteristics. *Fuel*, 2020, 279: 117984
 - 59 Barati-Harooni A, Najafi-Marghmaleki A, Mohammadi A H. Efficient estimation of acid gases (CO₂ and H₂S) absorption in ionic liquids. *Int J Greenh Gas Control*, 2017, 63: 338–349
 - 60 Zhang K, Wu J, Yoo H, et al. Machine learning-based approach for tailor-made design of ionic liquids: Application to CO₂ capture. *Sep Purif Technol*, 2021, 275: 119117
 - 61 Valeh-e-Sheyda P, Heidarian P, Rezvani A. A novel molecular structure-based model for prediction of CO₂ equilibrium absorption in blended imidazolium-based ionic liquids. *J Mol Liquids*, 2022, 360: 119420
 - 62 Valeh-e-Sheyda P, Faridi Masouleh M, Zarei-Kia P. Prediction of CO₂ solubility in pyridinium-based ionic liquids implementing new descriptor-based chemoinformatics models. *Fluid Phase Equilib*, 2021, 546: 113136
 - 63 Wu T, Li W L, Chen M Y, et al. Prediction of Henry's law constants of CO₂ in imidazole ionic liquids using machine learning methods based on empirical descriptors. *Chem Pap*, 2021, 75: 1619–1628
 - 64 Song Z, Shi H, Zhang X, et al. Prediction of CO₂ solubility in ionic liquids using machine learning methods. *Chem Eng Sci*, 2020, 223: 115752
 - 65 Jian Y, Wang Y, Barati Farimani A. Predicting CO₂ absorption in ionic liquids with molecular descriptors and explainable graph neural networks. *ACS Sustain Chem Eng*, 2022, 10: 16681–16691
 - 66 Tian Y, Wang X, Liu Y, et al. Prediction of CO₂ and N₂ solubility in ionic liquids using a combination of ionic fragments contribution and machine learning methods. *J Mol Liquids*, 2023, 383: 122066
 - 67 Landrum G. RDKit: Open-source cheminformatics. <http://www.rdkit.org>. Accessed November 11, 2022
 - 68 Zhang W, Wang Y, Ren S, et al. Novel strategy of machine learning for predicting Henry's law constants of CO₂ in ionic liquids. *ACS Sustain Chem Eng*, 2023, 11: 6090–6099
 - 69 Liu X, Chu J, Huang S, et al. Machine learning-based design of ionic liquids at the atomic scale for highly efficient CO₂ capture. *ACS Sustain Chem Eng*, 2023, 11: 8978–8987
 - 70 Mehraein I, Riahi S. The QSPR models to predict the solubility of CO₂ in ionic liquids based on least-squares support vector machines and genetic algorithm-multi linear regression. *J Mol Liquids*, 2017, 225: 521–530
 - 71 Ghaslani D, Gorji Z E, Gorji A E, et al. Descriptive and predictive models for Henry's law constant of CO₂ in ionic liquids: A QSPR study. *Chem Eng Res Des*, 2017, 120: 15–25
 - 72 Todeschini R, Consonni V, Mauri A, et al. DRAGON-Software for the Calculation of Molecular Descriptors, Version 6 for the Windows. 2010
 - 73 Venkatraman V, Alsberg B K. Predicting CO₂ capture of ionic liquids using machine learning. *J CO₂ Util*, 2017, 21: 162–168
 - 74 Kang X, Liu C, Zeng S, et al. Prediction of Henry's law constant of CO₂ in ionic liquids based on S_{EP} and $S_{g-profile}$ molecular descriptors. *J Mol Liquids*, 2018, 262: 139–147
 - 75 Kuroki N, Suzuki Y, Kodama D, et al. Machine learning-boosted design of ionic liquids for CO₂ absorption and experimental verification. *J Phys Chem B*, 2023, 127: 2022–2027

Summary for “人工智能机器学习助力离子液体CO₂捕集性能分析”

Artificial intelligence and machine learning for analysis of CO₂ capture by ionic liquids

Ruina Zhang^{1†}, Yuan Tian^{2†}, Chunliang Ge³, Wei Zhang³, Hanfeng Lu¹ & Guokai Cui^{1*}¹ Institute of Industrial Catalysis, College of Chemical Engineering, State Key Laboratory Breeding Base of Green Chemistry Synthesis Technology, Zhejiang University of Technology, Hangzhou 310014, China;² Beijing Key Laboratory of Ionic Liquids Clean Process, CAS Key Laboratory of Green Process and Engineering, State Key Laboratory of Multiphase Complex Systems, Institute of Process Engineering, Chinese Academy of Sciences, Beijing 100190, China;³ Zhejiang Zheneng Technology & Environment Group Co., Ltd., Hangzhou 310012, China

† Equally contributed to this work

* Corresponding author, E-mail: chemcgk@163.com

With the implementation of China's carbon peak and carbon neutrality strategy, CO₂ capture in industries such as coal-fired power, steel, and cement is urgent. To develop alternative absorbents without high energy consumption and volatility of amine-based absorbents, nonvolatile, highly stable, and designable ionic liquid-based CO₂ absorbents have received widespread attention. Machine learning (ML) is the core of artificial intelligence (AI) and a powerful tool that can search for patterns in high-dimensional data. In recent years, ML has been widely applied in fields such as energy and materials, chemistry and chemical engineering, and life science and medicine. By applying ML to the study of CO₂ capture by ionic liquids (ILs), computers can simulate the linear or nonlinear relationship between the structure of ILs and CO₂ capture performance, as well as predict the CO₂ capture performance by ILs with specific structures under specific conditions. Compared to traditional trial and error methods, AI and ML technology have become an important way to quickly predict the CO₂ capture performance of ILs and select the optimal IL-based absorbents. Based on a brief introduction to algorithms and models of ML including artificial neural network (ANN), support vector machine (SVM), tree-based models, multivariate linear regression (MLR), and gaussian processes regression (GPR), this article reviews the progress of AI and ML in the study of CO₂ capture by ILs in recent years from a viewpoint of descriptors of ILs. This review focuses on the types of molecular structure descriptors of ILs and the establishment of corresponding quantitative structure-property relationship (QSPR) or quantitative structure-activity relationship (QSAR) models. Thereby, ML models can be divided into two categories based on the types of descriptors used to describe the CO₂ dissolution performance of ILs. One is the ML models based on physical property descriptors, and another is the ML-based QSPR models based on molecular structure descriptors. The descriptors of former include molecular weight (*M_w*), acentric factor, (*ω*), critical temperature (*T_c*), critical pressure (*P_c*), critical compressibility factor (*Z_c*), operating temperature (*P*), and operating pressure (*T*). The descriptors of latter are based on the concept of “structure determines properties”, and the structure of anions and cations in ILs has a significant impact on their CO₂ capture properties. The descriptors of latter are generated based on atoms, ionic fragments contribution (IFC), group contribution (GC), fingerprint (FP), simplified molecular input line entry system (SMILES), molecular graph, quantum chemical descriptors. The development of ML models is gradually evolving from a single model to an integrated model. Generally, ML models have improved from “simple learning” to “deep learning”. ML-based QSPR models using molecular structural descriptors are more accurate and robust than ML models using physical property descriptors. Therefore, ML-based QSPR methods can be used to search and identify the optimal structures of ILs with the CO₂ absorption, solubility, or Henry constant as the aim. Finally, the perspective for further researches on the design of ML models for predicting CO₂ capture by ILs was also provided. It should be considered in the ML models that flue gas containing SO₂, NO_x, and moisture will affect the capture performance of ILs. Besides, functional groups in the functionalized ILs have a significant impact on capture performance, suggesting different ML models from models for physical solubility. In summary, the application of ML models in the prediction of CO₂ capture by ILs will expand the application of AI technology in the fields of energy and environmental protection.

artificial intelligence, machine learning, ionic liquid, carbon dioxide capture, quantitative structure-property relationshipdoi: [10.1360/TB-2024-0163](https://doi.org/10.1360/TB-2024-0163)