

# 基于深度学习预测 CO<sub>2</sub> 在离子液体中溶解度的研究

罗秉伟<sup>1,2</sup>, 杨玮婧<sup>1,2\*</sup>, 刘振盈<sup>1,2</sup>, 田敬晓<sup>1,2</sup>, 丁将敏<sup>1,2</sup>

(1. 国家能源集团宁夏煤业煤炭化学工业技术研究院, 宁夏 银川 750411;

2. 宁夏回族自治区煤炭间接液化产品高值化利用重点实验室, 宁夏 银川 750411)

**摘要:** 构建了深度神经网络(DNN)和随机森林(RF)技术用于预测 CO<sub>2</sub> 在 ILs 中的溶解度。收集 5 000 个不同温度和压力下 ILs 的 CO<sub>2</sub> 溶解度值, 采用分子描述符(MD)充分获取到 ILs 的复杂结构信息。以 MD 作为 DNN、RF 机器学习模型的输入特征。其中, MD-RF 模型以卓越的准确脱颖而出, 展示了 0.988 4 的决定系数( $R^2$ )和 0.000 5 的均方误差(MSE)。

**关键词:** 离子液体; CO<sub>2</sub>; 溶解度; 机器学习; 气体吸收

中图分类号: TH3

文献标志码: A

文章编号: 0253-4320(2026)02-0232-04

DOI: 10.16606/j.cnki.issn0253-4320.2026.02.036

## Prediction of solubility of CO<sub>2</sub> in ionic liquids based on deep learning

LUO Bing-wei<sup>1,2</sup>, YANG Wei-jing<sup>1,2\*</sup>, LIU Zhen-ying<sup>1,2</sup>, TIAN Jing-xiao<sup>1,2</sup>, DING Jiang-min<sup>1,2</sup>

(1. Institute of Coal Chemical and Industry Technology, Ningxia Coal Industry Co., Ltd., China Energy, Yinchuan 750

411, China; 2. Key Laboratory of High Value Utilization of Indirect Coal Liquefaction Products in Ningxia Hui

Autonomous Region, Yinchuan 750411, China)

**Abstract:** Deep neural networks (DNN) and random forest (RF) techniques were used to predict the solubility of CO<sub>2</sub> in ILs in this work. 5000 CO<sub>2</sub> solubility values of ILs at different temperatures and pressures were collected and the complex structural information of ILs using molecular descriptors (MD) were fully obtained. Using MD as the input feature for DNN and RF machine learning models. Among them, the MD-RF model stands out for its excellent accuracy, displaying a coefficient of determination ( $R^2$ ) of 0.988 4 and a mean squared error (MSE) of 0.000 5.

**Key words:** ionic liquids; CO<sub>2</sub>; solubility; machine learning; gas absorption

近年来,大气中 CO<sub>2</sub> 等温室气体排放量急剧增加,导致了干旱、气候异常、海平面上升等一系列严重问题,给社会各方面的发展带来了严重的负面影响<sup>[1]</sup>。为减少温室气体排放和缓解全球变暖,世界各地的研究团队正在努力开发新的高效温室气体减排和利用技术<sup>[2]</sup>。ILs 凭借独特的分子结构展现出卓越的性能,在吸收和捕集酸性气体方面显示出较高的效率和选择性,相较于传统有机溶剂,ILs 被认为是处理酸性气体问题的潜在解决方案。

ILs 是指全部由离子组成在室温或接近室温呈液态的盐。自从 ILs 在 1914 年首次报道以来<sup>[3]</sup>,在石油、化工、环境科学等领域引起广泛的关注。1975 年,Chum 等<sup>[4]</sup>合成 *N*-烷基吡啶的氯铝酸盐 ILs,同时利用 FT-IR、Raman 等技术手段对 ILs 的结构和

性质进行了深入的研究。该 ILs 的成功合成意味着 ILs 在有机、材料等方面进行工业应用有了理论基础。但是由于该 ILs 性质不够稳定,限制了氯铝酸盐 ILs 在化工中的应用。使得研究人员未来还需对合成性质稳定的 ILs 继续努力。随着相关领域的发展,ILs 的种类不断扩大,与此同时,科研人员开发了许多新功能的 ILs,引起了学术界和工业界的广泛研究。Nordness 等<sup>[5]</sup>探讨了阳离子溶液和阴离子烷基链的长度、结构以及阳离子和阴离子的相互作用基序、温度以及 ILs 溶液中溶剂强度的影响。证明了阳离子和阴离子的离解程度会影响黏度、密度、表面张力、挥发性以及许多其他重要的化学和物理性质,对 ILs 未来的继续研究具有重大意义。

目前,ILs 在气体吸收方面已经有了广泛的应

收稿日期:2025-05-20;修回日期:2025-12-24

基金项目:2019 年宁夏青年拔尖人才培养工程(宁人社发[2019]90 号);2024 年宁夏回族自治区重点研发计划项目(2024BEE02038);2023 年国家能源集团科技创新项目(GJNY-23-53)

作者简介:罗秉伟(1994-),男,硕士,研究方向为聚烯烃新产品开发与应用,364758026@qq.com;杨玮婧(1983-),女,本科,高级工程师,研究方向为聚烯烃新产品开发与应用,通讯联系人,87471628@qq.com。

用。Zhan 等<sup>[6]</sup>使用 ILs 搭建了一种包括热力学模型、流程模拟、多目标计算在内的含氨尾气分离和氨回收新工艺的系统策略,结果表明 ILs 的使用大大降低了净化成本和能耗,实现含氨气体的净化。Huang 等<sup>[7]</sup>使用 [Bmim][BF<sub>4</sub>] 溶液与 LiBF<sub>4</sub>、NaBF<sub>4</sub> 和 KBF<sub>4</sub> 3 种碱金属混合对 CO<sub>2</sub>/CH<sub>4</sub> 进行捕获和分离,金属离子都能与 ILs 缔合从而实现气体的纯化。Wang 等<sup>[8]</sup>通过分子动力学模拟探究纳米约束 ILs 杂化膜分离 CO<sub>2</sub> 和 CH<sub>4</sub> 的潜力,揭示了气体在石墨烯约束下结构、热力学性质和输运之间的定量关系,对高效低成本分离 CO<sub>2</sub> 具有重要意义。各研究人员对 ILs 分离捕捉气体作用规律的揭示为 ILs 气体分离的工业化应用提供了重要支撑。但是仍存在着很重大的问题:缺乏对 ILs 微观结构和宏观性质之间关系的研究,无法有效地预测 ILs 的性质,这使得 ILs 进一步的开发和应用受到很大的限制。同时,ILs 的复杂性导致科研人员和工程师在缺乏广泛的物性数据库的情况下,面临筛选合适 ILs 的困难。

随着越来越多关于 ILs 中气体溶解度的研究成果被公开报道,基于大数据驱动的人工智能模型预测气体在 ILs 中溶解度逐渐受到关注。这种方法主要利用已知的 ILs 结构和组成信息,通过挖掘和学习大量数据中的经验规律来进行性质的预测,从具有大量特征变量的大型数据集中提取独特见解。Khajeh 等<sup>[9]</sup>提出使用人工神经网络来预测 CO<sub>2</sub> 在聚合物中的溶解度,并表明所提出的模型是一种有效的方法。Song 等<sup>[10]</sup>使用基团贡献的方法关联 CO<sub>2</sub> 的溶解度与 ILs 结构、温度和压力之间的关系,开发了 ANN 和 SVM 2 种模型,分析了 2 种模型的预测误差分布。

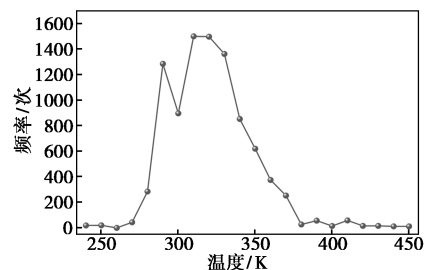
本工作提出了一种基于 MD 表征 ILs 结构的框架,并成功地与 DNN 和 RF 2 种不同的智能机器学习算法相结合,用于预测不同 ILs 中 CO<sub>2</sub> 溶解度。通过从文献中收集的 5 000 个不同温度和压力下 ILs 的 CO<sub>2</sub> 溶解度值,建立了 ILs 数据库,构建采用 RMSE、MSE 和 R<sup>2</sup> 衡量不同预测模型的效能。

## 1 基本理论与方法

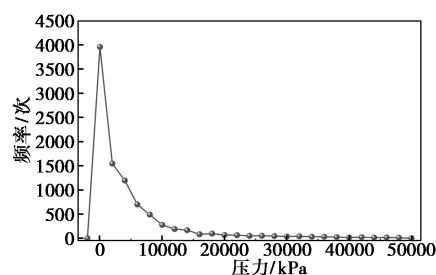
### 1.1 数据统计分析

本工作共收集了 5 000 个 CO<sub>2</sub> 在 ILs 中的溶解度数据点,涵盖了 243.2~453.15 K 的温度范围,涉及 103 种 ILs,数据来源于 Lei 等<sup>[11]</sup>早期的研究成果。图 1 展示了模型数据集在压力、温度和溶解度方面的分布情况。从图中可以观察到,温度和溶解

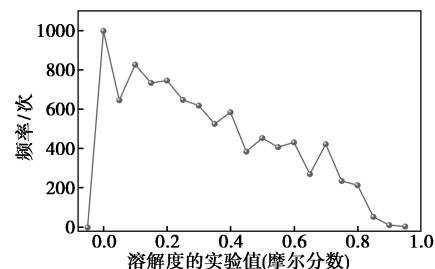
度数据呈现出均匀分布,而压力数据主要集中在 0~50 000 kPa 范围内。在这些 ILs 中,阳离子共有 43 种,包括吡啶、吡咯烷、咪唑、哌啶和铵等;阴离子则包括硫酸氢根、硝酸根、氯、四氟硼酸根、双(三氟甲基磺酰)胺根等共 28 种。



(a) 温度



(b) 压力



(c) CO<sub>2</sub> 溶解度分布

图 1 ILs 数据集中温度、压力和 CO<sub>2</sub> 溶解度分布

### 1.2 分子描述符

分子描述符(MD)可以用于量化 ILs 中特定分子结构的相关信息,这些描述符细分为定性和定量 2 种类型<sup>[12]</sup>。本工作使用开源化学信息 Python 软件包的 RDKit 软件获取阴离子和阳离子的分子描述符(MD),共计 168 个 MD。为了增加 MD 的丰富性,还添加了温度和压力作为附加描述符,使得 MD 的总数达到了 338 个。通过特征工程,成功地减少了数据中的冗余信息。根据特征选择器方法,删除了相关性大于 0.98 的 55 个描述符以及 151 个唯一值描述符。最终,得到了 132 个对模型贡献较大的描述符。由于原始数据集中不同描述符的值存在较大的量级差异,这可能会影响模型的正确训练。因此,利用 scikit-learn 模块对输入特征数据进行了归

一化处理。

### 1.3 深度神经网络

深度神经网络(DNN)是一种常用的机器学习模型。在 DNN 结构中,数据从输入层开始经过一系列隐藏层的处理,最终到达输出层,产生预测结果<sup>[13]</sup>。DNN 的架构通常包含 4 层,其中包括输入层和输出层,层与层之间是全连接的。输入层负责接收来自分子结构的特征数据,隐藏层则通过激活函数将输入信息转换成隐藏向量,再传递给输出层,最终输出层生成模型的预测结果。在网络训练过程中,反向传播算法用来调整神经元的权重和偏置。随着训练的推进,权重会在每次迭代中进行更新,从而提高模型的准确性,达到更好的数据拟合效果。为了让模型具有更强的表达能力,DNN 在隐藏层中引入了激活函数,这一设计大大增强了模型处理复杂数据的能力。

### 1.4 随机森林

随机森林(RF)是一种基于非参数技术的集成学习模型,是 Bagging 算法改进和扩展的一种由层次结构规则组成的算法,被广泛应用于解决回归和分类问题<sup>[14]</sup>。RF 的核心思想是构建多个决策树,并将它们整合在一起,以增强模型的稳定性和提高预测的准确性。每个决策树都是在数据集的一个随机子集上训练得到的。这种随机性体现在 2 个方面:一是通过对原始数据集进行放回抽样来构建训练集;二是在分裂决策树的节点时,从一个随机的特征子集中选择最优特征。这种方法既可以减少模型的过拟合风险,也可以提高模型在未知数据上的泛化能力。RF 的核心步骤包括:使用自主抽样法对原始数据进行可放回的随机抽样,形成新的样本集;利用决策树对这些数据进行建模并生成预测结果;最后,通过汇总多个决策树的预测结果来选择最佳的预测值。RF 的主要优势在于处理高维数据的能力强,对数据预处理要求低,能够有效地提供关键的特征选择信息。此外,还具有出色的并行性,使得算法可以轻松地多核处理器上进行并行计算。

## 2 结果与讨论

### 2.1 ILs 吸收 CO<sub>2</sub> 的模型性能讨论

利用  $R^2$ 、MSE 和 RMSE 3 种统计指标对已建立的 MD-DNN 和 MD-RF 模型进行预测和评估,发现所提出的机器学习模型均能有效预测 CO<sub>2</sub> 在 ILs 中的溶解度。对不同模型性能的评估表明,这些预测模型方案在训练和测试阶段都能产生卓越的价值,

表 1 显示了 2 种预测模型的评估性能。

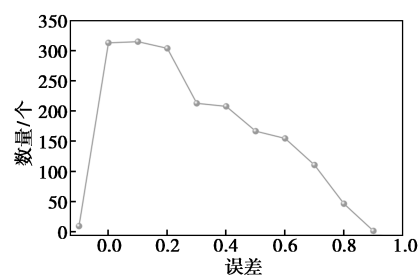
表 1 不同机器学习模型的性能指标

模型	数据	$R^2$	MSE	RMSE
MD-DNN	训练数据	0.9842	0.0015	0.039
	测试数据	0.9835	0.0011	0.033
	全部数据	0.9827	0.0009	0.030
MD-RF	训练数据	0.9926	0.0012	0.035
	测试数据	0.9813	0.0006	0.0245
	全部数据	0.9884	0.0005	0.0224

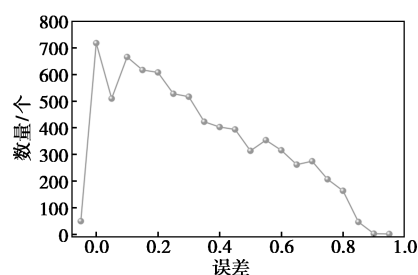
从表 1 中的统计指标数据可以得出结论:MD-DNN 和 MD-RF 2 个模型都能准确预测 CO<sub>2</sub> 在 ILs 中的溶解度,且  $R^2$  均大于 0.9。MD-RF 训练集的  $R^2$  略高于 MD-DNN,但测试集的  $R^2$  却低于 MD-DNN,这是因为 DNN 模型特别擅长处理高度非线性和高维度的数据,能够自动从数据中学习复杂的特征表示,并且 DNN 模型具有更好的泛化能力和对复杂数据模式的捕捉能力,对学习新数据有更好的适应性,因此,它对未经训练的测试集数据具有更高的预测精度。但从全部数据来看,MD-RF 模型的  $R^2$  更接近于 1,均方根误差和均方误差均比 MD-DNN 小,表明 MD-RF 模型具有更高的精度。

### 2.2 误差分析

为了深入探讨并比较不同模型的性能,图 2 提供了一种直观的方式,通过展示实验值与预测值之间的误差分布,来评估各模型的精度。通过误差直方图分析,可以观察到 2 个模型中大多数误差都聚



(a) MD-DNN



(b) MD-RF

图 2 模型的误差直方图

集在-0.1~0.1之间的区域,只有一小部分误差的绝对值超过了0.1。与MD-DNN模型相比,MD-RF模型显示出更窄的误差范围,这与之前的分析结果是一致的,进一步证明了其优越的预测能力,具有更高的预测精度。综合考虑所有因素,MD-RF模型的整体预测性能最为出色,能够准确地预测CO<sub>2</sub>在ILs中的溶解度,展现了在处理复杂非线性关系和高度特定领域问题时的强大能力。

### 3 结论

提出了一种基于MD量化ILs结构与DNN、RF 2种不同机器学习算法相结合的用于预测CO<sub>2</sub>在不同ILs中溶解度的方法。通过文献收集CO<sub>2</sub>在离子液体中不同条件下的溶解度,建立了离子液体MD结构信息数据库。将分子的MD输入DNN、RF机器学习模型,预测了CO<sub>2</sub>在离子液体中的溶解度。并利用R<sup>2</sup>、MSE和RMSE 3种统计指标对模型进行评估,且2个模型的误差都聚集在-0.1~0.1之间。但MD-RF模型的整体预测性能具有更高的精度,展现了在处理复杂非线性关系和高度特定领域问题时的强大能力,能够准确地预测CO<sub>2</sub>在ILs中的溶解度。

### 参考文献

- [1] Heinrich L, Koschinsky A, Markus T, *et al.* Quantifying the fuel consumption, greenhouse gas emissions and air pollution of a potential commercial manganese nodule mining operation [J]. *Marine Policy*, 2020, 114: 103678.
- [2] Dunn C A, Shi Z, Zhou R, *et al.* (cross-linked poly(ionic liquid)-ionic liquid-zeolite) Mixed-Matrix membranes for CO<sub>2</sub>/CH<sub>4</sub> gas separations based on curable ionic liquid prepolymers [J]. *Industrial & Engineering Chemistry Research*, 2019, 58(11): 4704-4708.
- [3] Plechkova N V, Seddon K R. Applications of ionic liquids in the chemical industry [J]. *Chem Soc Rev*, 2008, 37(1): 123-150.
- [4] Chum H L, Koch V R, Miller L L, *et al.* Electrochemical scrutiny of organometallic iron complexes and hexamethylbenzene in a room temperature molten salt [J]. *Journal of the American Chemical Society*, 1975, 97: 3264-3265.
- [5] Nordness O, Brennecke J F. Ion Dissociation in ionic liquids and ionic liquid solutions [J]. *Chemical Reviews*, 2020, 120(23): 12873-12902.
- [6] Zhan G, Cao F, Bai L, *et al.* Process Simulation and Optimization of ammonia-containing gas separation and ammonia recovery with ionic liquids [J]. *ACS Sustainable Chemistry & Engineering*, 2020, 9(1): 325312.
- [7] Huang X. Experimental and simulation study on the capture and separation of CO<sub>2</sub>/CH<sub>4</sub> by alkali metal complex ionic liquid [J]. *Fuel*, 2022, 329: 125444.
- [8] Wang C, Wang Y, Liu J, *et al.* Entropy driving highly selective CO<sub>2</sub> separation in nanoconfined ionic liquids [J]. *Chemical Engineering Journal*, 2022, 440: 135918.
- [9] Khajeh A, Modarress H, Rezaee B. Application of adaptive neuro-fuzzy inference system for solubility prediction of carbon dioxide in polymers [J]. *Expert Systems with Applications*, 2009, 36(3): 5728-5732.
- [10] Song Z, Shi H, Zhang X, *et al.* Prediction of CO<sub>2</sub> solubility in ionic liquids using machine learning methods [J]. *Chemical Engineering Science*, 2020, 223: 115752.
- [11] Lei Z, Dai C, Chen B. Gas Solubility in ionic liquids [J]. *Chemical Reviews*, 2014, 114(2): 1289-1326.
- [12] 张钰,魏世丞,董超芳,等.定量结构-性质关系(QSPR)中的计算方法研究进展[J].*科学通报*, 2021, 66(22): 2832-2844.
- [13] Stelzer F, Röhm A, Vicente R, *et al.* Deep neural networks using a single neuron: Folded-in-time architecture using feedback-modulated delay loops [J]. *Nature Communications*, 2021, 12: 5164.
- [14] 张镔滨,陈玉明,吴克寿,等.粒向量驱动的随机森林分类算法研究[J].*计算机工程与应用*, 2024, 60(3): 148-156. ■
- [15] 杨成玉,马军,李广玉,等.大型碱性电解水制氢装备多对一的应用与实践[J].*太阳能*, 2022, (5): 103-114.
- [16] Clift R, Grace J R, Weber M E. *Bubbles, drops, and particles* [M]. Mineola: Dover Publications, 2005.
- [17] Ullberg Ø. Modeling of advanced alkaline electrolyzers: A system simulation approach [J]. *International Journal of Hydrogen Energy*, 2003, 28(1): 21-33.
- [18] Sakas G, Ibanez-Rioja A, Ruuskanen V, *et al.* Dynamic energy and mass balance model for an industrial alkaline water electrolyzer plant process [J]. *International Journal of Hydrogen Energy*, 2022, 47(7): 4328-4345.
- [19] Dang J, Yang F, Li Y, *et al.* Experiments and microsimulation of high-pressure single-cell PEM electrolyzer [J]. *Applied Energy*, 2022, 321: 119351.
- [20] Khan M A, Golsorkhi M S, Ebel T. Modeling the operational dynamics of alkaline electrolyzer [C]. 2024 IEEE Design Methodologies Conference (DMC), IEEE, 2024: 1-6.
- [21] Trinke P, Haug P, Bensmann B, *et al.* Hydrogen crossover in PEM and alkaline water electrolysis: Mechanisms, direction comparison, and mitigation strategies [J]. *Journal of the Electrochemical Society*, 2018, 165(7): F502-F513.
- [22] Dieguez P M, Ursua A, Sanchis P, *et al.* Thermal performance of a commercial alkaline water electrolyzer [J]. *International Journal of Hydrogen Energy*, 2008, 33(24): 7338-7354.
- [23] Yang J, Zhang J, Liu M, *et al.* Dynamic simulation and performance analysis of alkaline electrolyzers for renewable energy-powered hydrogen production [J]. *Energies*, 2024, 17(19): 4915.
- [24] 王润东,黎静华,韦善阳.基于多物理场耦合模型的碱性水电解槽工作特性[J].*高压电技术*, 2024, 50(7): 3209-3220. ■

(上接第231页)