

# 机器学习筛选用于气体吸附分离和存储的金属有机骨架材料

赵晨 曹蓉 夏杰桢 吴琪\*

(西藏大学理学院; 西藏大学供氧研究院 珠峰研究院 拉萨 850000)

**摘要** 金属有机骨架(MOF)因其高孔隙率、高比表面积和结构可调性,在气体吸附分离领域广泛应用。随着 MOF 数量激增,传统分子模拟和实验方法验证 MOF 性能成本高且速度慢,因此目前 MOF 筛选工作已转向高通量计算辅助的机器学习(Machine-learning, ML)。机器学习作为一种高效的大数据处理方法,能够高通量筛选的基础上对数据进行拟合,从而快速而准确地筛选出气体吸附分离材料,并深入挖掘其结构与性能之间的关系。本文回顾了近年机器学习应用于 MOF 筛选的研究,重点讨论了一些运用机器学习从大量结构中筛选出可用于 CH<sub>4</sub>、H<sub>2</sub> 和 CO<sub>2</sub> 等气体吸附分离与存储的 MOF 材料的工作;同时,梳理了当前 MOF 材料筛选工作中的研究思路和进展,并指出了机器学习在筛选 MOF 材料工作中面临的一些瓶颈和挑战;最后,对该领域的未来发展前景进行了展望。

**关键词** 金属有机骨架 机器学习 高通量筛选 气体吸附分离

## Machine Learning Screening of MOF Materials for Gas Adsorption Separation

Zhao Chen, Cao Rong, Xia Jiezheng, Wu Qi\*

(Department of Physics, Tibet University; Institute of Oxygen Supply, Tibet University, Lhasa, 850000)

**Abstract** Metal-organic framework (MOF) is widely used in the field of gas adsorption separation due to their high porosity, high specific surface area and structural adjustability. As the number of MOF surges, traditional molecular simulation and experimental methods are costly and slow to verify the performance of MOF, so the current MOF screening efforts have turned to high-throughput computation-assisted machine learning (ML). As an efficient big data processing method, ML can fit data on the basis of high-throughput computational screening (HTCS), so as to quickly and accurately screen out gas adsorption separation materials, and deeply explore the relationship between their structure and performance. This paper reviews recent studies on the application of ML to MOF screening. In this paper, we focus on some work using ML to select MOF materials from a large number of structures that can be used for adsorption, separation and storage of CH<sub>4</sub>, H<sub>2</sub> and CO<sub>2</sub>. Moreover, we review the current research ideas and progress in MOF material screening, and point out some bottlenecks and challenges faced by ML in the screening of MOF materials. Finally, the future development prospect of this field is prospected.

**Keywords** MOF, Machine learning (ML), High-throughput computational screening (HTCS), Gas adsorption separation

## 1 背景

金属有机骨架(Metal-organic framework, MOF)是一种由金属节点和有机配体组成的多孔晶体材料<sup>[1]</sup>,与其他多孔材料(如活性炭、二氧化硅、硅胶、沸石等)相比,MOF 具有高孔隙率、大比表面积、热稳定性和可控结构等优势<sup>[2,3]</sup>。这使

得 MOF 在气体储存与分离<sup>[4-6]</sup>、催化剂<sup>[7,8]</sup>、超级电容器<sup>[9]</sup>、锂离子电池<sup>[10]</sup>等领域都有优异的表现,尤其是在气体吸附方向,图 1(a)展示了 Web of science 中关于 MOF 和 MOF + gas separation 相关文章的历年检索数量。

MOF 具有夸张的结构多样性,目前已发现的结构超过 90 万种,理论可设计的结构超过 500 万

\* 联系人,吴琪 女,博士,主要从事计算物理研究,E-mail: wuqi\_zangda@163.com

西藏自治区科技厅自然科学基金重点项目(XZ202301ZR0026G)资助

2023-09-05 收稿,2023-11-01 接受

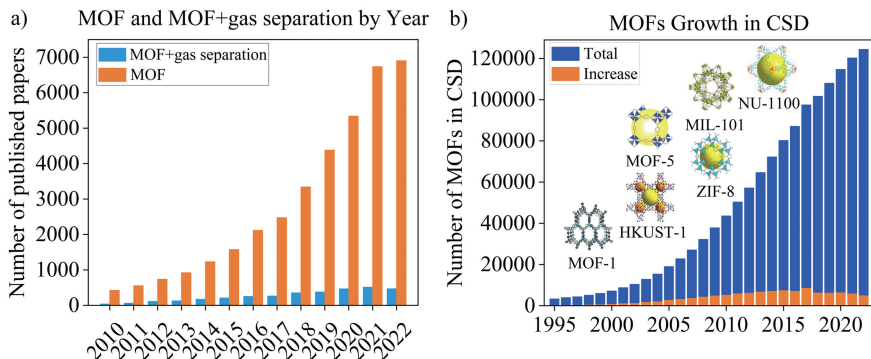


图 1 (a) 历年 Web of science 关于 MOF 和 MOF + gas separation 的文章数量 ;(b) CSD 数据库 MOF 收录情况和对应年份具有代表性的 MOF

Fig. 1 (a) The number of Web of science articles on the separation of MOF and MOF + gas separation; (b) The MOF included in the CSD database and the representative MOF of the corresponding year

种<sup>[11]</sup>。图 1(b)展示了历年来剑桥晶体结构数据库(Cambridge structural database, CSD)对 MOF 的收录情况以及 MOF 发展过程中出现的具有代表性的 MOF。例如,1995 年,首个被正式定义的金属有机骨架—MOF-1<sup>[12]</sup>;1999 年,《Nature》上报道的 MOF-5 是首个在气体吸附领域应用的 MOF;1999 年香港科技大学设计的 HKUST-1 在 298K  $6.5 \times 10^6 \text{ Pa}$  下具有  $267 \text{ cm}^3 (\text{STP}) \cdot \text{cm}^{-3}$  的  $\text{CH}_4$  存储量,成为了首个超过美国能源署规定的工作容量大于  $200 \text{ cm}^3 (\text{STP}) \cdot \text{cm}^{-3}$  标准的 MOF<sup>[13]</sup>;2005 年, Férey 研究团队<sup>[14]</sup>首次提出并利用计算机模拟辅助设计了具有超高比表面积( $5900 \text{ m}^2 \cdot \text{g}^{-1}$ )的 MIL-101;2006 年, Yaghi 团队<sup>[15]</sup>设计了具有极高热稳定性的 ZIF 系列。使用传统计算方法筛选 MOF 显然无法应对如此庞大数量的结构,而高通量筛选(High-throughput computational screening, HTCS)和机器学习能快速评估 MOF 的吸附性能<sup>[16,17]</sup>,并确定结构与性能间的构效关系。近年来, Alyintas 等<sup>[18]</sup>总结了用于氢气存储的 MOF 的设计和思路。Firooz 等<sup>[19]</sup>综述了 MOF 在气体分离方向的部分应用,讨论了包括气相色谱、液相色谱、膜分离等方向的最新进展。Lin 等<sup>[20]</sup>总结了预测 MOF 水稳定性和吸水性的 ML 工作,重点分析了微孔和介孔的作用。这些工作为 ML 在 MOF 筛选方向的研究提供了很多宝贵的建议。然而,不仅要分离和存储两个角度分别总结 ML 应用于 MOF 筛选的工作,还需要强调分离和存储工作间的异同点。

为了填补这方面的空白,在本文首先对数据库的发展过程进行了简单的介绍,然后讨论了机器学习在 MOF 筛选工作中的一般流程和常见算

法,最后重点介绍了利用机器学习方法从庞大的 MOF 数据库中筛选出适用于气体分离和储存相关工作的最新进展、机遇和可能遇到的挑战。

## 1.1 MOF 数据库的发展

数据库的发展是 HTCS 和 ML 工作应用于 MOF 领域的前提,随着 MOF 被大量制备和设计出来, MOF 结构的专项数据库也因此应运而生。

CSD 数据库是目前包含最多结构的材料数据库。2016 年剑桥晶体数据中心第一次划分出了两个 MOF 子集,目前该数据库已收录 119,927 种 MOF 结构。

“计算就绪、实验性(Computation-ready, experimental, CoRE) MOF 数据库是目前 HTCS 和 ML 计算中最常用的数据库。2013 年, Goldsmith 等<sup>[21]</sup>对 CSD 数据库中的 550,000 个结构进行了结构优化和溶剂去除,过程详见图 2。2014 年,该数据库正式命名为 CoRE MOF 数据库,包含 5,109 个 3D MOF<sup>[22]</sup>,如今该数据库已有 12,020 种计算就绪的 MOF<sup>[23]</sup>。

假设性 MOF(Hypothetical-MOF, hMOF)是一类模拟设计出来的结构。2011 年 Wilmer 等<sup>[24]</sup>分析 MOF 结构设计并重组了已有 MOF 中的 102 个次级构建单元和有机连接体“创造”了一个包含 137,953 个假设性的 hMOF 数据库,并通过计算  $\text{CH}_4$  的吸附等温线验证了该方法的可行性。次年, Qiao 等<sup>[20]</sup>使用该 hMOF 数据库模拟了  $\text{CO}_2/\text{N}_2/\text{CH}_4$  混合物的分离性能,进一步验证了其可信性。

此外,还有一些有创造性的关于数据库的研究值得被介绍。例如,2022 年, Gubsch 等<sup>[25]</sup>创建

了一个包含 MOF 的合成条件的 DigiMOF 数据库; Domingues 等<sup>[26]</sup>的工作证明了基于该数据库不但可以预测未知 MOF 的合成路径,并且能优化已有 MOF 的合成路径。

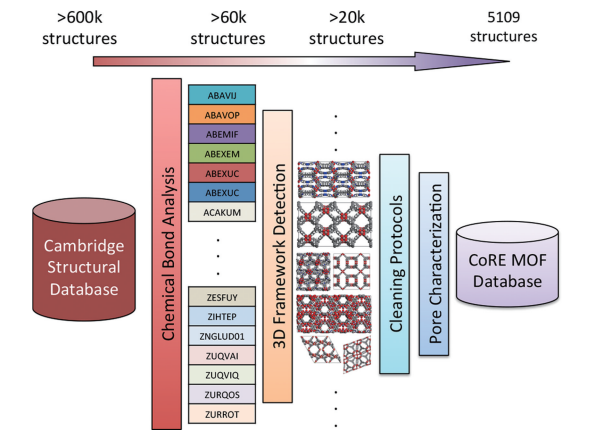


图 2 CoRE MOF 数据库构建流程示意图<sup>[22]</sup>

Fig. 2 CoRE MOF database construction process diagram screening<sup>[22]</sup>

## 1.2 机器学习应用于 MOF 筛选的一般流程

机器学习是一门多领域交叉形成的复合学科,该学科以计算机作为工具,通过模拟人类的学习方式来改善算法的性能,从而减少人工学习的时间成本,其一般流程如图 3 所示。

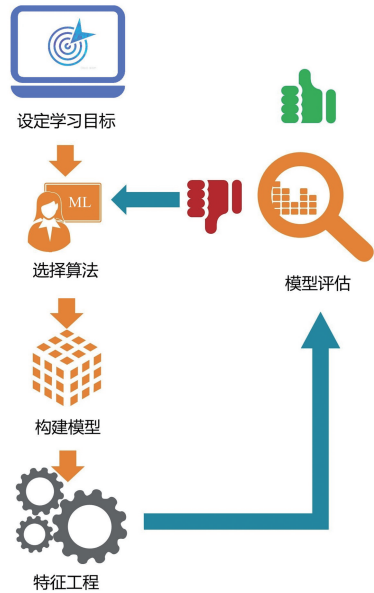


图 3 机器学习应用于 MOF 筛选的一般流程

Fig. 3 ML is applied to the general process of MOF

第一步,设定学习目标。其常见目标有:筛选 MOF 数据集中具有目标性能的 MOF;发掘描述符与性能指标间构效关系;预测尚未合成的 MOF。

第二步,选择算法。当选择合适的机器学习算法时,需要根据数据集的特性、可解释性和预测准确性需求进行综合考虑。表 1 是一些常见的数据集特性和相应的机器学习算法推荐。

表 1 常见的数据集特性和相应的机器学习算法推荐

Tab. 1 Common data set characteristics and corresponding ML algorithm recommendations

| 数据集    | 算法               | 特征                          |
|--------|------------------|-----------------------------|
| 小样本数据集 | 支持向量机            | 在小样本数据集上表现良好,能够处理高维数据。      |
|        | k 近邻算法           | 适用于小样本数据集,基于距离进行分类。         |
|        | 决策树              | 对于小样本数据集,决策树易于解释和可视化。       |
| 大规模数据集 | 随机森林             | 能够处理大规模数据集,并且对特征的重要性有较好的估计。 |
|        | 梯度提升树            | 在大规模数据集上具有高预测准确性,可以进行并行计算。  |
| 高维数据集  | 主成分分析            | 用于降维,减少高维数据的复杂性。            |
|        | Lasso 或 Ridge 回归 | 对于高维数据的特征选择和稀疏性处理。          |
|        | 神经网络             | 适用于高维数据集,可以学习复杂的特征表示。       |
| 文本数据集  | 朴素贝叶斯            | 适用于文本分类任务。                  |
|        | 词袋模型             | 用于将文本数据转化为向量形式。             |
| 图数据集   | 图神经网络            | 适用于图数据集,可以捕捉节点和边的复杂关系。      |
|        | 图嵌入方法            | 用于将图数据转化为向量表示,如节点嵌入、子图嵌入等。  |

第三步,构建模型。(1)数据收集,除使用现有数据集,如 CoRE 数据集、CSD 数据集等之外。还可以使用计算软件,如:Zeo++、RASPA 等获取相关数据。另外,使用遗传学算法、指纹算法<sup>[27]</sup>等方式生成 hMOF 结构也是一种可行的思路。(2)数据预处理,为解决数据集中的数据缺失、数据平衡等问题,通过删除异常值、去重复项和数据

规范化等手段可以有效提高模型的准确性。

第四步,特征工程。(1)特征选择,在 MOF 描述符中结构描述符,如孔隙率、比表面积等、化学描述符,如:吸附焓和亨利系数是目前分析较多的两类描述符。(2)通过特征转换或创建新特征也是不错的选择,这类描述符,如原子加权径向分布函数、能量直方图等在一些特定体系中往往有

不错的表现。

第五步,模型评估。从准确率、精确率、召回率三方面评估算法模型的性能,使用 F1 Score、ROC 曲线和 AUC 曲线、平均绝对误差 (Mean absolute error, MAE)、均方误差 (Mean square error, MSE) 和均方根误差 (Root mean square error, RMSE)、决定系数  $R^2$  等指标评估模型的表现。

## 2 ML 辅助筛选 MOF

对于 MOF 的气体储存和分离性能的 ML 研究通常侧重于构建结构-性能关系,对描述符的分析是这一过程的关键。

### 2.1 MOF 的气体分离工作

MOF 作为一种结晶性多孔材料,其在化学组

成、分子结构和材料结构等方面的确定性决定了其具有稳定的气体分离性能。

#### 2.1.1 描述符对气体分离工作影响的探讨

2012 年, Wilmer 等<sup>[28]</sup>首次通过 HTCS 的方法从 137,953 种 hMOF 的数据库中筛选出了适合分离甲烷中二氧化碳的 MOF。如图 4 所示的几种描述符是该工作中对  $\text{CO}_2$  分离影响最大的描述符,通过观察可以发现数据整体呈三角形分布,存在最优范围。具有较好分离效果的 MOF 往往其比表面积在  $1000 \sim 3000 \text{ m}^2 \cdot \text{g}^{-1}$ , 孔隙率在  $0.3 \sim 0.4$ , 吸附热在  $33 \text{ kJ}$  左右;再生性较高的 MOF 吸附热较低;选择性最高的 MOF 中,氯 (Cl) 和氟 (F) 官能团的出现非常频繁 (约  $60\% \sim 70\%$ )。

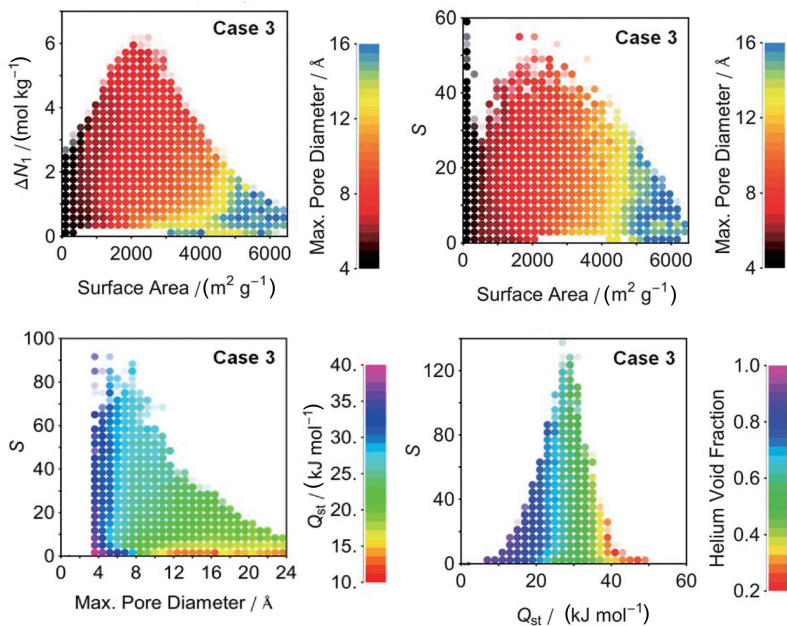


图 4 (a) 比表面积、最大孔径与工作能力的关系; (b) 比表面积、最大孔径与选择性的关系; (c) 最大孔径、吸附热与选择性的关系; (d) 吸附热、孔隙率与选择性的关系<sup>[28]</sup>

Fig. 4 (a) The relationship between specific surface area, maximum aperture and working capacity; (b) The relationship between specific surface area, maximum aperture and selectivity; (c) Relationship between maximum pore size, adsorption heat and selectivity; (d) Relationship between adsorption heat, porosity and selectivity<sup>[28]</sup>

基于 HTCS 的结论具有一定的普适性,但基于简单观察所获得的结论过于浅显,仍需要使用其他方法进行分析,机器学习方法是目前较为有效的高维数据分析方法。

结构描述符是气体分离中最重要的描述符,但其重要性却难以量化。假如仅使用结构描述符去预测  $\text{CO}_2/\text{N}_2$  分离体系其预测模型的  $R^2$  仅有  $0.71 \sim 0.75$ <sup>[29]</sup>,这一方面印证了结构描述符的重要性,另一方面也说明了还有其他影响 MOF 分离

性能的因素。

#### 2.1.2 寻找或开发分离工作中的新描述符

由于结构描述符在 MOF 气体分离工作中的重要性,探寻更为细致的结构特征成为了一种研究方向,基于官能团或拓扑结构的分析是一种更为贴近实验思路的研究方式。陈宇等<sup>[27]</sup>应用分子指纹技术将结构信息转化成二进制数列,提供了更加全面和准确的框架结构特征数据,提高了预测模型的预测精度和泛化能力。Gómez-

Gualdrón 等<sup>[30]</sup>使用决策树方法分析各个特征官能团对于 CO<sub>2</sub> 分离的影响,羟基、硫醇、氰基、氨基或硝基可以改善 MOF 对 CO<sub>2</sub> 的亲 and 性。Qiao 等<sup>[31]</sup>的工作说明了吡啶类官能团和唑类官能团对酸性气体有较好的分离效果。

同样,也有一些基于理论方法而非实验思路提出的描述符,例如,Haranczyk 等<sup>[32]</sup>在稀有气体氙/氪分离工作中提出:氙原子在可达空隙中的平均能量可以作为一种描述符,能够准确预测 MOF 的 Xe/Kr 分离性能;Woo 等<sup>[33]</sup>首次提出,原子加权径向分布函数(AP-RDF)能够很好地预测 MOF 的 CO<sub>2</sub> 吸附分离性能,其分析表明 MOF 骨架原子间距在 6~9 Å 范围内的 MOF 对 CO<sub>2</sub> 具有更高的亲和力。基于该描述符的构建的 QSPR (定量结构-属性关系) 分类器(SVM 模型)在 292,050 种 hMOF 的数据集中,仅需计算数据库的 10%,就能筛选出性能最优的 1000 个 MOF 中的 945 个。后续该课题组在 CO<sub>2</sub>/N<sub>2</sub> 体系<sup>[29]</sup>和 CO<sub>2</sub>/H<sub>2</sub> 体系<sup>[34]</sup>验证了该描述符在二氧化碳分离体系中的普适性。

## 2.2 MOF 的气体存储工作

MOF 作为最有潜力的存储材料,其性能远超沸石和 ZIF<sup>[35]</sup>,大多数的 ML 研究都集中在能源气体 CH<sub>4</sub> 和 H<sub>2</sub> 的低压存储。

### 2.2.1 描述符对气体存储工作影响的探讨

2011 年,Wilmer 等<sup>[24]</sup>首次采用 HTCS 方法从 137,953 种 hMOF 材料中,筛选出了 300 余种适宜存储甲烷的高性能 MOF 结构。并且成功通过

实验合成出了模拟吸附量最高的 MOF—PCN-14 (298K、3.5×10<sup>6</sup>Pa 条件下 267cm<sup>3</sup>(STP)·cm<sup>-3</sup>),打破了当时由 MOF—NOTT-107 保持的甲烷存储量记录。

预测新材料的性能能够为 MOF 的设计规划蓝图,Ahmed 等<sup>[36]</sup>通过极端随机树(ERT)算法,构建模型预测了 918,734 个 MOF 的 H<sub>2</sub> 储存容量。确定了存在 8,282 个 MOF 有可能超过目前最先进的储氢材料的单位体积氢储存量。并总结了高性能储氢 MOF 应具备的特征,包括低密度(<0.31g·cm<sup>-3</sup>)、高表面积(>5,300m<sup>2</sup>·g<sup>-1</sup>)、大空隙率(约为 0.90)以及孔体积(>3.3cm<sup>3</sup>·g<sup>-1</sup>)等特性。Pardakhti 等<sup>[37]</sup>运用 RF 模型探究了金属开放位点和不饱和度对 CH<sub>4</sub> 吸附量的影响。Fernandez 等<sup>[38]</sup>的研究使用 SVM 模型说明了在使用结构变量、化学变量和结构/化学混合变量来评估不同机器学习算法的预测能力时,使用结构/化学混合变量组合提高了算法的预测准确性。

### 2.2.2 寻找或开发气体存储工作中的新描述符

Bucior 等<sup>[39]</sup>提出了新的描述符——能量直方图,结合了能量直方图的 LASSO 回归模型成功预测了 54,776 种 MOF 的 H<sub>2</sub> 吸附性能(见图 5)。预测的最优 MOF—MFU-4/Zn 在 77K、1×10<sup>7</sup>Pa 的储存条件下,具有实验下 47g·L<sup>-1</sup>(模拟下 54g·L<sup>-1</sup>)的 H<sub>2</sub> 吸附量。Bobbitt 等<sup>[40]</sup>提出的一种可以快速筛选 MOF 储氢性能的参数——结合率,指 MOF 晶胞中与 H<sub>2</sub> 强相互作用区域的占比,结合率越高,H<sub>2</sub> 的储量越大。

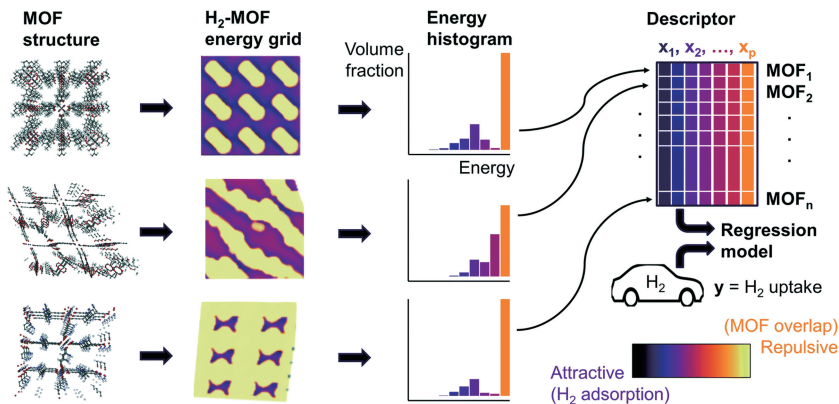


图 5 能量直方图构建流程示意图<sup>[39]</sup>

Fig. 5 Schematic diagram of energy histogram construction process<sup>[39]</sup>

## 2.3 分离与存储工作中描述符的异同性

MOF 的气体储存和分离性能的 ML 研究通常侧重于构建结构-性能关系,这些工作中存在一些

共性,也存在一些差异。Anderson 等<sup>[30]</sup>的工作向我们展示了常用描述符对于双组分混合体系下的吸附量、吸附选择性的影响。如图 6 所示,结构描

述符在确定绝对性能指标方面发挥着重要作用。如:最大孔径和限制孔径对所有的二氧化碳分离工作都非常重要。而与化学相关的因素中(如官能团中的 MNC 和 HBDM)对于  $\text{CO}_2/\text{NO}_2$  的分离体系更为重要。

通过观察不同体系和不同条件下的分离与存储工作,我们发现结构描述符中的密度、比表面积

和孔隙率等结构特征对于 MOF 的吸附和存储性能都很重要,它们共同反应了 MOF 结构中的吸附空间进而影响气体存储,同时也通过影响吸附位点的数量间接影响气体分离。非结构描述符中的吸附热在两种工作中都表现优异,并且无论是分离还是存储工作,压强越低,非结构描述符对于模型的贡献越大。

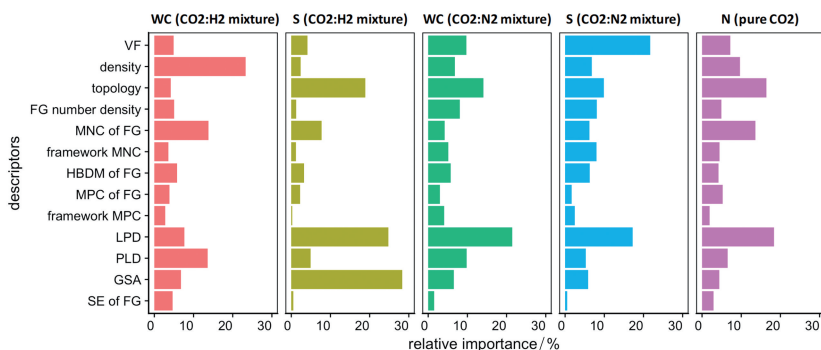


图 6 描述符相对五种性能指标的模型贡献度<sup>[30]</sup>

Fig. 6 The model contribution of descriptors to the five performance indicators<sup>[30]</sup>

同时,我们也分析了两种工作的差异性。在存储工作中,MOF 的体积表面积、孔体积密度是最为重要的描述符,虽然这些描述符对于分离工作也很重要,但二者所属的最佳区间却截然不同,选择性较高的 MOF 往往具有较小的孔隙率和比表面积,存储量较大的 MOF 对这些特征的要求则是在一定范围内越大越好。在分离工作中,限制孔径和吸附热是较为重要的描述符。前者杜绝了分子动力学直径大于限制孔径的分子进入 MOF 从而实现气体的分离,Zhou 等<sup>[41]</sup>在  $\text{D}_2\text{-H}_2$  分离工作中,说明了当 LCD 略大于  $\text{D}_2$  的尺寸时筛分效果最好,而当 PLD 接近  $\text{H}_2$  时出现排斥现象。后者从反应能量变化方面反应了吸附质分子与 MOF 骨架的亲水性,高吸附热通常意味着吸附剂对某种气体的选择性较高。例如,Qiao 等<sup>[42]</sup>在对天然气中有机硫气体的分离研究中,通过构建神经网络模型和偏最小二乘模型分析得出吸附热是该工作中具有最高权重的描述符。

### 3 气体的吸附分离机理

在 MOF 的气体吸附过程中,随着吸附压力增加,吸附过程会有以下四个阶段:(1)在极低压力状态下进行微孔填充,(2)在低压状态下进行单层吸附,(3)在中等压力下发生多层吸附,(4)在相对高压状态下发生毛细管凝聚作用。

对非高压状态下的吸附过程的研究是目前的主要方向,该过程主要由范德华力主导,同时氢键、卤键、 $\pi$ -络合等弱化学键起辅助作用。

影响范德华力的因素有很多,我们认为不饱和和金属位点是多数吸附工作的主要吸附位点<sup>[43~47]</sup>,而有机配体中的官能团则作为次级吸附位点。Lou 等<sup>[48]</sup>的研究通过对比具有不饱和位点的  $\text{Fe}_2(\text{DOBDC})$  和饱和的  $\text{Fe}_2(\text{O}_2)(\text{DOBDC})$  的  $\text{CO}_2$  吸附能力证明了不饱和金属位点对于  $\text{CO}_2$  吸附能力的影响。Hu 等<sup>[49]</sup>的研究说明将金属阳离子掺入大孔径 MOF 中可以有效提升  $\text{CO}_2$  吸附量,其中 Be、Ca、Ti、Mn 和 Fe 效果较好。涉及拓扑结构的 HTCS 和 ML 工作则很好地说明了官能团作对于吸附的影响,Fernandez 等<sup>[38]</sup>的研究发现,超过 75% 高性能  $\text{CH}_4$  存储 MOF 都含有甲基、乙基或丙基官能团,而卤素官能团表现出不利于  $\text{CH}_4$  吸附的性质。Giappa 等<sup>[50]</sup>的研究说明了  $-\text{OCONH}_2$ 、 $-\text{SO}_2\text{NH}_2$ 、 $-\text{OPO}_3\text{H}_2$ 、 $-\text{OSO}_3\text{H}$  官能团是 MOF 储氢工作中成键强度最高的。Jablonka 等<sup>[51]</sup>通过主成分分析方法对 MOF 的  $\text{CO}_2$  吸附性能进行了拓扑结构分析,图 7(a)展示了基于 MOF 属性相似度的等网状优化和全局优化优化过程;图 7(b,c)展示了不同官能团对于  $\text{CO}_2$  吸附的影响。

弱化学键在吸附过程中也发挥着一定作用:

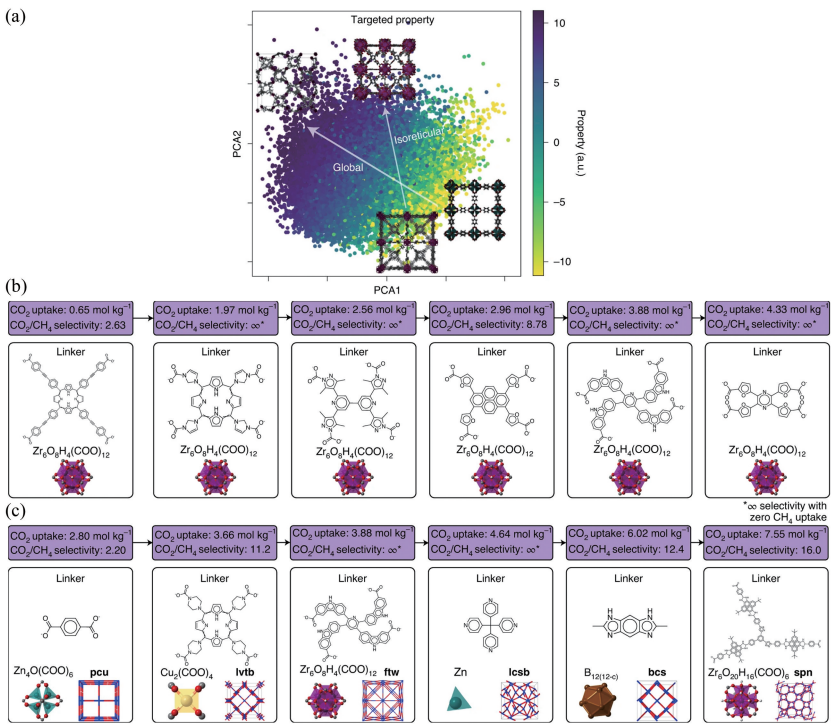


图 7 (a) 基于主成分分析的等网状设计和全局优化设计过程; (b) 基于 MOF-NU-1104 的等网状设计; (c) 基于 MOF-5 的全局优化设计<sup>[51]</sup>

Fig. 7 (a) Isomesh design and global optimization design process based on principal component analysis; (b) Isomesh design based on MOF-NU-1104; (c) Global optimization design based on MOF-5<sup>[51]</sup>

(1) 氢键吸附<sup>[47,52]</sup>, 在 Shimizu 等<sup>[53]</sup> 对 MOF—CALF-20 的研究中发现 CO<sub>2</sub> 与框架形成了氢键网络, 进而抑制了水的吸附; (2)  $\pi$ -络合吸附<sup>[54,55]</sup>, 络合物是指由中心原子(或离子)提供空轨道, 配体提供电子形成  $\pi$  键的络合物; (3) 酸碱吸附<sup>[56-58]</sup>, 在对天然气的纯化过程中, 使用偏碱性的 MOF 材料能与酸性气体发生酸碱吸附进而增强吸附效果。

除此之外, 有一项 MOF 所独有的特性也应该被说明, MOF 柔性骨架的呼吸效应<sup>[47,59-61]</sup>。例如, Long 等<sup>[62]</sup> 报道了两种柔性 MOF 结构 Co(bdp) 和 Fe(bdp), 其中的孔隙能够随压力变化在“坍塌状态”和“膨胀状态”之间切换。

## 4 总结与展望

进二十年来, 人工智能与 MOF 材料学两个领域的交叉合作取得了许多成果, 但也带来了一些挑战。以下是对未来发展的机遇和挑战的一些观点。

首先, 需要更紧密地融合理论工作与实验研究。目前, 大部分 ML 预测工作都依赖于模拟数据而不是实验数据。基于模拟数据而非实验数据

开发的机器学习模型, 即使在完全准确的情况下, 也仅能反映模拟的水平。使用模拟数据而非实验数据的原因主要有以下两点, 一方面是获取实验数据难度较大且质量不一, 另一方面是不同实验条件导致的数据差异使得数据难以进行量化分析。因此, 迫切需要将 ML 预测与实验相结合, 创建标准化的实验流程以构建可供计算的实验数据。

其次, 开发具有可解释性和通用性的模型。在 MOF 筛选领域, 机器学习方法通常使用难以解释的“黑盒”模型, 难以直接解释决策边界形成的原因。这对于理解模型的内部机理和决策边界的形成原因都不利。因此, 开发具有解释性的模型是一个重要方向。此外, 现有研究大多局限于特定数据集, 结论缺乏普适性。构建通用性的 ML 模型可以减少每个数据集的计算时间并提高模型的泛用性。

最后, 尽管目前已有大量成功的应用, 但这一机器学习与 MOF 形成的交叉领域仍处于起步阶段。随着相关工作(例如, 力场、电荷分布、柔性结构等)的不断推进, 可以预见 ML 在 MOF 功能材料的开发中将发挥越来越重要的作用。

参 考 文 献

[ 1 ] Kirchon A, Feng L, Drake H F, et al. Chem. Soc. Rev. , 2018, 47(23) : 8611~8638.

[ 2 ] Day G S, Drake H F, Zhou H C, et al. Commun. Chem. , 2021, 4(1) : 114.

[ 3 ] Furukawa H, Ko N, Go Y B, et al. Science, 2010, 329 (5990) : 424~428.

[ 4 ] Sahoo R, Das M C. Coord. Chem. Rev. , 2021, 442: 213998.

[ 5 ] Llewellyn P L, Bourrelly S, Serre C, et al. Langmuir, 2008, 24(14) : 7245~7250.

[ 6 ] Wang S, Wu D, Huang H, et al. Chin. J. Chem. Eng. , 2015, 23(8) : 1291~1299.

[ 7 ] Jiao L, Jiang H L. Chin. J. Catal. , 2023, 45: 1~5.

[ 8 ] Sun Q, Zhu K, Ji X, et al. Sci. China Mater. , 2022, 65 (6) : 1453~1462.

[ 9 ] Mohd Zain N K, Vijayan B L, Misnon I I, et al. Ind. Eng. Chem. Res. , 2019, 58(2) : 665~674.

[ 10 ] 陈跃颖, 盘盈滢, 杜文卿, 等. 材料研究与应用, 2022, 16 (1) : 68~80.

[ 11 ] Lee S, Kim B, Cho H, et al. ACS Appl. Mater. Interf. , 2021, 13(20) : 23647~23654.

[ 12 ] Yaghi O M, Li H. J. Am. Chem. Soc. , 1995, 117(41) : 10401~10402.

[ 13 ] Peng Y, Krungleviciute V, Eryazici I, et al. J. Am. Chem. Soc. , 2013, 135(32) : 11887~11894.

[ 14 ] Férey G, Mellot-Draznieks C, Serre C, et al. Science, 2005, 309(5743) : 2040~2042.

[ 15 ] Park K S, Ni Z, Côté A P, et al. PNAS, 2006, 103(27) : 10186~10191.

[ 16 ] Himanen L, Geurts A, Foster A S, et al. Adv. Sci. , 2019, 6(21) : 1900808.

[ 17 ] Demir H, Daglar H, Gulbalkan H C, et al. Coord. Chem. Rev. , 2023, 484: 215112.

[ 18 ] Altintas C, Keskin S. Mater. Today Energy, 2023: 101426.

[ 19 ] Firooz S K, Armstrong D W. Anal. Chim. Acta, 2022, 1234: 340208.

[ 20 ] Lin H, Yang Y, Hsu Y, et al. Adv. Mater. , 2023: 2209073.

[ 21 ] Goldsmith J, Wong-Foy A G, Cafarella M J, et al. Chem. Mater. , 2013, 25(16) : 3373~3382.

[ 22 ] Chung Y G, Camp J, Haranczyk M, et al. Chem. Mater. , 2014, 26(21) : 6185~6192.

[ 23 ] Altintas C, Avci G, Daglar H, et al. J. Mater. Chem. A, 2019, 7(16) : 9593~9608.

[ 24 ] Wilmer C E, Leaf M, Lee C Y, et al. Nat. Chem. , 2012, 4 (2) : 83~89.

[ 25 ] Glasby L T, Gubsch K, Bence R, et al. Chem. Mater. , 2023, 35(11) : 4510~4524.

[ 26 ] Domingues N P, Moosavi S M, Talirz L, et al. Commun. Chem. , 2022, 5(1) : 170.

[ 27 ] 陈宇. 面向 H<sub>2</sub>S 捕获的金属-有机框架的分子指纹算法研 究. 广州大学硕士学位论文, 2023.

[ 28 ] Wilmer C E, Farha O K, Bae Y S, et al. Energy Environ. Sci. , 2012, 5(12) : 9849.

[ 29 ] Burner J, Schwiedrzik L, Krykunov M, et al. J. Phys. Chem. C, 2020, 124(51) : 27996~28005.

[ 30 ] Anderson R, Rodgers J, Argueta E, et al. Chem. Mater. , 2018, 30(18) : 6325~6337.

[ 31 ] Qiao Z, Xu Q, Jiang J. J. Mater. Chem. A, 2018, 6(39) : 18898~18905.

[ 32 ] Simon C M, Mercado R, Schnell S K, et al. Chem. Mater. , 2015, 27(12) : 4459~4475.

[ 33 ] Fernandez M, Boyd P G, Daff T D, et al. J. Phys. Chem. Lett. , 2014, 5(17) : 3056~3060.

[ 34 ] Dureckova H, Krykunov M, Aghaji M Z, et al. J. Phys. Chem. C, 2019, 123(7) : 4133~4139.

[ 35 ] Li B, Wen H M, Zhou W, et al. Chem, 2016, 1(4) : 557~ 580.

[ 36 ] Ahmed A, Siegel D J. Patterns, 2021, 2(7) : 100291.

[ 37 ] Sezginel K B, Uzun A, Keskin S. Chem. Eng. Sci. , 2015, 124: 125~134.

[ 38 ] Fernandez M, Woo T K, Wilmer C E, et al. J. Phys. Chem. C, 2013, 117(15) : 7681~7689.

[ 39 ] Bucior B J, Bobbitt N S, Islamoglu T, et al. Mol. Syst. Design Eng. , 2019, 4(1) : 162~174.

[ 40 ] Bobbitt N S, Chen J, Snurr R Q. J. Phys. Chem. C, 2016, 120(48) : 27328~27341.

[ 41 ] Zhou M, Vassallo A, Wu J. J. Membrane Sci. , 2020, 598: 117675.

[ 42 ] Liang H, Yang W, Peng F, et al. APL Materials, 2019, 7 (9) : 091101.

[ 43 ] Grant Glover T, Peterson G W, Schindler B J, et al. Chem. Eng. Sci. , 2011, 66(2) : 163~170.

[ 44 ] Blanco-Brieva G, Campos-Martin J M, Al-Zahrani S M, et al. Fuel, 2011, 90(1) : 190~197.

[ 45 ] Huo S H, Yan X P. J. Mater. Chem. , 2012, 22(15) : 7449~7455.

[ 46 ] Karra J R, Walton K S. Langmuir, 2008, 24(16) : 8620 ~8626.

[ 47 ] Britt D, Tranchemontagne D, Yaghi O M. PNAS, 2008, 105 (33) : 11623~11627.

[ 48 ] Lou W, Yang J, Li L, et al. J. Solid State Chem. , 2014, 213: 224~228.

[ 49 ] Hu J, Liu Y, Liu J, et al. J. Phys. Chem. C, 2018, 122 (33) : 19015~19024.

[ 50 ] Giappa R M, Tylanakis E, Di Gennaro M, et al. Int. J. Hydrogen Energy, 2021, 46(54) : 27612~27621.

[ 51 ] Yao Z, Sánchez-Lengeling B, Bobbitt N S, et al. Nat. Mach. Intell. , 2021, 3(1) : 76~86.

[ 52 ] Lomig Hamon, Hervé Leclerc, Laetitia Oliviero, et al. J. Phys. Chem. C, 2011, 115(5) : 2047~2056.

设计和合成策略十分广泛,在结构修饰中主要是引入具有较高活性的烷烃长链、酰胺、酯键、含氮杂环等药效团。对具有 HDACs 抑制活性的上市药物进行修饰改造,保留锌结合基团的羧酸结构,改造帽端和连接基团,是开发新型 HDACi 的热点。HDACs 与 BRD4、p53、JAK 等多个靶点具有抗肿瘤协同作用,以 HDACs 为靶点,开发具有协同增效作用的多靶点抑制剂,可扩大 HDACi 的应用范围,增强抗肿瘤疗效,改善 HDACs 亚型选择性低、生物利用度低等缺点。

综上所述,引入具有活性的药效官能团,合成出一些结构新颖、亚型选择性高且具有高抑制活性、高效性的双靶点/多靶点羧酸类化合物,以及如何让双靶点或者多靶点化合物发挥协同增效作用,对于 HDACi 的新药研发具有深远意义。

参 考 文 献

[ 1 ] Li Y, Seto E. Cold Spring Harb. Perspect. Med. , 2016, 6 ( 10 ) : 1 ~ 34.

[ 2 ] Waldmann H, Janning P. Chemical Biology, Germany: Wiley-VCH, 2009, 135 ~ 137.

[ 3 ] De R A J M, Van G A H, C H N, et al. Biochem. J. , 2003, 370( 3 ) : 737 ~ 749.

[ 4 ] Liu J, Zhou J X, He F J, et al. Eur. J. Med. Chem. , 2020, 192: 112189.

[ 5 ] Noonepalle S, Shen S, Ptáček J, et al. J. Med. Chem. , 2020, 63( 18 ) : 10246 ~ 10262.

[ 6 ] Gawel J M, Shouksmith A E, Raouf Y S, et al. Eur. J. Med. Chem. , 2020, 201: 112411.

[ 7 ] Toutah K, Nawar N, Timonen S, et al. J. Med. Chem. , 2021, 64( 12 ) : 8486 ~ 8509.

[ 8 ] Olaoye O O, Watson P R, Nawar N, et al. J. Med. Chem. , 2021, 64: 2691 ~ 2704.

[ 9 ] Shen S, Picci C, Ustinova K, et al. J. Med. Chem. , 2021, 64( 8 ) : 4810 ~ 4840.

[ 10 ] Guo Z, Zhang Z, Zhang Y, et al. Eur. J. Med. Chem. , 2021, 218: 113383.

[ 11 ] Reßing N, Sönnichsen M, Osko J D, et al. J. Med. Chem. , 2020, 63( 18 ) : 10339 ~ 10351.

[ 12 ] Chen X, Chen X Y, Steimbach R R, et al. Eur. J. Med. Chem. , 2020, 187: 111950.

[ 13 ] Relitti N, Saraswati A P, Chemi G, et al. Eur. J. Med. Chem. , 2021, 212: 112998.

[ 14 ] Saraswati A P, Relitti N, Brindisi M, et al. ACS Med. Chem. Lett. , 2020, 11( 11 ) : 2268 ~ 2276.

[ 15 ] Tseng H J, Lin M H, Shiao Y J, et al. Eur. J. Med. Chem. , 2020, 192: 112193.

[ 16 ] Hsu K C, Chu J C, Tseng H J, et al. Eur. J. Med. Chem. , 2021, 219: 113419.

[ 17 ] Yao D H, Li C Y, Jiang J, et al. Eur. J. Med. Chem. , 2020, 205: 112648.

[ 18 ] Wei M M, Xie M D, Zhang Z, et al. Eur. J. Med. Chem. , 2020, 206: 112677.

[ 19 ] Mak J Y W, Wu K C, Gupta P K, et al. J. Med. Chem. , 2021, 64( 4 ) : 2186 ~ 2204.

[ 20 ] Liu Q Q, Zhang B, Wang Y J, et al. Eur. J. Med. Chem. , 2022, 229: 114058.

[ 21 ] Duan Y C, Jin L F, Ren H M, et al. Eur. J. Med. Chem. , 2021, 220: 113453.

[ 22 ] Zhai S Y, Zhang H M, Chen R, et al. Eur. J. Med. Chem. , 2021, 225: 113824.

[ 23 ] Zhao L, Fan T T, Shi Z C, et al. Eur. J. Med. Chem. , 2021, 213: 113173.

[ 24 ] Luo G S, Lin X, Ren S N, et al. Eur. J. Med. Chem. , 2021, 226: 113870.

[ 25 ] Zhao C X, Tang C, Li C H, et al. Bioorg. Med. Chem. , 2021, 40: 116185.

[ 26 ] Huang Y H, Liu N, Pan Z Z, et al. J. Med. Chem. , 2023, 66: 1239 ~ 1253.

(上接第 324 页)

[ 53 ] Lin Y, Kong C, Zhang Q, et al. Adv. Energy Mater. , 2017, 7( 4 ) : 1601296.

[ 54 ] Khan N A, Jhung S H. Angew. Chem. Int. Ed. , 2012, 51 ( 5 ) : 1198 ~ 1201.

[ 55 ] Khan N A, Jhung S H. J. Hazard. Mater. , 2012, 237-238: 180 ~ 185.

[ 56 ] Khan N A, Jun J W, Jeong J H, et al. Chem. Commun. , 2011, 47( 4 ) : 1306 ~ 1308.

[ 57 ] Khan N A, Jhung S H. Fuel Proc. Technol. , 2012, 100: 49 ~ 54.

[ 58 ] Ahmed I, Hasan Z, Khan N A, et al. Appl. Catal. B, 2013, 129: 123 ~ 129.

[ 59 ] Ke F, Qiu L G, Yuan Y P, et al. J. Hazard. Mater. , 2011, 196: 36 ~ 43.

[ 60 ] Haque E, Lee J E, Jang I T, et al. J. Hazard. Mater. , 2010, 181( 1-3 ) : 535 ~ 542.

[ 61 ] Arstad B, Fjellvåg H, Kongshaug K O, et al. Adsorption, 2008, 14( 6 ) : 755 ~ 762.

[ 62 ] Mason J A, Oktawiec J, Taylor M K, et al. Nature, 2015, 527( 7578 ) : 357 ~ 361.