

# 机器学习辅助金属有机框架的高通量筛选：丙烷/丙烯分离

薛小雨<sup>1</sup>，程敏<sup>1</sup>，王诗慧<sup>1</sup>，周利<sup>1\*</sup>，吉旭<sup>1\*</sup>

(<sup>1</sup> 四川大学化学工程学院，四川省 成都市 610065)

## 摘要：

金属-有机骨架 (MOFs) 作为分离丙烷 (C<sub>3</sub>H<sub>8</sub>) / 丙烯 (C<sub>3</sub>H<sub>6</sub>) 的潜在吸附材料引起了人们的极大兴趣，基于分子模拟的高通量筛选可以加速发现新的吸附材料，但需耗费大量计算资源探索可能的材料空间。本文提出一种将分子模拟与机器学习算法相结合的高通量筛选方法，以吸附剂性能得分 (APS) 为评价指标，从真实和假想 MOFs 共同构成的约 14 万个 MOFs 的混合数据集中筛选出可用于分离 C<sub>3</sub>H<sub>8</sub> / C<sub>3</sub>H<sub>6</sub> 的最佳 MOFs。随机从混合数据集中选取 6000 个 MOFs，以结构和化学描述符为输入，构建随机森林 (RF) 模型，预测 APS。用训练好的模型预测剩余 MOFs 的 APS，筛选得到分离性能最好的 5 个真实 MOFs 和 5 个假想 MOFs。此外，通过计算 RF 模型中决策树回归器的平均特征重要度，发现重量比表面积和密度对预测结果影响最大。

**关键词：**丙烷/丙烯分离；金属-有机框架；高通量筛选；机器学习；随机森林

**中图分类号：**□□□□□□

**文献标识码：**A

## High-throughput screening of Metal-organic frameworks assisted by machine learning: Propane/propylene separation

Xiaoyu Xue<sup>1</sup>, Min Cheng<sup>1</sup>, Shihui Wang<sup>1</sup>, Li Zhou<sup>1\*</sup>, Xu Ji<sup>1\*</sup>

(<sup>1</sup>Sichuan University, Chengdu 610065, Sichuan, China)

### Abstract:

Metal-organic frameworks (MOFs) have garnered interest as potential adsorbents for the separation of propane/propylene mixture. High-throughput screening based on molecular simulation can accelerate the discovery of new adsorbents, but requires a lot of computational resources to explore space of the possible materials. In this paper, taking the adsorbent performance score (APS) as the performance assessment metric, a hybrid high-throughput screening approach combining molecular simulation and machine learning was proposed to screen the best MOFs for separating C<sub>3</sub>H<sub>8</sub> / C<sub>3</sub>H<sub>6</sub> mixture from the data set of more than 140000 MOFs composed of real MOFs and hypothetical MOFs. 6000 MOFs were chosen randomly from the mixing dataset as a training set. The random forest (RF) model was built using chemical and structural descriptors to predict APS based on the training set. the APS of the remaining MOFs were predicted with the trained RF model, and top five real MOFs and top five hypothetical MOFs were screened. Further, by calculating the average feature importance among the decision tree regressors in the RF model, the gravimetric surface area and density play a significant role in predicting APS.

**Key words:** Propane/propylene separation; Metal-organic frameworks; High-throughput screening; Machine learning; Random forest

## 引言

烯烃/烷烃的分离是石油化工产业至关重要的过程<sup>[1-3]</sup>，在其众多的烯烃产品中，丙烯 (C<sub>3</sub>H<sub>6</sub>) 作为仅次于乙烯的第二重要产品，是许多重要化工原料的上游预料，如聚丙烯、丙烯腈等。工业上，C<sub>3</sub>H<sub>6</sub> 主要从蒸汽裂解或催化裂化过程中产生<sup>[4]</sup>，过程同

时还会伴随丙烷 (C<sub>3</sub>H<sub>8</sub>) 的生成，因此，为了生产聚合物等级（纯度大于 99.5%）的 C<sub>3</sub>H<sub>6</sub> 就必须从中分离出 C<sub>3</sub>H<sub>8</sub>。传统的分离 C<sub>3</sub>H<sub>6</sub>/C<sub>3</sub>H<sub>8</sub> 混合物的方法是利用二者在沸点上的微小差异采用低温蒸馏的方式分离，而这种由热驱动的分离方式耗能很高，大约占全球能源消耗的 0.3%<sup>[5]</sup>。因此，为了降低能耗，迫切需要更节能的分离 C<sub>3</sub>H<sub>6</sub>/C<sub>3</sub>H<sub>8</sub> 混合物

**联系人：**周利（1987—），女，四川大学化学工程学院特聘副研究员，邮箱：chezli@scu.edu.cn **第一作者：**薛小雨（1998—），女，硕士研究生。  
**联系人：**吉旭（1968—），男，四川大学化学工程学院教授，邮箱：jixu@scu.edu.cn

工艺,这将对降低能源成本和碳排放方面对有很大的帮助<sup>[6]</sup>。

目前,已经有很多种工艺应用于  $C_3H_6/C_3H_8$  的分离,基于吸附的分离方法由于其耗能低,操作条件简单的特点成为其中最具有前景的方法之一<sup>[7]</sup>。用于  $C_2-C_3$  分离的吸附剂有沸石、活性炭、氧化铝和二氧化硅等,但它们的分离能力和选择性普遍较低<sup>[8]</sup>。近年来,金属有机框架 (Metal Organic Frameworks, MOFs) 因其结构和化学性质的高度可调性受到了人们越来越多的关注,已成为分离  $C_3H_6/C_3H_8$  的新的潜在候选材料<sup>[9-12]</sup>。MOFs 内孔几何形状和表面功能的无限化学可调性可以促进对吸附选择性的控制,而高表面积/孔体积通常会产生高的分离工作能力<sup>[13]</sup>。目前已经有许多 MOFs 被合成可以用于分离  $C_3H_6/C_3H_8$ <sup>[14]</sup>。M-MOF-74 (M=Co、Mn、Mg) 是一种具有高密度开放金属位点 (Open Metal Sites, OMS) MOFs, 展现出了较高的  $C_3H_6/C_3H_8$  选择性<sup>[15]</sup>; Cadiau<sup>[16]</sup>等人合成了一种氟化的 MOFs 材料 NbOFFIVE-1-Ni, 它只吸附  $C_3H_6$ , 不吸附  $C_3H_8$ 。

然而,几乎所有可用于  $C_3H_6/C_3H_8$  分离的 MOFs 都是选择性吸附  $C_3H_6$  的。由于这些 MOFs 含有 OMS 可以与  $C_3H_6$  分子的碳碳双键产生  $\pi$  络合作用,均展现出了对  $C_3H_6$  的高选择性。此外,由于  $C_3H_8$  的分子尺寸大于  $C_3H_6$  ( $C_3H_8$ : 5.1 Å,  $C_3H_6$ : 4.68 Å), 基于分子筛效应或动力学机制分离更有利于  $C_3H_6$  在 MOFs 上选择性吸附<sup>[17]</sup>。在实际工业中,  $C_3H_8$  作为二者混合物中的杂质,如果优先被去除,那么, MOFs 作为吸附剂用于变压吸附

(Pressure Swing Adsorption, PSA) 的过程生产聚合物等级的  $C_3H_6$ , 不仅可以降低床层体积,还可以第一个吸收塔的出口就获得  $C_3H_6$ , 同时烷烃选择性吸附剂在一次突破性操作中使用,节能高达 40%<sup>[18]</sup>。目前,学者对  $C_3H_8$  选择性吸附剂的研究很少, Wang<sup>[19]</sup>等人首次通过表面化学设计合成了  $C_3H_8$  选择性吸附 2 种锆基 MOFs, g-C3N4@Zr-BPDC 和 Zr-BPYDC, 并研究了其在无限稀释条件下  $C_3H_8/C_3H_6$  选择性; Yang<sup>[17]</sup>等人通过精确调节孔隙空间实现了  $C_3H_8/C_3H_6$  分离。综上, MOFs 在选择性吸附  $C_3H_8$  的分离性能上还有很大的提升空间。

近年来,实验室已经合成的 MOFs 超过了 10 万个<sup>[20]</sup>,而假想 MOFs 数量级更是高达百万级别,

人们迫切需要确定可用于  $C_3H_8$  选择性吸附分离的最佳候选 MOFs。实验上合成  $C_3H_8$  选择性吸附剂 MOFs 通常需要花费大量的时间,高通量筛选的方法由于高效的特点成为了一种发掘高性能 MOFs 材料的有效手段。目前,高通量计算筛选的方法已经成功应用于  $CO_2$  捕捉<sup>[21-24]</sup>, Xe/Kr 分离<sup>[25-27]</sup>,  $CH_4/H_2$ <sup>[28, 29]</sup> 分离等领域,  $C_3$  分离的高通量计算筛选研究鲜见报道。Yeo<sup>[30]</sup>等人利用高通量筛选的方法,从近百万种多孔材料筛选出 60000 余个,模拟了  $C_3H_6/C_3H_8$  分离性能,分析了结构与性能之间的关系,最终发现了吸附选择性与工作能力之间的权衡。Solanki<sup>[31]</sup>等人利用分子模拟和高通量筛选的方法用于  $C_3H_8/C_3H_6$  分离的 12351 个真实 MOFs 进行筛选,最后得到的 MOFs 的吸附剂性能得分 (Adsorbent Performance Score, APS) 最高为  $15.62 \text{ mol} \cdot \text{kg}^{-1}$ , 再生能力 (Regenerability, R%) > 80。然而,由于 MOFs 结构的多样性以及 MOFs 数据库的不断扩大,高通量筛选同样也会耗费大量的时间。

近年来,随着机器学习 (Machine Learning, ML) 技术的不断发展,基于数据驱动的 ML 技术越来越多的用于 MOFs 的定量结构-性质关系 (Quantitative Structure-Property Relation, QSPR) 的探索,同时,分子模拟与机器学习相结合也成为了筛选 MOFs 有效手段<sup>[32]</sup>。区别于传统大批量分子模拟方法,ML 仅需要小部分样本构建模型便可在几分甚至几秒钟内获得备选材料的吸附性能指标,从而指导分子筛选。因此准确的机器学习模型通常用于高通量筛选过程的预筛选,以此来过滤性能较差的材料,从而加速筛选。目前,已有许多机器学习算法应用于 MOFs 的气体分离和储存领域,如  $C_2H_6/C_2H_4$  分离<sup>[33, 34]</sup>、Xe/Kr 分离<sup>[35, 36]</sup>、 $CO_2/N_2$  分离<sup>[37]</sup>、 $CO_2/H_2$  分离<sup>[38]</sup>、 $CH_4$  储存<sup>[39-41]</sup>等,然而在  $C_3H_8/C_3H_6$  分离领域,ML 技术尚未应用。

以往的研究发现,大多数 MOFs 应用于气体分离时,选择性和工作能力之间存在着权衡,即高选择性的 MOFs 通常工作能力较低,这在  $C_3H_8/C_3H_6$ <sup>[30, 31]</sup>、 $C_2H_6/C_2H_4$ <sup>[33]</sup>、 $CH_4/H_2$ <sup>[42]</sup>、 $CO_2/H_2$ <sup>[43]</sup> 的分离中都存在这样的权衡。APS 综合考虑了吸附剂的选择性和工作能力对 PSA 工艺回收纯度和回收率的影响,解释工作能力和选择性二者之间对变压吸附装置性能影响,是一种很好的评定吸附剂吸附性能的方法<sup>[43]</sup>。目前,许多研究都以

APS 为指标评价分离性能<sup>[31, 42, 44-46]</sup>, 然而在机器学习辅助的高通量筛选中尚未使用。因此, 选择 APS 作为性能指标用于机器学习辅助的高通量筛选是一种很有前景的方法。

本研究提出利用分子模拟结合机器学习算法从包含 14 余万个 MOFs 数据库中筛选可用于  $C_3H_8/C_3H_6$  分离的最佳 MOFs。利用结构描述符和化学描述符构建预测模型的输入, APS 作为输出, 基于随机森林算法构建从 MOFs 的三维结构到分离性能指标的预测模型。文章的第二章介绍了筛选流程以及具体的模拟方法, 第三章对机器学习的结果进行了讨论, 并对筛选出的最好的 MOFs 进行了分析。

## 1 研究方法

### 1.1 筛选流程

本研究所用的 MOFs 主要来自于两个数据库, 分别是 CoRE MOFs (Computational Ready, Experimental MOFs) 数据库<sup>[25]</sup>和 hMOFs (Hypothetical MOFs) 数据库<sup>[47]</sup>, 共计约 140000 余个 MOFs。图 1 展示了本研究的筛选流程, 共计 7 步。(1) 挑选出不包含 OMS 的 MOFs: MOFs 中的 OMS 可以和  $C_3H_6$  中的碳碳双键形成  $\pi$  络合作用, 因此比起  $C_3H_8$ , MOFs 与  $C_3H_6$  的相互作用更强, 因此含有 OMS 的 MOFs, 并不是  $C_3H_8$  选择性吸附剂的候选材料, 同时, 巨正则蒙特卡洛 (grand canonical Monte Carlo, GCMC) 模拟所用的通用力场 (UFF) 对于  $C_3H_6$  双键与 MOFs 的 OMS 的相互作用的模拟是不准确的<sup>[48, 49]</sup>。对每个 MOFs 是否具有 OMS 进行了统计计算<sup>[25]</sup>, 计算结果显示, 约有 13 万 5000 个 MOFs 没有 OMS。后续的研究将基于不包含 OMS 的 MOFs 进行。(2) 特征化: 计算候选 MOFs 的描述符, 包括化学描述符和结构描述符, 见表 1; (3) 随机选择: 随机从候选样本中选出 7500 个 MOFs 用于模型的构建, 选择 6000 个 MOFs 作为训练集, 用于机器学习模型的训练, 用 1500 个 MOFs 测试模型; (4) 模拟  $C_3H_8/C_3H_6$  分离: 用 GCMC 模拟  $C_3H_8/C_3H_6$  在这 7500 个 MOFs 上的吸附, 并基于模拟结果计算对应的性能指标。(5) 建立预测模型: 用训练集的描述符与分离性能指标建立随机森林预测模型, 并用测试集对模型进行评价与验证; (6) 预筛选: 用训练好的随机森林模型预测剩余 MOFs 的吸附性能, 选择随机森林的预测结果中 top 2500 MOFs 针对这些 MOFs 进行

GCMC 模拟来修正预测结果; (7) 发现新 MOFs: 从 top 2500 MOFs 筛选出性能最好的 MOFs。

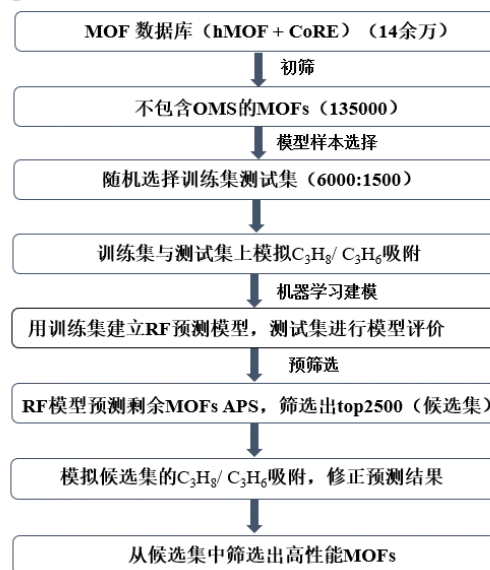


图 1 筛选流程

Figure 1: Screening workflow

### 1.2 分子模拟

使用 RASPA 软件<sup>[50]</sup>通过 GCMC 方法模拟  $C_3H_8/C_3H_6$  在 MOFs 上的平衡吸附分离。 $C_3H_8/C_3H_6$  气体混合物中  $C_3H_8$  与  $C_3H_6$  的摩尔比为 1: 1, 假定二者在 298K, 1Bar 时吸附, 在 298K, 0.1bar 时解吸。对于吸附质, 分子间相互作用的 Lennard-Jones 势能参数取自 TraPPE 力场<sup>[51, 52]</sup>。对于吸附剂, 假定 MOFs 是刚性结构, MOFs 原子的 Lennard-Jones (LJ) 势能参数取自 UFF 力场<sup>[53]</sup>。吸附剂原子与吸附质组分之间的 LJ 参数通过 Lorentz - Berthelot 混合规则计算。对每一个 MOFs, 共计运行了两个 10000 次循环, 前一个循环用于平衡系统, 后一个循环用于获得平均热力学性质, 在每个循环由 n 个 MC 试验移动组成(n: 被吸附分子的数量), MC 移动包括平移、旋转、再生和交换。三维体系中施加周期性边界, 模拟时的截断距离设置为 12.5 nm, 模拟晶胞沿三维方向分别扩展到至少 25 nm。

### 1.3 描述符

本研究共考虑了两类描述符, 分别是结构描述符和化学描述符。其中结构描述符包括密度( $\rho$ ), 重量比表面积 (GSA), 体积比表面积 (VSA), 最大空腔直径 (LCD), 孔隙极限直径 (PLD) 和氩孔隙率 (VF)。前五个结构描述符使用开源软件 Zeo++<sup>[54]</sup>来计算, 最后一个结构描述符使用 RASPA 软件计

算。由于化学描述符已成功应用于 C<sub>2</sub>H<sub>6</sub>/C<sub>2</sub>H<sub>4</sub> 分离的机器学习模型的建模中<sup>[33,34]</sup>，考虑到 C<sub>3</sub>H<sub>8</sub>/C<sub>3</sub>H<sub>6</sub> 分离与 C<sub>2</sub>H<sub>6</sub>/C<sub>2</sub>H<sub>4</sub> 分离的相似处，本研究还计算了化学描述符，列于表 1 中。

表 1：化学及结构描述符汇总

Table 1: Summary of chemical and structural descriptors		
种类	描述符	单位
化学描述符	no. of H atoms (H)	—
	no. of C atoms (C)	—
	no. of N atoms (N)	—
	oxygen (O)	—
	total degree of unsaturation (TDU)	$[(C \times 2) + 2 - H]/2$
	Metallic percentage(MP)	(number of metal atoms/C) $\times$ 100
	Oxygen to-metal ratio(OMR)	$[2 \times O]/\text{number of metal atoms}$
	Electronegative to total ratio(ETR)	—
	weighted electronegativity per atom (WEPA)	—
	nitrogen to oxygen (N-O)	N/O
	Henry's coefficient (HC)	mol·kg <sup>-1</sup> ·Pa <sup>-1</sup>
结构描述符	crystal density ( $\rho$ )	g·cm <sup>-3</sup>
	void fraction (VF)	—
	gravimetric surface area(GSA)	m <sup>2</sup> ·g <sup>-1</sup>
	volumetric surface area(VSA)	m <sup>2</sup> ·cm <sup>-3</sup>
	largest cavity diameter (LCD)	Å
	pore limiting diameter(PLD)	Å

## 1.4 随机森林 (Random Forest, RF)

本研究采用 scikit-learn 库<sup>[55]</sup> 中的随机森林算法来构建模型预测 APS。为了更好地研究结构描述符与化学描述符对 MOFs 的 APS 的影响规律，本研究分别使用结构描述符、化学描述符及二者耦合下的描述符作为预测模型的输入，APS 作为模型的输出构建回归模型。随机森林算法基于自助法 (bootstrap) 对样本有放回地进行采样，使得有大约 36.8% 的样本未被用于建模，这样的样本成为袋外数据 (out-of-bag)，可用于模型的验证。同时，本研究还计算了特征重要度，讨论了各个描述符对预测模型的影响程度。

## 1.5 吸附性能评价指标

### 1.5.1 吸附选择性 (Selectivity, $S$ )

吸附选择性是一个特定分子从它与其他分子的混合物中优先被吸附的量度。在 C<sub>3</sub>H<sub>8</sub>/C<sub>3</sub>H<sub>6</sub> 分离中，C<sub>3</sub>H<sub>8</sub> 的吸附选择性定义为下式：

$$S_{ads} = \frac{q_{C_3H_8}/q_{C_3H_6}}{f_{C_3H_8}/f_{C_3H_6}}$$

其中， $q$  为吸附量，单位为 mol·kg<sup>-1</sup>； $f$  为气体的逸度，单位为 Pa。

### 1.5.2 工作能力 (Working Capacity, $\Delta N$ )

在变压吸附 (PSA) 分离过程中，工作能力  $\Delta N$  (mol·kg<sup>-1</sup>) 是在吸附和解吸压力下模拟的气体吸附量之间的差值，工作能力决定了将压力降低到解吸压力后能输送多少被吸附的分子。本研究中，设定吸附压力为 1bar，解吸压力为 0.1 bar。

$$\Delta N = q_{C_3H_8}(1bar) - q_{C_3H_6}(0.1bar)$$

### 1.5.3 吸附剂性能得分 (Adsorbent Performance Score, APS)

一直以来，人们致力于寻找具有高选择性和高容量的吸附剂。然而，选择性和工作能力之间通常存在权衡：孔径较小的 MOFs 通常具有较高的吸附选择性，但其工作能力较低，反之亦然。为了确定具有高选择性和高容量的 MOFs，吸附剂性能得分 (mol·kg<sup>-1</sup>) 定义为选择性和工作能力的乘积。

$$APS = S_{ads} * \Delta N$$

### 1.5.4 再生能力 (Regenerability, $R\%$ )

再生能力是在 PSA 过程的吸附气体解吸过程中，吸附剂的吸附位置有多少可以再生的度量，是评价 MOFs 作为吸附剂重复使用的一个指标。在任何实际应用中，吸附剂应易于再生，再生能力大于 80%。再生能力极低的吸附剂在实际中无法使用，因为吸附剂会被气体分子永久堵塞。对于强吸附组分，其数学表达式如下：

$$R\% = \frac{\Delta N}{q} \times 100\%$$

## 2 结果与讨论

### 2.1 选择性-工作能力关系

由图 2 可知，选择性和工作能力之间有个很明显的权衡，同时对于相同的 MOFs 结构，这两个指标并没有表现出最高的值。在另一项研究中，Yeo<sup>[30]</sup> 等人调查了剑桥结构数据库和无机晶体结构数据库中的数百万多孔材料，以筛选出 C<sub>3</sub>H<sub>8</sub>/C<sub>3</sub>H<sub>6</sub> 混合

物分离的最佳候选 MOFs。该研究同样观察到吸附选择性和工作能力之间的类似权衡,即表现出高选择性的结构通常与低或零工作能力相关。而结合吸附选择性和工作能力的 APS 是一种更好的评价  $C_3H_8/C_3H_6$  分离的性能指标。因此,本研究选择 APS 为评价指标,进行后续的工作研究。

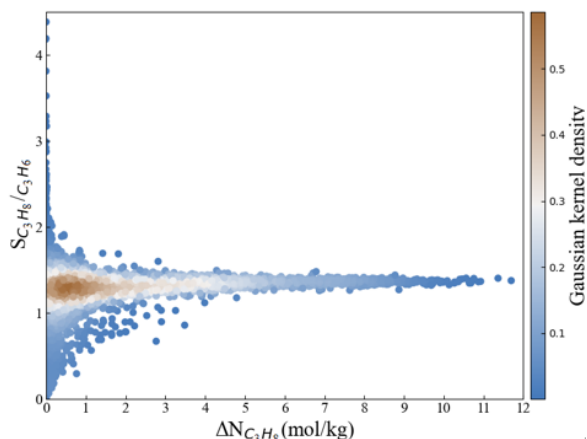


图 2: 选择性、工作能力之间的关系

Figure 2: The relationship between selectivity and working capacity.

## 2.2 随机森林模型性能

利用 6 个结构描述符和 10 个化学描述符构建随机森林机器学习模型来预测 APS, 训练集包含 6000 个 MOFs, 测试集包含 1500 个 MOFs (训练集:测试集=4: 1), 随机森林采用 250 棵数建立预测模型, 利用 out-of-bag 数据对模型进行调参。表 2 总结了随机森林模型的性能指标, 包括  $R^2$ 、RMSE、MAE。

表 2: 具有不同类型描述符的模型的性能指标

Table 2: Performance metrics of models developed by different types of descriptors.

Descriptor types	Validation set		Test set	
	$R^2$	$R^2$	RMSE	MAE
6 STRU.	0.815	0.804	1.492	0.851
12 CHEM.	0.841	0.852	1.295	0.796
6 STRU & 12 CHEM	0.958	0.960	0.677	0.373

从表 2 中可以看出, 在仅使用结构描述符的 RF 模型的表现最差, 测试集的  $R^2$  仅有 0.804, 相对地, 在仅用化学描述符的 RF 模型测试集的  $R^2$  达到了 0.852, 较前者有一定的提升, 两种描述符的  $R^2$  均高于 0.80, 说明两种模型均具备不错的预测

效果, 但是要应用于高通量筛选, 模型的预测效果还是不够好。因此为了提升 RF 模型的性能, 本研究结合两种描述符共同建立预测模型, 最后模型测试集的  $R^2$  提升到了 0.960, RMSE 为 0.677, MAE 为 0.373, 模型性能得到了较大的提升, 这说明二者结合能很好地预测 APS。

图 3 对比了测试集结合两种描述符的 RF 模型预测的 APS 值和 GCMC 模拟得到的 APS 值, 图上不同的颜色对应于不同区域数据点的高斯核密度。从图中可以看出, 预测值和模拟值有很好的—致性。

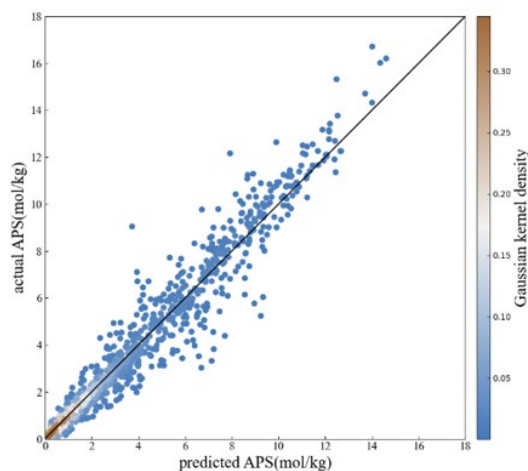


图 3: 测试集上预测值和模拟值对比

Figure3: Comparison of predicted APS values and simulated values on the test set.

图 4 展示了结合化学和结构描述符的 RF 模型中决策树回归器的平均特征重要度。特征的重要度越高, 说明该特征对预测模型的贡献程度越高。可以看出, GSA 是所有的描述符中重要度最高的, 它对提升 RF 模型的性能有着很重要的影响, 密度、 $C_3H_8$  的亨利系数也对模型有较大的影响。



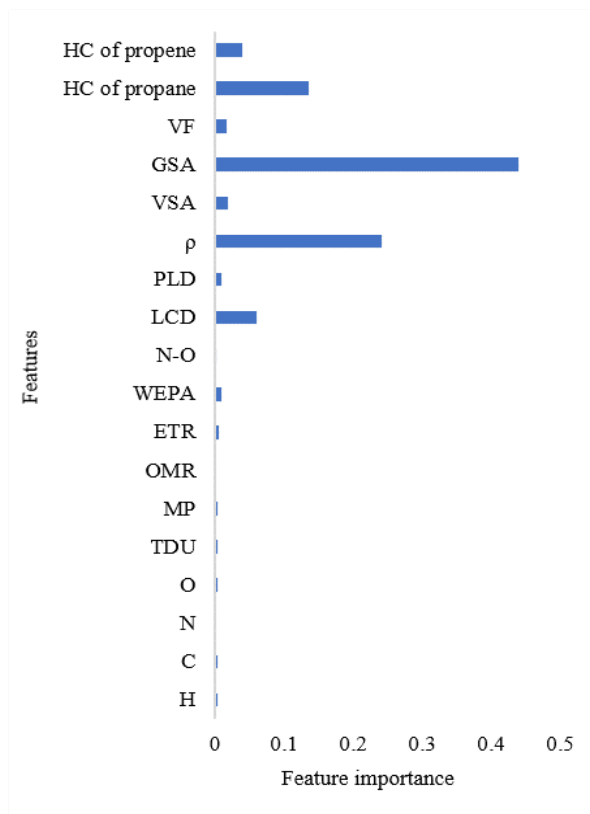


图 4: 随机森林中决策树回归器的平均特征重要度

Figure 4: The average feature importance among the decision tree regressors in the random forest

### 2.3 筛选剩余 MOFs

从上述分析中可以看出, 结合结构描述符和化学描述符建立 RF 模型预测的 APS 和实际模拟的 APS 有很好的 consistency。因此选择上述 RF 模型来预测剩余的 MOFs 的 APS, 用预测值对模型进行预筛选。同样, 计算剩余 MOFs 的结构和化学描述符作为模型的输入。以 ML 模型的预测结果为预筛选基础, 选择预测结果中 top 2500 的 MOFs 对它们进行 GCMC 来对预测结果进行修正, 这 2500 个 MOFs 构成的数据集称作候选集。

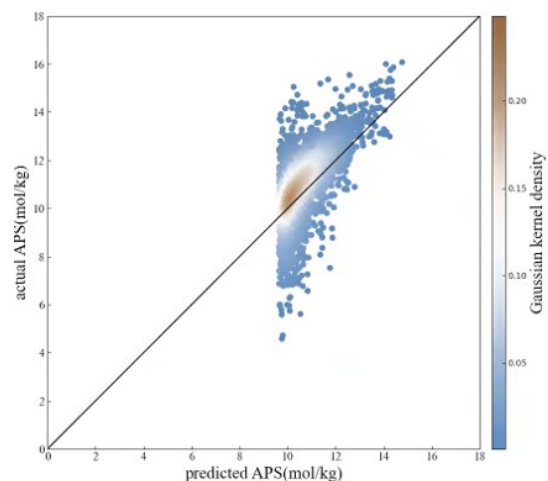


图 5: 候选集中 APS 预测值和模拟值对比图

Figure 5: Comparison of APS predicted values and simulated values in promising set

图 5 展示了候选集的预测值和模拟值的对比图, 从图中可以看出, 大部分的 MOFs 的 APS 预测值和实验值有比较好的一致性, 而其中一些 APS 高的 MOFs 机器学习模型显然低估了它们的性能。正如对比模型性能时多提到的 (图 3), 该模型在预测低性能 MOFs 时有很好的表现, 然而在预测高性能 MOFs 时有一定的偏差, 因此在使用机器学习模型进行预筛选时更多关注的是该模型能否找出性能好的 MOFs, 即 APS 高的 MOFs。在机器学习预测的 top 2500 MOFs 中, 性能最差的 MOFs 其 APS 值为  $9.637 \text{ mol} \cdot \text{kg}^{-1}$ , 以该 MOFs 预测结果作为阈值, 发现这 2500 个 MOFs 在实际 GCMC 模拟值中有 2120 个 MOFs 的模拟结果高于 9.637 这个阈值, 比例达到了 84.8%, 因此, 本研究认为, 所构建的 RF 模型作为预筛选手段是可靠的。

### 2.4 发现高性能 MOFs

在候选集中, 绝大多数的高性能 MOFs 来自于 hMOFs 数据库, 仅有 8 个 MOFs 来自 CoRE 数据库, 虽然大多数的 hMOFs 表现出高 APS 值, 但 CoRE 数据库中的 MOFs 因其已被合成从而具备更实际的应用价值。最后, 从候选集中筛选出了 5 个 hMOFs 和 5 个 CoRE MOFs, 它们的参数列于表 3 中。从表中可以看出, 筛选出的 MOFs 的 APS 均高于  $11 \text{ mol} \cdot \text{kg}^{-1}$ , 同时, 可再生能力均大于 85%, 说明这些吸附剂易于再生, 这对实际的 PSA 过程是很理想的。值得注意的是, Solanki<sup>[31]</sup>等人同样以 APS 为评价指标筛选出 20 个 CoRE 数据库中的 MOFs, 这其中仅有 5 个 MOFs 不包含 OMS, 而本

研究所筛选出的 5 个 CoRE 数据库的 MOFs 中有 4 个与该研究筛选出的不含 OMS 的 MOFs 相同，这再次说明了本研究筛选工作的有效性。

表 3: R% 大于 80% 且具有最高的 APS 的前十个 MOFs

Table 3: Top 10 MOFs that possess both R% > 80 and the highest APS values.

MOF ID	$\Delta N_{C_3H_8}$ (mol·kg <sup>-1</sup> )	$S_{C_3H_8/C_3H_6}$	APS (mol·kg <sup>-1</sup> )	R %
hMOF-5081969	11.68	1.38	16.08	91
hMOF-5083374	11.66	1.37	16.00	91
hMOF-5057347	11.53	1.38	15.87	91
hMOF-5058692	11.62	1.36	15.75	90
hMOF-5033500	11.48	1.35	15.54	92
FUYCIN*	10.01	1.39	13.91	95
ZILBUT*	9.01	1.44	12.94	86
ICAQIO*	9.45	1.34	12.68	92
ICAQOU*	9.26	1.36	12.58	92
CAVPEW	8.74	1.36	11.91	93

\*Screened MOFs same as in the literature.

### 3 结 论

本研究提出一种将机器学习和分子模拟相结合的新的高通量筛选策略，用于从大型多孔材料数据库中快速高效地筛选出具有高 APS 值的金属有机框架材料，从而选择性的从 C<sub>3</sub>H<sub>6</sub> 中分离出 C<sub>3</sub>H<sub>8</sub>。为了展示所提出的方法的有效性，本研究考虑了一个由 CoRE 数据库和 hMOFs 数据库构成的包含超过 14 万个 MOFs 数据集。从该数据集中随机选择 6000 个 MOFs，以其结构描述符和化学描述符作为模型输入，APS 值作为模型输出，基于随机森林算法构建预测模型，该模型在测试集上表现较好 ( $R^2=0.960$ )。使用该模型对剩余 MOFs 的 APS 进行预测，随后从预测结果中挑选出前 2500 个表现最好的 MOFs，并进一步使用 GCMC 模拟验证。结果表明，本研究所开发的预测模型能较准确的预测外来 MOFs 的 APS。最后，从这 2500 个 MOFs 中筛选出性能最好的 5 个真实 MOFs 和 5 个假想 MOFs，而 5 个最佳真实 MOFs 和文献中报道的 MOFs 有 4 个相同，进一步证明所开发的模型的准确性和高效性。此外，本文最后对建模所用的描述符的重要性进行讨论，发现重量比表面积和密度是两个对预测结果影响最大的结构特征。本研究所提出的方法不仅可以用于 C<sub>3</sub>H<sub>8</sub>/C<sub>3</sub>H<sub>6</sub> 的性能参数预

测，也可迁移应用于其他气体分离体系，从而高效地筛选出表现优异的多孔材料。

### 参考文献

- [1] REGE S U, PADIN J, YANG R T. Olefin/paraffin separations by adsorption: pi-complexation vs. kinetic separation [J]. Aiche J, 1998, 44(4): 799-809.
- [2] SAFARIK D J, ELDRIDGE R B. Olefin/paraffin separations by reactive absorption: A review [J]. Industrial & Engineering Chemistry Research, 1998, 37(7): 2571-2581.
- [3] WANG Y X, PEH S B, ZHAO D. Alternatives to Cryogenic Distillation: Advanced Porous Materials in Adsorptive Light Olefin/Paraffin Separations [J]. Small, 2019, 15(25): 38.
- [4] KORZYNSKI M D, DINCA M. Oxidative Dehydrogenation of Propane in the Realm of Metal-Organic Frameworks [J]. ACS Central Sci, 2017, 3(1): 10-12.
- [5] SHOLL D S, LIVELY R P. Seven chemical separations to change the world [J]. Nature, 2016, 532(7600): 435-437.
- [6] KOROS W J, LIVELY R P. Water and beyond: Expanding the spectrum of large-scale energy efficient separation processes [J]. Aiche J, 2012, 58(9): 2624-2633.
- [7] YANG R. Gas separation by adsorption processes [J]. Chemical Engineering Science, 1997, 43(4):
- [8] TANG H, JIANG J. In silico screening and design strategies of ethane - selective metal - organic frameworks for ethane/ethylene separation [J]. Aiche J, 2020, 67(3):
- [9] BLOCH E D, QUEEN W L, KRISHNA R, et al. Hydrocarbon Separations in a Metal-Organic Framework with Open Iron(II) Coordination Sites [J]. Science, 2012, 335(6076): 1606-1610.
- [10] BOHME U, BARTH B, PAULA C, et al. Ethene/Ethane and Propene/Propane Separation via the Olefin and Paraffin Selective Metal-Organic Framework Adsorbents CPO-27 and ZIF-8 [J]. Langmuir, 2013, 29(27): 8592-8600.
- [11] PENG J J, WANG H, OLSON D H, et al. Efficient kinetic separation of propene and propane using two microporous metal organic frameworks [J]. Chem Commun, 2017, 53(67): 9332-9335.
- [12] WANG H, DONG X L, COLOMBO V, et al. Tailor-Made Microporous Metal-Organic Frameworks for the Full Separation of Propane from Propylene Through Selective Size Exclusion [J]. Adv Mater, 2018, 30(49): 9.
- [13] BAO Z B, CHANG G G, XING H B, et al. Potential of microporous metal-organic frameworks for separation of hydrocarbon mixtures [J]. Energy Environ Sci, 2016, 9(12): 3612-3641.
- [14] CUI W G, HU T L, BU X H. Metal-Organic Framework Materials

for the Separation and Purification of Light Hydrocarbons [J]. *Adv Mater*, 2020, 32(3): 24.

[15] BAE Y S, LEE C Y, KIM K C, et al. High Propene/Propane Selectivity in Isostructural Metal-Organic Frameworks with High Densities of Open Metal Sites [J]. *Angew Chem-Int Edit*, 2012, 51(8): 1857-1860.

[16] CADIAU A, ADIL K, BHATT P M, et al. A metal-organic framework-based splitter for separating propylene from propane [J]. *Science*, 2016, 353(6295): 137-140.

[17] YANG L, CUI X, DING Q, et al. Polycatenated Molecular Cage-Based Propane Trap for Propylene Purification with Recorded Selectivity [J]. *ACS Appl Mater Interfaces*, 2020, 12(2): 2525-2530.

[18] MERSMANN A, FILL B, HARTMANN R, et al. The potential of energy saving by gas-phase adsorption processes [J]. *Chem Eng Technol*, 2000, 23(11): 937-944.

[19] WANG S, ZHANG Y, TANG Y, et al. Propane-selective design of zirconium-based MOFs for propylene purification [J]. *Chemical Engineering Science*, 2020, 219.

[20] MOGHADAM P Z, LI A, WIGGIN S B, et al. Development of a Cambridge Structural Database Subset: A Collection of Metal-Organic Frameworks for Past, Present, and Future [J]. *Chemistry of Materials*, 2017, 29(7): 2618-2625.

[21] ALTINTAS C, AVCI G, DAGLAR H, et al. Database for CO<sub>2</sub> Separation Performances of MOFs Based on Computational Materials Screening [J]. *ACS Appl Mater Interfaces*, 2018, 10(20): 17257-17268.

[22] ALTINTAS C, KESKIN S. Role of partial charge assignment methods in high-throughput screening of MOF adsorbents and membranes for CO<sub>2</sub>/CH<sub>4</sub> separation [J]. *Mol Syst Des Eng*, 2020, 5(2): 532-543.

[23] HALDOUPIS E, NAIR S, SHOLL D S. Finding MOFs for Highly Selective CO<sub>2</sub>/N<sub>2</sub> Adsorption Using Materials Screening Based on Efficient Assignment of Atomic Point Charges [J]. *J Am Chem Soc*, 2012, 134(9): 4313-4323.

[24] QIAO Z W, ZHANG K, JIANG J W. In silico screening of 4764 computation-ready, experimental metal-organic frameworks for CO<sub>2</sub> separation [J]. *Journal of Materials Chemistry A*, 2016, 4(6): 2105-2114.

[25] CHUNG Y G, HALDOUPIS E, BUCIOR B J, et al. Advances, Updates, and Analytics for the Computation-Ready, Experimental Metal-Organic Framework Database: CoRE MOF 2019 [J]. *J Chem Eng Data*, 2019, 64(12): 5985-5998.

[26] SIKORA B J, WILMER C E, GREENFIELD M L, et al. Thermodynamic analysis of Xe/Kr selectivity in over 137 000 hypothetical metal-organic frameworks [J]. *Chem Sci*, 2012, 3(7): 2217-2223.

[27] SUMER Z, KESKIN S. Molecular simulations of MOF adsorbents and membranes for noble gas separations [J]. *Chemical Engineering*

*Science*, 2017, 164(108-121).

[28] ALTINTAS C, AVCI G, DAGLAR H, et al. Computer simulations of 4240 MOF membranes for H<sub>2</sub>/CH<sub>4</sub> separations: insights into structure-performance relations [J]. *Journal of Materials Chemistry A*, 2018, 6(14): 5836-5847.

[29] CHIAU M J, WANG Y G, WU X J, et al. Computational screening of metal-organic frameworks with open copper sites for hydrogen purification [J]. *Int J Hydrog Energy*, 2020, 45(51): 27320-27330.

[30] YEO B C, KIM D, KIM H, et al. High-Throughput Screening to Investigate the Relationship between the Selectivity and Working Capacity of Porous Materials for Propylene/Propane Adsorptive Separation [J]. *The Journal of Physical Chemistry C*, 2016, 120(42): 24224-24230.

[31] SOLANKI V A, BORAH B. Computational screening of metal-organic framework structures for separation of propane/propene mixture [J]. *Molecular Simulation*, 2020, 46(17): 1362-1372.

[32] ALTINTAS C, ALTUNDAL O F, KESKIN S, et al. Machine Learning Meets with Metal Organic Frameworks for Gas Storage and Separation [J]. *J Chem Inf Model*, 2021, 61(5): 2131-2146.

[33] HALDER P, SINGH J K. High-Throughput Screening of Metal-Organic Frameworks for Ethane-Ethylene Separation Using the Machine Learning Technique [J]. *Energy & Fuels*, 2020, 34(11): 14591-14597.

[34] WU Y, DUAN H, XI H. Machine Learning-Driven Insights into Defects of Zirconium Metal-Organic Frameworks for Enhanced Ethane-Ethylene Separation [J]. *Chemistry of Materials*, 2020, 32(7): 2986-2997.

[35] LIANG H, JIANG K, YAN T-A, et al. XGBoost: An Optimal Machine Learning Model with Just Structural Features to Discover MOF Adsorbents of Xe/Kr [J]. *ACS Omega*, 2021, 6(13): 9066-9076.

[36] SIMON C M, MERCADO R, SCHNELL S K, et al. What Are the Best Materials To Separate a Xenon/Krypton Mixture? [J]. *Chemistry of Materials*, 2015, 27(12): 4459-4475.

[37] BURNER J, SCHWIEDRZIK L, KRYKUNOV M, et al. High-Performing Deep Learning Regression Models for Predicting Low-Pressure CO<sub>2</sub> Adsorption Properties of Metal-Organic Frameworks [J]. *The Journal of Physical Chemistry C*, 2020, 124(51): 27996-28005.

[38] DURECKOVA H, KRYKUNOV M, AGHAJI M Z, et al. Robust Machine Learning Models for Predicting High CO<sub>2</sub> Working Capacity and CO<sub>2</sub>/H<sub>2</sub> Selectivity of Gas Adsorption in Metal Organic Frameworks for Precombustion Carbon Capture [J]. *The Journal of Physical Chemistry C*, 2019, 123(7): 4133-4139.

[39] FERNANDEZ M, WOO T K, WILMER C E, et al. Large-Scale Quantitative Structure-Property Relationship (QSPR) Analysis of Methane Storage in Metal-Organic Frameworks [J]. *The Journal of Physical Chemistry C*, 2013, 117(15): 7681-7689.



- [40] MA R M, COLON Y J, LUO T F. Transfer Learning Study of Gas Adsorption in Metal-Organic Frameworks [J]. ACS Appl Mater Interfaces, 2020, 12(30): 34041-34048.
- [41] WANG R, ZHONG Y, BI L, et al. Accelerating Discovery of Metal-Organic Frameworks for Methane Adsorption with Hierarchical Screening and Deep Learning [J]. ACS Appl Mater Interfaces, 2020, 12(47): 52797-52807.
- [42] ALTINTAS C, ERUCAR I, KESKIN S. High-Throughput Computational Screening of the Metal Organic Framework Database for CH<sub>4</sub>/H<sub>2</sub> Separations [J]. ACS Appl Mater Interfaces, 2018, 10(4): 3668-3679.
- [43] CHUNG Y G, GOMEZ-GUALDRON D A, LI P, et al. In silico discovery of metal-organic frameworks for precombustion CO<sub>2</sub> capture using a genetic algorithm [J]. Sci Adv, 2016, 2(10): 9.
- [44] AVCI G, VELIOGLU S, KESKIN S. In Silico Design of Metal Organic Frameworks with Enhanced CO<sub>2</sub> Separation Performances: Role of Metal Sites [J]. Journal of Physical Chemistry C, 2019, 123(46): 28255-28265.
- [45] PARK J, LANDA H O R, KAWAJIRI Y, et al. How Well Do Approximate Models of Adsorption-Based CO<sub>2</sub> Capture Processes Predict Results of Detailed Process Models? [J]. Industrial & Engineering Chemistry Research, 2020, 59(15): 7097-7108.
- [46] POLAT H M, KAVAK S, KULAK H, et al. CO<sub>2</sub> separation from flue gas mixture using [BMIM] [BF<sub>4</sub>] /MOF composites: Linking high-throughput computational screening with experiments [J]. Chemical Engineering Journal, 2020, 394(10).
- [47] WILMER C E, LEAF M, LEE C Y, et al. Large-scale screening of hypothetical metal-organic frameworks [J]. Nature Chemistry, 2012, 4(2): 83-89.
- [48] ALTINTAS C, KESKIN S. Computational screening of MOFs for C<sub>2</sub>H<sub>6</sub>/C<sub>2</sub>H<sub>4</sub> and C<sub>2</sub>H<sub>6</sub>/CH<sub>4</sub> separations [J]. Chemical Engineering Science, 2016, 139(49-60).
- [49] ALTINTAS C, KESKIN S. Molecular simulations of MOF membranes for separation of ethane/ethene and ethane/methane mixtures [J]. RSC Adv, 2017, 7(82): 52283-52295.
- [50] DUBBELDAM D, CALERO S, ELLIS D E, et al. RASPA: molecular simulation software for adsorption and diffusion in flexible nanoporous materials [J]. Molecular Simulation, 2016, 42(2): 81-101.
- [51] MARTIN M G, SIEPMANN J I. Transferable potentials for phase equilibria. 1. United-atom description of n-alkanes [J]. Journal of Physical Chemistry B, 1998, 102(14): 2569-2577.
- [52] WICK C D, MARTIN M G, SIEPMANN J I. Transferable potentials for phase equilibria. 4. United-atom description of linear and branched alkenes and alkylbenzenes [J]. Journal of Physical Chemistry B, 2000, 104(33): 8008-8016.
- [53] RAPPE A K, CASEWIT C J, COLWELL K S, et al. UFF, A FULL PERIODIC-TABLE FORCE-FIELD FOR MOLECULAR MECHANICS AND MOLECULAR-DDYNAMICS SIMULATIONS [J]. J Am Chem Soc, 1992, 114(25): 10024-10035.
- [54] WILLEMS T F, RYCROFT C, KAZI M, et al. Algorithms and tools for high-throughput geometry-based analysis of crystalline porous materials [J]. Microporous Mesoporous Mat, 2012, 149(1): 134-141.
- [55] PEDREGOSA F, VAROQUAUX G, GRAMFORT A, et al. Scikit-learn: Machine Learning in Python [J]. J Mach Learn Res, 2011, 12(2825-2830).