Finding advanced MOFs to capture argon from the air through machine-learning and molecular simulation

Xiaoyi Xu, a Xu Ji,b Yiyang Daia\*

aSchool of Chemical Engineering, Sichuan University, Chengdu 610065, China

\*daiyy@scu.edu.cn

Abstract

Argon, widely used in industry as inert gas, is colorless and oderless with extremely low chemical reactivity. Traditional method of obtaining this rare gas is fractionation which has some shortcomings such as high energy consumption, high cost and complex production process. Metal-organic frameworks(MOFs), as a class of emerging organic-inorganic hybrid porous materials, have great potential to become new adsorbents of argon. Until now, Computation-ready experimental metal-organic framework(CoRE MOF) database contains 12,020 experimentally synthesized MOFs so that it is like finding a needle in a haystack to discover advanced MOFs through only methods of experimental. This study proposes a high-throughput screening strategy based on machine learning(ML) and grand canonical Monte Carlo(GCMC), which could quickly and precisely discover the MOFs with excellent properties for separating argon and air. At the same time, the addition of ML greatly saved computing power. Firstly, the adsorbent performance scores(APS) of randomly selected 25 % candidates in CoRE database after pre-screening were given by simulating pressure swing adsorption(PSA) process of air on MOF using GCMC. Then, with 23 descriptors characterized structure and chemical properties of MOFs as features and APS as target values, ML models were built to predict the APS of remaining 75 % MOFs. Finally, TOP5 MOFs with excellent adsorption performance were known according to result of sorting all APS. This screening strategy could greatly save calculation cost and improved efficiency of finding high-performance MOFs. Comparing with traditional materials discovery approaches, this work could be faster, more focused and possess more guiding insight for new MOFs adsorbents of argon ahead synthesized in the laboratory.

**Keywords**: Argon separation, MOF, Grand Canonical Monte Carlo,machine-learning, high-throughput screening, Structure-performance relationship.

* 1. Introduction

Argon is one of the most abundant noble gases in the air(About 0.932 %). It is colorless, odorless, non-flammable, extremely chemically inactive and innocuous, which has a very wide range of applications.To be specific, argon is used as protective gas for welding processes to prevent welding defects(Phakpeetinan et al.,2016 ). The argon knife is used for clinical minimally invasive surgical treatment(Camille et al.,2019) and the argon laser is used for the treatment of various eye and skin diseases(Korkmaz et al., 2015). In research and experiments related to the semiconductor industry, argon also has an extremely broad range of applications. To summarize, Argon plays an indispensable role in various fields such as industry, medicine, environmental protection, mining, geology, manufacturing and etc.

Currently, the main method for manufacturing argon is air separation, which involves using low-temperature liquid air distillation to obtain large-scale argon separation based on different boiling points of each component. This production method's drawbacks comprise high energy consumption, high cost, complex and hazardous production process, and an enormous coolant and media requirement throughout(Xu Jin et al.,2006). Pressure swing adsorption separation(PSA) is a gas separation technique known for its excellent low energy consumption, easy operation, cheap regeneration cost, good environmental benefits and great dependability. Thus, through PSA process to capture argon from the air, may offer a fresh approach to resolving the aforementioned issues and enabling argon production.So far, there have been several studies on the application of PSA for argon separation from air, such like Saburo Hayashi et al. developed a secondary pressure swing adsorption system consisting of two molecular sieve zeolite columns which is capable of producing argon from a ternary gas mixture that contains argon, oxygen and nitrogen. The most critical point of this process is the selection of adsorbent. As a new class of super porous nanomaterials connected by metal atoms, organic ligands, and topological structures, metal-organic framework materials (MOFs) have a lot of promise and appealing prospects,which have notable properties like large specific surface area, high porosity, adjustable pore size, robust functionality, and exceptional adsorption activity. With 12,020 MOFs in computationally ready experimental MOF database (CoRE MOF), built by Snurr et al.(2014), it is nearly impossible to find high-performance MOF adsorbents using only methods of experimental synthesis and characterization for all MOFs, which is similar to trying to find a needle in a haystack. Consequently, a rapid screening strategy for superior-performing MOF adsorbents holds immense significance in achieving efficient recovery and capture of argon from air. Molecular simulation-based high throughput computational screening (HTCS) techniques have become powerful tools for accelerating material discovery due to the remarkable advances in computational chemistry and increased computational power(Ramprasad et al.,2017). Experts and academics have conducted a number of studies on the HTCS of MOF materials, including ethane/ethylene separation, hydrogen purification, separation of Cs/Sr, and natural gas purification (CO2/CH4).

However, there aren't many studies on inert gases. The majority concentrate on the gas separation of Xe and Kr. Compared to the vast number of MOFs available, existing studies on Ar adsorption and separation mainly focus on specific types of MOFs. Anastasios I. et al.(2004), utilized grand canonical Monte Carlo(GCMC) and equilibrium molecular dynamics to compute the adsorption and diffusion performance of argon in CuBTC MOF. The reports about HTCS on MOFs for argon capture from air are lacking in the relevant literature, which may indicate that promising MOF adsorbent candidates are still to be found. Even so, it would be extremely expensive and time-consuming to perform a molecular simulation for all data. Machine learning(ML) is routinely combined with GCMC in the MOF domain to identify MOFs with exceptional performance from large databases.

Given this, this work presents a high-throughput screening strategy based on ML and GCMC simulation, which is used to efficiently and precisely screen MOFs with the best argon/air separation performance from CoRE MOF database. In order to facilitate calculation, the system is regarded as a three-component mixture of argon, oxygen and nitrogen (0.01:0.78:0.21). The simulated temperature is 298 K, and the adsorption and desorption pressure are 1 bar and 0.1 bar respectively.

* 1. Model and Method

This section is used to detail the workflow for high-throughput screening of MOFs adsorbents to separate argon from air adsorbents.



Figure 1 Workflow for high throughput screening of MOFs to separate argon from air

* + 1. MOF Database and Pre-screening

The database utilized in this study was CoRE MOF database, which comprised the Crystallographic Information File (CIF) of 12,020 structurally diverse authentic MOFs.

In order to comprehensively characterize MOFs, this research considered both structural and chemical descriptors that had been successfully applied to ML models with great predictive accuracy of various gas systems. The free source software Zeo++ 0.3 calculated six structural descriptors: largest cavity diameter(LCD), pore limiting diameter(PLD), accessible surface area(GSA), volume surface area(VSA), porosity(Φ), and density(ρ). The code proposed by barisoo and ibarisorhan calculated seventeen chemical characteristics: number of atoms per unit cell, metal type, unsaturation, metallic percentage, oxygen to metal ratio, electronegative atoms to total atoms ratio, weighted electronegative per atom, and nitrogen to oxygen ratio. During pre-screening stage, MOFs that clearly could not adsorb argon were eliminated based on the geometric structure analysis; that is those with PLD<0.3405 nm or GSA=0 (the molecular dynamic diameter of Ar is 0.3405 nm). As a result, there were 7,328 MOFs remained in the pre-screened database.

* + 1. Grand Canonical Monte Carlo Simulate

Based on RASPA 2.0 software, the GCMC simulation method was used to calculate performance indexes for capturing argon from air. To reduce the time of simulation, all MOFs skeletons were considered rigid structures. The Lennard-Jones(LJ) potential energy parameter of the MOF skeleton atom was taken from the UFF force field, The LJ truncation distance was set to 17 Å. The performance evaluation indexes included working capacity ($∆$N, mol/kg), selectivity (S), regeneration performance (R%) and adsorbent performance score (APS), and the calculation formula is shown as follows.

|  |  |
| --- | --- |
| $$∆N\_{Ar}=∆N\_{Ar}^{ads}−∆N\_{Ar}^{des}$$ | (1) |
| $$S\_{Ar/(O\_{2}+N\_{2})}=\frac{∆N\_{Ar}^{ads}/(∆N\_{N\_{2}}^{ads}+∆N\_{O\_{2} }^{ads})}{f\_{Ar}/(f\_{N\_{2}}+f\_{O\_{2}})}$$ | (2) |
| $$APS = ∆N\_{Ar}×S\_{Ar/(O\_{2}+N\_{2})}$$ | (3) |
| $R\%=\frac{∆N\_{Ar}}{∆N\_{Ar}^{ads}}×1$00 | (4) |

* + 1. Model

We constructed five ML regression-prediction models to describe the structure-activity relationship model between MOF descriptors and the adsorption properties of argon. Each model took 6 structural descriptors and 17 chemical descriptors as inputs, and the values of APS as targets. A series of feature engineering operations such as data cleaning, outlier processing, dimensionality reduction, standardization, and normalization were performed on dataset to improve the effect of models. Detailed parameters of the machine learning models are shown in Table .

* 1. Result and Discussion
		1. Validation with Literature Data

The applicability and accuracy of the method utilized in this work must be validated before large-scale molecular simulation. The argon single-component adsorption isotherm of IRMOF74 was calculated and compared to experimental isotherm. Besides, we chose 10 % MOFs in CoRE MOF database at random and calculated their working capabilities(including several specific Ar adsorbents with potential) at 298K and 1bar. As shown in Figure 2 and Table 1, the simulation results are in good agreement with the experimental and simulation data reported in the literature, which indicates that the force field and parameter settings used in this article are reliable.

Table 1 Several potential argon MOF adsorbents

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| MOFname | GUXLIU | BUSQIQ | WOWGEU | CUNXIS | PARMIG | BEKSAM |
| Simulation | 3.624 | 4.758 | 4.137 | 4.837 | 4.224 | 5.546 |
| Literature | 3.9 | 4.58 | 3.89 | 3.82 | 3.94 | 4.98 |



Figure 2 Comparison of 10%CoREMOFs working capabilities with literature

* + 1. Simulation Results

25 % MOFs were randomly selected to simulate their adsorption property by GCMC, and APS was sorted to obtain the top 5 MOFs, as shown in Table 2. According to simulation results, the APS of MOF adsorbents in this system is rather minimal. Some probable reasons are as follows: 1. Compared to oxygen and nitrogen, argon level in air is exceptional low. The extremely low adsorption capacity strongly hinders adsorption performance improvement. 2. Under normal conditions, the sequence of preferential adsorption of MOFs for these three mixed gases is: O2>Ar>N2. 3. Oxygen and argon have very similar physical properties, including size, boiling point, and isothermal adsorption heat, resulting in their separation is more difficult.

Table 2 TOP 5 argon MOF adsorbents

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| name | $$∆N\_{Ar}$$ | S | R % | APS |
| CUHLUO | 0.06565 | 2.305 | 86.03 | 0.1513 |
| AFEKAX | 0.03959 | 2.099 | 88.45 | 0.0831 |
| BUFPIC | 0.04203 | 1.551 | 57.54 | 0.0652 |
| BIBYAO | 0.04604 | 1.350 | 79.84 | 0.0621 |
| EKADIF | 0.02093 | 2.595 | 88.16 | 0.0543 |

* + 1. Structure-activity Relationship

The structure-activity relationship of MOFs was drawn based on PLD. As shown in Figure 1, higher performance is more likely to arise in smaller PLD intervals.



Figure 3 Structure-activity relationship analysis based on PLD

* + 1. ML Model Results

Using 6 geometric descriptors and 17 chemical descriptors as input variables, and APS value of MOFs as the target variable, five machine learning regression-prediction models were constructed. The detailed parameters are shown in Table 3.

Table 3 Detailed parameters of the machine learning model

*R2: R-Square of model on training data; MAE: mean absolute error*

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| model | n\_estimators | max\_features | R2 | MAE |
| XgBoost | 500 | - | 0.999 | 0.00323 |
| ExtraTrees | 100 | 0.7 | 0.9782 | 0.00310 |
| Gradient Boost | 500 | - | 0.9522 | 0.00355 |
| Random Forest | 100 | 10 | 0.5095 | 0.00326 |
| Voting | RF | 350 | 0.75 | 0.8252 | 0.00308 |
| gbdt | 100 | 0.5 |
| XgB | 25 | - |
| ET | 100 | 0.75 |

* 1. Conclusion

In this paper, a method that combines ML and GCMC with large-scale calculation for screening 12,020 real MOFs as Ar/ air separation adsorbents is proposed. Firstly, 7,328 MOFs were pre-screened through geometric analysis. 25 % candidates were extracted for molecular simulation calculations, and resulting data was used to build ML models which predicted APS of remaining 75 % MOFs. The results show that simulation data based on GCMC are consistent with experimental and literature data, and are reliable and inspiring. This study found that the argon adsorption capacity of BIBYAO is 5.9803 mol/L, which is much higher than the current working capacity of argon adsorbents. Therefore, our work is highly crucial for future identification of high-performance argon adsorbents. But ML models built on 25 % MOFs data performed poorly on remaining 75% MOFs performance predictions. To improve model’s accuracy, we can further screen the modeling features, enlarge the dataset, or adjust the model structure.

References

1. Phakpeetinan, A. Chianpairot, E. Viyanit, F. Hartung, and G. Lothongkum, 2016, Effects of nitrogen and hydrogen in argon shielding gas on bead profile, delta-ferrite and nitrogen contents of the pulsed GTAW welds of AISI 316L stainless steel. Materials Testing, Vol. 58 (Issue 6), pp. 489-494.

P.Camille, H. Eduard, F. Pierre, N. Lars, 2019, Effect of argon and nitrogen atmospheres on the properties of stainless steel 316 L parts produced by laser-powder bed fusion, Materials & Design,Vol. 179, ISSN 0264-1275

S. Korkmaz, F. Ekici, S. Sül, 2015, Argon laser-assisted treatment of benign eyelid lesions, Lasers in Medical Science, 1435-604X

J. Xu, A. Malek, S. Farooq, 2006, Production of Argon from an Oxygen−Argon Mixture by Pressure Swing Adsorption, Industrial & Engineering Chemistry Research,5775-5787

H. C. Gulbalkan, Z. P. Haslak, C. Altintas, 2022, Assessing CH4/N2 separation potential of MOFs, COFs, IL/MOF, MOF/polymer, and COF/polymer composites, Chemical Engineering Journal, 428: 131239

R. Ramprasad, R. Batra, G. Pilania, A. Mannodi-Kanakkithodi, C. Kim, 2017, Machine learning in materials informatics: recent applications and prospects, npj Computational Materials, 2057-3960

Barisoo, Ibarisorhan, 2020, MofFeatures, <https://github.com/davidtangGT/MOF-Features>