



Finding the optimal CO₂ adsorption material: Prediction of multi-properties of metal-organic frameworks (MOFs) based on DeepFM

Minggao Feng^a, Min Cheng^a, Xu Ji^a, Li Zhou^a, Yagu Dang^a, Kexin Bi^a, Zhongde Dai^{b,c,*}, Yiyang Dai^{a,*}

^a School of Chemical Engineering, Sichuan University, Chengdu 610065, PR China

^b College of Architecture & Environment, Sichuan University, Chengdu 610065, PR China

^c School of Carbon Neutrality Future Technology, Sichuan University, Chengdu 610065, PR China

ARTICLE INFO

Keywords:

Carbon dioxide
MOFs
Adsorption
Recommendation system
Property prediction
Deep learning

ABSTRACT

Metal-organic frameworks (MOFs) has been widely considered as promising candidates for CO₂ adsorption due to their high porosity and high structural adjustability. By combining the various properties of MOFs, an MOFs material-property matrix can be obtained to find the best MOFs for different applications, but this matrix is often incomplete in practice. In this work, a DeepFM model was developed to predict the multi properties of CO₂ absorbents with limited start-up training data, but with high prediction accuracy. The DeepFM model contains deep neural network, with better non-linear fitting ability, thus significantly improved the prediction accuracy. Meanwhile, by adding the descriptors of MOFs as the input data of new features, the DM model also alleviates the cold start problem. By predicting 28 adsorption properties of 8206 screened hypothetical MOFs, 7 high-performance MOFs for CO₂ adsorption were selected. In addition, the model also helps to find out the relative importance degree of various descriptors on the CO₂ capture capabilities of MOFs, which is of great help in future MOFs synthesis.

1. Introduction

Carbon Capture, Utilization and Storage (CCUS) is a vital and potentially effective technology to substantially decrease CO₂ emissions [1–3]. As a class of solid materials with high porosity, high internal surface area and high structural adjustability, nanoporous materials (NPMs) [4] have been widely used in the adsorption and separation of CO₂ [5–7]. However, due to the fact that water competes with CO₂ for the same adsorption sites, many NPMs tuned to separate CO₂ do not function well when used with realistic flue gases containing water [8]. Most studies using MOFs as CO₂ adsorbents were carried out under dry conditions [9,10]. In practical applications, the water vapor in the flue gas will be needed to be removed before the adsorption process [11]. NPMs, such as metal-organic frameworks (MOFs) [12], covalent-organic frameworks (COFs) [13], porous organic cages (POCs) [14], porous polymer networks (PPNs) [15] and metal-organic polyhedra (MOPs) [16], can be constructed modularly from molecular building blocks. The diversity of building blocks and topological structures, as well as the modifiability after synthesis, all contribute to the enormous

variety of NPMs that are feasible [17]. Compared to COFs, POCs, PPNS and MOFs, MOFs have been intensively studied for various applications [18]. Up to now, more than 100 thousand different kinds of MOFs have been developed and the number is still increasing, which offers abundant data for machine learning [19]. Grand canonical Monte Carlo (GCMC), molecular dynamics (MD), and other computational techniques have been proven helpful in assessing the performance of NPMs in a wide range of applications [20–24]. Even though these methods have substantially reduced the amount of time needed to evaluate a given material, they may still be insufficiently quick enough when hundreds or thousands of CO₂ adsorption materials are being taken into consideration. It is becoming clear that machine learning (ML), which is becoming increasingly common in this field, is a useful supplement to the existing technologies [25].

ML plays an essential part in the research of MOFs [26–32]. The traditional workflow is to first compute descriptors for MOFs, features in ML, utilize them to train a model on a certain amount of data set, and then test the model on the remaining MOFs. Descriptors include but are not limited to structural descriptors [33], chemical descriptors [33] and

* Corresponding authors.

E-mail addresses: zhongde.dai@scu.edu.cn (Z. Dai), daiyy@scu.edu.cn (Y. Dai).

topological descriptors [34], and their number is still increasing to improve the accuracy of the model. Pardakhti et al. [33] applied multiple ML models to predict the methane adsorption properties of 130,398 MOFs, and they used structural descriptors such as void fraction, surface area, and density, as well as chemical descriptors such as degree of unsaturation, metallic percentage, and nitrogen to oxygen ratio. The result was that the random forest model achieved the highest determination coefficient (R^2) of 0.98. However, since NPMs have been widely utilized in gas storage [35], separation [36] and catalysis [37], researchers are interested in learning about a variety of their properties. Traditional workflows may not work well on the prediction of certain properties due to the limitation of the amount of training data.

Transfer learning (TL), as a branch of ML, can take advantage of “knowledge” learned in large datasets to migrate to small datasets, so TL was also used to perform property prediction of NPMs. Ma et al. [38] used H₂ adsorption data of 13,506 MOFs to train the neural network, and the prediction accuracy of the target task was effectively improved when they migrated to both H₂ adsorption data and CH₄ adsorption data at different temperatures. However, when using the adsorption data of the Xe/Kr (20:80) mixture as the target task for TL, TL could not improve the prediction accuracy of the model, and this difference should be attributed to the descriptors that play a dominant role in the two data sets are different, which eventually leads to the insufficient performance of TL. Therefore, using TL requires a certain degree of expertise to identify the transferable fields, and requires a large amount of manual attempts [38].

However, it is found that although the construction of the blocks brought great variety to the NPMs, it also brought the possible structure resemblance to the NPMs. Fig. 1 shows two different COFs with similar structures [39], which results in extremely similar absorption and diffusion performances. The specific details were listed in Table S1.

In previous studies, similarity has been used as the weight to predict the missing data in commercial recommendation systems [40]. It has also been applied to the property prediction of NPMs. Sturluson et al. [41] have built a COF recommendation system using Singular Value Decomposition (SVD) [42] to predict missing values in the NPM-property matrix. Good predictive results can be obtained from their COF recommendation system, but their model ignores features of materials, and overall, forecast accuracy is not high, and it also suffers from a cold-start problem [43], i.e. limited data will lead to inadequate extraction of similarities in the model, resulting in low accuracy of the model in predicting new NPMs or properties. Furthermore, the model of matrix decomposition is linear, and the absence of nonlinearity will result in a poor fitting ability of the model for complex relationships between materials and properties [44].

For cold-start problems, the experience of the movie recommendation system shows that adding features of users and movies can greatly improve the prediction accuracy of the model [45]. Similarly, in a material recommendation system, cold-start problems can be partially combatted by adding material features. Deep neural network (DNN) was employed to improve the non-linear fitting ability, which has shown good results in the fields of computer vision [46] and natural language processing [47].

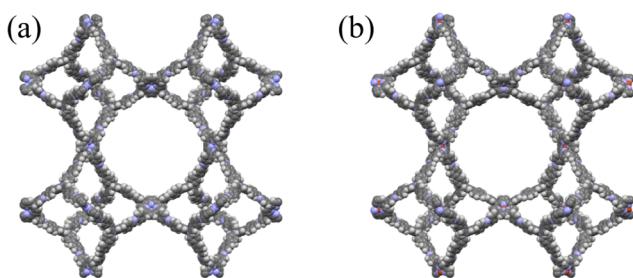


Fig. 1. Structures of (a) COF-1-tbo [39] and (b) COF-2-tbo [39].

Herein, a material recommendation system based on the DeepFM model was developed [48] which incorporates descriptors (e.g., porosity) to predict the various adsorption properties of hypothetical MOFs [49]. 28 adsorption properties of 8206 hypothetical MOFs were selected as a data set. Fortunately, unlike what may be encountered in practice, the MOFs material-property matrix in this work is complete, allowing us to study how imputation performance varies with the percentage of missing values. In addition, the performance of the model for solving cold-start problems was studied. Last but not least, an analysis of the relative importance of different descriptors on CO₂ capture capability of MOFs was conducted, which should be useful for the design of high-performance MOFs for CO₂ capture.

2. DeepFM MOF recommendation system

In this work, a material recommendation system based on DeepFM model is demonstrated, and a material recommendation system based on SVD model is implemented as a comparison model. From the perspective of model construction, the descriptors of MOFs were added into the input data as new features for DeepFM, and the cold-start problem is alleviated from the root. In addition, the DeepFM model contains DNN, which is employed to improve the non-linear fitting ability. Detailed information about these two models can be found in the supporting information.

2.1. Data set

The CO₂/N₂ adsorption property dataset of 8206 pre-screened hypothetical MOFs obtained from Materials Cloud [50] was employed to construct the input data of the MOF recommendation system. In principle, the recommendation system can find more similarities between properties when the quantity of properties is larger, which means that more properties will offer better accuracy. Thus, all the adsorption properties (28 properties) available in the dataset were selected, as shown in Table 1, including the adsorption capacity and adsorption heat of CO₂ and N₂ under various conditions. After the data processing mentioned above, 229,768 (8206 × 28) pieces of adsorption data were finally obtained. In addition, the data of each property was normalized separately for model training and data display.

Fig. 2 shows the scatter diagram matrix of 28 adsorption properties, which shows the distribution of adsorption properties (diagonal) and the pairwise relationship of adsorption properties (off-diagonal). Each point in the scatter diagram represents a MOF, and all adsorption properties are normalized. It can be seen that there are strong similarities between the distributions of some adsorption properties, such as property-1 and property-5. This data distribution is very similar to the movie dataset because the movie dataset also contains similar movies and dissimilar movies. Similar movies will be rated similarly by the same user, thus they will have similar data distribution of ratings. The role of SVD model and FM component in DeepFM is to automatically capture the correlation between these properties to predict missing values.

2.2. Features

As shown in Table 2, a number of descriptors were added to the data set as the new features, including 4 continuous descriptors such as specific surface area and porosity and 7 discrete descriptors such as functional groups and topological types. In terms of descriptor types, they include structural, chemical and topological descriptors. These descriptors are easy to calculate, and their calculation time is significantly shorter.

2.3. Evaluation metrics

Two evaluation metrics, namely the coefficient of determination (R^2) and the mean absolute error (MAE) were employed to evaluate the

Table 1

List of adsorption properties included in our MOF recommendation system.

propertyID	Property	Thermodynamic condition	Components	Units
1	CO ₂ uptake	0.15 bar, 298 K	single	mmol/g
2	CO ₂ uptake error	0.15 bar, 298 K	single	mmol/g
3	CO ₂ heat adsorption	0.15 bar, 298 K	single	kcal/mol
4	CO ₂ heat adsorption error	0.15 bar, 298 K	single	kcal/mol
5	CO ₂ excess uptake	0.15 bar, 298 K	single	mmol/g
6	CO ₂ uptake	0.10 bar, 363 K	single	mmol/g
7	CO ₂ uptake error	0.10 bar, 363 K	single	mmol/g
8	CO ₂ heat adsorption	0.10 bar, 363 K	single	kcal/mol
9	CO ₂ heat adsorption error	0.10 bar, 363 K	single	kcal/mol
10	CO ₂ excess uptake	0.10 bar, 363 K	single	mmol/g
11	CO ₂ uptake	0.70 bar, 413 K	single	mmol/g
12	CO ₂ uptake error	0.70 bar, 413 K	single	mmol/g
13	CO ₂ heat adsorption	0.70 bar, 413 K	single	kcal/mol
14	CO ₂ heat adsorption error	0.70 bar, 413 K	single	kcal/mol
15	CO ₂ excess uptake	0.70 bar, 413 K	single	mmol/g
16	Working capacity vacuum swing	–	single	mmol/g
17	Working capacity temperature swing	–	single	mmol/g
18	CO ₂ uptake	0.15 bar, 298 K	binary	mmol/g
19	CO ₂ uptake error	0.15 bar, 298 K	binary	mmol/g
20	CO ₂ heat adsorption	0.15 bar, 298 K	binary	kcal/mol
21	CO ₂ heat adsorption error	0.15 bar, 298 K	binary	kcal/mol
22	CO ₂ excess uptake	0.15 bar, 298 K	binary	mmol/g
23	N ₂ uptake	0.85 bar, 298 K	binary	mmol/g
24	N ₂ uptake error	0.85 bar, 298 K	binary	mmol/g
25	N ₂ heat adsorption	0.85 bar, 298 K	binary	kcal/mol
26	N ₂ heat adsorption error	0.85 bar, 298 K	binary	kcal/mol
27	N ₂ excess uptake	0.85 bar, 298 K	binary	mmol/g
28	CO ₂ /N ₂ selectivity	1 bar, 298 K	binary	–

performance of each model [35].

The calculation formula of R² is shown in eq (6), where m is the number of samples, y_i is the real value, $f(x_i)$ is the predicted value, \bar{y} is the average of the real values. The value of R² is between 0 and 1. The closer it is to 1(0), the better(worse) the prediction accuracy of the model. Of course, there is also a negative value, which means the accuracy of the model is very poor.

$$R^2 = 1 - \frac{\sum_{i=1}^m (y_i - f(x_i))^2}{\sum_{i=1}^m (y_i - \bar{y})^2} \quad (6)$$

R² is the most well-known metric, but sometimes it will fail to evaluate the model. Therefore, MAE was also added to better evaluate the model [51]. The calculation formula of MAE is shown in eq (7). MAE calculates the absolute value of the difference between the predicted value and the

real value of each sample, and then sums it and takes the average value. It is used to evaluate the closeness between the prediction results and the real data set. The lower MAE value means a better fitting effect of the model.

$$MAE = \frac{1}{m} \sum_{i=1}^m (|y_i - f(x_i)|) \quad (7)$$

3. Results and discussion

3.1. Model performances evaluation

3.1.1. Overall prediction performance

80 % of 229,768 data was randomly selected as the training set, on which the SVD model and the DeepFM model were trained, and the remaining 20 % of data were used as the test set to test the prediction accuracy of the two models. The hyper-parameters of these models are listed in Table S2 of SM. After three independent experiments, the results of the test set are shown in Fig. 3. The color of the points in the figure indicates the density. The warmer the color, the higher the density of the points in the area. It can be seen from the scatter diagram of SVD that there is a systematic deviation in the prediction results of the SVD model. When the test value is <0.4, it tends to overestimate the adsorption properties of MOFs, while when the test value is greater than 0.4, it tends to underestimate the adsorption properties. Moreover, the dense areas of the SVD diagram are scattered, indicating that the prediction accuracy of SVD is relatively low. In contrast, the points of DeepFM are distributed along the diagonal, indicating that there is no systematic deviation in DeepFM. In addition, the dense area of the DeepFM diagram is very narrow, which means that DeepFM has a very accurate prediction of most adsorption properties. From the perspective of the evaluation metrics, the R² of DeepFM reaches 0.91, which is much higher than that of SVD. Similarly, it can be seen from the MAE of the two models that the prediction error of DeepFM is much smaller than that of SVD.

3.1.2. Analysis of observed fraction

In order to study the influence of the observed fraction of material-property matrix on the prediction accuracy of the model, analysis experiments on different observed fractions were carried out, and the results are shown in Fig. 4. From the perspective of R², DeepFM is higher than SVD on any observed fraction, and the R² of DeepFM when the observed fraction is only 0.4 exceeds that of SVD when the observed fraction is 0.8, which means that DeepFM can achieve very good prediction performance when the material property matrix is sparse, which is impossible for SVD. The accurate prediction results of DeepFM should be attributed to the adding of the descriptors and the improved nonlinear fitting ability. In practical application, the experimental or simulated data are often collected from different work, which means that the materials and properties studied are not organized, so the material-property matrix will be relatively sparse. In this actual case, DeepFM will have a much better prediction effect than SVD. From the perspective of MAE, similar conclusions can be still drawn.

3.1.3. Comparative experiment of DeepOnly

In the previous description, it is found that the deep component is responsible for extracting high-order feature interaction, while the FM component automatically generates second-order feature combinations to make up for the lack of the model's ability to extract low-order feature interaction, so that the DeepFM model has a better ability to extract the correlation between features, that is, FM component improves the ability of DeepFM to extract the similarity between materials or properties. Reflected in the prediction results is to improve the prediction accuracy of the model. The FM component in the DeepFM model was removed and only keep the deep components to obtain the DeepOnly model. Then it was compared with the DeepFM in order to check

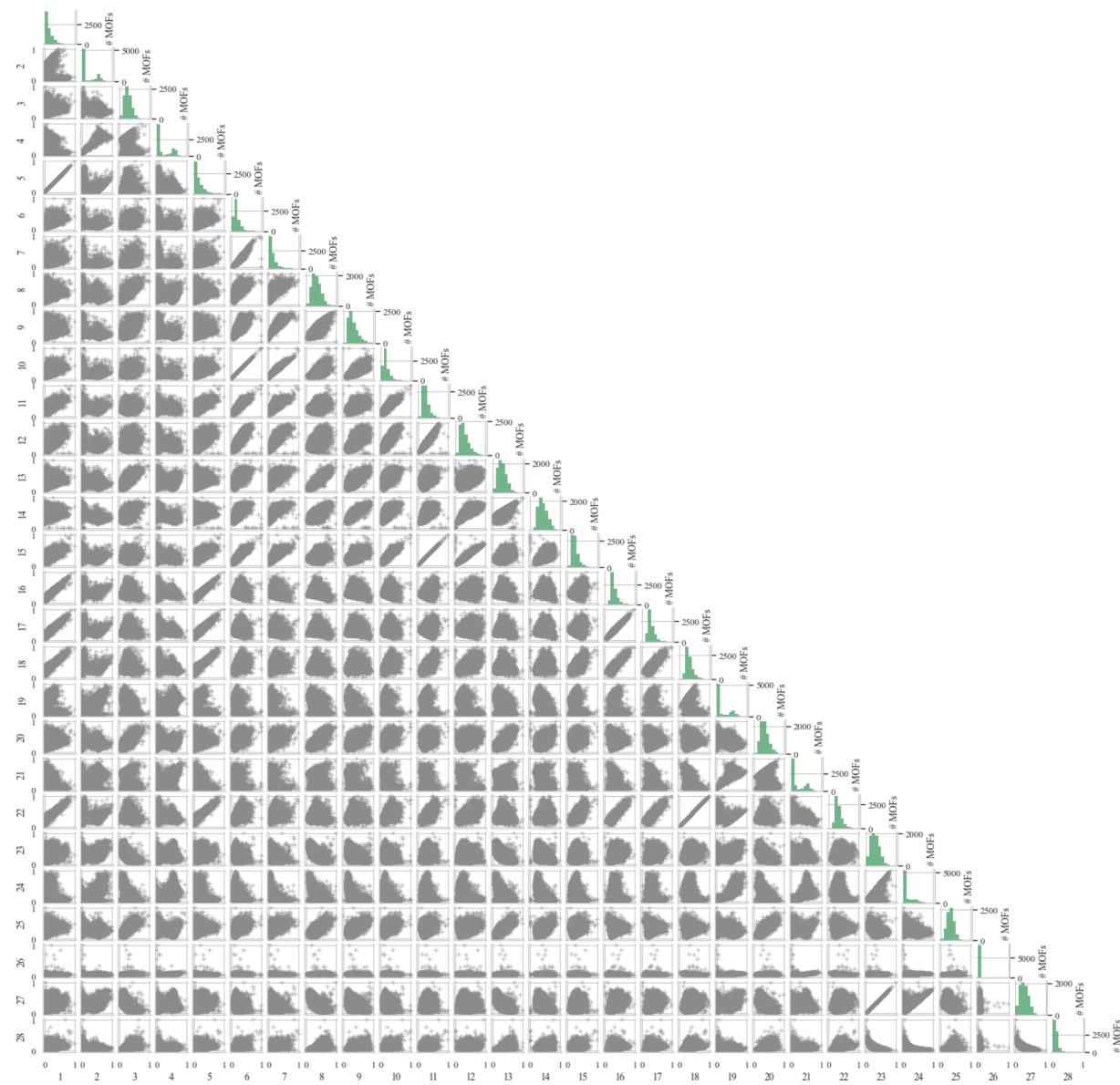


Fig. 2. Scatter matrix of 28 adsorption properties.

Table 2
List of features included in our MOF recommendation system.

Features	Descriptor	Feature	Units	Min.	Medium	Max.	Mean
surface area (SA)	structural	continuous	m^2/g	0	676	4118	774
porosity	structural	continuous	—	0	0.152	0.554	0.165
largest cavity diameter (LCD)	structural	continuous	\AA	3.39	5.33	18.80	5.61
pore limiting diameter (PLD)	structural	continuous	\AA	2.79	5.20	18.80	5.48
metal linker	chemical	discrete	—	1	3	12	5.10
organic linker1	chemical	discrete	—	1	10	41	11.39
organic linker2	chemical	discrete	—	1	18	41	19.82
functional group	chemical	discrete	—	—	—	—	—
topology	topological	discrete	—	—	—	—	—
MOFID	—	discrete	—	1	4103	8206	4103.5
propertyID	—	discrete	—	1	14	28	14.5

whether the FM component improved the performance of the model. The results are shown in Fig. 5. From the figure, it can be seen that although the points of DeepOnly are also distributed along the diagonal, its dense area (density) is slightly larger (smaller) than that of DeepFM, which means that DeepFM has a more accurate prediction of most adsorption properties. In addition, the R^2 and MAE value also show that

DeepFM presents better prediction, indicating that the addition of FM components does improve the prediction accuracy of the model.

3.1.4. Cold-start

In practical application, the cold-start sometimes can be a problem, that is, a new NPM may have only a small amount of property data, or a

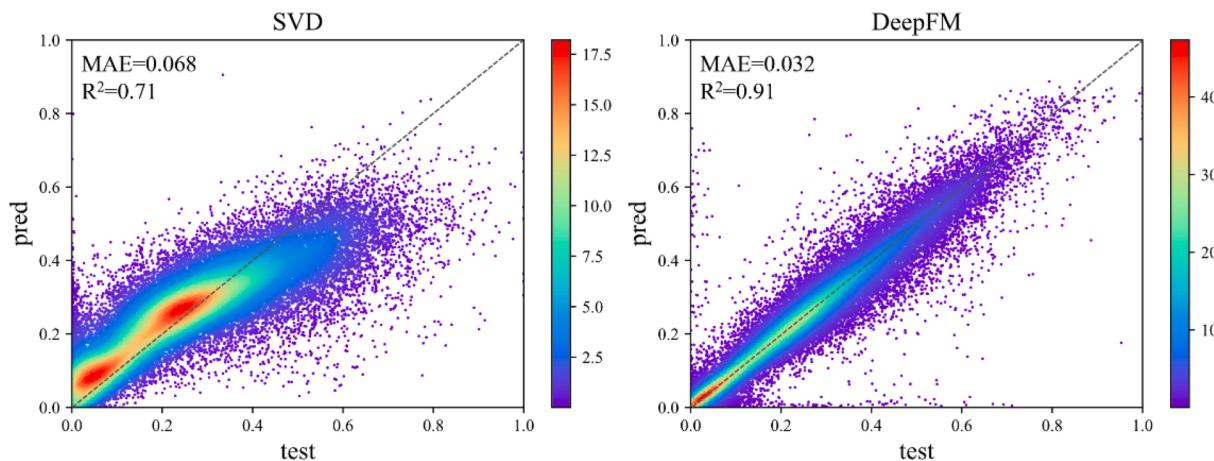


Fig. 3. Prediction results of SVD and DeepFM. The proportion of the training set is 80%, which means that the observed fraction of material-property matrix is 0.8. The diagonal represents the perfect prediction.

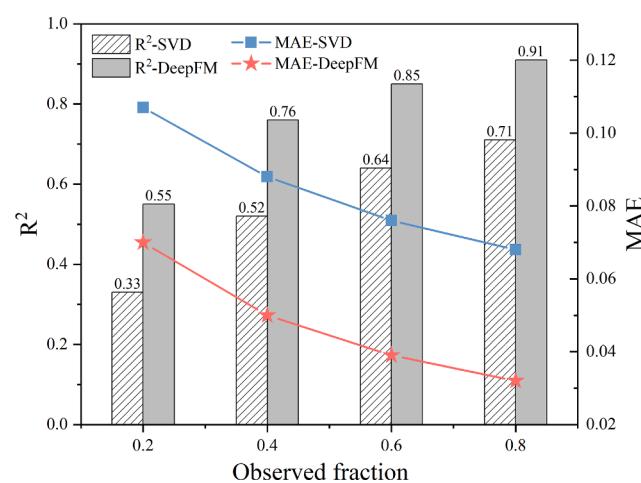


Fig. 4. Analysis of observed fraction.

new property may have only a small amount of NPMs data. Both situations may lead to the insufficient extraction of similarity in the model, which may result in low accuracy of the model in predicting new NPM or property. In this section, cold-start experiments on both SVD and DeepFM were carried out to evaluate the ability of these two models to combat cold-start problems.

As shown in Fig. 6, the data set was divided into two parts. Most data of a target property is extracted from the data set as the test set, the remaining data is used as the training and validation set. Therefore, for this target property, the model will face the problem of cold-start. In the training and validation set, 90 % of the data is used as the training model, and the remaining 10 % of the data is used to validate the model. Finally, the test set is used to check the prediction performance of the model for the target properties.

These experiments have been done on the first five adsorption properties, and the test set is 80 % of the data of the target properties. As shown in Fig. 7, the prediction accuracy of the DeepFM model is very good. Except for property-3, other properties investigated have achieved R^2 close to 1 and very low Mae, while the prediction accuracy of SVD for each property is still significantly lower than that of DeepFM. This shows that our DeepFM model with descriptors has strong cold-start resistance. The only exception is the prediction effect of the two models on property-3. As can be seen from the scatter diagram between property-3 and other properties in Fig. 2, the correlation between property-3 and other properties is very low, which is possibly the reason for low prediction accuracy.

In addition, since the retention of 20 % of the target property data has made DeepFM achieve excellent results, the target property data was further reduced to 10 %. The experimental results are shown in Fig. 8. It can be seen that even with a target property retained only to 10 %, the prediction accuracy of DeepFM is still in a high region (>95 % for four properties), demonstrating that DeepFM can accurately predict the

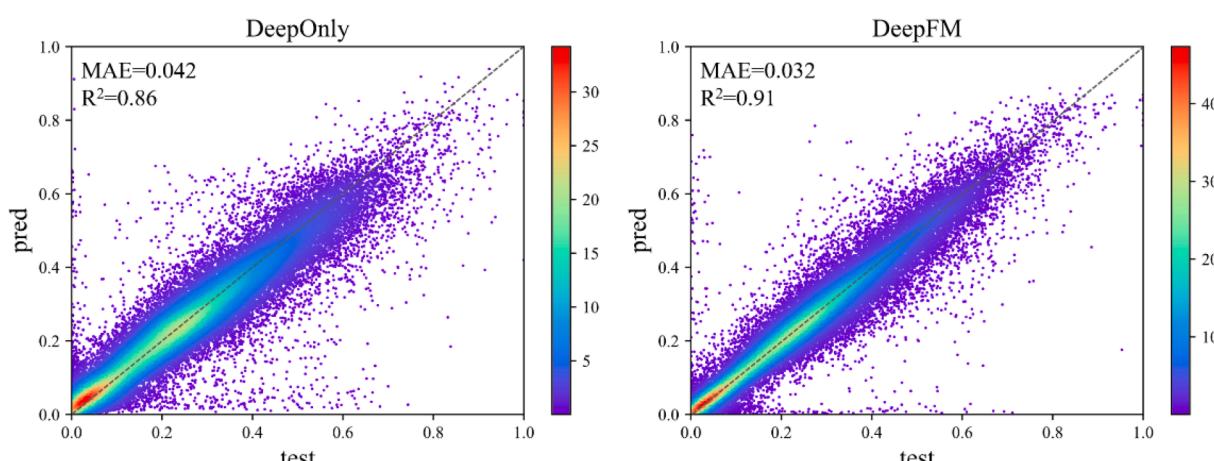


Fig. 5. Prediction results of DeepOnly and DeepFM. The observed fraction is 0.8.

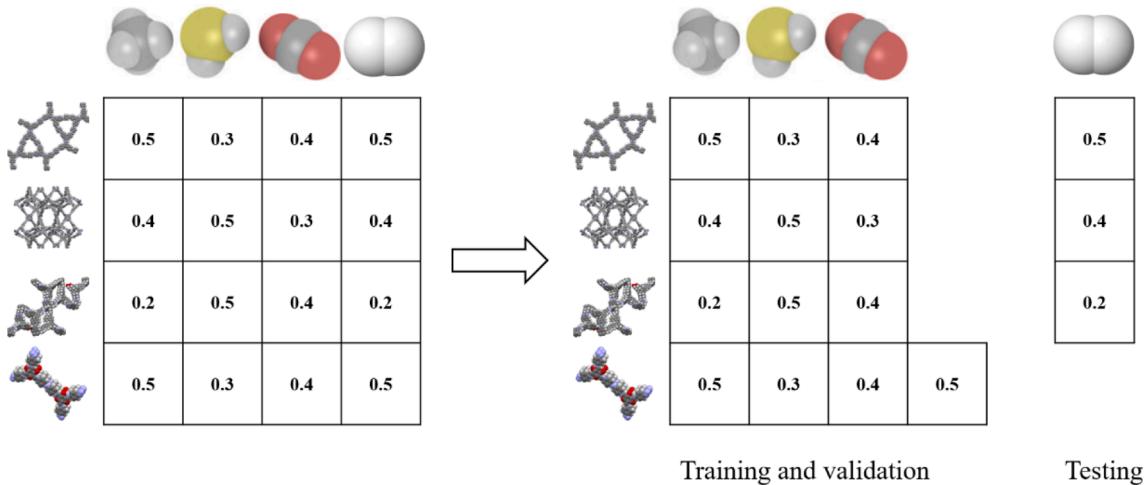


Fig. 6. Data division of cold-start experiments.

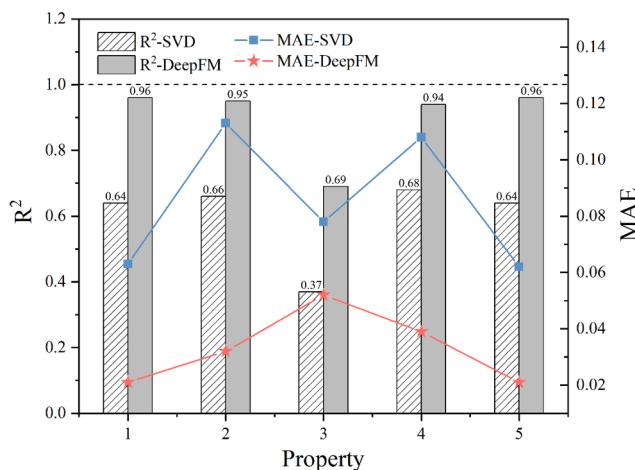


Fig. 7. Cold-start experimental results of the first five properties. The test set is 80% of the target property data. The dotted line indicates the upper limit of R^2 .

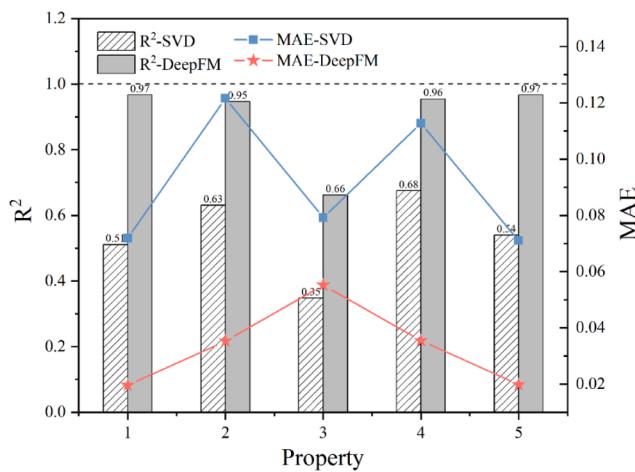


Fig. 8. Cold-start experimental results of the first five properties. The test set is 90% of the target property data. The dotted line indicates the upper limit of R^2 .

remaining target properties with only a small amount of target property data, that is, DeepFM has a strong ability to resist the cold-start problem.

3.1.5. Modeling with single property

Compared with the material recommendation system based on matrix decomposition [41], the material recommendation system based on DeepFM not only makes use of the similarity between materials or properties, but also integrates the descriptors of materials, so it achieves much better prediction accuracy. However, in theory, there is a possibility that very good prediction accuracy of NPMs properties can be obtained by using descriptors without similarities. Therefore, the performance of the model developed in this study may be attributed to the addition of descriptors. In order to explore this possibility and prove the superiority of using the material recommendation system to predict the properties of NPMs, in this section, only descriptors were used to model the properties.

The first five features were modeled separately. DeepFM was used to control the variables. Similar to previous sections, 80 % of data was used for training while the remaining 20 % was for testing. R^2 and MAE of the test set are shown in Fig. 9. It can be seen that except for property-3, the accuracy of separate modeling of other properties is much lower than that of DeepFM in the cold-start simulation. It also shows that the addition of descriptors does not play a decisive role in the prediction of these properties. However, for property-3, its single property modeling accuracy is very close to that of DeepFM in the cold-start simulation.

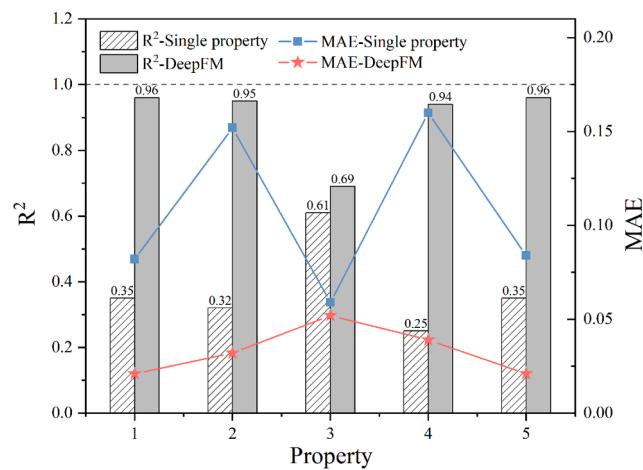


Fig. 9. The single property modeling results of the first five properties. The dotted line indicates the upper limit of R^2 .

which may indicate that descriptors are very important for the prediction of this property.

3.1.6. Analysis of Property-3

In the previous experiments, property-3 shows different prediction results from several other properties. In this section, the previous experimental results about property-3 were sorted out. As shown in Table 3, in the test of SVD based on similarity, the prediction performance of property-3 is significantly lower than that of other properties, which is due to the lack of high similarity between property-3 and other properties. The scatter diagram in Fig. 2 can support this view. When using descriptors to model single properties separately, the prediction performance of property-3 is significantly higher than that of other properties, indicating that descriptors have a strong positive effect on the prediction of property-3. However, in the cold-start experiment of DeepFM which considers similarity and descriptors at the same time, the prediction performance of property-3 is significantly lower than that of other properties.

The possible explanation is that although the DeepFM model considers both similarity and descriptors, it is controlled by similarity, which means that similarity is the cornerstone, and the descriptors further improve the prediction accuracy on this basis. Therefore, when the similarity between the target property and other properties is lacking, although the descriptors can better predict the target property, DeepFM is still unable to make a satisfactory prediction. The DeepFM model has been applied in several different fields, such as disease prediction [52], stock market prediction [53] and taxi pick-up area recommendation [54]. However, few works pay attention to the evaluation of cold-start resistance of DeepFM. Consequently, there is still no literature to verify our conjecture so far.

3.2. Predicting high-performance MOFs

3.2.1. CO₂ uptake and selectivity

The uptake of CO₂ (Property-18) and N₂ (Property-23) of a CO₂/N₂ mixture (3:17 mol ratio) at 298 K and 1 bar in these MOFs were predicted by the MOF recommendation system. As shown in Fig. 10, DeepFM has made a very accurate prediction of the uptake of the two gases, which could be attributed to the high similarity between these two properties and other properties. In addition, 90 of the top 100 MOFs with the highest CO₂ uptake were correctly predicted by DeepFM.

In the adsorption separation processes, high selectivity is strongly desired because highly selective adsorbents lead to high product purity as well as low energy costs. Therefore, the CO₂/N₂ selectivity (Property-28) of a CO₂/N₂ mixture at 298 K and 1 bar in these MOFs were also predicted by DeepFM. As shown in Fig. 11, when using DeepFM to predict CO₂/N₂ selectivity, due to the lack of similarity with other properties, the selectivity prediction was not very accurate. R² could only reach 0.69, and only 45 of the top 100 MOFs with the highest selectivity were predicted correctly. However, when the uptake of CO₂ (Property-18) and N₂ (Property-23) predicted by DeepFM were used to calculate the CO₂/N₂ selectivity, the predicted results were in good agreement with the experimental value, R² value of 0.95 can be obtained, and 80 of the top 100 MOFs with the highest selectivity were predicted correctly. These results show that when a property lacks

similarity with other properties, it is more effective to predict it indirectly, if feasible, using other properties that could be predicted more accurately.

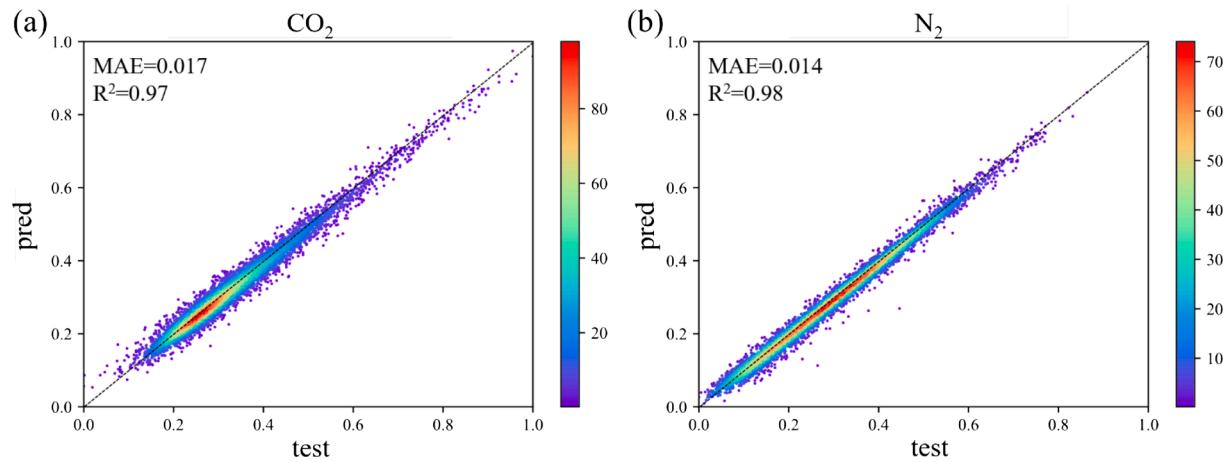
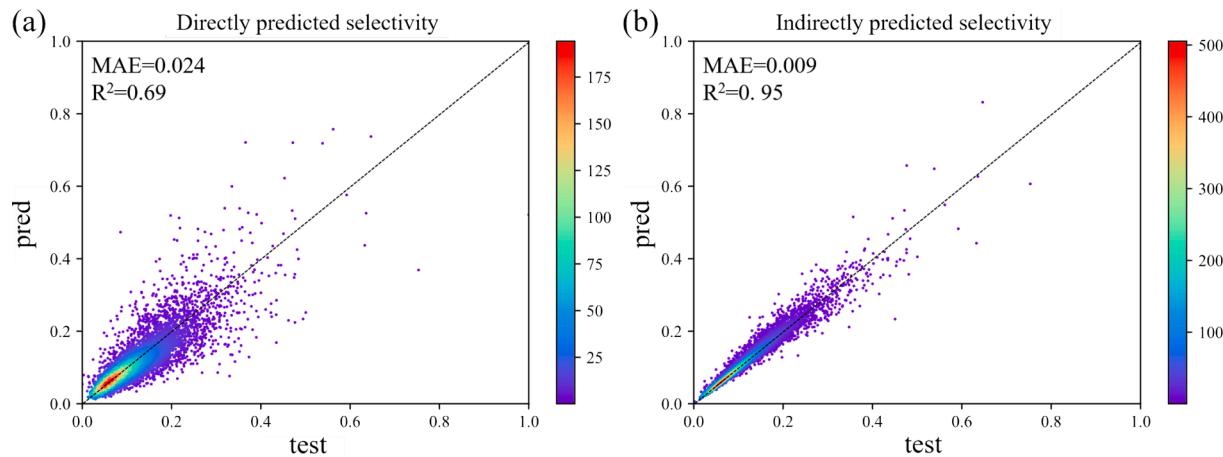
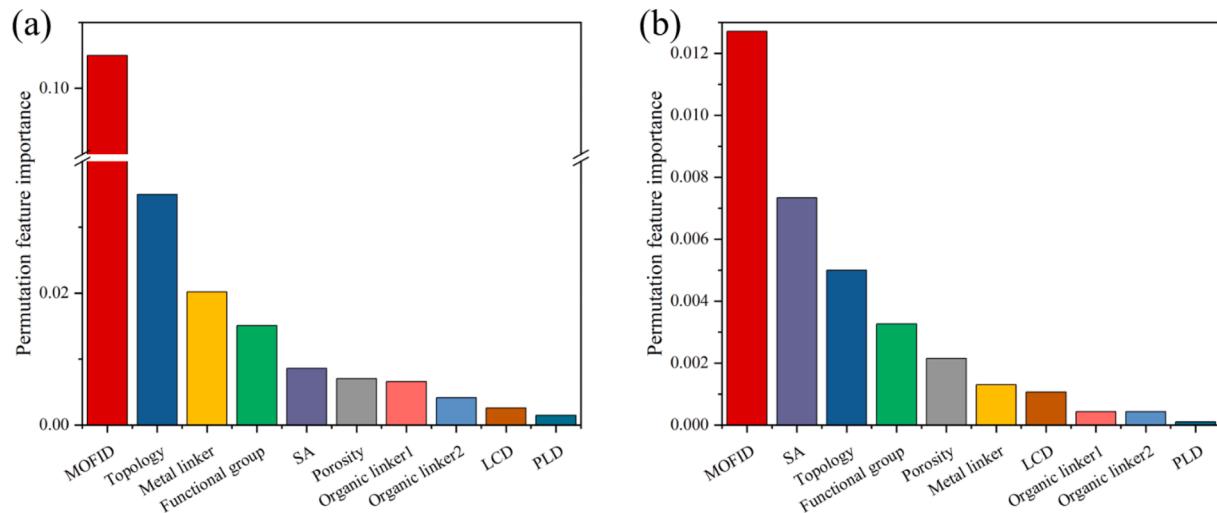
3.2.2. Importance of descriptors

The above results show that DeepFM can be used to predict adsorption properties using simple, intuitive descriptors, but another focus of this work is on understanding the relative importance of these descriptors on the CO₂ capture capabilities of MOFs. To this end, by calculating permutation feature importance based on the DeepFM models trained for the CO₂ uptake and CO₂/N₂ selectivity [55], the relative importance of descriptors of the studied MOFs can be determined. As shown in Fig. 12, since the MOFID identifies the MOF to be predicted, it is foreseeable that MOFID is extremely crucial. Excluding the MOFID, despite the differences in ranking, the top 5 important descriptors for CO₂ uptake and CO₂/N₂ selectivity are all the same, namely topology, metal linker, functional group, SA and porosity. Notably, consistent with the conclusions of multiple experimental and simulation works, topology is of relatively high importance for both CO₂ uptake and CO₂/N₂ selectivity, which should be attributed to the topology “encodes” a lot of information about the structure (e.g., pore shape and pore polydispersity) [56–61]. As chemical descriptors, metal linker and functional group play a crucial role in affecting the affinity of the MOFs towards CO₂, which indicate that the selection of proper metal ions and metal groups are of great significance for CO₂ binding [57,62]. Besides, both SA and porosity have long been recognized as important structural descriptors influencing CO₂ adsorption [33,63]. In the experimental study, SA was found to have a strong positive correlation with CO₂ uptake rather than CO₂/N₂ selectivity [64], but it emerged as the most important descriptor for CO₂/N₂ selectivity. This abnormal result may be attributed to the presence of a positive correlation between SA and pore size (shown in Figure S6), while pore size is well recognized as an important factor affecting selectivity [65]. According to the literature, many MOFs’ molecular sieving effects are exploited to adsorb a variety of gas molecules that imply only those molecules can crossover the pores with diameters well-matched to the pore window [66–70]. And the insignificant importance of pore size (PLD and LCD) should be attributed to the information it contains has been captured by SA.

Fig. 13 presents the effects of 5 important descriptors on the CO₂/N₂ separation performances of MOFs, in which only the top 10 out of 399 functional groups that can simultaneously lead to high CO₂ uptake and CO₂/N₂ selectivity are exhibited. The structures corresponding to the number of the metal linkers are listed in Figure S7. From Fig. 13, it can be seen that high-performance MOFs have a clear preference for five important descriptors. For the topology, which is the most important descriptor, the 100 MOFs with the highest CO₂/N₂ selectivity are all derived from MOFs with sra, pcu or acs topology. Figure S8 shows that although their average SA and porosity are not outstanding, their average pore size (PLD and LCD) is the first three smallest, so appropriate pore size results in higher CO₂/N₂ selectivity of MOFs with sra, pcu or acs topology. With regard to metal linkers, MOFs having metal linker-5 with vanadium ions tend to have higher CO₂ uptake and CO₂/N₂ selectivity. According to the literature [49], it should be attributed to CO₂ binds selectively near the bridging oxygen of the metal linker-5. For functional groups, because CO₂ has a high quadrupole moment, in general, more strongly polarizing groups will influence CO₂ adsorption more favorably. However, weakly polarizing groups such as -H, -Me and -Cl also show a beneficial effect on CO₂ adsorption, which may be due to the variability of adsorption mechanisms. Anderson et al. [57] found through simulation that for the MOFs with these functional groups, CO₂ preferentially adsorbs at the nodes of the framework. There, the linkers “converge”, and the overlap of potentials creates adsorption sites (π pockets) stronger than the sites adjacent to the functional groups [71]. With regard to SA, similar to topology, since smaller SA tends to appear with smaller pore size (PLD and LCD), MOFs with the highest CO₂/N₂ selectivity were all found to have SA < 1000 m²/g. Finally, with regard

Table 3
List of R² in previous experiments.

Principle	SVD Similarity	Single property Descriptors	DeepFM Similarity + Descriptors
R ² of Property-3	0.37	0.61	0.69
R ² of Property-1	0.64	0.35	0.96
R ² of Property-2	0.66	0.32	0.95
R ² of Property-4	0.68	0.25	0.94
R ² of Property-5	0.64	0.35	0.96

Fig. 10. Prediction results of (a) CO₂ and (b) N₂ uptake.Fig. 11. Prediction results of CO₂/N₂ selectivity (a) directly and (b) indirectly.Fig. 12. Relative importance of material descriptors for the (a) CO₂ uptake and (b) CO₂/N₂ selectivity obtained from DeepFM training.

to porosity, MOFs with porosity between 0 and 0.1 tend to have low CO₂ uptake and high CO₂/N₂ selectivity, while those with porosity between 0.2 and 0.3 tend to have high CO₂ adsorption and low CO₂/N₂ selectivity. Owing to the trade-off between CO₂ uptake and CO₂/N₂ selectivity, MOFs with the highest performance were all found to have the

porosity < 0.2. It may be attributed to that MOFs with low porosity tend to have smaller cages and channels, which often provide relatively stronger van der Waals interaction since the gas molecules can interact with multiple "surfaces" [72].

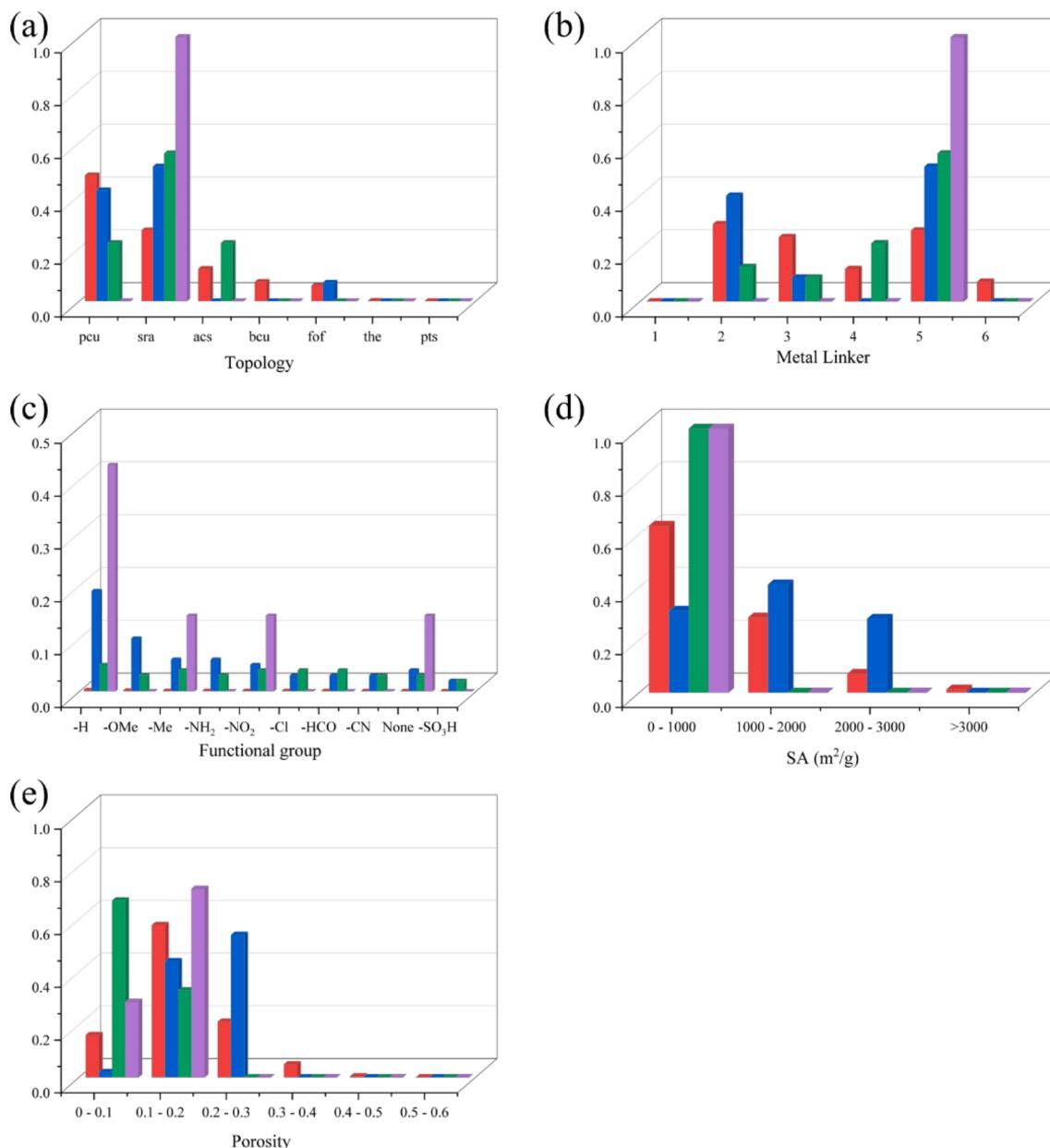


Fig. 13. Effects of (a) topology, (b) metal linker, (c) functional group, (d) SA and (e) porosity on the CO₂/N₂ separation performances of MOFs. Red columns represent all 8206 MOFs, blue columns represent the top 100 MOFs with the highest uptake of CO₂, green columns represent the top 100 MOFs with the highest CO₂/N₂ selectivity, and purple columns represent the intersection (7 MOFs) of the top 100 MOFs with the highest uptake of CO₂ and the top 100 MOFs with the highest CO₂/N₂ selectivity.

4. Conclusions

Material-property matrices are useful tools in developing new CO₂ adsorption materials. However, many times, there are a lot of missing values in the corresponding material-property matrix. In this paper, the DeepFM model is employed to predict the missing values to help determine the optimal material for CO₂ adsorption. The material recommendation system based on DeepFM not only takes into account the correlation between materials and properties, but also integrates descriptors that have been proved to play an important role in predicting the missing values [26–30].

In this system, the input data was constructed by the 28 adsorption properties of over 8000 hypothetical MOFs obtained from materials cloud [50]. Then, the prediction performances of the DeepFM model and matrix decomposition (SVD) were evaluated under different observed fractions of the material-property matrix. The results showed that

DeepFM shows much higher prediction accuracy than SVD at observed fractions from 0.2 to 0.8. In addition, the DeepFM model also presents better cold-start resistance and can conduct very accurate predictions when there is only a small amount of available data. What's more, while the DeepFM model simultaneously takes both similarity and descriptors into account, similarity may still be the primary factor. Finally, the relative importance of various descriptors on the CO₂ capture capabilities of MOFs was analyzed, which can be helpful in designing new MOFs with high CO₂ capture performances.

In the future, the features of properties (e.g., kinetic diameters of gas molecules) can also be added to the MOF recommendation system to further improve the prediction accuracy. In addition, other properties (e.g., diffusion properties) could also be introduced to extend the application of this system. However, more descriptors related to new properties may be required. We surmise that as more and more properties are involved, more underlying similarities between properties would be

captured by DeepFM, which could be helpful to improve the accuracy of the system. In addition, as the negative effect of water vapor on CO₂ adsorption processes is of critical importance for successful CO₂ adsorption application, in the future, more effort can be also dedicated to develop MOFs with good water vapor resistance.

CRediT authorship contribution statement

Minggao Feng: Conceptualization, Methodology, Software, Writing – original draft. **Min Cheng:** Writing – review & editing. **Xu Ji:** Resources. **Li Zhou:** Resources. **Yagu Dang:** Resources. **Kexin Bi:** Formal analysis. **Zhongde Dai:** Supervision, Writing – review & editing. **Yiyang Dai:** Funding acquisition, Project administration, Supervision.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

Data will be made available on request.

Acknowledgements

The authors gratefully acknowledge financial support from the National Key Research and Development Program of China (2021YFB4000505), National Natural Science Foundation of China (Grant No. 52170112) Sichuan Science and Technology Program (2021YFS0301, and 2021YFH0116).

Appendix A. Supplementary material

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.seppur.2022.122111>.

References

- [1] Z. Zhang, et al., Advances in carbon capture, utilization and storage, *Appl. Energy* 278 (2020).
- [2] Á.A. Ramírez-Santos, C. Castel, E. Favre, A review of gas separation technologies within emission reduction programs in the iron and steel sector: Current application and development perspectives, *Sep. Purif. Technol.* 194 (2018) 425–442.
- [3] H. Wang, et al., Ion exchange membrane related processes towards carbon capture, utilization and storage: current trends and perspectives, *Sep. Purif. Technol.* (2022), 121390.
- [4] A.G. Slater, A.I. Cooper, Porous materials. Function-led design of new porous materials, *Science* 348 (6238) (2015) p. aaa8075.
- [5] S.R. Venna, M.A. Carreón, Metal organic framework membranes for carbon dioxide separation, *Chem. Eng. Sci.* 124 (2015) 3–19.
- [6] E. Adatoz, A.K. Avci, S. Keskin, Opportunities and challenges of MOF-based membranes in gas separations, *Sep. Purif. Technol.* 152 (2015) 207–237.
- [7] M. Vinoba, et al., Recent progress of fillers in mixed matrix membranes for CO₂ separation: a review, *Sep. Purif. Technol.* 188 (2017) 431–450.
- [8] K. Sumida, et al., Carbon dioxide capture in metal-organic frameworks, *Chem. Rev.* 112 (2) (2012) 724–781.
- [9] S.B. Reddy, M., et al., *Carbon dioxide adsorption based on porous materials*. RSC, Advances 11 (21) (2021) 12658–12681.
- [10] R.-S. Liu, et al., Advances in post-combustion CO₂ capture by physical adsorption: from materials innovation to separation, *Practice*. 14 (6) (2021) 1428–1471.
- [11] G. Li, et al., Capture of CO₂ from high humidity flue gas by vacuum swing adsorption with zeolite 13X, *Adsorption* 14 (2) (2008) 415–422.
- [12] H.C. Zhou, J.R. Long, O.M. Yaghi, Introduction to metal-organic frameworks, *Chem. Rev.* 112 (2) (2012) 673–674.
- [13] M.S. Lohse, T. Bein, Covalent organic frameworks: structures, synthesis, and applications, *Adv. Funct. Mater.* 28 (33) (2018).
- [14] A.I. Cooper, Porous molecular solids and liquids, *ACS Cent. Sci.* 3 (6) (2017) 544–553.
- [15] W. Lu, et al., Porous polymer networks: synthesis, porosity, and applications in gas storage/separation, *Chem. Mater.* 22 (21) (2010) 5964–5972.
- [16] D.J. Tranchemontagne, et al., Reticular chemistry of metal-organic polyhedra, *Angew. Chem. Int. Ed.* 47 (28) (2008) 5136–5147.
- [17] P.G. Boyd, Y. Lee, B. Smit, Computational development of the nanoporous materials genome, *Nat. Rev. Mater.* 2 (8) (2017) 1–15.
- [18] Q. Wang, D. Astruc, State of the art and prospects in metal-organic framework (MOF)-based and MOF-derived nanocatalysis, *Chem. Rev.* 120 (2) (2020) 1438–1511.
- [19] R. Freund, et al., The Current Status of MOF and COF Applications. *60*(45) (2021) 23975–24001.
- [20] A. Sharma, et al., Computational materials chemistry for carbon capture using porous materials, *J. Phys. D Appl. Phys.* 50 (46) (2017), 463002.
- [21] M. Feng, et al., High-throughput computational screening of Covalent– Organic framework membranes for helium purification. *Results in Engineering*, 2022: p. 100538.
- [22] M. Cheng, et al., High-throughput virtual screening of metal-organic frameworks for xenon recovery from exhaled anesthetic gas mixture, *Chem. Eng. J.* (2022), 138218.
- [23] H. Daglar, S. Aydin, S. Keskin, MOF-based MMMs breaking the upper bounds of polymers for a large variety of gas separations, *Sep. Purif. Technol.* 281 (2022), 119811.
- [24] S. Wang, et al., High-throughput screening of metal-organic frameworks for hydrogen purification, *Chem. Eng. J.* (2022), 138436.
- [25] K.M. Jablonka, et al., Big-data science in porous materials: materials genomics and machine learning, *Chem. Rev.* 120 (16) (2020) 8066–8129.
- [26] K. Mukherjee, Y.J. Colón, Machine learning and descriptor selection for the computational discovery of metal-organic frameworks, *Mol. Simul.* 47 (10–11) (2021) 857–877.
- [27] Z. Shi, et al., Machine-learning-assisted high-throughput computational screening of high performance metal-organic frameworks, *Mol. Syst. Des.* 5 (4) (2020) 725–742.
- [28] Janet, J.P. and H.J. Kulik, *Machine Learning in Chemistry*. 2020: American Chemical Society.
- [29] S.M. Moosavi, et al., Understanding the diversity of the metal-organic framework ecosystem, *Nat. Commun.* 11 (1) (2020) 4068.
- [30] S.M. Moosavi, K.M. Jablonka, B. Smit, The role of machine learning in the understanding and design of materials, *J. Am. Chem. Soc.* (2020).
- [31] K.N. Pai, V. Prasad, A. Rajendran, Experimentally validated machine learning frameworks for accelerated prediction of cyclic steady state and optimization of pressure swing adsorption processes, *Sep. Purif. Technol.* 241 (2020), 116651.
- [32] H. Tang, et al., Rapid screening of metal-organic frameworks for propane/ propylene separation by synergizing molecular simulation and machine learning, *ACS Appl. Mater. Interfaces* 13 (45) (2021) 53454–53467.
- [33] M. Pardakhti, et al., Machine learning using combined structural and chemical descriptors for prediction of methane adsorption performance of metal organic frameworks (MOFs), *ACS Comb. Sci.* 19 (10) (2017) 640–645.
- [34] P.Z. Moghadam, et al., Structure-mechanical stability relations of metal-organic frameworks via machine learning, *Matter* 1 (1) (2019) 219–234.
- [35] A. Schoedel, Z. Ji, O.M. Yaghi, The role of metal-organic frameworks in a carbon-neutral energy cycle, *Nat. Energy* 1 (4) (2016) 1–13.
- [36] J.-R. Li, R.J. Kuppler, H.-C. Zhou, Selective gas adsorption and separation in metal-organic frameworks, *Chem. Soc. Rev.* 38 (5) (2009) 1477–1504.
- [37] D. Farrusseng, S. Aguado, C. Pinel, Metal-organic frameworks: opportunities for catalysis, *Angew. Chem. Int. Ed.* 48 (41) (2009) 7502–7513.
- [38] R. Ma, Y.J. Colon, T. Luo, Transfer learning study of gas adsorption in metal-organic frameworks, *ACS Appl. Mater. Interfaces* 12 (30) (2020) 34041–34048.
- [39] X. Kang, et al., Reticular synthesis of two topology covalent organic frameworks, *J. Am. Chem. Soc.* 142 (38) (2020) 16346–16356.
- [40] J.L. Herlocker, J.A. Konstan, J. Riedl, Explaining collaborative filtering recommendations. In Proceedings of the 2000 ACM conference on Computer supported cooperative work, 2000.
- [41] A. Sturluson, et al., Recommendation system to predict missing adsorption properties of nanoporous materials, *Chem. Mater.* 33 (18) (2021) 7203–7216.
- [42] H. Polat, W. Du, SVD-based collaborative filtering with privacy. Proceedings of the 2005 ACM symposium on Applied computing, 2005.
- [43] X.N. Lam, et al., Addressing cold-start problem in recommendation systems. Proceedings of the 2nd international conference on Ubiquitous information management and communication, 2008.
- [44] N.D. Lawrence, R. Urtasun, Non-linear matrix factorization with Gaussian processes. Proceedings of the 26th annual international conference on machine learning, 2009.
- [45] B. Liká, K. Kolomvatsos, S. Hadjicostyliades, Facing the cold start problem in recommender systems, *Expert Syst. Appl.* 41 (4) (2014) 2065–2073.
- [46] A. Voulodimos, et al., Deep learning for computer vision: a brief review, *Comput. Intell.* (2018).
- [47] E. Cambria, B. White, Jumping NLP curves: a review of natural language processing research, *IEEE Comput. Intell. Mag.* 9 (2) (2014) 48–57.
- [48] H. Guo, et al., DeepFM: a factorization-machine based neural network for CTR prediction. arXiv preprint arXiv:04247, 2017.
- [49] P.G. Boyd, et al., Data-driven design of metal-organic frameworks for wet flue gas CO₂ capture, *Nature* 576 (7786) (2019) 253–256.
- [50] L. Talirz, et al., Materials Cloud, a platform for open computational science. *Scientific data* 7(1) (2020) 1–12.
- [51] T.O. Kvålseth, Cautionary note about R 2, *Am. Statist.* 39 (4) (1985) 279–285.
- [52] Z. Yu, et al., Research on disease prediction based on improved DeepFM and IoMT, *IEEE Access* 9 (2021) 39043–39054.
- [53] J. Huang, X. Zhang, B. Fang, Costock: a DeepFM model for stock market prediction with attentional embeddings, *IEEE*, 2019.

- [54] X. Wang, et al., DeepFM-Based Taxi Pick-Up Area Recommendation. International Conference on Pattern Recognition, Springer, 2021.
- [55] A. Altmann, et al., Permutation importance: a corrected feature importance measure, *Bioinformatics* 26 (10) (2010) 1340–1347.
- [56] V. Guillerm, et al., A supermolecular building approach for the design and construction of metal–organic frameworks, *Chem. Soc. Rev.* 43 (16) (2014) 6141–6172.
- [57] R. Anderson, et al., Role of pore chemistry and topology in the CO₂ capture capabilities of MOFs: from molecular simulation to machine learning, *Chem. Mater.* 30 (18) (2018) 6325–6337.
- [58] M.-H. Yu, et al., Construction of a multi-cage-based MOF with a unique network for efficient CO₂ capture, *ACS Appl. Mater. Interfaces* 9 (31) (2017) 26177–26183.
- [59] P.M. Bhatt, et al., Isoreticular rare earth fcu-MOFs for the selective removal of H₂S from CO₂ containing gases, *Chem. Eng. J.* 324 (2017) 392–396.
- [60] D.-X. Xue, et al., Tunable rare-earth fcu-MOFs: a platform for systematic enhancement of CO₂ adsorption energetics and uptake, *J. Am. Chem. Soc.* 135 (20) (2013) 7660–7667.
- [61] L. Zhang, et al., Construction of hexanuclear Ce (III) metal–porphyrin frameworks through linker induce strategy for CO₂ capture and conversion, *Catal. Today* 374 (2021) 38–43.
- [62] Y. Liu, Z.U. Wang, H.C. Zhou, Recent advances in carbon dioxide capture with metal–organic frameworks, *Greenhouse Gases: Science Technology* 2 (4) (2012) 239–259.
- [63] R. Ma, Y.J. Colon, T. Luo, Transfer learning study of gas adsorption in metal–organic frameworks, *ACS Appl. Mater. Interfaces* 12 (30) (2020) 34041–34048.
- [64] A.R. Millward, O.M. Yaghi, Metal– organic frameworks with exceptionally high capacity for storage of carbon dioxide at room temperature, *J. Am. Chem. Soc.* 127 (51) (2005) 17998–17999.
- [65] Y. Li, R.T. Yang, Gas adsorption and storage in metal– organic framework MOF-177, *Langmuir* 23 (26) (2007) 12937–12944.
- [66] W. Zhuang, et al., Robust metal–organic framework with an octatopic ligand for gas adsorption and separation: combined characterization by experiments and molecular simulation, *Chem. Mater.* 24 (1) (2012) 18–25.
- [67] B. Chen, et al., Rationally designed micropores within a metal– organic framework for selective sorption of gas molecules, *Inorg. Chem.* 46 (4) (2007) 1233–1236.
- [68] D.N. Dybtsev, et al., Microporous manganese formate: a simple metal– organic porous material with high framework stability and highly selective gas sorption properties, *J. Am. Chem. Soc.* 126 (1) (2004) 32–33.
- [69] T. Loiseau, et al., MIL-96, a porous aluminum trimesate 3D structure constructed from a hexagonal network of 18-membered rings and μ 3-oxo-centered trinuclear units, *J. Am. Chem. Soc.* 128 (31) (2006) 10223–10230.
- [70] M. Xue, et al., Robust metal– organic framework enforced by triple-framework interpenetration exhibiting high H₂ storage density, *Inorg. Chem.* 47 (15) (2008) 6825–6828.
- [71] S. Chen, et al., Understanding the fascinating origins of CO₂ adsorption and dynamics in MOFs, *Chem. Mater.* 28 (16) (2016) 5829–5846.
- [72] H. Wu, et al., Metal–organic frameworks with exceptionally high methane uptake: where and how is methane stored? *Chem.–A Eur. J.* 16 (17) (2010) 5205–5214.