Microporous and Mesoporous Materials 231 (2016) 82–91

Contents lists available at ScienceDirect

Microporous and Mesoporous Materials

journal homepage: www.elsevier.com/locate/micromeso

The rational synthetic parameter analysis for subclasses of microporous aluminophosphates based on hierarchical feature selection model

Yuting Guo [a], Zhenhua Tian [b], Na Gao [c, d], Miao Qi [a], Jianzhong Wang [a, *]

[a] College of Computer Science and Information Technology, Northeast Normal University, Changchun 130117, China
[b] Changchun Institute of Optics, Fine Mechanics and Physics, Chinese Academy of Sciences, Changchun 130033, China
[c] State Key Laboratory of Inorganic Synthesis and Preparative Chemistry, Changchun 130012, China
[d] Institute of Modern Agriculture, Jilin Economic Management Cadre College, Changchun 130012, China

ARTICLE INFO

Article history:
Received 25 January 2016
Received in revised form
20 April 2016
Accepted 14 May 2016
Available online 14 May 2016

Keywords:
AlPOs database
Aluminophosphates
Subclass
Rational synthesis
Feature selection

ABSTRACT

Open-framework aluminophosphates (AlPOs) is an important family of porous crystal materials. But the synthetic chemistry of this kind of materials is very complicated, and the synthesis mechanism has not been clearly understood yet. In this paper, we propose a Hierarchical Feature Selection Model (HFSM) composed of two layers to analyze the rational synthetic parameters for the subclass of microporous aluminophosphates (AlPOs) containing (6,8)-rings. In the first layer, we select a feature subset that could separate the (6,8)-ring-containing microporous AlPOs from other AlPOs. In the second layer, we further analyze which of these selected features are critical for the formation of each special subclass in (6,8)-ring-containing microporous AlPOs. With the optimal feature subset selected by the proposed model, we can obtain the highest accuracy rates as 94.28%, 94.03%, 91.27% and 92.20% for the classification of AEN, AWO, CHA and ERI, respectively. Extensive analysis is presented for the synthetic parameters selected by the hierarchical model, which could provide a useful guidance to the rational synthesis of such materials.

1. Introduction

Zeolites and related microporous materials have been widely applied in petroleum industry for catalysis, separation and ion-exchange [1,2]. These materials are formed by TO_4 tetrahedra (T infers to Si, P, Al, Ge, Ga, etc.) with a well defined regular pore system. As an important member of Zeolites and related microporous materials, open-framework aluminophosphate molecular sieve has caught much attention for the past 20 years [3–7]. However the process for synthesizing such materials is complicated and influenced by many parameters, such as gel composition, PH value, solvent, template etc. In order to provide guidance to rational synthesis of microporous inorganic materials, the group of State Key Laboratory of Inorganic Synthesis and Preparative Chemistry of Jilin University established AlPOs synthesis database internationally [8]. This database contains about 1700 synthetic records which belong to 17 classes according to the sizes of the pore rings.

Data mining is the transformation bridge from data and information to knowledge. With the rapid development of computer technology and artificial intelligence, data mining plays an important role in more and more application fields. In chemical researches, data mining has been widely applied to the original data processing and retrieval [9–11], statistical analysis [12–14] and parameterization of the molecular descriptors [15–17]. The establishment of the AlPOs synthesis database makes it possible that we can use data mining technology to study the synthetic parameters for the rational synthesis of AlPOs. Recently, several researchers have already adopted data mining techniques to analyze the impact of the synthetic parameters on the resulting structures based on the database established by Jilin University. In Ref. [18], Li et al. studied the relationship between synthetic parameters and rational formation of (6,12)-ring-containing AlPOs. The feature subset to be evaluated in Ref. [18] was obtained through an exhaustive searching strategy. They evaluated the classification performance of all the combinations of synthetic parameters using Support Vector Machines. The combined parameters which influenced most for distinguishing (6,12)-ring-containing AlPOs from non-(6,12)-ring-containing AlPOs were deemed as the optimal feature subset. In

* Corresponding author. School of Computer Science and Information Technology, Northeast Normal University, Changchun 130117, China.
E-mail address: wangjz019@nenu.edu.cn (J. Wang).

http://dx.doi.org/10.1016/j.micromeso.2016.05.020
1387-1811/

Ref. [19], Yao et al. analyzed the affect of different synthetic parameters on the production of (6,12)-ring-containing AlPOs. In order to take the discriminant information of features into account, they calculated the importance degree of each feature by a fusing method which fused Fisher Score [20] and Mutual Information [21]. In Ref. [22], Qi et al. explored the relationship between the synthetic parameters and the specific resulting structure containing (6,12)-ring. In their method, the random subspace technique was first employed to pre-rank the synthetic parameters. Then, the fusion weights of synthetic parameters were obtained by Fisher score. At last, a sequential forward searching algorithm was utilized to select the most significant synthetic parameters based on the fusing results of previous two steps. In Ref. [23], Gao et al. discussed the impact of the parameters on the formation of (6,8)-ring-containing AlPOs and the subclasses of this ring-type based on their professional knowledge, and they validated their conclusions by Support Vector Machine (SVM).

Although the pioneering works mentioned above have made some achievements, there are still some limitations in them. In Refs. [18,19 and 22], the researchers only took the (6,12)-ring-containing AlPOs as prediction target to analyze the synthetic parameters. However, the (6,12)-ring-containing AlPOs actually consist of several subclasses. For example, both ATO and AFI are (6,12)-ring-containing AlPOs, but they are two kinds of molecular sieves since they are different in topological structures, and the relevant synthetic parameters which are important for their formation may be also different. Therefore, in order to analyze the synthetic parameters more reasonable, the diversity of the subclasses for a specific class of AlPOs must be taken into account. In Ref. [23], although the subclasses of (6,8)-ring-containing AlPOs have been considered, the optimal feature subsets for them was analyzed by professional knowledge of the domain experts rather than fully mining the data. Furthermore, the correlation among the selected features was neglected in all aforementioned works [18,19,22,23]. This may cause information redundancy since a good feature subset should be the one that contains features highly correlated with the class, while uncorrelated with the features each other [24].

Among various microporous aluminophosphates, AlPO molecular sieve with 8-ring channel is a typical kind of small pore materials used for gas separation and catalysis. For example, SAPO-34 (CHA zeotype) is an effective catalyst for the conversion of methanol to light olefins [25,26]. Thus, in this paper we focus on analyzing the formation parameters of the four important subclasses (AEN, AWO, CHA and ERI) of (6,8)-ring-containing AlPOs. In order to overcome the limitations of the previous works and better analyze the formation parameters of subclasses, we propose a Hierarchical Feature Selection Model (HFSM). Compared with the previous works in Refs. [18,19,22 and 23], the proposed model possesses two advantages: (1) The proposed model takes the subclasses of (6,8)-ring-containing AlPOs into consideration. Thus, with the proposed HFSM, the parameters which are critical for the formation of each specific subclass of (6,8)-ring-containing AlPOs can be well analyzed. (2) Since the correlations among the selected features are considered in our model, the feature subset selected by HFSM is more optimal than the previous works.

This paper is organized as follows. The material and method are presented in Section 2. The experimental results and analysis are shown in Section 3 and the conclusions are given in Section 4.

2. Materials and method

2.1. Data sets

Like other works in Refs. [18,19,22 and 23], we also use the microporous AlPOs database established by State Key Laboratory of Inorganic Synthesis and Preparative Chemistry of Jilin University in this study(http://zeobank.jlu.edu.cn/). By removing records which include missing values, we use the remainder 1279 records as the experimental samples. The experimental samples contain 332 (6,8)-ring-containing AlPOs which are composed of 18 zeotypes (or subclass). The number of four subclasses (AEN, AWO, CHA and ERI, as shown in Fig. 1) that possess the largest sample number of the (6,8)-ring-containing AlPOs are 26, 77, 100 and 36 respectively. The descriptions of the input synthetic parameters are shown in Table 1. The gel chemistry is crucial for the formation of microporous AlPOs. Therefore, four important features related to the molar concentrations of Al_2O_3, P_2O_5, solvent and the organic template in the starting gel are used as part of the input features for training the classifier [18]. As a result, the remaining seventeen synthetic parameters (or features) belonging to two classes (solvent and organic template) in Table 1 are analyzed in this work. Here, it should be pointed out that some other parameters such as synthetic temperature, time and pressure which may also crucial for the synthesis of zeolites are not considered in our paper and previous works in Refs. [18,19,22,23]. This is due to that some records in the microporous AlPOs database did not provide the values for them. Even for the records contain these parameters, most of them are given in a range form rather than precise value, which makes them hardly to be exactly analyzed by the data mining techniques.

2.2. The Hierarchical Feature Selection Model

AEN, AWO, CHA and ERI are four subclasses of (6,8)-ring-containing aluminophosphate molecular sieve. Although they all contain (6,8)-rings, their topological structures are very different from each other, as shown in Fig. 1. Thus, in order to better understand the rational synthesis of the four subclasses, it is necessary to select the important synthetic parameters for each of them.

Feature selection is one of the key steps in machine learning and pattern recognition problems. The aim of feature selection is to find the optimal feature subset that is necessary and sufficient for a specific task. Feature selection has several potential benefits, such as improving the accuracy of classification, avoiding the well-known “curse of dimensionality” problem, speeding up the training process and reducing storage demands. Specially, it can provide a better understanding and interpret ability for a domain expert [27,28].

In our study, let$X = [x_1, x_2,\ldots, x_n] \in \Re^{D\times n}$ be the entire AlPOs dataset containing n samples in D dimensional space. Suppose that these samples belong to C classes, we can denote the dataset as $\{X_1,\ldots,X_C\}$, in which X_i $(i = 1, \ldots,C)$ is the samples in the ith class. The original feature set is denoted by F, and each feature in the data set is F_m $(m = 1, \ldots, D)$. With the aim of analyzing the relationship between the synthetic parameters and the (6,8)-ring-containing AlPOs subclasses, we propose a Hierarchical Feature Selection Model as shown in Fig. 2.

From the flowchart in Fig. 2, we could see that the proposed feature selection model consists of two layers. In the first layer, we find the optimal feature subset Q which could separate the (6,8)-ring-containing AlPOs form non-(6,8)-ring-containing AlPOs. Then, in the second layer of the model, we further analyze which of the features in Q are important for the formation of AEN, AWO, CHA and ERI respectively.

2.2.1. The first layer of the Hierarchical Feature Selection Model

As discussed in Ref. [24], the correlation among features is a critical factor which should be taken into consideration in the feature selection process. For the AlPOs synthesis database analyzed in our study, there are some serious correlation relationships between the synthetic parameters. Taking the synthetic