Microporous and Mesoporous Materials 195 (2014) 174–179

Contents lists available at ScienceDirect

Microporous and Mesoporous Materials

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

Prediction of the formation of microporous aluminophosphates containing (6, 8)-rings using support vector machines

Na Gao [a], Yan Yan [a], Jinsong Li [b,c], Jiyang Li [a,*]

[a] State Key Laboratory of Inorganic Synthesis and Preparative Chemistry, College of Chemistry, Jilin University, Changchun 130012, PR China
[b] School of Computer Science and Information Technology, Key Laboratory of Intelligent Information Processing of Jilin Universities, Northeast Normal University, Changchun 130117, PR China
[c] Faculty of Chemistry, Northeast Normal University, Changchun 130024, PR China

ARTICLE INFO

Article history:
Received 10 December 2013
Received in revised form 11 April 2014
Accepted 15 April 2014
Available online 24 April 2014

Keywords:
Database
Aluminophosphate
Synthesis
Structure
Support vector machine

ABSTRACT

A support vector machine (SVM) method has been used on the database of aluminophosphate (AlPO) synthesis with ca. 1700 reaction data to predict the formation of AlPOs with (6, 8)-rings based on 20 synthetic parameters associated with the gel composition, solvent and templates. Furthermore, four zeotype structures with different channel systems, such as AEN, AWO, ERI and CHA have been computational classified in the family of (6, 8)-ring-containing AlPOs. Analysis results reveal that several synthetic parameters, particular the molar ratio of solvent/Al_2O_3 (F3), and the Van der Waals (VDW) volume (F14), the greatest length (F11), and the number of free rotated single bond (F20) of templates have significant effects in directing the formation of target structure. Using single or combinations of several parameters, a better classifier can be trained to distinguish the (6, 8)-ring-containing AlPOs with desired zeotype structure or not, and the classification accuracy is about 85%. This work demonstrates that the computational method together with database of AlPO synthesis provides a possibility to establish the relationship between the synthetic factors and the resulting structures, which is important in guiding the rational synthesis of microporous aluminophosphate materials.

1. Introduction

Zeolites and related inorganic microporous materials have widespread applications in catalysis, adsorption, and ion exchange [1]. Aluminophosphate (AlPO) molecular sieves constitute an important family of such materials since the first discovery of aluminophosphate molecular sieves $AlPO_4$-*n* in 1982 by U.C.C [2]. The rational synthesis of AlPO molecular sieves with specified structure and property is desired for the practical application of zeolitic materials. However, the crystallization kinetics of these materials is very complex, which is dependent on many variables including the source materials, the gel composition, the pH value, the templates or structure-directing agents (SDAs), the solvent, the crystallization temperature and time, etc. [3–8]. How these factors affect the crystallization of a target structure is unclear. Hence, the rational synthesis of AlPO molecular sieves is of great challenge.

Better understanding the synthesis and structure chemistry, as well as the formation mechanism of AlPO molecular sieves is great helpful to rational synthesis of such materials. For this purpose, the database of AlPO synthesis has been established by the research group of Prof. Yu and Xu [9–11], which contains ca.1700 reaction data including detailed synthesis conditions and resulting structure information. By using this database, the formation of microporous AlPOs with (6, 12)-rings can be predicted by support vector machine (SVM) classification technique. The prediction accuracy is 82.44% and the geometric size of the organic templates has been found to play a vital role in the prediction [12]. In addition, Partial Least Squares (PLS) regression and Logistic Discrimination (LD) have been employed to predict the formation of microporous aluminophosphate $AlPO_4$-5 based on the database of AlPO synthesis [13]. Moreover, four computational methods have been successfully used to treat the missing data problem in the database of AlPO synthesis [14], and target products as pure phases of aluminophosphate $AlPO_4$-5 (AFI zeotype) have been synthesized based on the results of missing value estimation [15]. These works indicate that the data-mining techniques can derive useful information between synthesis and structure from a large number of synthetic data, which can be used to guide the designed synthesis of microporous materials.

Among various microporous aluminophosphates, AlPO molecular sieves with 8-ring channels are a typical kind of small pore

* Corresponding author. Tel./fax: +86 431 85168614.
E-mail address: lijiyang@jlu.edu.cn (J. Li).

http://dx.doi.org/10.1016/j.micromeso.2014.04.032
1387-1811/

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 [16,17]. $AlPO_4$-14 (AFN zeotype) as a sorbent has been proposed for the separation of propane/propylene mixtures [18,19]. In the database of AlPO synthesis, 17 kinds of zeotypes belong to the (6, 8)-ring-containing AlPO molecular sieves, while these compounds are synthesized under different conditions. In this work, the support vector machine (SVM) classification method has been employed to study the synthesis of AlPO molecular sieves with (6, 8)-rings, aiming to predict the formation of such materials with specified zeotypes including CHA, AWO, ERI and AEN, as well as to reveal the prominent synthetic parameters affected the synthesis of these materials.

2. Experiments

2.1. Data selection

The database of AlPO synthesis contains about 1700 reaction data for ca. 230 AlPO structures. Excluding the missing data in the database, 1279 integrated synthetic data is selected for computational experiments [10,11]. These data include 372 records associated with (6, 8)-ring-containing AlPOs as the positive data, and the remaining 907 records used as the negative data. As shown in Fig. 1, 17 different zeotypes are found in the subset of AlPOs containing (6, 8)-rings, and four zeotypes CHA, AWO, ERI and AEN with more synthesis entries and different channel systems are selected as the target for the prediction.

2.2. Data encoding and feature selection

As we previous mentioned, many synthetic factors have influence on the formation of AlPO molecular sieves. In this work, we have selected 20 synthetic attributes as the input parameters for training our classifier, which include three parts: (1) three attributes of the gel compositions in the reaction; (2) six attributes of the solvent and (3) 11 attributes associated with the organic templates. The detailed information of input attributes is given in Table 1. It is noted that the synthetic parameters such as the crystallization temperature and time are ignored in the classification because they are easily controlled in the synthesis.

2.3. Computational experiments

2.3.1. SVM method

Support vector machine (SVM) is a typical supervised classification technique. This method constructs a hyperplane or set of hyperplanes in a high- or infinite-dimensional space, which can be used for classification, regression analysis, or other tasks. Intuitively, a good separation is achieved by the hyperplane that has the largest distance to the nearest training data point of any class (so-called functional margin), since in general the larger the margin, the lower the generalization error of the classifier [20,21].

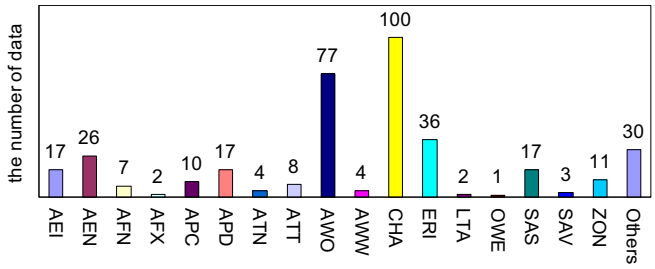

Fig. 1. Classification of AlPOs containing (6, 8)-rings in the database according to their topological structures (others indicate the open-framework AlPOs).

Table 1
Description of the input synthetic parameters in the classification.[a]

	Code	Description of parameters
Gel composition	F2	The molar ratio of P_2O_5/Al_2O_3
	F3	The molar ratio of solvent/Al_2O_3
	F4	The molar ratio of template/Al_2O_3
Solvent	F5	The density
	F6	The melting point
	F7	The boiling point
	F8	The dielectric constant
	F9	The dipole moment
	F10	The polarity
Organic template	F11	The greatest length of organic template
	F12	The second greatest length of organic template
	F13	The shortest length of organic template
	F14	The Van der Waals (VDW) volume
	F15	The dipole moment
	F16	The ratio of C/N
	F17	The ratio of N/(C + N)
	F18	The ratio of N/Van der Waals volume
	F19	The Sanderson electronegativity
	F20	The number of free rotated single bond
	F21	The maximum number of protonated H atoms

[a] F1 means the molar ratio of Al_2O_3/Al_2O_3.

Supervised learning models with associated learning algorithms that analyze data and recognize patterns have been used for classification and regression. LIBSVM is a library for support vector machines. It may help users to apply easily SVM to their applications. LIBSVM has gained wide applications in machine learning and many other areas [22,23]. Here, the LIBSVM provided in Weka 3.7 [24] has been employed in our classifier, in which the radial basis function (RBF) is used as the basic kernel function of SVM.

2.3.2. Classification of AlPOs with (6, 8)-rings or not

The goal of computational experiments is distinguishing microporous AlPOs containing (6, 8)-rings or not based on the initial input synthetic factors. In the selected synthetic records from database of AlPO synthesis, ratio of positive and negative samples (about 1:3) is of the imbalance, which may lead to undesired classification performance. According to previously reported data processing method [12,14], we have randomly selected 372 samples from the negative data together with 372 positive samples to construct a new dataset for SVM classification. A 10-fold cross-validation (CV) procedure is used to assess the prediction performance of the classifier. Namely, a completely training data is randomly partitioned into ten subsets of approximately the same size, and nine subsets used as training model and the remaining one for testing. The process is repeated for 20 times, in which different test subset is chosen. The performance of the SVM classification is evaluated by the average classification accuracy, sensitivity and specificity, defined as follows:

$$\text{Accuracy } (\%) = \frac{TP + TN}{TP + FP + FN + TN} \times 100$$

$$\text{Sensitivity } (\%) = \frac{TP}{TP + FN} \times 100$$

$$\text{Specificity } (\%) = \frac{TN}{TN + FP} \times 100$$

where TP and TN are the number of correctly classified positive and negative samples, respectively, FP is the number of (6, 8)-ring structures incorrectly classified as non-(6, 8)-ring-containing structures, and FN is the number of non-(6, 8)-ring-containing structures incorrectly classified as (6, 8)-ring structures. In addition, the area under the receiver operating characteristic curve (AUC) is also used for assessing the classification results. Usually, an area of 1.00 means