Computational Materials Science 136 (2017) 29–35

Contents lists available at ScienceDirect

Computational Materials Science

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

Data mining assisted materials design of layered double hydroxide with desired specific surface area

Biao Hu [a], Kailiang Lu [a], Qing Zhang [b], Xiaobo Ji [a], Wencong Lu [a,c,*]

[a] Department of Chemistry, College of Sciences, Shanghai University, Shanghai 200444, China
[b] College of Materials Science and Engineering, Shanghai University, Shanghai 200072, China
[c] Materials Genome Institute, Shanghai University, Shanghai 200444, China

ARTICLE INFO

Article history:
Received 19 December 2016
Received in revised form 17 March 2017
Accepted 18 March 2017

Keywords:
Layered double hydroxides
Data mining
Specific surface area

ABSTRACT

The value of specific surface area (SSA) of layered double hydroxide (LDH) can be employed as the main parameters of adsorbent evaluation. In this work, data mining methods were used to explore the correlations of the SSA (ranged 10–90 $m^2 g^{-1}$) with their chemical compositions and technical parameters in search of the certain LDH material with desired SSA. The genetic algorithm (GA)-support vector regression (SVR) method was used to filter the main molecular descriptors for modeling. The related coefficient (R) between predicted SSA and experimental SSA reached as high as 0.937 for training set and 0.892 for leave-one-out cross validation (LOOCV), respectively. The optimal model was then applied to 9 independent samples to test the prediction ability with the mean relative error equal to 14.7%. A case study of controllable synthesis predicted by the model was also carried out, and the new LDH material (Ni-Fe CO_3 LDH) with desired SSA was verified by our experiments with the relative error equal to 13.8%. The method outlined here can provide valuable hints into the exploration of materials design with the assistance of machine learning.

1. Introduction

In recent years, the presence of harmful ions such as F^-, Br^-, Pb^{2+} in drinking water above permissible level has increased many serious incidents around the world [1]. Keeping the view of toxic effects of fluoride on human health, there is an urgent need to find out an effective and robust technology for the removal of excess fluoride from drinking water [2]. It is evident that the adsorbent plays an important role in the pollution treatment. Among various adsorbents employed for defluoridation of waste water, the adsorbent with advantages in terms of cost, simplicity of design and operation seems to be more attractive. From the perspective of practical applications, a well-designed adsorbent should involve the following issues [3]: (1) How to tune a high dispersion of adsorption sites for great activity. (2) How to stabilize the designed adsorbent with eco-friendly for long-term usage and easy to be recycled. (3) How to establish a pattern which can quickly design a new adsorbent meeting various requirements rather than the traditional trial-and-error method and obtain the product technical parameters to facilitate mass production in specific technical process. Indeed, many endeavors have been devoted to assessing the removal of excess fluoride from drinking water by various types of adsorbents.

The metal and metal-based materials have emerged as an important adsorbent for many years. For example, Farrah et al. [4] studied the adsorption of $Al(OH)_3$, gibbsite and Al_2O_3 for F^- removal, while the efficiency in bringing fluoride to desired concentration levels was not satisfactory. To enhance the adsorption capacity, a variety of modified activated alumina with higher adsorption has been developed [5–7]. However, the drawbacks related to high residual aluminum concentration in the treated water have also been reported. Other metal-based adsorbents such as calcium-based [8], iron-based [9] and other metal oxides have been extensively studied, though it is high efficient for defluoridation, these adsorbents are mainly limiting for the accompanying precipitation of fluorite. They are relatively expensive to install, operate and prone to fouling, scaling, or membrane degradation. Also, the residual water may face the secondary pollution for heavy metals introducing.

Layered double hydroxides [10] with the abundant cheap raw materials, easy synthesis, without toxicity and unique layered structure have attracted wide attention in catalysis [11], adsorption [12], medicine [13], optical materials [14] and super- capacitors [15] in the last decades. Their outstanding behavior for the

* Corresponding author at: Department of Chemistry, College of Sciences, Shanghai University, Shanghai 200444, China.
E-mail address: wclu@shu.edu.cn (W. Lu).

http://dx.doi.org/10.1016/j.commatsci.2017.03.027
0927-0256/

adsorption of inorganic anions, heavy metal ions and organic anions, make it promising material as an attractive adsorbent. On one hand, the traits of the layered structure and active sites in the layer board can increase the reaction area obviously. On the other hand, the tunable composition and the memory effect, which means the broken LDH can return its original layered structure under the certain conditions, make it possible to tailor the LDH with long-term use and recycling. As a kind of 2D anion clay consisting of the positively-charged host layers and exchangeable interlayer anions (as shown in Fig. 1), the metal cations occupy the centers of edge sharing octahedra, whose vertexes contain hydroxide ions that connect to form infinite 2D sheets. The general formula can be described as $[M^{2+}_{1-x}M^{3+}_{x}(OH)_2][A^{n-}]_{x/n}\cdot zH_2O$, where M^{3+} is trivalent cation such as Al^{3+}, Fe^{3+}, $Cr^{3+}$and Mn^{3+}. M^{2+} is divalent metal such as Mg^{2+}, Zn^{2+} and Ni^{2+}. The A^{n-} is exchangeable inorganic anion in the interlayer (CO_3^{2-}, SO_4^{2-}, NO_3^-, Cl^-, etc.) or the organic anions, and the x is normally between 0.2 and 0.4.

However, different LDHs have a huge difference in anion adsorptions since the various composition, complex process preparation and topography. How to find a simple, efficient way to design the LDH with desired property is still to be a challenge. In June of 2011, the Materials Genome Initiative (MGI) [16,17] for global competitiveness was proposed by the National Science and Technology Council of America for the development of an infrastructure to shorten the materials development cycle. The most important and fundamental goal of MGI is to accelerate materials design through the use of computational capabilities, data management and an integrated approach to materials science and engineering [18,19]. Recently, the researches about machine learning by adaptive design to assist materials discovery have been widely concerned [20]. Xue et al. demonstrated how to accelerate search for materials with targeted properties by adaptive design, where the inference and global optimization were simultaneously considered to find the lowest thermal hysteresis NiTi-based shape memory alloys [21]. Raccuglia et al. reported that machine-learning models of support vector machine and decision tree were used to predict reaction outcomes for the crystallization of templated vanadium selenites with satisfactory results [22]. As many studies reported, the LDHs with various compositions, process preparation and topography have different SSA. Therefore, we think there is a pattern among the SSA, composition and process preparation. The reason we concerned about the SSA of LDHs is their close contact with the adsorption properties. Different adsorption performances of LDHs are often attributed to the diversities in SSA. In general, the LDH with a larger SSA has a better behavior [23]. So, the SSA can be used as an indicator for adsorption behavior. In all, the exploration of the rules among SSA, composition, structure and other parameters may be an efficient approach to develop a new LDH with desired properties.

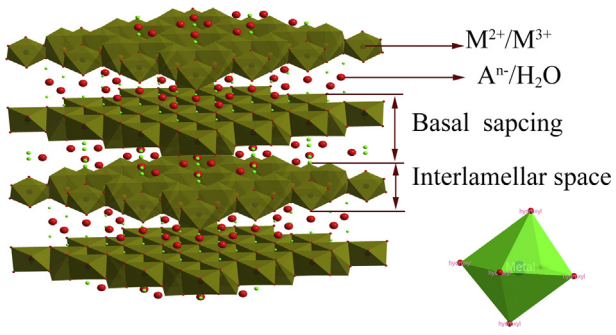

Fig. 1. Structure of layered double hydroxides.

Herein, we reported a novel method based on data mining to establish a quantity-structure property relationship (QSPR) model for discovering the correlation with the SSA, the composition and the process preparation. Moreover, according to the model we established, the SSA of Ni-Fe CO_3 LDH, which had seldom reported yet, was also obtained. To verify the model prediction, a series of experimental characterization had been employed.

2. Methods

2.1. Data set and descriptors available

The data set collected from the reference [23–45] were composed of 37 layered double hydroxides with various SSA ranged from 10 $m^2\,g^{-1}$ to 90 $m^2\,g^{-1}$. It was divided into 2 parts: (1) 28 LDHs, used as the internal training set. (2) 9 LDHs, employed as the external test set for testing the predictive power of the model. The training set and test set were divided randomly with the ratio about 3. Table 1 lists a total of 15 candidate descriptors including 12 atomic parameters (The data were quoted from Pauling[a] [24] and Lange's Handbook of Chemistry [25]) and 3 technical parameters [26–48].

2.2. Computational software

In this article, the entire computational process using SVM software package named ChemSVM has been programmed in our laboratory. The free version of ChemSVM can also be downloaded on the websites of Laboratory of *Materials Data Mining* in Shanghai University. (http://chemdata.shu.edu.cn:8080/MyLab/Lab/download.jsp).

3. Results and discussion

3.1. Selection of molecular descriptors

To get rid of redundant, noisy, or irrelevant descriptors in the model building, selection of molecular descriptors is an efficient means to reduce the dimensionality of input space without loss of important information.

The genetic algorithm (GA)-support vector regression (SVR) method [26] was employed to screen the subset of descriptors for modeling. GA is a highly parallel, random and adaptive search algorithm based on the natural selection and evolution mechanism of the biological world. It can be used to find the optimal sub set of features. Compared with other optimization algorithms, GA has an ability to move from local optima present on the response surface, and it can work out a wide variety of optimizations with the requirement of no knowledge or gradient present about the response surface.

In this model, RMSE were employed as evaluation for molecular descriptors selection. The RMSE [27] is defined as follows:

$$\mathrm{RMSE} = \sqrt{\frac{\sum_{i=1}^{n}(p_i - e_i)^2}{n}}$$

where e_i and p_i are the measured and predicted values of i sample, respectively. n is the number of the whole samples in the training set. Usually, the smaller the RMSE is, the better the set of descriptors gets.

Fig. 2 illustrates how GA can be used to select the materials descriptors. It is found that the smallest RMSE after 10 generations of evolution corresponds to the optimal sub set of features including two atomic parameters and three technical parameters. Two atomic parameters are R (the ratio of the number of atoms of the divalent metal element to the trivalent metal element) and R_a/R_b