Computational Materials Science 151 (2018) 41–48

Contents lists available at ScienceDirect

Computational Materials Science

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

Accelerated search for perovskite materials with higher Curie temperature based on the machine learning methods

Xiuyun Zhai[a,b], Mingtong Chen[c], Wencong Lu[d,*]

[a] College of Materials Science and Engineering, Shanghai University, Shanghai 200444, China
[b] School of Mechanical Engineering, Panzhihua University, Panzhihua 617000, China
[c] Material Engineering School, Panzhihua University, Panzhihua 617000, China
[d] Department of Chemistry, College of Sciences, Shanghai University, Shanghai 200444, China

ARTICLE INFO

Keywords:
Perovskite materials
Curie temperature
Machine learning
Support vector machine
Relevance vector machine
Random forest

ABSTRACT

Curie temperature (Tc), the second order phase transition temperature, is also one of the important physical properties of perovskite materials. It is a meaningful work to quickly and efficiently predict Tc of new perovskite materials before doing a considerable amount of experimental work. In the work, SVM (support vector machine), RVM (relevance vector machine) and RF (random forest) were employed to establish the prediction models of Tc with the physicochemical parameters, respectively. The results reveal that the three models all have high precision and reliability. According to K-fold cross validation, the SVR model had better prediction performance than the RVM and RF models. Meanwhile, the potential perovskite material with higher Tc was found by using the SVR model integrated with the search strategy of genetic algorithm from the virtual samples. The methods outlined here can provide valuable hints into the exploration of materials with desired property and can accelerate the process of materials design.

1. Introduction

Perovskite-type oxides, commonly represented by ABO_3, have been considered as one of the most promising materials due to the application of electronic and magnetic components such as multilayer capacitors and sensors [1,2]. In the properties of perovskite materials, Curie temperature (Tc), also called Curie point, is the phase transition temperature of ferroelectrics from ferroelectric phase to paraelectric phase. So, it has an important influence on many applications of perovskite materials such as erasing and writing new data of magneto-optical storage medium, temperature control of soldering irons, and stabilizing the magnetic field of tachometer generators against temperature variation. In recent years, many researchers attempt to synthesize perovskite materials with high Tc or higher Tc than room temperature. Therefore, the effects of different doping elements on Tc of perovskite materials have been widely reported [3,4]. Yu et al. [5] synthesized a very complex perovskite material ($Pb_{0.6}Bi_{0.4}Ti_{0.75}Zn_{0.15}Fe_{0.1}O_3$) with Tc of 978 K. However, it is a challenge to break through existing Tc in that the compositions of perovskite materials and different doping ratios of elements are highly complex.

At present, materials design with assistance of machine learning methods, promoted by efforts such as the Materials Genome Initiative, has become a research hotspot and an alternative approach to trial-and-error experiments. Pilania et al. [6] constructed a model to predict the bandgaps of double perovskites with the help of the machine learning methods. Raccuglia et al. [7] used the resulting data of failed experiments to train a machine-learning model to predict reaction success. Xue et al. [8] provided an adaptive approach and employed the machine learning regression algorithms to find very low thermal hysteresis (ΔT) NiTi-based shape memory alloys. Accordingly, it is no doubt that machine learning methods can shorten the cycle of materials design and realize controllable synthesis of materials.

In this work, a slew of machine learning methods was employed to forage for the model with the optimal regression performance to predict Tc of perovskite materials. To develop a really useful machine-learning predictor for a material or biological system as reported in a series of recent publications [9–23], one should observe the Chou's 5-step rule [24]; i.e., making the following five steps very clear: (i) how to construct or select a valid benchmark dataset to train and test the predictor; (ii) how to formulate the samples with an effective mathematical expression that can truly reflect their intrinsic correlation with the target to be predicted; (iii) how to introduce or develop a powerful algorithm (or engine) to operate the prediction; (iv) how to properly perform cross-validation tests to objectively evaluate the anticipated accuracy of

* Corresponding author.
E-mail address: wclu@shu.edu.cn (W. Lu).

https://doi.org/10.1016/j.commatsci.2018.04.031
Received 31 March 2018; Received in revised form 13 April 2018; Accepted 15 April 2018

Table 1
The twenty-one descriptors of perovskite materials.

No.	Meanings	Features
1	Weighted ionic radii of A-site (Å)	R_a
2	Weighted ionic radii of B-site (Å)	R_b
3	Weighted electronegativity Pauling of A-site	χ_{pa}
4	Weighted electronegativity Pauling of B-site	χ_{pb}
5	Tolerance factor	t
6	Unit cell lattice edge (Å)	$\alpha_O{}^3$
7	Critical radii (Å)	r_c
8	Weighted ionization energy of A-site (kJ/mol)	I_{1a}
9	Weighted ionization energy of B-site (kJ/mol)	I_{1b}
10	Molecular mass (g/mol)	M
11	Ratio of ionic radii of A-site to B-site	R_a/R_b
12	Weighted electron affinity of A-site (eV)	EA_a
13	Weighted electron affinity of B-site (eV)	EA_b
14	The melt point of A-site metal (°C)	t_{ma}
15	The melt point of B-site metal (°C)	t_{mb}
16	The boil point of A-site metal (°C)	t_a
17	The boil point of B-site metal (°C)	t_b
18	The enthalpy of fusion of A-site (kJ/mol)	$\Delta_{fus}H_a$
19	The enthalpy of fusion of B-site (kJ/mol)	$\Delta_{fus}H_b$
20	The density of A-site metal (g/cm^3)	ρ_a
21	The density of B-site metal (g/cm^3)	ρ_b

the predictor; (v) how to establish a user-friendly web-server for the predictor that is accessible to the public. The fifth step is the direction of our future work that we will provide a web-server for the prediction tools presented in this paper. Below, we are to describe how to deal with these steps one-by-one.

The main outcomes of the paper are: the support vector regression (SVR), relevance vector machine (RVM) and random forest (RF) models were constructed with the better regression and generalization performances according to the K-fold cross validation; meanwhile, the SVR model has been verified that it has the stunning performance; finally, by the means of the SVR model and genetic algorithm (GA), the candidate perovskite material with perhaps higher Tc was provided to guide future researches and experiments, and then accelerate search for perovskite materials with higher Tc.

2. Methods

2.1. Dataset

We collected forty-seven perovskite materials from nine references [25–33] as the dataset in Tab S1 of supplementary information. There are several types of elements with different doping ratio both in A-site and B-site of the samples in the dataset to provide the conditions for following screening perovskite materials. The range of the target value (Tc) is from 170 K to 380 K. Besides, there are twenty-one physico-chemical parameters [34] (in Table 1) as the descriptors of perovskite materials and the candidate inputs of the models.

Table 2
The results of GA feature selection.

Algorithm	The selected features
SVR	χ_{pb}, r_c, R_a/R_b, EA_a, t_{mb}, t_a
RVM	R_a, χ_{pa}, t, r_c, I_{1a}, R_a/R_b, EA_a, t_a, $\Delta_{fus}H_a$, ρ_b

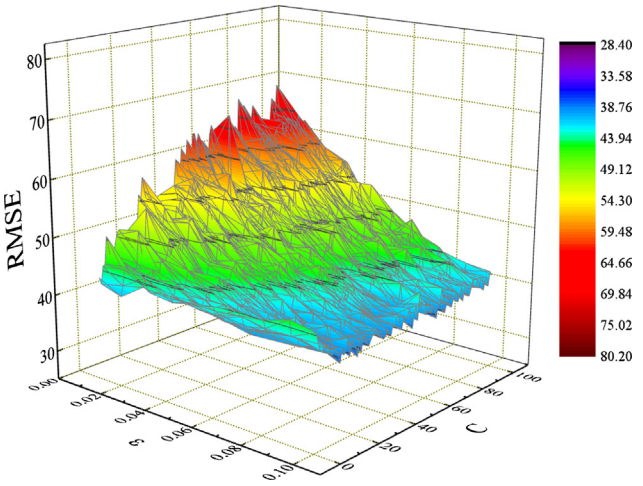

Fig. 2. The optimization process of hyper-parameter of the SVR model in GA.

Table 3
The list of hyper-parameters of SVR and RVM.

Algorithm	Hyper-parameters
SVR	C = 2; σ = 0.2; ε = 0.01
RVM	γ = 0.171

2.2. SVR

SVR [35–37], a powerful methodology for solving problems in nonlinear classification and regression, is also a supervised learning algorithm that has been widely applied to various fields. It considers the balance between empirical risk and expected risk, and then makes computational model have the good prediction and generalization performances.

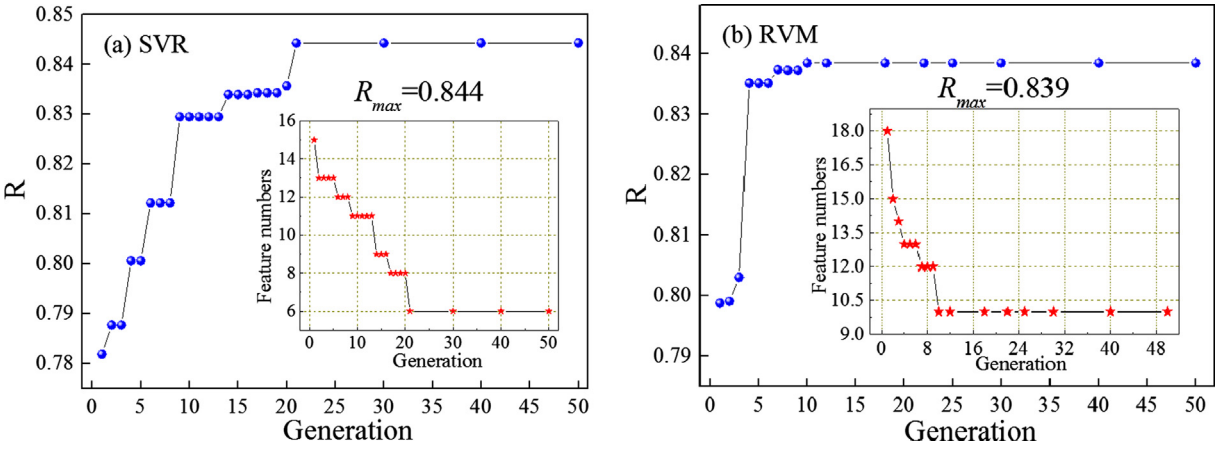

Fig. 1. The R versus generation of the evolution process in GA.