Materials and Design 92 (2016) 371–377

Contents lists available at ScienceDirect

Materials and Design

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

Modelling and optimization of the superconducting transition temperature

Hui-ran Zhang [a,b,*], Yan Zhang [a], Dong-bo Dai [a], Min Cao [a,b], Wen-feng Shen [a,b]

[a] School of Computer Engineering and Science, Shanghai University, Shanghai 200444, China
[b] Materials Genome Institute of Shanghai University, Shanghai 200444, China

ARTICLE INFO

Article history:
Received 1 September 2015
Received in revised form 8 December 2015
Accepted 14 December 2015
Available online 17 December 2015

Keywords:
Superconducting transition temperature
Prediction model
Rough set theory
Support vector regression
Data mining

ABSTRACT

Previous researches show that the superconducting transition temperature T_{C0} of high temperature superconductors can be calculated approximately by the algebraic relation $T_{C0} = K_B^{-1}\beta/(\ell\zeta)$. To predict T_{C0} more accurately, we propose a data mining approach called RS-PSO-SVR combining Rough Set theory, Particle Swarm Optimization with Support Vector Regression method. Based on the prior experimental data,the optimized model was established for predicting the T_{C0}. The analyses show that the interlayer Coulomb interaction is an effective descriptor for predicting T_{C0}. By our experiment, the proposed algorithm successfully predicted the T_{C0} of high temperature superconductors. These results show that our model provide theoretical guidance for physical experiments by reducing arbitrary experiments.

1. Introduction

The high temperature superconductors [1] are characterized by a two-dimensional (2D) layered superconducting condensate with unique features [2]. Superconducting properties can be optimized by element doping or applied pressure to yield higher transition temperature and bulk Meissner effect [3]. Thus, various high temperature superconductors have been the interests of extensive research [4–7]. Many investigations show that the superconducting transition temperature (T_{C0}) of high temperature superconductors depends on its crystal structure, cell parameters, ionic valences, and Coulomb coupling between electronic bands in adjacent, spatially separated layers [5]. Some researchers analyzed more than thirty high temperature materials with five structural and chemical family types, such as cuprate, ruthenate, rutheno-cuprate, iron-pnictide, iron-chalcogenide, and organic. It is shown that T_{C0} can be given by the following algebraic expression [6,7].

$$T_{C0} = K_B^{-1}\beta/(\ell\zeta) \tag{1}$$

Here, ℓ is related to the mean spacing between interacting charges in the layers, ζ is the distance between interacting electronic layers, β is a universal constant, and k_B is Boltzmann's constant.

It is critical to predict T_{C0} of various superconductors. As one of the data analysis methods, Rough Set (RS) theory was firstly proposed by Pawlak et al. [8]. Drawing lessons from various definitions of uncertainty and ambiguity in logic and philosophy, Pawlak proposed a concept of imprecise category for knowledge base. Then it developed into a complete RS theory. As an effective mathematical tool for manipulating imprecise, incomplete, and incompatible data, it has been widely applied in the fields of intelligent information systems and achieves great success.

Recently, in order to accelerate the process of discovery and deployment of new materials at a fraction of cost, Materials Genome Initiative (MGI) was proposed in the United States in 2011 [9]. Under the framework of MGI, the concept of materials design is emphasized by utilizing the database with big data characteristics, computational simulation, optimization, and prediction methods. In this paper, based on experimental data in literatures, we established an optimized model using Support Vector Regression (SVR) [10] and RS theory [8,11,12] to predict T_{C0} of high temperature superconductors more accurately than the methods studied in this work. Data mining technology is also used to extract effective information from the available experimental results. Here, we use the predicted T_{C0} by equation method and the experimental value (target value) as prior experimental data, and RS is adopted as a data preprocessing method to get a normalized dataset suitable for the machine learning algorithm [12]. Our method is applied to predict the T_{C0} through the interlayer Coulomb interaction, leading to more accurate results than the T_{C0} estimated by the Eq. (1). Our model provides a theoretical tool guiding further experiments by reducing the uncertainty of prediction.

2. Theory and methods

2.1. Theory of PSO-SVR

Support Vector Machine (SVM) used in the present work is a machine learning algorithm based on statistical learning theory and firstly

* Corresponding author at: School of Computer Engineering and Science, Shanghai University, Shanghai, 200444, China.
E-mail address: hrzhangsh@shu.edu.cn (H. Zhang).

http://dx.doi.org/10.1016/j.matdes.2015.12.081
0264-1275/

proposed by Vapnik et al. [13]. Comparing with traditional learning machines such as genetic algorithm and artificial neural network, SVM has a better generalization precision and nonlinear processing ability. SVM has been successfully applied to solve classification and regression problems in many computing researches [14–18]. SVR is an extension of SVM [13,19–22]. It is suitable for handling nonlinear problems with the aid of nonlinear mapping function generally known as kernel function that helps in mapping descriptors to high-dimensional feature space ***F*** where the linear regression is conducted [12]. The complete representation of regression function for a training dataset is presented as below:

$$f(x) = \sum_{i=1}^{l} (\alpha_i - \alpha_i^*) k(x, x_i) + b_i \tag{2}$$

where l is the number of support vectors, α_i and α_i^* are Lagrange multipliers, $k(\boldsymbol{x}, \boldsymbol{x_i}) = \Phi(\boldsymbol{x}) \cdot \Phi(\boldsymbol{x_i})$ is a kernel function, and b is a bias. In this study, radial basis kernel is adopted as the kernel function. The detailed principle of SVR can be found in Ref. [13].

As an optimization technique, the Particle Swarm Optimization (PSO) method was proposed by Kennedy and Eberhart [23,24]. It was motivated by social behavior of organisms such as bird flocking and fish schooling. In the SVR method, it is important to determine the key parameters including regularized constant C, and the kernel function parameter γ. The Grid algorithm can be used to find the optimal C and γ, but it is time consuming and cannot converge at the global optimum. Therefore, in order to improve the efficiency of prediction, PSO is utilized to search the optimal parameters (C, γ) of SVR in this work [9]. Root mean square error (*RMSE*) serves as the fitness function:

$$RMSE = \sqrt{\frac{1}{m} \sum_{i=1}^{m} (\hat{y}_i - y_i)^2} \tag{3}$$

where m denotes the number of training samples, y_i and $\hat{y}_i$ represent the measured and estimated values for the *ith* training sample, respectively.

Table 1
The main parameters, including measured T_{c0}, the distance between interacting layers ζ, the calculated spacing between interacting charges within layers ℓ, and the theoretical T_{c0} [6–7].

No.	Superconducting compounds	ζ(Å)	l (Å)	Equation.val (K)	Meas.val (K)
1	$YBa_2Cu_3O_{6.92}$	2.2677	5.7085	96.36	93.7
2	$YBa_2Cu_3O_{6.60}$	2.2324	8.6271	64.77	63
3	$LaBa_2Cu_3O_{7-\delta}$	2.1952	5.7983	98	97
4	$YBa_2Cu_4O_8$(12GPa)	2.1658	5.5815	103.19	104
5	$Tl_2Ba_2CuO_6$	1.9291	8.0965	79.86	80
6	$Tl_2Ba_2CaCu_2O_8$	2.0139	5.7088	108.5	110
7	$Tl_2Ba_2Ca_2Cu_3O_{10}$	2.0559	4.6555	130.33	130
8	$TlBa_2CaCu_2O_{7-\delta}$	2.0815	5.7111	104.93	103
9	$TlBa_2Ca_2Cu_3O_{9+\delta}$	2.0315	4.6467	132.14	133.5
10	$HgBa_2Ca_2Cu_3O_{8+\delta}$	1.9959	4.6525	134.33	135
11	$HgBa_2Ca_2Cu_3O_{8+\delta}$(25GPa)	1.9326	4.4664	144.51	145
12	$HgBa_2CuO_{4.15}$	1.9214	7.0445	92.16	95
13	$HgBa_2CaCu_2O_{6.22}$	2.039	4.8616	125.84	127
14	$La_{1.837}Sr_{0.163}CuO_{4-\delta}$	1.7828	18.6734	37.47	38
15	$La_{1.8}Sr_{0.2}CaCu_2O_{6\pm\delta}$	1.7829	11.99	58.35	58
16	$(Sr_{0.9}La_{0.1})CuO_2$	1.7051	17.6668	41.41	43
17	$Ba_2YRu_{0.9}Cu_{0.1}O_6$	2.0809	18.6123	32.21	35
18	$(Pb_{0.5}Cu_{0.5})Sr_2(Y,Ca)Cu_2O_{7-\delta}$	1.9967	9.2329	67.66	67
19	$Bi_2Sr_2CaCu_2O_{8+\delta}$	1.795	8.0204	89.32	89
20	$(Bi,Pb)_2Sr_2Ca_2Cu_3O_{10+\delta}$	1.6872	6.5414	113.02	112
21	$Pb_2Sr_2(Y,Ca)Cu_3O_8$	2.028	8.0147	76.74	75
22	$Bi_2(Sr_{1.6}La_{0.4})CuO_{6+\delta}$	1.488	24.0797	34.81	34
23	$RuSr_2GdCu_2O_8$	2.182	11.3699	50.28	50
24	$La(O_{0.92-y}F_{0.08})FeAs$	1.7677	28.4271	24.82	26
25	$Ce(O_{0.84-y}F_{0.16})FeAs$	1.6819	19.9235	37.23	35
26	$Tb(O_{0.80-y}F_{0.20})FeAs$	1.5822	17.2624	45.67	45
27	$Sm(O_{0.65-y}F_{0.35})FeAs$	1.667	13.2895	56.31	55
28	$(Sm_{0.7}Th_{0.3})OFeAs$	1.671	14.3711	51.94	51.5
29	$(Ba_{0.6}K_{0.4})Fe_2As_2$	1.932	17.4816	36.93	37
30	$Ba(Fe_{1.84}Co_{0.16})As_2$	1.892	28.0043	23.54	22
31	$FeSe_{0.977}$(7.5GPa)	1.424	23.8828	36.68	36.5
32	$Fe_{1.03}Se_{0.57}Te_{0.43}$(2.3GPa)	1.597	30.4467	25.65	23.3
33	$K_{0.83}Fe_{1.66}Se_2$	2.0241	20.4923	30.07	29.5
34	$Rb_{0.83}Fe_{1.70}Se_2$	2.1463	18.2889	31.78	31.5
35	$Cs_{0.83}Fe_{1.71}Se_2$	2.3298	18.1873	29.44	28.5
36	κ–[BEDT-TTF]$_2$Cu[N(CN)$_2$]Br	2.4579	43.7194	11.61	10.5

Table 2
Statistics of variables for the prediction model.

Variables	Min	Max	Mean	Standard deviation
ζ(Å)	1.424	2.4579	1.9323	0.24
l (Å)	4.4664	43.7194	13.9872	9.26
Equation.val (K)	11.61	144.51	68.2739	38.58
Meas.val (K)	10.5	145	68.0139	38.88

2.2. The algorithm of rough set preprocessing

RS theory is one of the data analysis methods based on the concept of imprecise category. In this research, it is adopted as a data preprocessing method to deal with the literature data. T_{c0} can be estimated by the Eq. (1) within an accuracy of ± 1.31 K. Comparing the T_{c0} estimated by Eq. (1) and the target value measured experimentally, the dataset can be classified as two categories: overestimated and underestimated values, respectively. Applying the RS preprocessing algorithm, the weight of each feature can be calculated. The weights are associated with the balance of the equation and the target values, affecting the predicted values. In the processing of training and testing, the weights are adjusted to achieve the optimal predicted values.

Meantime, the application of RS theory for a given set of sample data preprocessing makes the characteristic values of all data more general and the weight more remarkable [12]. It means that all the dimensions of the data set are scaled to [0, 1] and the specific characteristics attributes of each dimension are kept. The normalization makes the data processing easy. Moreover, the normalized data can improve the prediction abilities of the model.

Base on the discussions above, the proposed algorithm are briefly described below.

2.2.1. RS preprocessing algorithm

Input: training set and test set.

Output: weight of each feature.

Step 1: Scale all the dimensions in both training set and test set to [0, 1].

Step 2: Make a copy of training set. Each data in the copied training set, round to two decimal places. Then magnify them 100 times so that all the data are integer and range in [0, 100].

Step 3: For each dimension in the copied training set, count the times of each integer i range in [0, 100] in the copied training set and belong to the j class, denoted by T_{ji}.

Step 4: Calculate the percentage classification P by the following formula.

$$P_i = max(T_{ji}) / \sum_{j=1}^{m} T_{ji} \tag{4}$$

Table 3
The calculated feature weights.

Feature	ξ	L	Equation.val
Weight	0.53	0.44	0.50