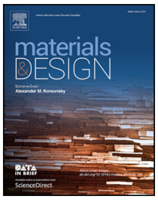
Contents lists available at ScienceDirect

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/© 2015 Elsevier Ltd. All rights reserved.

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 $\boldsymbol{F}$ 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, $k(\boldsymbol{x}, \boldsymbol{x_i}) = \Phi(\boldsymbol{x}) \bullet \Phi(\boldsymbol{x_i})$ 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:

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

where m denotes the number of training samples, y_i and $\widehat{y}_i$ 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	$La(O_{0.92}-yF_{0.08})FeAs$	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–$[BEDT\text{-}TTF]_2Cu[N(CN)_2]Br$	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

Table 4
Comparison between the actual (experimental) and predicted values of T_{C0}.

No.	exp value	equation value	absolute error	RS-PSO-SVR value	absolute error	PSO-SVR value	absolute error	BPNN value	absolute error
1	93.7	96.36	2.66	94.43	0.73	95.76	2.06	95.72	2.02
2	63	64.77	1.77	63.09	0.09	63.02	0.02	64.33	1.33
3	97	98	1	96.99	−0.01	97.21	0.21	98.83	1.83
4	104	103.19	−0.81	103.04	−0.96	103.15	−0.85	101.47	−2.53
5	80	79.86	−0.14	79.84	−0.16	80.07	0.07	82.87	2.87
6	110	108.5	−1.5	109.75	−0.25	109.12	−0.88	108.26	−1.74
7	130	130.33	0.33	130.17	0.17	131.18	1.18	130.77	0.77
8	103	104.93	1.93	105.03	2.03	104.96	1.96	103.51	0.51
9	133.5	132.14	−1.36	132.67	−0.83	131.84	−1.66	132.49	−1.01
10	135	134.33	−0.67	135.07	0.07	135.29	0.29	134.86	−0.14
11	145	144.51	−0.49	144.63	−0.37	144.03	−0.97	145.87	0.87
12	95	92.16	−2.84	92.66	−2.34	92.40	−2.60	90.84	−4.16
13	127	125.84	−1.16	125.02	−1.98	125.40	−1.60	126.16	−0.84
14	38	37.47	−0.53	37.51	−0.49	37.46	−0.54	37.91	−0.09
15	58	58.35	0.35	58.16	0.16	58.36	0.36	57.35	−0.65
16	43	41.41	−1.59	41.29	−1.71	41.13	−1.87	43.60	0.60
17	35	32.21	−2.79	32.28	−2.72	31.84	−3.16	33.02	−1.98
18	67	67.66	0.66	67.27	0.27	67.64	0.64	66.78	−0.22
19	89	89.32	0.32	88.91	−0.09	88.98	−0.02	89.13	0.13
20	112	113.02	1.02	112.53	0.53	113.58	1.58	115.93	3.93
21	75	76.74	1.74	76.80	1.80	77.40	2.40	76.00	1.00
22	34	34.81	0.81	33.97	−0.03	33.87	−0.13	34.17	0.17
23	50	50.28	0.28	49.50	−0.50	49.37	−0.63	49.51	−0.49
24	26	24.82	−1.18	23.77	−2.23	23.54	−2.46	22.65	−3.35
25	35	34.23	−0.77	37.02	2.02	37.26	2.26	37.42	2.42
26	45	45.67	0.67	45.00	0.00	45.09	0.09	45.68	0.68
27	55	56.31	1.31	56.24	1.24	56.18	1.18	55.77	0.77
28	51.5	51.94	0.44	51.63	0.13	51.98	0.48	52.02	0.52
29	37	36.93	−0.07	36.80	−0.20	36.77	−0.23	36.82	−0.18
30	22	23.54	1.54	22.31	0.31	22.70	0.70	23.46	1.46
31	36.5	36.68	0.18	35.90	−0.60	35.19	−1.31	35.39	−1.11
32	23.3	25.65	2.35	25.37	2.07	25.95	2.65	24.17	0.87
33	29.5	30.07	0.57	29.62	0.12	29.70	0.20	30.59	1.09
34	31.5	31.78	0.28	31.47	−0.03	31.22	−0.28	32.14	0.64
35	28.5	29.44	0.94	29.04	0.54	29.20	0.70	31.69	3.19
36	10.5	11.61	1.11	10.30	−0.20	11.04	0.54	22.07	11.57

where m represents the classes that the dataset should be totally divided. In this study, according to the prior experimental data (the value of equation and target T_{C0}), the dataset are divided into over and underestimated groups.

Step 5: Scale P_i to [0, 1], using the formula as below:

$$P_i = \left(\frac{m}{m-1}\right)\cdot\left(P_i - \frac{1}{m}\right) \tag{5}$$

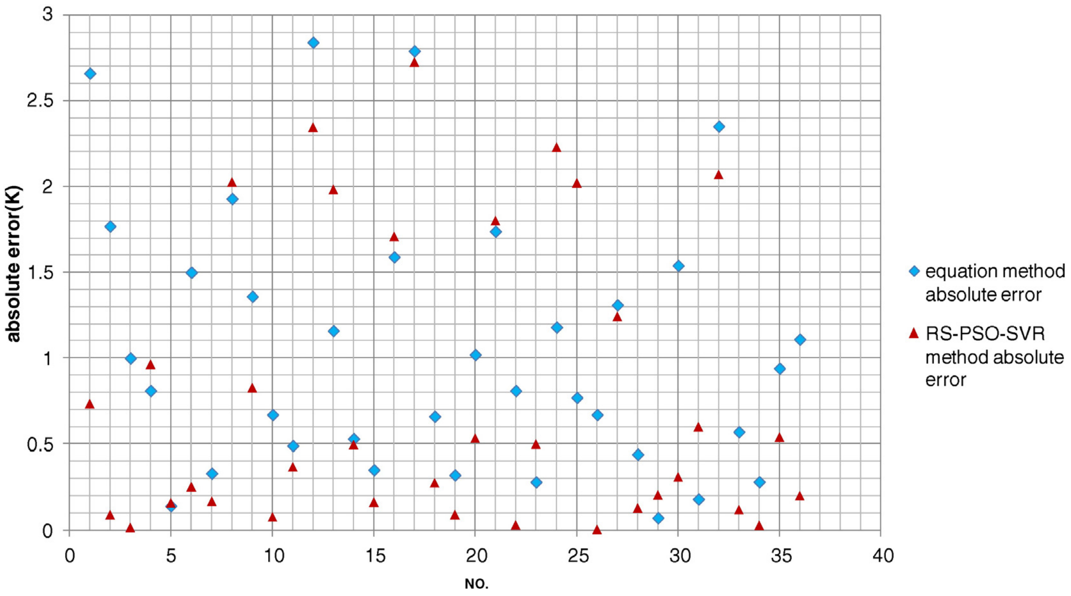


Fig. 1. Pair wise comparison of testing absolute error by using the equation method and RS-PSO-SVR method.

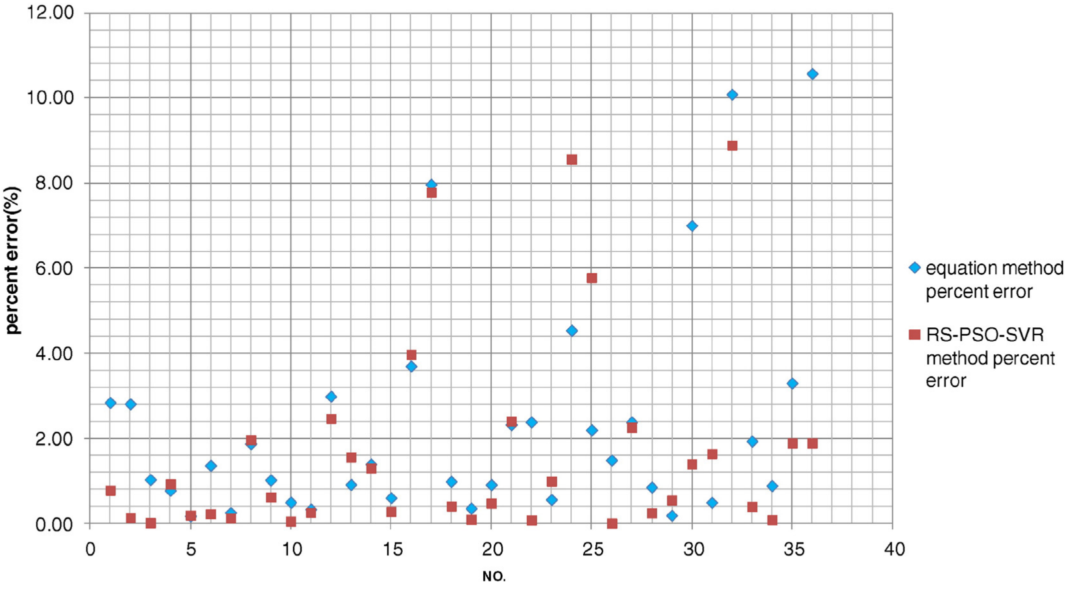


Fig. 2. Pair wise comparison of testing percent error by using the equation method and RS-PSO-SVR method.

Step 6: For each dimension in the copied training set, calculate the value S through the formula below:

$$S = \left[\sum_{i=0}^{100} \sum_{j=1}^{m} T_{ji} \cdot P_i \right] / \sum_{i=0}^{100} \sum_{j=1}^{m} T_{ji} \tag{6}$$

Step 7: Deal with each dimension through the following function:

$$S' = \frac{S}{100} \tag{7}$$

Step 8: For each dimension in both training set and test set, multiply all the data by its weight S'.

Step 9: Train the new training set with PSO-SVR and predict T_{C0}'s by the test set.

2.3. Generalization performance evaluation

Three indexes, i.e., mean absolute error (MAE), mean absolute percentage error ($MAPE$) and correlation coefficients (R) are adopted for generalization performance evaluation. They are formulated as following:

$$MAE = \frac{1}{n} \sum_{j=1}^{n} |\,\hat{y}_j - y_j|, \tag{8}$$

$$MAPE = \frac{1}{n} \sum_{j=1}^{n} \left| \frac{\hat{y}_j - y_j}{y_j} \right| \tag{9}$$

Table 5
Comparison of the prediction performance.

Regression method	Min Absolute Error	Max Absolute Error	RMSE/K	MAE/K	MAPE/%	R
Equation	0.07	2.84	1.31	1.06	2.33	0.9994
PSO-SVR	0.02	3.26	1.42	1.08	2.30	0.9993
RS-PSO-SVR	0	2.72	1.14	0.78	1.69	0.9996
BPNN	0.09	11.57	2.54	1.60	5.59	0.9978

$$R = \sum_{j=1}^{n} \left(\hat{y}_j - y\right)\left(y_j - \bar{y}\right) \Bigg/ \sqrt{\sum_{j=1}^{n} \left(\hat{y}_j - y\right)^2 \sum_{j=1}^{n} \left(y_j - \bar{y}\right)^2} \tag{10}$$

where, n denotes the number of test samples, y_j and $\hat{y}_j$ stand for the target (measured) and predicted values of the jth test sample respectively, and $\bar{y}$ is the mean target value for all test samples.

3. Establishment of prediction model

The dataset used in this study is derived from the literatures [6–7]. It contains four vector data on 36 samples: the measured T_{C0}, the equation T_{C0}, the distance between interacting layersζ, and the calculated spacing between interacting charges within layers ℓ. The target value is the measured T_{C0}. The relevant electronic and structural parameters for these compounds are given in Table 1. Table 2 lists the fundamental statistic information of the input variables for the prediction model.

The RS-PSO-SVR prediction method is used to establish a prediction model. The 36 sample data are used to train the model as well as to validate the model. In the beginning, the vectors of ζ and ℓ are used as the input vectors of PSO-SVR training model. The test result indicates that MAE is 3.96 K; $RMSE$ is 5.28 K and $MAPE$ is 7.74%. Although the $RMSE$

Table 6
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} [26,27].

No.	Superconducting compounds	ζ(Å)	l (Å)	Equation.val (K)	Meas.val (K)
1	$(Ca_{0.45}La_{0.55})(Ba_{1.30}La_{0.70})Cu_3O_y$	2.1297	7.1176	82.3	80.5
2	$Na_{0.16}(PC)_yTiNCl$	7.6735	25.528	6.37	6.3
3	$Na_{0.16}(BC)_yTiNCl$	7.7803	25.528	6.28	6.9
4	$Li_{0.08}ZrNCl$	1.5817	54.95	14.35	15.1
5	$Li_{0.13}(DMF)_yZrNCl$	3.4	26.397	13.9	13.7
6	$Na_{0.25}HfNCl$	1.658	29.864	25.19	24
7	$Li_{0.2}HfNCl$	1.595	38.505	20.31	20
8	$Eu_{0.08}(NH_3)_yHfNCl$	2.6686	19.246	24.29	23.6
9	$Ca_{0.11}(NH_3)_yHfNCl$	2.7366	20.113	22.66	23
10	$Li_{0.2}(NH_3)_yHfNCl$	2.7616	21.082	21.43	22.5
11	$TlBa_{1.2}La_{0.8}CuO_5$	1.9038	14.684	44.62	45.4
12	$Tl_{0.7}LaSrCuO_5$	1.8368	17.138	39.63	37

Table 7
The results of the RS-PSO-SVR prediction model.

No.	Meas.val (K)	Equation.val (K)	Prediction.val 1(K)	Prediction.val 2(K)	Prediction.val 3(K)	Prediction.val 4(K)	Prediction.val 5(K)	Average Prediction.val (K)
1	80.5	82.3	82.03	81.89	82.10	81.33	81.31	81.73
2	6.3	6.37	7.23	7.48	7.34	7.66	7.45	7.43
3	6.9	6.28	7.18	7.20	7.20	7.20	7.16	7.19
4	15.1	14.35	14.78	14.57	14.81	14.76	14.71	14.73
5	13.7	13.9	13.88	13.77	13.77	13.67	13.72	13.76
6	24	25.19	24.20	24.14	24.20	24.20	24.14	24.17
7	20	20.31	18.86	18.99	18.88	18.89	18.88	18.90
8	23.6	24.29	24.35	24.48	23.95	24.49	24.02	24.26
9	23	22.66	22.67	22.68	22.48	22.88	22.77	22.69
10	22.5	21.43	21.93	21.60	21.57	21.63	21.43	21.63
11	45.4	44.62	44.35	44.28	44.29	44.28	44.37	44.32
12	37	39.63	39.38	39.39	39.25	39.31	39.32	39.33

5.28 K is larger than that of the equation value 1.31 K, the performance denotes that ζ and ℓ have great correlation with T_{C0}.

To take advantage of the prior experimental data, the T_{C0} can be predicted through the three vectors: the T_{C0} calculated by the Eq. (1)(equation value), the distance between interacting layers ζ and the calculated spacing between interacting charges within layers ℓ. The weights of these three vectors are obtained through RS theory. Then, the combination of the weights and the vectors are acted as the input vectors of the proposed methods. The predicted result is more persuasive, and the generalization ability is stronger, as discussed in the next section.

4. Results and discussion

The calculated feature weights of three input vectors based on RS are shown in Table 3. The weight indicates the importance of the vector and the predicted value. From Table 3, it can be seen that the ζ makes the greatest influence on the predicted result, followed by the equation value and ℓ. The application of RS theory on preprocessing the given sample data set makes the characteristic values of all data more general. The normalization makes the data processing easy. Moreover, the normalized data can improve the prediction abilities of the model.

Here, the given data set is divided into two categories: training dataset and test dataset. The modeling and prediction are conducted by using all the 36 samples based on back propagation neural network (BPNN), PSO-SVR and RS-PSO-SVR approaches via leave-one-out cross validation (LOOCV) [25]. Once the prediction model is trained, it can be used to test its performance on test data. The test data is given as an input to the trained prediction model. The generated results are then used to compute overall performance. Table 4 shows the measured and estimated T_{C0} and the absolute errors by using Eq. (1), BPNN, PSO-SVR and RS-PSO-SVR method. It can be seen that, the error of almost every sample is small. The proposed RS-PSO-SVR method outperforms the equation method in 25 out of 36 samples. But for sample 4,5,8,13,16,21,23,24,25,29,31, RS-PSO-SVR method achieves larger errors than equation method. However, comparing with equation method, RS-PSO-SVR method obtains more accurate results in majority samples. Figs. 1 and 2 intuitively indicate that RS-PSO-SVR method has better performances of predicting T_{C0} than that of equation method. The vast majority points (RS-PSO-SVR method absolute error and percent error) are much closer to the zero-error-line than the majority points (equation method absolute error and percent error). It means that the proposed method RS-PSO-SVR improves the prediction abilities of the T_{C0}. Table 4 also shows the comparison of the PSO-SVR model and BPNN model on the prediction result. In majority samples, PSO-SVR method achieves smaller errors than BPNN method. Especially, the absolute error of 36th sample is up to 11.57 K by BPNN model, while it is only 0.54 by PSO-SVR model. The 36th sample is organic κ–[BEDT-TTF]$_2$Cu[N(CN)$_2$]Br, which is very different from other superconductors. Probably, there are no other organic superconductors to be trained by the BPNN model, which leads to the big prediction error. It offers another perspective on the higher generalization performance of PSO-SVR model than that of BPNN model.

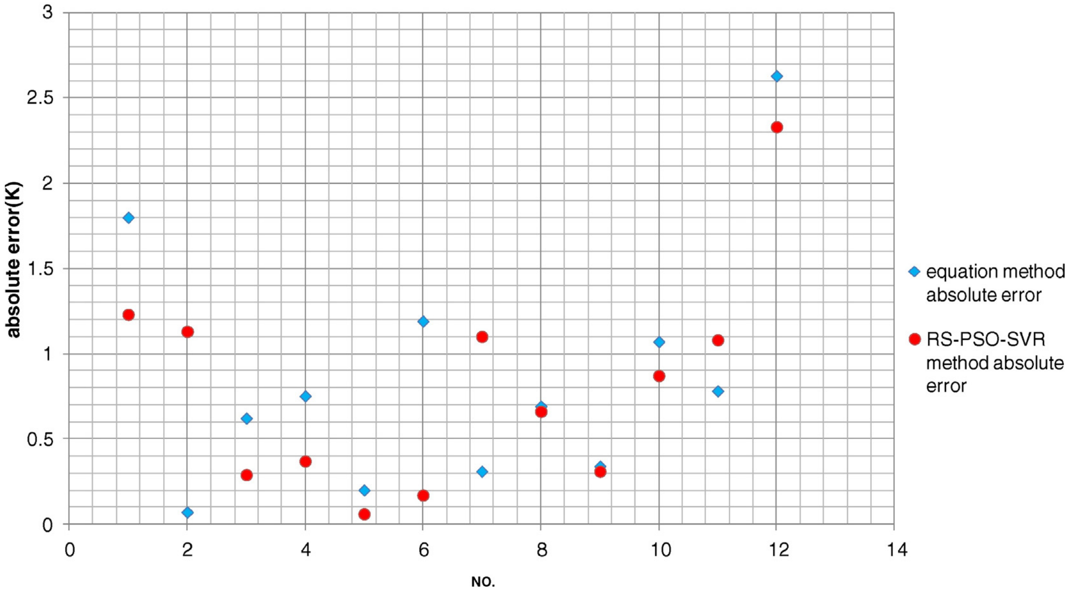


Fig. 3. Pair wise comparison of validation absolute error by using the equation method and RS-PSO-SVR method.

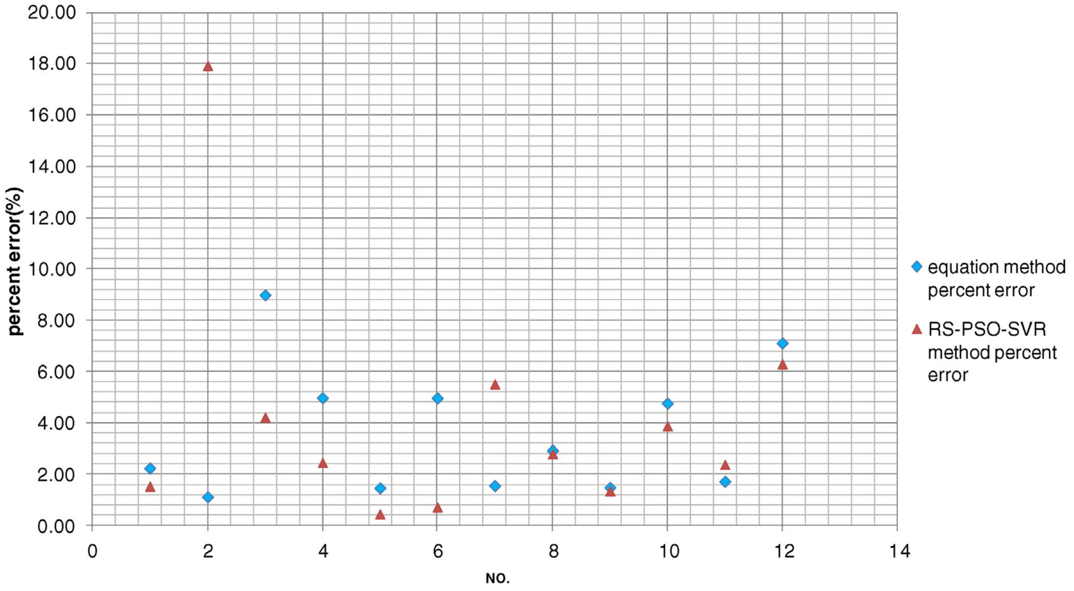


Fig. 4. Pair wise comparison of validation percent error by using the equation method and RS-PSO-SVR method.

Table 5 gives a comparison of the prediction performance of RS-PSO-SVR, BPNN, PSO-SVR, and equation method. The R ($R = 0.9996$) of the RS-PSO-SVR model is the greatest, followed by the equation model ($R = 0.9994$) and PSO-SVR model($R = 0.9993$). But the *MAPE* of the RS-PSO-SVR model is 1.69%, which is smaller than those (2.33% and 2.30%) for the equation model and PSO-SVR model calculated results. The *MAE* and *RMSE* ($MAE = 0.78$ and $RMSE = 1.14$) for the predicted results estimated by RS-PSO-SVR are also less than those of equation ($MAE = 1.06$ and $RMSE = 1.31$) and PSO-SVR ($MAE = 1.08$ and $RMSE = 1.42$). Thus, the prediction errors for the calculations based on the RS-PSO-SVR model are smaller than those of equation model and PSO-SVR model. It means that the prediction result of RS-PSO-SVR model is more accurate than that of the equation model and PSO-SVR model. Comparing the PSO-SVR and the BPNN method, The *MAE MAPE* and *R* ($MAE = 1.08$ $MAPE = 2.30$ and $R = 0.9993$) for the predicted results estimated by PSO-SVR are better than those of BPNN method ($MAE = 1.60$ $MAPE = 5.59$ and $R = 0.9978$). It indicates that the generalization performance of PSO-SVR model is better than that of BPNN method.

The performance of RS-PSO-SVR is better than that of equation method and PSO-SVR method. The reason will be briefly explained. The high temperature superconductors display remarkable variety in their crystalline, magnetic,and electronic band structures. They obey the algebraic relation, Eq. (1), within accuracy of ± 1.31 K. The reason why the *RMSE* and *MAE* for the predicted results estimated by PSO-SVR are larger than that by equation method is that PSO-SVR does not use the prior experimental data. On the other hand, the *MAPE* of the PSO-SVR model is 2.30%, which is smaller than those (2.33%) for the equation model calculated results. It denotes that machine learning methods have their own advantages. Using the prior experimental data, RS theory makes the characteristic values of all data more general. In the processing of training and testing, the weights are adjusted to achieve the optimal predicted values. RS-PSO-SVR can achieve the best *RMSE MAE* and *MAPE*. Overall, the results above suggest that the prediction accuracy of RS-PSO-SVR is quite superior to that of PSO-SVR and equation method.

With addition of the 12 compounds [26,27], the list of the compounds behaving in a manner consistent with the Eq. (1) has grown to 48 in total. Eq. (1) has been validated through extensive study of experimental results for high temperature superconductors based on cuprate, ruthenate, rutheno-cuprate, iron-pnictide, iron-chalcogenide, organic, and intercalated group-5-metal nitride chloride structures, with T_{C0} ranging from 6.3 K to 145 K [27]. Here, these additional samples are used to validate the prediction model. The relevant electronic and structural parameters of these compounds are given in Table 6.

Now, we use the 36 samples in Table 1 to train the RS-PSO-SVR prediction model and the data of Table 6 to validate the prediction model. The results of the prediction model are given in Table 7. The proposed RS-PSO-SVR method outperforms the equation method in 9 out of 12 samples. But for sample 2,7,11, RS-PSO-SVR method leads to larger errors than equation method. However, comparing with equation method, RS-PSO-SVR method obtains great result in most samples, which indicates that the proposed method improves the prediction abilities of the T_{C0}. It is intuitively showed in Figs. 3 and 4.

Table 8 gives a comparison of the prediction performance of RS-PSO-SVR and equation method. The R($R = 0.9988$) of the RS-PSO-SVR model is equal to the R of the equation model. The *MAPE* of the RS-PSO-SVR model is 4.12%, which is larger than that (3.6%) of the equation model. The main reason is that the Meas.val of the second (6.3 K) and the seventh (20 K) sample are small, while the absolute errors (1.13 K and −1.1 K) are large. Despite all these, the *MAE* of the RS-PSO-SVR model is 0.80, smaller than 0.87 K obtained by the equation model. And the RMSE of the RS-PSO-SVR model is 1.03 K, which is also smaller than 1.13 K from the equation model. The new data provide additional support for the universality of the high temperature superconductor prediction model described here. It denotes that RS-PSO-SVR is one kind of more effective techniques than the equation method to predict the transition temperature T_{C0} for high temperature superconducting compounds.

Table 8
Comparison of the prediction performance.

Regression method	Min Absolute Error	Max Absolute Error	RMSE/K	MAE/K	MAPE/%	R
Equation	0.07	2.63	1.13	0.87	3.60	0.9988
RS-PSO-SVR	0.06	2.33	1.03	0.80	4.12	0.9988

5. Conclusion

The RS-PSO-SVR method, using the prior experimental data, was used to predict the transition temperature T_{C0} for the high temperature superconductors in this work. Firstly, the characteristic values of all data were normalized through the preprocessing of RS theory. Secondly, we compared RS-PSO-SVR with the equation method, PSO-SVR and BPPN. The predicted results with 36 samples suggested that RS-PSO-SVR is a practical technique to predict the T_{C0} for high temperature superconducting compounds more effective than the other methods studied in this work. Finally, we conducted validation on the 12 samples from the recent works of Harshman and Fiory [26,27], and the results indicated the effectiveness of the proposed prediction model. Generally speaking, the machine learning methods can be applied in some advanced materials area and RS-PSO-SVR may be used to predict the T_{C0} of new superconductors.

Acknowledgements

This work is supported by Shanghai Institute of Materials Genome (Project No. 14DZ2261200) from the Shanghai Municipal Science and Technology Commission.

References

[1] J.G. Bednorz, K.A. Müller, Possible highT_C superconductivity in the Ba — La — Cu — O system, Z. Phys. B: Condens. Matter 64 (2) (1986) 189–193.

[2] D.R. Harshman, A.P. Mills Jr., Concerning the nature of high-T_C superconductivity: survey of experimental properties and implications for interlayer coupling, Phys. Rev. B 45 (18) (1992) 10684.

[3] J.J. Ying, X.F. Wang, X.G. Luo, et al., Superconductivity and magnetic properties of single crystals of $K_{0.75}Fe_{1.66}Se_2$ and $Cs_{0.81}Fe_{1.61}Se_2$, Phys. Rev. B 83 (21) (2011) 212502.

[4] J. Guo, S. Jin, G. Wang, et al., Superconductivity in the iron selenide $K_xFe_2Se_2$ ($0 \leq x \leq 1.0$), Phys. Rev. B 82 (18) (2010) 180520.

[5] D.R. Harshman, A.T. Fiory, Charge compensation and optimal stoichiometry in superconducting $(Ca_x La_{1-x})(Ba_{1.75-x}La_{0.25+x}) Cu_3 O_y$, Phys. Rev. B 86 (14) (2012) 144533.

[6] D.R. Harshman, A.T. Fiory, J.D. Dow, Theory of high-T_C superconductivity: transition temperature, J. Phys. Condens. Matter 23 (29) (2011) 295701.

[7] D.R. Harshman, A.T. Fiory, The superconducting transition temperatures of $Fe_{1+x}Se_{1-y}$, $Fe_{1+x}Se_{1-y}Te_y$ and $(K/Rb/Cs)_zFe_{2-x}Se_2$, J. Phys. Condens. Matter 24 (13) (2012) 135701.

[8] Z. Pawlak, Rough sets, Int. J. Comput. Inform. Sci. 11 (5) (1982) 341–356.

[9] Executive office of the president national science and technology council, Materials Genome Initiative for Global Competitiveness, June 2011.

[10] Y.F. Wen, C.Z. Cai, X.H. Liu, et al., Corrosion rate prediction of 3C steel under different seawater environment by using support vector regression, Corros. Sci. 51 (2) (2009) 349–355.

[11] F. Jinsong, F. Tingiian, A method of pattern classification on line based on rough set and SVM algorithm, Pattern Recognit. Artif. Intell. 4 (2000) 419–423.

[12] Huang Z, Guo J. A data preprocessing algorithm based on rough set for SVM classifier//Control System, Computing and Engineering (ICCSCE), 2013 IEEE international conference on. IEEE, 2013: 441–444.

[13] V.N. Vapnik, The Nature of Statistical Learning Theory, Springer-Verlag, New York, 1995.

[14] C.Z. Cai, W.L. Wang, L.Z. Sun, et al., Protein function classification via support vector machine approach, Math. Biosci. 185 (2) (2003) 111–122.

[15] C.Z. Cai, L.Y. Han, Z.L. Ji, et al., SVM-prot: web-based support vector machine software for functional classification of a protein from its primary sequence, Nucleic Acids Res. 31 (13) (2003) 3692–3697.

[16] C.Z. Cai, G.L. Wang, Y.F. Wen, et al., Superconducting transition temperature T_C estimation for superconductors of the doped MgB_2 system using topological index via support vector regression, J. Supercond. Nov. Magn. 23 (5) (2010) 745–748.

[17] T.O. Owolabi, K.O. Akande, S.O. Olatunji, Support vector machines approach for estimating work function of semiconductors: addressing the limitation of metallic plasma model, Appl. Phys. Res. 6 (5) (2014) p122.

[18] O. Abuomar, S. Nouranian, R. King, et al., Comprehensive mechanical property classification of vapor-grown carbon nanofiber/vinyl ester nanocomposites using support vector machines, Comput. Mater. Sci. 99 (2015) 316–325.

[19] A. Majid, A. Khan, G. Javed, et al., Lattice constant prediction of cubic and monoclinic perovskites using neural networks and support vector regression, Comput. Mater. Sci. 50 (2) (2010) 363–372.

[20] T.O. Owolabi, K.O. Akande, S.O. Olatunji, Development and validation of surface energies estimator (SEE) using computational intelligence technique, Comput. Mater. Sci. 101 (2015) 143–151.

[21] C. Cortes, V. Vapnik, Support-vector networks, Mach. Learn. 20 (3) (1995) 273–297.

[22] S.F. Fang, M.P. Wang, W.H. Qi, et al., Hybrid genetic algorithms and support vector regression in forecasting atmospheric corrosion of metallic materials, Comput. Mater. Sci. 44 (2) (2008) 647–655.

[23] R.C. Eberhart, J. Kennedy, A new optimizer using particle swarm theory, Proceedings of the sixth international symposium on micro machine and human science, 1 1995, pp. 39–43.

[24] X.G. Shao, H.Z. Yang, G. Chen, Parameters selection and application of support vector machines based on particle swarm optimization algorithm, Kongzhi Lilun yu Yingyong 23 (5) (2006) 740–743.

[25] S. Peng, Q. Xu, X.B. Ling, et al., Molecular classification of cancer types from microarray data using the combination of genetic algorithms and support vector machines, FEBS Lett. 555 (2) (2003) 358–362.

[26] D.R. Harshman, A.T. Fiory, Comment on "superconductivity in electron-doped layered TiNCl with variable interlayer coupling", Phys. Rev. B 90 (18) (2014) 186501.

[27] D.R. Harshman, A.T. Fiory, Superconducting interaction charge in thallium-based high-T_C cuprates: roles of cation oxidation state and electronegativity, J. Phys. Chem. Solids 85 (2015) 106–116.