CrossMark

ORIGINAL ARTICLE

Silicon content prediction and industrial analysis on blast furnace using support vector regression combined with clustering algorithms

Changchun Hua[1] · Jinhua Wu[1] · Junpeng Li[1] · Xinping Guan[1,2]

Abstract Silicon content prediction is quite significant for supervising the state of blast furnace and is usually selected as the indicator to represent the thermal state. In practical industry, the fluctuation exists in the operation of blast furnace all the time. What's worse, it is inaccurate to build the predictive model with many outliers. To solve these problems, this paper has developed a model to predict the silicon content using support vector regression (SVR) combined with clustering algorithms, including hard C-means (HCM) clustering and fuzzy C-means (FCM) clustering. Through data processing, the data points are clustered based on the similarity, and then different SVR models are established. In order to make full use of FCM, a new method using multiple SVRs and FCM based on membership degree (MFCM-SVRs) is proposed where the membership degree is applied to eliminate the outliers. Simulation results verify that the multiple SVRs based on HCM (HCM-SVRs) and MFCM-SVRs possess superiority in terms of accuracy and speed, which makes the method serve better for practical production.

Keywords Silicon content prediction · Fuzzy C-means clustering · Support vector regression · Blast furnace

✉ Jinhua Wu
 wjh_ysu@163.com

1 Institute of Electrical Engineering, Yanshan University, Qinhuangdao 066004, Hebei, China

2 Institute of Electronic, Information, and Electrical Engineering, Shanghai Jiaotong University, Shanghai 200240, China

1 Introduction

Iron and steel industry occupies an important position in national economy, with the production and quality being one of the most crucial signs of a national economic development. Figure 1 displays the ironmaking process from iron ore to steel. Solid charge, such as coke, iron and solvent are fed from the top of blast furnace. When the coke arrives in the tuyere raceway, it will be burned by the oxygen combined with other auxiliary fuels in the hot air, and then blast furnace gas will be produced. The gas rises from bottom to top, with temperature decreasing and oxygen content increasing. At last, the gas comes out from the top of blast furnace and will be purified in the gas cleaning system. In the process of gas rising, the iron ore is falling from top to bottom. As a result, the iron ore will be heated and reduced by the gas. Contrary with gas transformation, the iron ore changes with temperature increasing and oxygen decreasing. Based on the characteristics of temperature, chemical component and physical form of furnace burden, blast furnace can be divided into five regions, which are lump zone, cohesive zone, dripping zone, raceway and hearth region [1]. Since the state of reaction cannot be described in detail and the content of internal components cannot be calculated accurately either, black-box models have attracted more scientists' attention and achieved further process [2, 3].

It becomes necessary to construct mathematical models to reflect the operational state, as there are hundreds of physical changes and chemical reactions happening in gas–liquid–solid phase under high pressure and high temperature when a blast furnace runs. The thermal state is one of the most critical state factors as it reflects much information about the running state of blast furnace [4, 5]. Since the thermal state cannot be obtained directly, the silicon
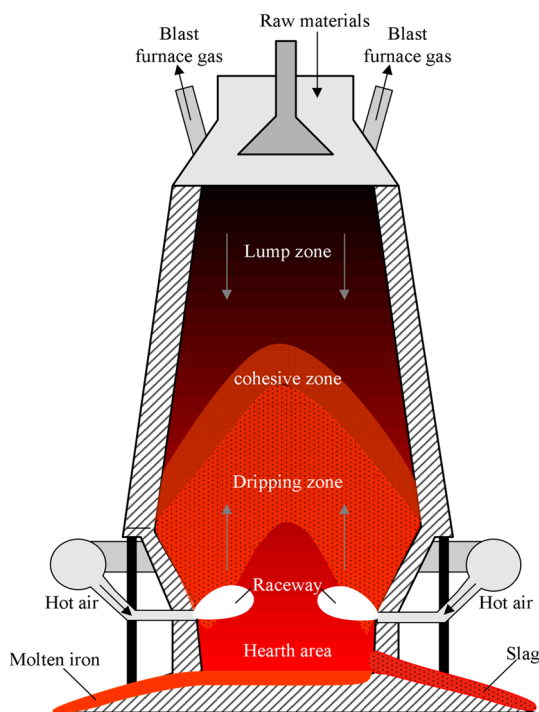
Fig. 1 Blast furnace structure diagram

content is usually selected as the indicator to supervise the state of blast furnace. It has been demonstrated that the main silicon behavior at the blast furnace process is reduction reaction [6]. Firstly, a part of silicon goes to hot metal at liquid phase during SiO_2 reduction by coke carbon or the carbon dissolved in hot metal, which is displayed as follows

$$SiO_2^1 + 2[C] = [Si] + 2CO^g \tag{1}$$
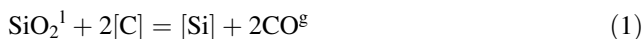
While most of the silicon transforms into gaseous SiO, and the corresponding SiO_2 is from coke ash and slag.

$$SiO_2^1 + [C] = SiO^g + CO^g \tag{2}$$

Then SiO rises up with blast furnace gas and will be dissolved by slag and hot metal from the cohesive zone. The dissolved SiO will react with coke in metal.

$$SiO^g + [C] = [Si] + CO^g \tag{3}$$

According to researches based on silicon behavior at blast furnace, the silicon content is associated with the cohesive zone position, the hearth temperature and many other blast furnace variables [6]. The correlation makes the silicon content a research focus that numbers of mathematical models have been developed, including neural networks [7, 8], partial least squares [9], Bayesian networks [10] and support vector machines [11, 12]. Some of these data-driven models have been applied into practice, which can

provide guidance on modulating input parameters in advance to control the silicon content in a proper range.

Support vector machine (SVM), developed by Vapnic [13], is a kind of kernel-based black-box modeling method. The basic idea of SVM is mapping inputs into a high-dimensional feature space where two classes can be separated to the utmost extent by an optimal hyperplane. With the structural risk minimization, SVM possesses obverse superiority on generalization ability [14]. As a result, SVM has been widely applied on nonlinear systems forecasting [15], data mining [16] and document classification [17]. When SVM is applied to tackle the problems of function approximation and regression estimation, SVR is proposed. SVR is capable of approximating nonlinear function and effectively avoids over fitting. Thus, growing efforts have been made to apply SVR to various domains [18–20].

Based on the industrial data, the fluctuation exists in the operation of the ironmaking progress just as Fig. 2 shows. Both of gas permeability and silicon content vary in a large range, which can also be demonstrated by other variables. This variation would cause difficulty in modeling process inevitably. What's worse, the existence of noises and outliers can result in inaccuracy, which would decrease modeling effectiveness [21]. In view of this fact, the data should be processed at first.

Clustering analysis, based on the idea that things are together by their attribute, is a multivariate statistical analysis method. As a means of unsupervised classification, clustering analysis has been widely applied in pattern recognition [22], data mining [23] and computer vision [24], etc. The clustering algorithm partitions a data set into several clusters, by which the data points in one cluster are similar to each other, while the data points in different clusters have some different properties. Hard C-means clustering (HCM) partitions the data set by defining a hard boundary. If the boundaries between subpartition are vague, which means that each data points belong to different clusters with different membership degrees, then fuzzy C-means clustering (FCM) is developed. Duan et al. [25] introduced the weighted SVM-based fuzzy C-means clustering algorithm, where the best training sample set is obtained by FCM clustering. Multiobjective fuzzy clustering combined with SVM was proposed by Mukhopadhyay et al. [26]. In this approach, some high-confidence points are selected to train SVM classifier. Unfortunately, multiple support vector machines were not constructed in these two papers. Yang et al. [27] have proposed a kernel fuzzy C-means clustering-based fuzzy support vector machine algorithm, where the data are clustered in a high-dimensional feature space and the farthest pair of clusters is selected to degrade the effects the outliers or noises have on the decision function. In practice, as the data collected
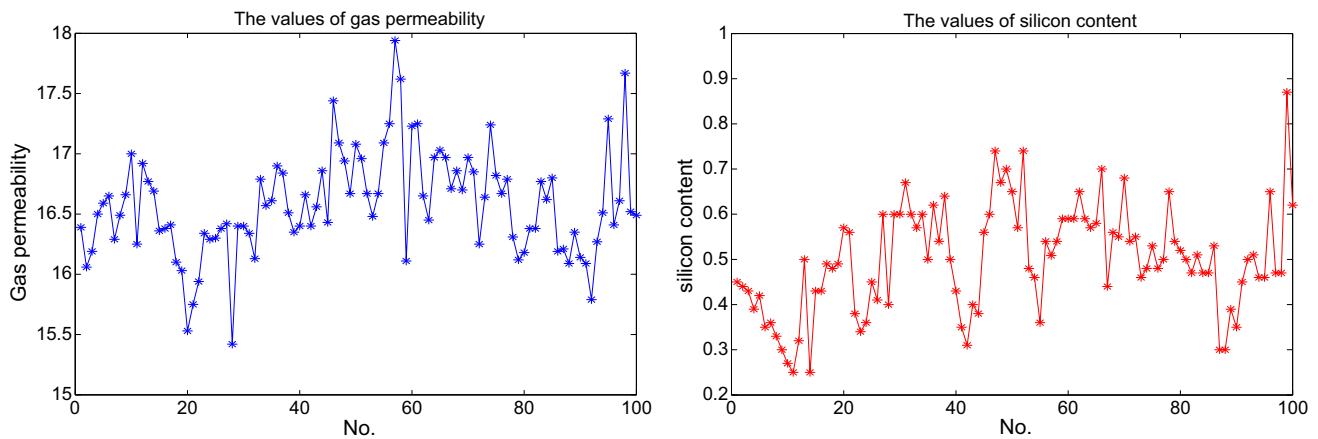
Fig. 2 The fluctuations exist in the operation of the ironmaking progress

have homogenous characteristics to some extent, clustering in the original space becomes quite meaningful.

In this paper, some models based on SVR combined with clustering algorithms are presented. Firstly, HCM-SVRs model is constructed where a hard boundary is defined with which the data are partitioned into different clusters exactly, and then SVR is introduced to predict the silicon content. Different from HCM, FCM clustering can obtain the fuzzy partition with membership degrees, by which MFCM-SVRs are proposed. In this method, the outliers are eliminated twice based on FCM clustering and membership degrees. Multiple SVRs are constructed on each cluster to predict the next silicon content.

The remaining parts of this paper are arranged as follows. Section 2 gives a brief introduction on SVR. The silicon content prediction using support vector regression and clustering algorithms is introduced in Sect. 3. Section 4 displays the predictive procedures and results analysis. Finally, Sect. 5 concludes this paper.

2 Support vector regression

Support vector regression (SVR), a regression method [28], is based on SVM which improved the capability of generalization by searching the structural risk minimization. Different from SVM, SVR attempts to predict the distribution of information and is applied for regression estimation. As a result, SVR is used to acquire the silicon content precisely in this paper.

For a set of training data $T = \{(x_i, y_i), x_i \in R^n, y_i \in R, i = 1, 2, \ldots, l\}$, where x_i is the input vector, y_i is the corresponding output, and l is the number of training samples. SVR maps a low-dimensional input space to a high-dimensional feature space, and the nonlinear mapping can be expressed as

$$f(x) = \omega^T \psi(x) + b \tag{4}$$

where the weight vector ω and threshold value b can be obtained by solving an optimization problem.

$$\min_{\omega, b, \xi, \xi^*} \frac{1}{2} \omega^T \omega + C \sum_{i=1}^{l} \xi_i + C \sum_{i=1}^{l} \xi_i^*$$

$$\text{s.t.} \begin{cases} \omega^T \psi(x_i) + b - y_i \leq \varepsilon + \xi_i \\ y_i - (\omega^T \psi(x_i) + b) \leq \varepsilon + \xi_i^* \\ \xi_i, \xi_i^* \geq 0, \quad i = 1, 2, \ldots, l \end{cases} \tag{5}$$

where ε is the insensitive loss function that determined the admissible uncertainty of the data points, C is the penalization parameter, and ξ_i, ξ_i^* are slack variables that denote the training error.

Lagrange function and kernel function are introduced to transform Eq. (4) into quadratic optimization problem.

$$\max \sum_{i=1}^{l} (a_i^* - a_i) y_i - \varepsilon \sum_{i=1}^{l} (a_i^* + a_i) - \frac{1}{2} \sum_{i,j=1}^{l} (a_i^* - a_i)(a_j^* - a_j) k(x_i, x_j)$$

$$\text{s.t. } 0 \leq a_i, a_i^* \leq C \quad (i = 1, 2, \ldots, l)$$

$$\sum_{i=1}^{l} (a_i^* - a_i) = 0 \tag{6}$$

The decision regression function can be written as

$$f(x, a_i, a_i^*) = \sum_{i=1}^{l} (a_i - a_i^*) k(x_i, x) + b \tag{7}$$

where $k(x_i, x_j) = \psi(x_i) \cdot \psi(x_j)$ is kernel function. In this paper, the RBF function is adopted.

$$k(x_i, x_j) = \exp\left(-\frac{\|x_i - x_j\|^2}{2\sigma^2}\right) \tag{8}$$

where σ is the adjustable parameter of the kernel function.

However, when the SVR model is used to predict the silicon content, the fluctuation of the operation of blast furnace cannot be ignored. The data vary in a large range which would result in wrong support vectors selecting. This is an adverse factor for model precision. What's worse, if the training data are large, it would cost considerable amount of time to search the optimal parameters by grid search method. As a result, it is indispensable to process the input data into different classes to improve the prediction accuracy and the speed of modeling.

3 Silicon content prediction using support vector regression and clustering algorithms

This section presents the clustering analysis and a new silicon content prediction method using SVR and FCM clustering technique.

Clustering algorithm is used to partition the data into several clusters, so that the similar data can be divided into the same class [29]. And then, the data with similar properties are used for modeling. This method has two advantages. One is that the data set is divided into smaller ones which can improve the speed when modeling, especially searching the best parameters for SVR. The other advantage is that the accuracy can be increased because similar data are produced from similar blast furnace operational conditions, and they are more reasonable to be clustered in one set to build a model. Simulations also verify the increasing of accuracy in Sect. 4. The FCM algorithm is defined as minimizing the objective function

$$J_m = \sum_{i=1}^{n} \sum_{j=1}^{c} u_{ij}^m \parallel x_i - c_j \parallel^2$$
$$\text{s.t.} \sum_{j=1}^{c} u_{ij} = 1, \quad 0 \le u_{ij} \le 1 \tag{9}$$

where u_{ij} is the degree of membership of x_i to cluster j, m is a weighting parameter greater than 1, $X = \{x_1, x_2, \ldots, x_n,\} \subseteq R^d$ is the d-dimensional data set, c_j represents the jth center of the cluster, n and c are the numbers of samples and clusters, respectively.

Then the question can be translated into an iterative optimization of the objective function through the updates of membership u_{ij} and cluster center c_j

$$u_{ij} = \frac{1}{\sum_{k=1}^{c} \left[\frac{\|x_i - c_j\|}{\|x_i - c_k\|} \right]^{\frac{2}{m-1}}} \tag{10}$$

$$c_j = \frac{\sum_{i=1}^{n} u_{ij}^m x_i}{\sum_{i=1}^{n} u_{ij}^m} \tag{11}$$

The iteration will be stopped when $| u_{ij}^k - u_{ij}^{k-1} | < \varepsilon$, where ε is a parameter that defines the termination condition, and k is the iteration step.

While unlike FCM clustering, HCM defines the objective function as

$$J_m = \sum_{i=1}^{n} \sum_{j=1}^{c} \parallel x_i - c_j \parallel^2 \tag{12}$$

And then the membership and cluster center change accordingly.

$$u_{ij} = \begin{cases} 1, & \text{If } \parallel x_i - c_j \parallel = \min_{1 \le r \le c} \{x_i - c_r\} \\ 0, & \text{Others} \end{cases} \tag{13}$$

$$c_j = \frac{\sum_{i=1}^{n} u_{ij} x_i}{\sum_{i=1}^{n} u_{ij}} \tag{14}$$

From above, the main idea of silicon content prediction can be concluded as follows: (1) Use the HCM clustering and FCM clustering algorithms to cluster the input data. When using FCM, the data are clustered based on the largest membership degree. (2) Set the threshold of the membership degree after FCM clustering. If the membership degree of an input set is lower than the threshold, the input data will be eliminated as the distance to each center is long. These input data cannot be partitioned into the cluster with the largest membership, as it does not appear to be much different in membership values. (3) Construct the silicon content prediction model using SVR algorithm for each cluster.

This method can reduce the number of training samples and improve the accuracy and accelerate the speed. The details for this method are illustrated as follows. (1) Normalize the input variables. As different input variables have different orders of magnitude, these data cannot be used directly. So all the input variables should be normalized to the range 0–1 by Eq. (15).

$$\overline{x_i(k)} = \frac{x_i(k) - \min(x_i(k))}{\max(x_i(k)) - \min(x_i(k))} \tag{15}$$

where $\overline{x_i(k)}$ is the normalized value of the ith input. (2) Select more influential input variables and determine the time delay of every selected input. The methods will be illustrated in Sect. 4.1. Combined with the guidance of expert knowledge, the final input variables can be singled out. (3) Cluster the input data using HCM clustering algorithm considering the unstable operation of blast furnace and the number of experimental data. Unlike hard clustering method, FCM processes an advantage that each sample can belong to various clusters with different

membership degree. Taking advantage of this superiority, the input data will be processed for a second time. In this paper, select the largest membership of each sample to different clusters, and then the data can be separated into classes based on the largest membership. What's more, we propose to set a threshold when clustering the training set. If the largest membership of each sample is smaller than the threshold, the data will be eliminated. (4) Determine the parameters ε, C, and γ. The grid search method is introduced to search the optimal (C, γ), respectively. Tenfold cross-validation is applied to partition the training set into ten subsamples, and then the cross-validation process is repeated ten times with each of the ten subsamples used once as the testing data. The optimal (C, γ) is produced, while the MSE of the train set is the lowest which is defined as

$$\text{MSE} = \sqrt{\frac{1}{l} \sum_{i=1}^{l} (\widehat{y_i} - y_i)^2} \tag{16}$$

(5) Construct the silicon content prediction model using SVR algorithm for each cluster.

4 Experimental results and analysis

The data are collected from No. 2 blast furnace of Liuzhou Steel in China, whose volume is $2650 \, \text{m}^3$. In total, 500 group data are available to validate the method, and 400 data points are for the training set and 100 data points for test set. MATLAB 2010 is used to write source code. All experiments are carried out under the MATLAB 2010 environment with 2 GB memory. The CPU is Intel Pentium 32 G2030 with 3 GHz dominant frequency. In this paper, clustering method is introduced to classify the data based on the similarity. Considering the number of samples, the data will be divided into four clusters. The simulation results of HCM-SVRs and MFCM-SVRs will be compared with results of SVR and SVRs, which will verify the superiority of HCM-SVRs and MFCM-SVRs models.

4.1 Input variables selecting

In the report forms of blast furnace main parameters, 23 candidate variables are displayed with 1-h time interval. The last silicon content values are also selected as input vectors. However, not all the candidate variables have a significant influence on the fluctuation of silicon content. What's worse, too many inputs would add model complexity and reduce prediction accuracy. So input variables selecting becomes quite indispensable. Gray correlation

method is introduced to determine the influence degree that the input variables have on silicon content.

The gray correlation degree is a quantitative value calculating the correlation between factors. The higher the gray correlation degree is, the more relevant the variables are. The gray correlation method can be summarized as follows:

Step 1: Get the reference sequence $X_0 = (x_0(1), x_0(2), \ldots, x_0(n))$ and contrast series $X_i = (x_i(1), x_i(2), \ldots, x_i(n)), i = 1, 2, \ldots, m$, where X_0 is the silicon content series and X_i are the candidate input variables.

Step 2: The gray relational coefficient can be calculated as

$$\xi_i(k) = \frac{\Delta_{\min} + \rho \Delta_{\max}}{|y_0(k) - y_i(k)| + \rho \Delta_{\max}} \tag{17}$$

where the so-called management coefficient $\xi_i(k)$ is the relative difference value between the reference and contract sequences. $\rho \in [0, 1)$ is the distinguishing coefficient with function of improving the difference among the gray relational coefficients. In this paper, ρ is set as 0.5, and $\Delta_{\min}$ and $\Delta_{\max}$ are defined as follows

$$\Delta_{\min} = \min_i (\min_k |y_0(k) - y_i(k)|) \tag{18}$$

$$\Delta_{\max} = \max_i (\max_k |y_0(k) - y_i(k)|) \tag{19}$$

Step 3: After obtaining the gray relational coefficients, the mean of the coefficients has been always adopted as the gray correlation degree.

$$\gamma(x_0, x_i) = \frac{1}{n} \sum_{k=1}^{n} \xi_i(k) \tag{20}$$

where $\gamma(x_0, x_i)$ represents the gray correlation degree of the ith comparison series X_i to the reference series X_0.

Step 4: Table 1 displays the input variables and corresponding gray correlation degrees. Considering the number of inputs with expert guidance, the variables whose gray correlation degree is larger than 0.870 are selected as the final inputs.

For these input variables selected by gray correlation method, there exists time delay in the influence the inputs make on silicon content. Considering the accuracy and real production requirements, the time delay should not be ignored. Here, correlation analysis is introduced to obtain the more important inputs with different time delays. The correlation coefficient is defined as

$$R_i = \frac{\sum_{j=1}^{n} (x_{ij} - \overline{x_i})(y_j - \overline{y})}{\sqrt{\sum_{j=1}^{n} (x_{ij} - \overline{x_i})^2} \sqrt{\sum_{j=1}^{n} (y_j - \overline{y})^2}} \tag{21}$$

Table 1 The gray correlation degrees of candidate variables

Input variable	Gray correlation
Cold wind flow	0.8861
Feed wind ratio	0.8860
Hot blast pressure	0.8641
Furnace top pressure	0.8814
Pressure difference	0.8666
Top pressure blast volume ratio	0.8804
Gas permeability	0.8842
Drag coefficient	0.8827
Hot blast temperature	0.8737
Oxygen enrichment flow	0.8199
Oxygen enrichment percentage	0.8195
Pulverized coal injection	0.8702
Blast humidity	0.8864
Theoretical burning temperature	0.8832
Standard wind speed	0.8862
Actual wind speed	0.8772
Blast momentum	0.8689
Bosh gas volume	0.8848
Bosh gas index	0.8853
Top temperature (northeast)	0.8852
Top temperature (southwest)	0.8860
Top temperature (northwest)	0.8896
Top temperature (southeast)	0.8811
Silicon content $Si_{[n-1]}$	0.9160

Table 2 The correlation coefficients with different time delays

Input variable	q^{-3}	q^{-2}	q^{-1}	q^{0}
Cold wind flow	0.1150	0.0813	0.0107	0.0576
Feed wind ratio	0.1162	0.0824	0.0130	0.0552
Furnace top pressure	0.1072	0.0961	0.0569	0.0401
Top pressure blast volume ratio	0.0293	0.0252	0.0240	0.0731
Gas permeability	0.1356	0.1265	0.0626	0.0282
Drag coefficient	0.1151	0.1365	0.1366	0.0382
Hot blast temperature	0.2574	0.3250	0.4049	0.4489
Pulverized coal injection	0.1737	0.1972	0.1961	0.2281
Blast humidity	0.1007	0.1050	0.1134	0.1237
Theoretical burning temperature	0.2599	0.2774	0.2845	0.2276
Standard wind speed	0.1155	0.0810	0.0110	0.0570
Actual wind speed	0.2436	0.2611	0.2485	0.2205
Blast momentum	0.3257	0.3556	0.3452	0.2830
Bosh gas volume	0.1496	0.1196	0.0511	0.0205
Bosh gas index	0.1514	0.1247	0.0528	0.0176
Top temperature (northeast)	0.2874	0.2913	0.3136	0.2623
Top temperature (southwest)	0.2833	0.2881	0.3151	0.2932
Top temperature (northwest)	0.3575	0.3587	0.3745	0.3345
Top temperature (southeast)	0.3385	0.3549	0.3650	0.3349
Silicon content	0.4250	0.4723	0.5252	0.6157

Table 3 The numbers of training and test data of each cluster

	Cluster 1	Cluster 2	Cluster 3	Cluster 4	In total
Training set	86	188	123	3	400
Test set	24	42	34	0	100

where R_i is the ith correlation coefficient of input variables, $\overline{x_i}$, $\overline{y}$ represent the mean values of input and output variables, respectively. The time series are set as (0, 1, 2, 3 h). Table 2 shows the correlation coefficients of input variables with different time delays. After repeated experiments with different thresholds, the inputs with underline under the corresponding values in Table 2 are the final inputs, where q^{-1} represents the time an hour before.

4.2 Simulation results and analysis

Here four experiments are conducted based on these normalized input variables, one is to predict silicon content only by SVR, and then for further comparison with HCM-SVRs and MFCM-SVRs, the training data are separated into 3 classes based on the time sequence. To keep the consistency of numbers of each class with HCM clustering in Table 3, the training data from 1 to 86, 86 samples in total are partitioned into class 1; the training data from 87 to 274, 188 samples in total are partitioned into class 2; the training data from 275 to 397, 123 samples in total are partitioned into class 3. The test data are partitioned using

the same method as training data, and then multiple SVRs are constructed. The third one is using HCM-SVRs, and the last one is MFCM-SVRs.

Based on the number of available sample, the input data are classified into four clusters using HCM clustering. The numbers of training and test data of each cluster are displayed in Table 3.

It can be seen from Table 3 that there are only 3 groups of samples of training set and no test data in cluster 4. After checking the data in cluster 4, though the silicon content values are in normal range, the cold wind flow, feed wind ratio, furnace top pressure, gas permeability and wind speed and so on are obversely much lower than normal values. Based on the input data and clustering result, these samples in cluster 4 can be treated as outliers.

Taking advantage of membership degree in FCM clustering, the input data will be separated into classes based on the largest membership. The clustering result is the same

with HCM clustering which is displayed in Table 3. In this paper, the threshold of FCM clustering is set as 0.4. If the largest membership of each sample is smaller than 0.4, the data will be eliminated as outliers. It demonstrates that the sample cannot be put into the cluster with the largest membership smaller than 0.4. If the largest membership is too small, it means that the samples's memberships to the whole clusters do not differentiate much. In other words, the property of membership is not obvious. The numbers of samples after elimination are exhibited in Table 4.

The parameter ε in SVR model is supposed as 0.01 in the whole experiments, and then the grid search method is introduced to search the optimal (C, γ) in the grid set $\{2^{-10}, 2^{-9.5}, \ldots, 2^{9.5}, 2^{10}\} \times \{2^{-10}, 2^{-9.5}, \ldots, 2^{9.5}, 2^{10}\}$, respectively. Tenfold cross-validation is applied to partition the training set into ten subsamples, and then the cross-validation process is repeated ten times with each of the ten subsamples used once as the testing data. The optimal (C, γ) is produced, while the MSE of the train set is the lowest. Figure 3 displays the grid searching result for the optimal (C, γ) of SVR model. The simulation result demonstrates that the optimal (C, γ) is (0.5,1) for one SVR model. Table 5 displays the grid searching results of each method.

Table 5 demonstrates that after classification based on time sequence, the searching time for the optimal (C, γ) decreases quite a lot. What's more, when clustering algorithm is used to cluster the data, the time consumption of grid searching has been reduced even further, which is quite significant for industry production. To compare the effectiveness of HCM-SVRs and MFCM-SVRs exactly, the optimal (C, γ) are set the same which is searched just based on HCM-SVRs. Indeed, the time consumption of the two models is more or less the same.

Train SVRs on the processed data as well as the unprocessed data with the optimal parameters, respectively, and the predictive results are displayed as follows (Figs. 4, 5, 6, 7).

The MAPEs (mean absolute percentage errors) and MSEs of the test set based on each method are displayed in Table 6, where absolute percentage errors (APEs) are displayed as follows.

Table 4 The numbers of samples after elimination

	Cluster 1	Cluster 2	Cluster 3	In total
Training set	82	186	118	386
Test set	24	42	34	100

The parameter selection results for SVR [GridSearchMethod]
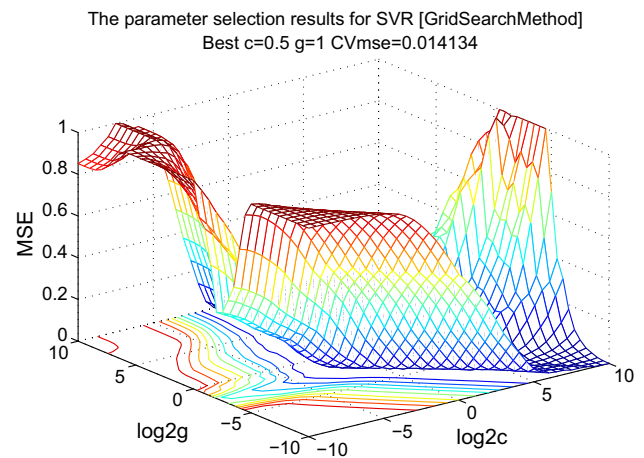Best c=0.5 g=1 CVmse=0.014134

Fig. 3 The grid searching result for the optimal (C, γ) of SVR model

Table 5 The grid searching results of each experiment

Model	Class	(C, γ)	Time/s
SVR	–	(0.5000, 1.0000)	666.5036
SVRs	Class 1	(0.1250, 0.7071)	28.6450
	Class 2	(1.0000, 0.5000)	136.1708
	Class 3	(0.7071, 0.0442)	53.4116
	In total	–	218.2274
HCM-SVRs /MFCM-SVRs	Cluster 1	(0.6250, 0.2500)	5.0873
	Cluster 2	(0.2500, 0.2500)	15.5661
	Cluster 3	(0.1250, 0.2500)	7.4034
	In total	–	28.0568

$$\mathrm{APE}_i = \frac{|\hat{y}_i - y_i|}{y_i} \times 100\% \tag{22}$$

Table 6 exhibits the comparison of SVR predictive errors based on SVR, SVRs, HCM-SVRs and MFCM-SVRs. It is obvious that the HCM-SVRs model and the MFCM-SVRs model achieve higher accuracy in test set than the SVRs model. The improvement is meaningful, and the classification model according to the fluctuation of blast furnace is practical. In conclusion, the HCM-SVRs and MFCM-SVRs models can forecast the trend of silicon content quite well.
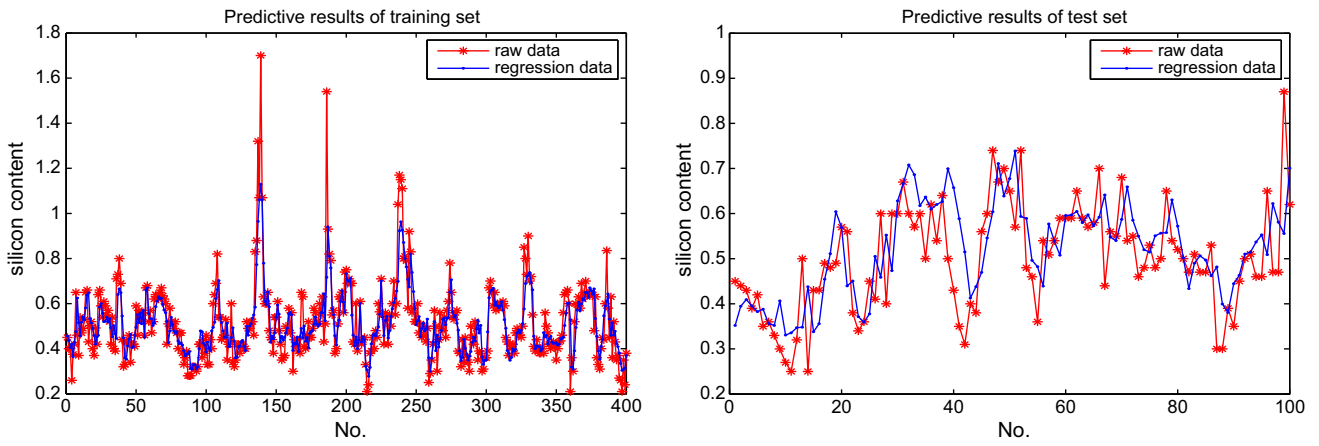
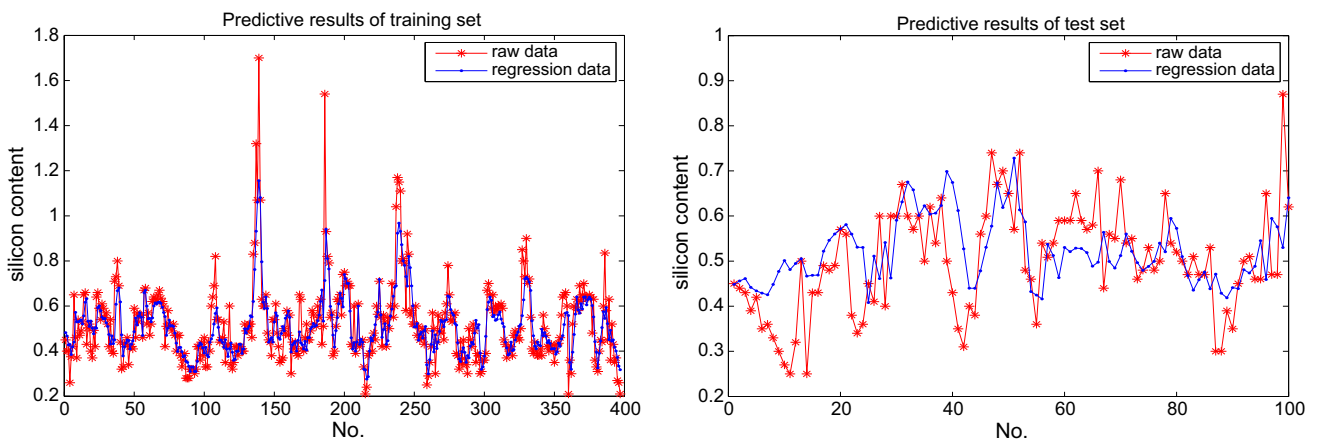**Fig. 4** Predictive results of training set and test set based on SVR

Fig. 5 Predictive results of training set and test set based on SVRs
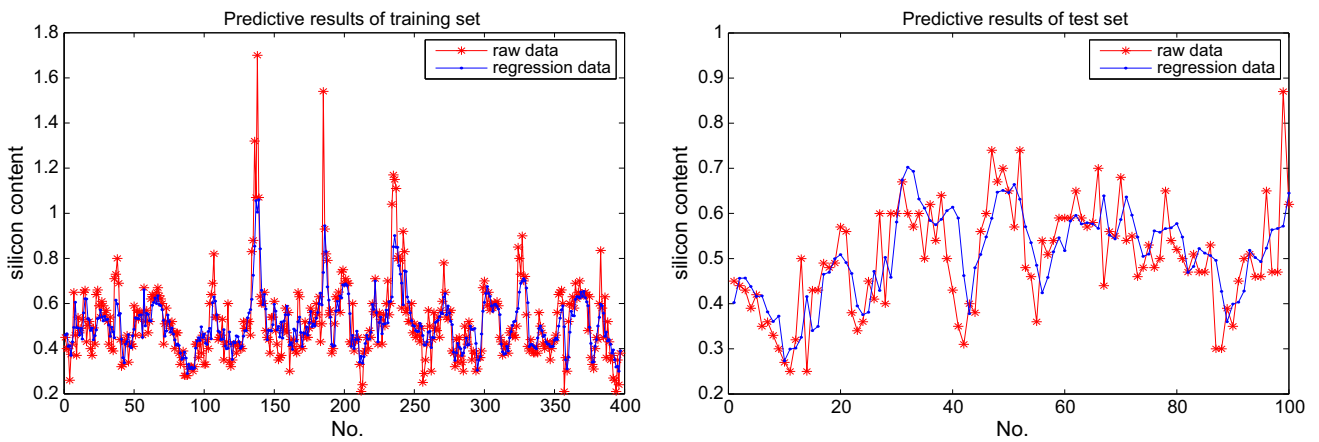

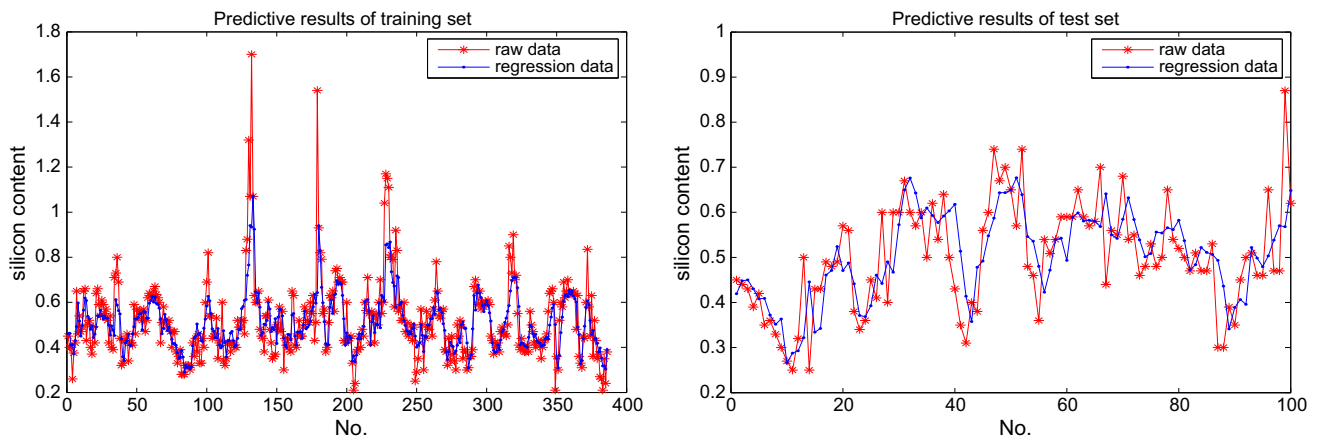**Fig. 6** Predictive results of training set and test set based on HCM-SVRs

Fig. 7 Predictive results of training set and test set based on MFCM-SVRs

Table 6 The MAPEs and MSEs of each method

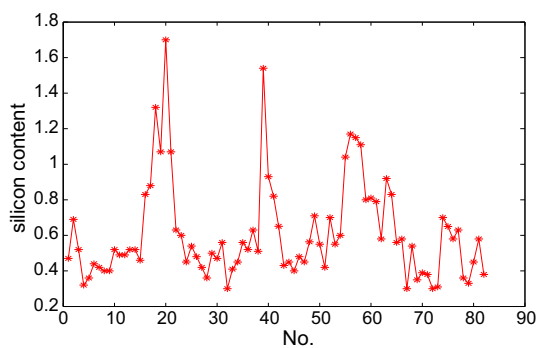| Model	Class	MAPEs	MSEs
SVR	–	0.1634	0.0977
SVRs	Class 1	0.2882	0.1218
	Class 2	0.1837	0.1133
	Class 3	0.1420	0.0998
	Mean	0.2046	0.1120
HCM-SVRs	Cluster 1	0.1989	0.1005
	Cluster 2	0.1268	0.0761
	Cluster 3	0.1357	0.0927
	Mean	0.1538	0.0903
MFCM-SVRs	Cluster 1	0.1891	0.0970
	Cluster 2	0.1261	0.0749
	Cluster 3	0.1354	0.0945
	Mean	0.1449	0.0893

Fig. 8 The silicon values of training set of cluster 1

4.3 Industrial analysis

In the previous discussion based on Table 6, the cluster 1 gets the lowest accuracy. The silicon values of training set of cluster 1 are displayed in Fig. 8

Figure 8 demonstrates that the higher silicon content values are almost classified into the same cluster, which proves that the data from one cluster have similar furnace status. The fuzzy partition of the data points in the original input space is quite meaningful in practical industry. Due to the limit of data points, the data set is just clustered into four classes in this experiment. However, the data are excessively plentiful in industry. As a result, through processing the big data, different data sets reflecting different furnace status would be clustered, which is pretty suggestive for the furnace operators. These clustered data can reflect different blast furnace operational conditions. The operators can make useful summaries and suitable decisions based on different conditions to adjust the operational state.

In this paper, we proposed to cluster the training set and test set as a whole, and then after clustering, the data are to train and test the model separately. Compared with clustering the training data, this method possesses an advantage that all data in the training set similar to test data can be used to train the prediction model. However, if the training data are clustered and the centers are determined, and the test data are partitioned based on the minimum distance to each center, some training points similar to test data may not make contributions to improve the prediction accuracy. To verify this superiority, another simulation is conducted. In this simulation, the training data are clustered by MFCM clustering and the test samples are partitioned based on the minimum distance to each center and then different SVRs are conducted. The results are displayed in Fig. 9. The MAPE and MSE are 0.1641 and 0.0925, respectively, which are larger than the results of HCM-SVRs and MFCM-SVRs.

The method MFCM-SVRs with continuous clustering is quite meaningful in the beginning of modeling in practice. Before conducting SVR models, the data are clustered
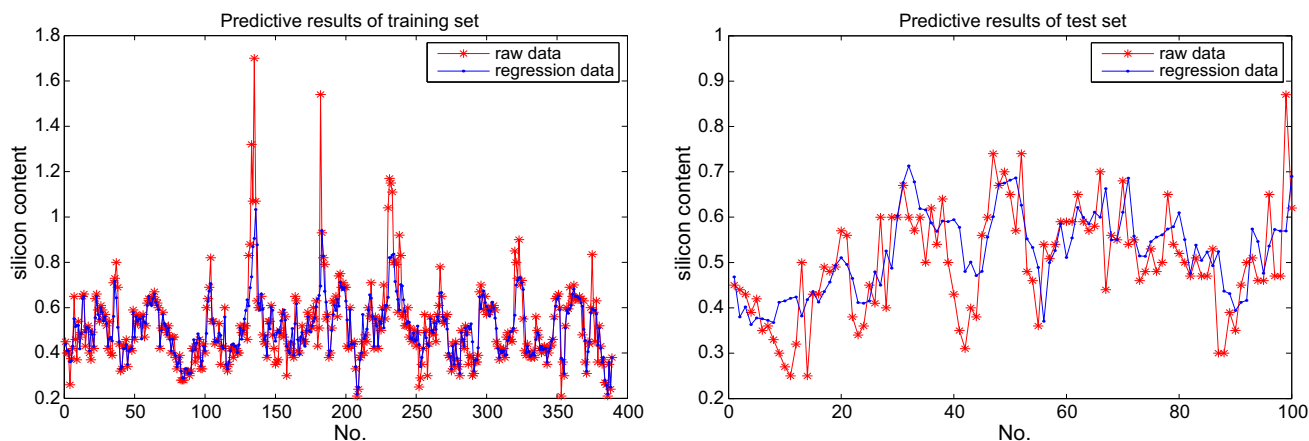
Fig. 9 Predictive results of training set and test set

again including new data to predict. The simulation results based on industrial data demonstrate the effectiveness when applying the method in industry.

For real-time silicon content prediction, a discriminant analysis method is essential to identify which cluster the new data belong to. The discriminant method can be based on the minimum distance just as discussed above. And then the silicon content can be predicted by the corresponding SVR model with the current input parameters. With accurate prediction of silicon content, blast furnace operators can judge the trend of blast furnace temperature. If there is deviation between the predicted temperature and the ideal temperature, or the predicted data are out of the controlled range, the operators can change the inputs which have influence on silicon content in advance. In this way, furnace cooling or other problems can be prevented effectively, and then the quality of smelting will be improved. In general, the silicon content prediction using support vector regression combined with fuzzy C-means clustering is quite encouraging for stable operation of blast furnace.

5 Conclusion

This paper has developed HCM-SVRs and MFCM-SVRs to predict the silicon content. These two models cluster the data using clustering algorithms, and the outliers are eliminated in the process of clustering. Through clustering, the data are separated into smaller groups, which makes the time decreased to one tenth of the total time. By defining a threshold of membership in FCM clustering, MFCM-SVRs are proposed for data purification. Through this method, the outliers are eliminated once again, which offers more accurate forecasting results. The HCM-SVRs show better results than SVR with 0.0096 lower of MAPE and 0.0074

of MSE. What's more, MFCM-SVRs are the best in these methods with 0.0185 lower of MAPE and 0.0084 lower of MSE than SVR. These two models, including gray correlation method, are introduced to predict the next silicon content, and the experimental results show the effectiveness both in accuracy and in speed. This modeling method makes it more efficient to use industrial data and closer to industrial practice, which will provide decision support for supervising and controlling blast furnace temperature.

Future efforts will be made to analyze blast furnace ironmaking process and modify modeling approach to improve accuracy.

References

1. Jindal A, Pujari S, Sandilya P, Ganguly S (2007) A reduced order thermo-chemical model for blast furnace for real time simulation. Comput Chem Eng 31(11):1484–1495
2. Su HY, Zeng JS, Gao CH (2010) Data-driven predictive control for blast furnace ironmaking process. Comput Chem Eng 34(11):1854–1862
3. Jian L, Gao CH, Ge QH (2014) Rule extraction from fuzzy based blast furnace SVM multiclassifier for decision-making. IEEE Trans Fuzzy Syst 22(3):586–596
4. Du SW, Cheng WT, Huang EN (2014) Numerical analysis on transient thermal flow of the blast furnace hearth in tapping process through CFD. Int J Heat Mass Transf 57:13–21
5. Zhang Y, Rohit D, Huang D, Pinakin C, Zhou CQ (2008) Numerical analysis of blast furnace hearth inner profile by using CFD and heat transfer model for different time periods. Int J Heat Mass Transf 51(1–2):186–197
6. Das SK, Kumari A, Bandopadhay D, Akbar SA, Mondal GK (2011) A mathematical model to characterise effects of liquid hold-up on bosh silicon transport in the dripping zone of a blast furnace. Appl Math Model 35(9):4208–4221
7. Chen J (2001) A predictive system for blast furnaces by integrating a neural network with qualitative analysis. Eng Appl Artif Intell 14(1):77–85

8. Yang YL, Zhang S, Yin YX (2016) A modified ELM algorithm for the prediction of silicon content in hot metal. Neural Comput Appl 27(1):241–247

9. Yu T, Li JP, Shi L, Li ZL (2011) Model of hot metal silicon content in blast furnace based on principal component analysis application and partial least square. J Iron Steel Res Int 18(10):13–16

10. Wang WH, Liu XY, Wang YK (2007) Prediction of silicon content in hot metal based on Bayesian network. In: Third international conference on natural computation, Haikou, China, vol 5, pp 446–450

11. Luo SH, Gao CH, Jian L (2012) Modeling of the thermal state change of blast furnace hearth with support vector machines. IEEE Trans Ind Electron 59(2):1134–1145

12. Gao CH, Jian L (2013) Binary coding SVMs for the multiclass problem of blast furnace system. IEEE Trans Ind Electron 60(9):3846–3856

13. Vapnic VN (1998) The nature of statistical learning theory. Wiley, New York

14. Guo ZK, Guan XP (2015) Nonlinear generalized predictive control based on online least squares support vector machines. Nonlinear Dyn 79(2):1163–1168

15. Sudheer C, Maheswaran R, Panigrahi BK, Shashi M (2014) A hybrid SVM-PSO model for forecasting monthly streamflow. Neural Comput Appl 24:1381–1389

16. Wei LW, Xu WX, Shi Y, Chen ZY, Li JP (2011) Multiple-kernel SVM based multiple-task oriented data mining system for gene expression data analysis. Expert Syst Appl 38(10):12151–12159

17. Lee SL, Fu JH (2012) A multi-class SVM classification system based on learning methods from indistinguishable Chinese official documents. Expert Syst Appl 39(3):3127–3134

18. Lu CJ (2013) Hybridizing nonlinear independent component analysis and support vector regression with particle swarm optimization for stock index forecasting. Neural Comput Appl 23:2417–2427

19. Vanhoucke M, Wauters M (2014) Support vector machine regression for project control forecasting original research article. Autom Constr 47:92–106

20. Jin C, Jin SW (2014) Software reliability prediction model based on support vector regression with improved estimation of distribution algorithms. Appl Soft Comput 15:113–120

21. Shakouri HG, Nadimi R (2013) Outlier detection in fuzzy linear regression with crisp input–output by linguistic variable view. Appl Soft Comput 13(1):734–742

22. Pierpaolo D, Riccardo M (2013) Fuzzy clustering of human activity patterns. Fuzzy Sets Syst 215:29–54

23. Velmurugan T (2014) Performance based analysis between k-means and fuzzy C-means clustering algorithms for connection oriented telecommunication data. Appl Soft Comput 19(3):134–146

24. Zhao ZX, Cheng LZ, Cheng GQ (2014) Neighbourhood weighted fuzzy c-means clustering algorithm for image segmentation. IEEE Trans Image Process 8(3):150–161

25. Duan P, Xie KG, Guo TG, Huang XG (2011) Shot-term load forecasting for electric power systems using the PSO-SVR and FCM clustering techniques. Energies 4(1):173–184

26. Mukhopadhyay A, Maulik U (2009) Unsupervised pixel classification in satellite imagery using multiobjective fuzzy clustering combined with SVM classifier. IEEE Trans Geosci Remote Sens 47(4):1132–1138

27. Yang XW, Zhang GQ, Lu J, Ma J (2011) A kernel fuzzy c-means clustering-based fuzzy support vector machine algorithm for classification problems with outliers or noises. IEEE Trans Fuzzy Syst 19(1):105–115

28. Chen JQ, Witold P, Ha MH, Ma LT (2015) Set-valued samples based support vector regression and its applications. Expert Syst Appl 42(5):2502–2509

29. Lin KP (2014) A novel evolutionary kernel intuitionistic fuzzy C-means clustering algorithm. IEEE Trans Fuzzy Syst 22(5):1074–1087