

# Using Principal Component Analysis and Least Squares Support Vector Machine to Predict the Silicon Content in Blast Furnace System

<https://doi.org/10.3991/ijoe.v14i04.8397>

Shihua Luo, Tianxin Chen<sup>(✉)</sup>  
Jiangxi University of Finance & Economics, Nanchang, China  
526971826@qq.com

Ling Jian  
China University of Petroleum, Qingdao, China

**Abstract**—Blast furnace system is a typical example of complex industrial system. The silicon ([Si]) content in blast furnace system is an important index to reflect the temperature of furnace. Therefore, it is significant to carry out an accurate predictive control of furnace temperature. In this paper a composite model combining Principal Component Analysis (PCA) and Least Squares Support Vector Machine (LSSVM) is established to predict the furnace temperature. At the very beginning, in order to avoid redundancy and excessive noise pollution, PCA method is applied to reduce the dimensionality of original input variables. Secondly, the dimension-reduced variables are introduced to predict the silicon content by applying the LSSVM model. Finally, the result is compared with direct multivariable LSSVM prediction. The simulation results show that the new algorithm has positive significance as it achieves more obvious prediction hit rate (more than 80%) than direct multivariable LSSVM (with rate lower than 75%).

**Keywords**—blast furnace system; principal component analysis; least squares support vector machine; silicon content prediction

## 1 Introduction

Blast furnace temperature is an important index during operating. Thus keeping it stationary in a reasonable range is essential. Unexpected temperature fluctuation in the operation may cause a series of problems such as: unnecessary heat loss, furnace cool down and blast furnace abnormal nodulation, etc. The reasonable temperature control of blast furnace causes the blast furnace with high efficiency, low cost and stable operation in the process of blast furnace iron making. Furthermore, the control of the silicon content ([Si]) of molten iron is closely related to the furnace condition, stability, production efficiency (utilization coefficient), energy consumption (ratio) and the quality of molten iron ([S]) in the process of blast furnace smelting. In addition, ([Si]) is an effective characterization of the physical temperature of blast furnace

hearth, so it is usually used as the main index of smelting process control. The prediction and control of blast furnace temperature ([Si]) has always been a hot and difficult problem in this field[1-2].

The accurate prediction of furnace temperature and the linkage relationship between furnace temperature, coal injection, air volume, air temperature and coke load are the key and difficult points in the modeling of blast furnace smelting process. Since entering the new century, due to truly understand the internal dynamics of physical and chemical reactions in blast furnace is very difficult. The model research based on the mechanism of temperature prediction has gradually stagnated. What's more, the research on Intelligent Furnace Temperature Prediction Control based on intelligent algorithm is developing rapidly. Many domestic and foreign research teams have established many models such as Bayesian network model[3], chaotic prediction model[4], neural network model based on genetic algorithm[5], fuzzy data generation rule control model[6-7], partial least squares model[8], mathematical model of multifluid theory[9], support vector machine and intelligent algorithm cross model[10], wavelet analysis model etc. These models obtain satisfactory results in different aspects; however, due to the complexity of the blast furnace system, it is difficult to achieve closed-loop predictive control in this field.

Analyzing the existing prediction model, we found that some models only used single silicon content of ([Si]) sequences without many key state and control variables, this is not reality. Some multivariable models have too many variables, which lead to the shortage such as: modeling variables, information redundancy, noise pollution and computational complexity. It also restricts the practical application of the model. Based on this, this paper concludes the advantages and disadvantages of principal component analysis (PCA) [11-15] and least squares support vector machine (LSSVM) [15-25] model, and constructs a new multivariable combination forecasting model for furnace temperature prediction.

In order to rationalize the multivariate information modeling, we use PCA method to take a plurality of variables influencing the silicon content integration become the main variable which is a linear combination of the initial variables. On the one hand, the number of input variables can be reduced and the model dimension reduction can be realized. Using this model, the number of input variables can be reduced and the model dimension reduction can be realized, on the other hand, information redundancy and noise pollution can be avoided, and the computational complexity can be significantly reduced. Then the LSSVM algorithm is used to predict the silicon content by introducing these composite variables. The simulation results show that the prediction algorithm based on PCA and LSSVM is better than LSSVM's direct multivariate modeling prediction.

Section 2 will introduce the basic methods of PCA and LSSVM, section 3 is the empirical analysis and section 4 is the conclusion.

## 2 Basic methods

### 2.1 Principal component analysis (PCA)

PCA is a commonly used dimension reduction method. Consider  $X = (x_1, x_2, \dots, x_p)^T$ ,  $X$  stands for a random vector of  $p$  dimensions, the covariance matrix can be written as:

$$\sum X = \begin{bmatrix} \sigma_{11} & \sigma_{12} & \cdots & \sigma_{1p} \\ \sigma_{21} & \sigma_{22} & \cdots & \sigma_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{p1} & \sigma_{p2} & \cdots & \sigma_{pp} \end{bmatrix}_{p \times p} \quad (1)$$

$\lambda_1, \lambda_2, \dots, \lambda_p$  are the  $p$  nonzero eigenvalues of the covariance matrices. Generally speaking,  $\lambda_1 > \lambda_2 > \dots > \lambda_p$ , according to the knowledge of linear algebra, there must be an orthogonal matrix  $U$ ,  $U$  satisfies the following equations:

$$U^T \sum X U = \begin{bmatrix} \lambda_1 & & & \\ & \lambda_2 & & \\ & & \ddots & \\ & & & \lambda_p \end{bmatrix}_{p \times p} \quad (2)$$

Among them,  $U$  is exactly the orthogonal matrix of the  $p$  eigenvectors corresponding to the  $p$  characteristic roots of the covariance matrix,  $U$  can be rewritten as:

$$U = \begin{bmatrix} U_{11} & U_{12} & \dots & U_{1p} \\ U_{21} & U_{22} & \dots & U_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ U_{p1} & U_{p2} & \dots & U_{pp} \end{bmatrix}_{p \times p} \quad (3)$$

$U_i = (U_{i1}, U_{i2}, \dots, U_{ip})^T$ ,  $Y = (Y_1, Y_2, \dots, Y_m)^T$  as a principal component vector,  $Y_1, Y_2, \dots, Y_m$  can be regarded as the  $m$  principal component ( $m < p$ ). Then  $Y$  can be expressed as  $Y = U^T X$ . Furthermore, the new principal component variables can be illustrated linearly by the original variables, which can be described as:

$$\begin{cases} Y_1 = U_{11}X_1 + U_{12}X_2 + \dots + U_{1p}X_p; \\ Y_2 = U_{21}X_1 + U_{22}X_2 + \dots + U_{2p}X_p; \\ \vdots \\ Y_m = U_{m1}X_1 + U_{m2}X_2 + \dots + U_{mp}X_p \end{cases} \quad (4)$$

$Y$  need satisfy two conditions:

(1)  $Y_i, Y_j (i \neq j; i, j = 1, 2, \dots, m)$  should be mutual independence, which means that  $\text{cov}(Y_i, Y_j) = 0, (i \neq j)$ ;

(2)  $Y_1, Y_2, \dots, Y_m$  should satisfy the diminishing of variance, which meant to be  $\text{Var}(Y_1) > \text{Var}(Y_2) > \dots > \text{Var}(Y_m)$ . It reflects the amount of information contained in the principal component variable is gradually reduced.

As demonstrated above, the variance of principal components are equal to their respective eigenvalues. It is obvious that  $\text{Var}(Y_k) = \lambda_k (k = 1, 2, \dots, m)$ . In addition, the variance contribution rate of the principal component can be depicted as:

$$\alpha_k = \frac{\lambda_k}{\sum_{k=1}^m \lambda_k} (k = 1, 2, \dots, m) \tag{5}$$

If  $\alpha_k$  is increasing, the variance contribution rate of the principal component  $Y_k$  is greater than the other principal components therefore  $Y_k$  integrates  $X$  information better. Otherwise,  $Y_1, Y_2, \dots, Y_m$ , their ability to integrate  $X$  information is progressively decreasing. In order to choose component  $m (m < p)$ , the cumulative contribution

rate can be applied. Consider  $\frac{\sum_{i=1}^m \lambda_i}{\sum_{i=1}^p \lambda_i}$  as the cumulative contribution rate of  $m$  prin-

cipal components, popularly, we assume that the value of  $m$  makes the cumulative contribution rate greater than 85%, and the effect is better. The  $m$  principal component can contain the information of most of the  $p$  variables  $X$ . Thus, we use PCA to eliminate the noise of original variable dimension most of which not only can keep the original variable information, but also can reduce the number of variables, the sample properties and the amount of computation.

## 2.2 Least squares support vector machine (LSSVM)

The LSSVM uses kernel function to map the low dimensional linear non-separable samples into the high-dimensional space, so that the samples are linearly separable and can be predicted more accurately. Assume the training data set is  $T = \{(x_i, y_i | i = 1, 2, \dots, n)\}$ , where  $n$  is the number of training concentrated samples.

$x_i \in R^n$  is the input variables while  $y_i \in R^n$  is the output variables. Under linear conditions, we construct the optimal decision function:

$$y = \omega^T x + b \tag{6}$$

$\omega = (\omega_1, \omega_2, \dots, \omega_n)$  is an unknown parameter vector which can be considered as the weight of all samples,  $b \in R$  is the critical value. When input variables are nonlinear, we need to build a map  $\phi, \phi: R^n \rightarrow R^{n\phi}$  expressed by kernel function which is able to linearization X by lifting the its dimension. In high-dimensional space, the optimal decision function can be expressed as:

$$y = \omega \cdot \phi(x) + b \tag{7}$$

which means  $\phi(\cdot)$  can use kernel function to calculate, such that  $k(x, z) = \phi(x) \cdot \phi(z)$ , as the kernel functions of variables X and Z in high dimensional spaces, where  $x, z \in R^n$ , a commonly used kernel function is the Radial Basis Function (RBF) kernel given by

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

$\|\cdot\|$  is the usual Euclid norm and  $\sigma$  is hyper parameter which can respond to the non-linear mapping  $\phi(\cdot)$ . In fact, the optimal decision function problem can be transformed into a dual form. In order to solve  $\omega$  and  $b$  in the equation, the following two constraints can be used to optimize the problem, which can be depicted as:

$$\min_{\omega, b, \xi} J(\omega, \xi) = \frac{1}{2} \omega^T \omega + \frac{\gamma}{2} \sum_{i=1}^N \xi_i^2 \tag{9}$$

The constraint condition is  $y_i = \omega \cdot \phi(x_i) + b + \xi_i$ ,  $\gamma$  is the penalty factor which represents the degree of penalty beyond the error sample and belongs to the adjustable parameter. In fact,  $\gamma > 0$ , the smaller the  $\gamma$ , the better sample that indicates the wrong fit;  $\xi_i$  is the loss function. For the sake of transform the constrained optimization problem into an unconstrained problem, we can solve it by Lagrange multipliers, as follows this equation:

$$L(\omega, b, \xi, \alpha) = \frac{1}{2} \omega^T \omega + \frac{\gamma}{2} \sum_{i=1}^N \xi_i^2 - \sum_{i=1}^N \alpha_i [\omega \cdot \phi(x_i) + b + \xi_i - y_i] \tag{10}$$

$\alpha_i \in R, i = 1, 2, \dots, N$  is the Lagrange multipliers. According to the Karush-Kuhn-Tucker conditions<sup>[26]</sup>, the unknown parameters  $\omega, b, \xi, \alpha$  can be estimated. We can derivative these parameters as follows:

$$\begin{cases} \frac{\partial L}{\partial \omega} = 0 \rightarrow \omega = \sum_{i=1}^N \alpha_i \phi(x_i) \\ \frac{\partial L}{\partial b} = 0 \rightarrow \sum_{i=1}^N \alpha_i = 0 \\ \frac{\partial L}{\partial \xi} = 0 \rightarrow \alpha_i = \xi_i, i = 1, 2, \dots, N \\ \frac{\partial L}{\partial \alpha} = 0 \rightarrow y_j = \omega \cdot \phi(x_j) + b + \xi_j, j = 1, 2, \dots, N \end{cases} \quad (11)$$

Elimination  $\omega$  and  $\xi$  will acquire:

$$\begin{cases} \sum_{i=1}^N \alpha_i = 0 \\ y_j = \sum_{i=1}^N \alpha_i \cdot k(x_i, x_j) + b + \frac{\alpha_j}{\gamma}, j = 1, 2, \dots, N \end{cases} \quad (12)$$

Transformation into matrix form, the former equation can be rewritten as

$$\begin{bmatrix} 0 \\ y \end{bmatrix} = \begin{bmatrix} 0 & 1_N^T \\ 1_N & \Omega + \frac{1}{\gamma} I \end{bmatrix} \begin{bmatrix} b \\ \alpha \end{bmatrix} \quad (13)$$

Denote

$$y = (y_1, y_2, \dots, y_N)^T; 1^N = (1, 1, \dots, 1)^T; \alpha = (\alpha_1, \alpha_2, \dots, \alpha_N)^T; \Omega_{ij} = k(x_i, x_j);$$

$$I = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & \ddots & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}_{N \times N}$$

Therefore

$$\begin{bmatrix} b \\ \alpha \end{bmatrix} = \begin{bmatrix} 0 & 1_N^T \\ 1_N & \Omega + \frac{1}{\gamma} I \end{bmatrix}^{-1} \begin{bmatrix} 0 \\ y \end{bmatrix} \quad (14)$$

Finally, the decision function can be shown as:

$$y = \sum_{i=1}^N \alpha_i k(x_i, x) + b \quad (15)$$

To sum up, the decision functions of LSSVM can be represented by Lagrange multipliers and kernel functions. In fact, besides the dual variables  $\alpha$ , the two parameters of the kernel function  $\sigma^2$  and the penalty factor  $\gamma$  need to be determined. The model

can show good performance only when the parameter is appropriately selected. The choice of the two parameters can be achieved by  $k$  cross validation. The concrete steps are as follows:

1. Assume the initial value of the parameter  $\gamma$  and the  $\sigma^2$ .
2. Divide the training set into  $k$  classes, selects part of them as the test set, and the rest as the training set.
3. Set the model and the square error then calculate the parameters  $\gamma$  and  $\sigma^2$ .
4. Apply the test set into the model and repeats  $k$  times. (Note: at least one test set for each collection);
5. Calculate the  $k$  mean square error by adding up the sum.
6. Upload the value of the super parameter and repeats steps 2 through 5 until the mean square error is acceptable.

### 3 Empirical analysis

#### 3.1 Analysis idea

PCA is used to reduce the dimension of the input variables of  $X$ , which can reduce the number of inputs for LSSVM. In this paper, firstly 1108 blast furnace data were selected as sample space, 8 variables were selected as input variables, which the current ([Si]) as output variables. The empirical analysis is as follows:

1. Use PCA to process the input variables and determine the number of principal components and the number of input variables of LSSVM.
2. Set the training set and test set, then normalize the data and find the best parameters according to the cross validation. So, we can get the LSSVM model.
3. The sample data of the test set are brought into the training model to predict ([Si]).

You may mention here granted financial support or acknowledge the help you got from others during your research work.

#### 3.2 PCA's analysis

In this paper, 8 variables which represent the primary nature of the blast furnace system are selected from 1108 blast furnace data sample space as the research objects respectively, these variables are air volume ( $x_1$ ), wind temperature ( $x_2$ ), wind pressure ( $x_3$ ), top pressure ( $x_4$ ), material speed ( $x_5$ ), permeability ( $x_6$ ), coal injection ( $x_7$ ) and oxygen enrichment rate ( $x_8$ ). The PCA was carried out by using SPSS 22, and the results were shown in table 1.

As demonstrated in table 1, the cumulative contribution rate of the variance of the first 4 principal components is 89.966%, which is greater than 85%, therefore it is well represented. The first 4 principal components are extracted as input variables. The input dimension of the sample is reduced by half, which can greatly improve the

training speed of the LSSVM. Extracting those 4 principal components, the resulting component matrix is shown in table 2.

**Table 1.** principal component analysis table

Component	Initial eigenvalue		
	Total	Percentage variance	Accumulate %
1	4.373	54.658	54.658
2	1.161	14.515	69.173
3	.916	11.450	80.623
4	.747	9.344	89.966
5	.384	4.801	94.767
6	.280	3.498	98.265
7	.124	1.549	99.814
8	.015	.186	100.000

**Table 2.** component matrix table

Variables	Component			
	1	2	3	4
air volume	.907	.159	.009	-.248
wind temperature	.486	-.248	.629	.537
wind pressure	.913	-.144	-.179	.148
top pressure	.933	-.131	-.184	.104
material speed	.857	.039	-.159	-.190
permeability	.462	.565	.580	-.318
coal injection	.332	.736	-.304	.453
oxygen enrichment rate	.742	-.417	-.014	-.147

From the component matrix of Table 2, we can derive 4 principal component equations:

$$\begin{cases} F_1 = 0.907x_1 + 0.486x_2 + 0.913x_3 + L + 0.332x_7 + 0.742x_8 \\ F_2 = 0.159x_1 - 0.248x_2 - 0.144x_3 + L + 0.736x_7 - 0.417x_8 \\ F_3 = 0.009x_1 + 0.629x_2 - 0.179x_3 + L - 0.304x_7 - 0.014x_8 \\ F_4 = -0.248x_1 + 0.537x_2 + 0.148x_3 + L + 0.453x_7 - 0.147x_8 \end{cases} \quad (16)$$

From the component matrix, the PCA method has a good effect on reducing dimension. Among them, the first principal component is a linear combination of five variables, which is highly related to air volume, air pressure, top pressure, material speed and oxygen enrichment rate. Besides, the other three principal components are highly correlated with two variables. The second principal component is highly related to coal injection and permeability. The third principal component and the fourth principal components explained mainly two variables, air temperature and permeability. In summary, we can understand that the first principal component is a comprehensive factor, which can be used in the blast furnace top pressure, air volume, air



pressure, feed rate and oxygen enrichment rate linear representation. The second principal component is a mixing factor relating to coal injection and permeability. The third and fourth principal components are the mixing factors of wind temperature and coal injection. To sum up, we can redefine the new four variables:  $Z_1=F_1$ ;  $Z_2=F_2$ ;  $Z_3=F_3$ ;  $Z_4=F_4$ , we set the new four variables as input variables of LSSVM algorithm.

### 3.3 LSSVM's analysis

In this paper, two sets of training sets are selected from 1108 blast furnace samples. The first training set is blast furnace numbered from 1 to 450, and the second one is 631 to 980. The corresponding test set is from 461 to 500 and from 981 to 1020, respectively. At the same time, we set 12 times cross validation to confirm the parameters  $\gamma$  and  $\sigma^2$ . It can ensure each subset has enough space to illustrate its effectiveness in cross validation and the reliability of the study. Figure 1 shows the observation of ([Si]) in 1108 original blast furnace in the blast furnace system number value.

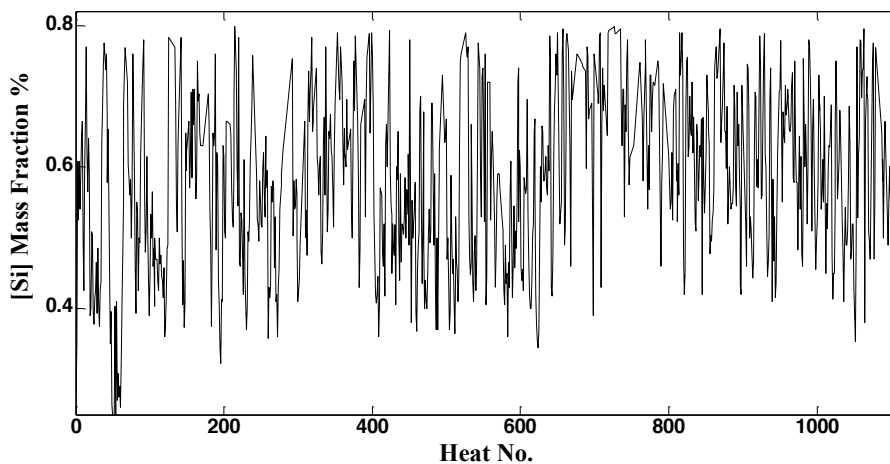


Fig. 1. Observation of ([Si]) in blast furnace

In the previous PCA, we use four principal components instead of the original eight variables as input variables. These four input variables can be defined as

$$(Z_1, Z_2, Z_3, Z_4) = z = (\tilde{z}^1, \tilde{z}^2, \tilde{z}^3, \tilde{z}^4) \tag{17}$$

The PCA shows that the principal components are independent of each other, so the correlation coefficient between the four principal components is 0 and linearly independent. Therefore the relationship between variables can be eliminated.

With different dimension, these data need to be manipulated before being analyzed. In general, we use the normalization method to fix all variables between 0 and 1, and the formula of normalization is as follows:

$$\begin{cases} z_i^j = \frac{\tilde{z}_i^j - \min(\tilde{z}^j)}{\max(\tilde{z}^j) - \min(\tilde{z}^j)}, \\ i = 1, 2, \dots, 1108; j = 1, 2, \dots, 4 \\ \max(\tilde{z}^j) = \max\{\tilde{z}_i^j | i = 1, 2, \dots, 1108\} \\ \min(\tilde{z}^j) = \min\{\tilde{z}_i^j | i = 1, 2, \dots, 1108\} \end{cases} \quad (18)$$

$\tilde{z}_i^j$  can be performed as observations.

Thus, all the data can be normalized and the dimensional influence eliminated by the above formula.

After we processed treatment  $z_i = (z_i^1, z_i^2, z_i^3, z_i^4)$ , which used as input variables of LSSVM, then the ([Si]) content can be seen as the output variable. We Select the RBF kernel as the kernel function  $k(x_i, x_j)$  and use 12-fold cross validation for choosing the hyper parameters according to the former procedure. This is done through the LSSVM-lab 1.8 package Toolkit. After this, we estimated two groups of parameters for training set:  $(\gamma_1, \sigma_1) = (10.35, 0.58)$  and  $(\gamma_2, \sigma_2) = (8.38, 0.35)$ . Finally, we predict the ([Si]) of the two sets of tests, which is shown in Figure 2. Figure 2 shows the comparison between the observed and predicted values of ([Si]) in the two sets of tests. It illustrates that the prediction of LSSVM algorithm based on PCA is better than using LSSVM algorithm.

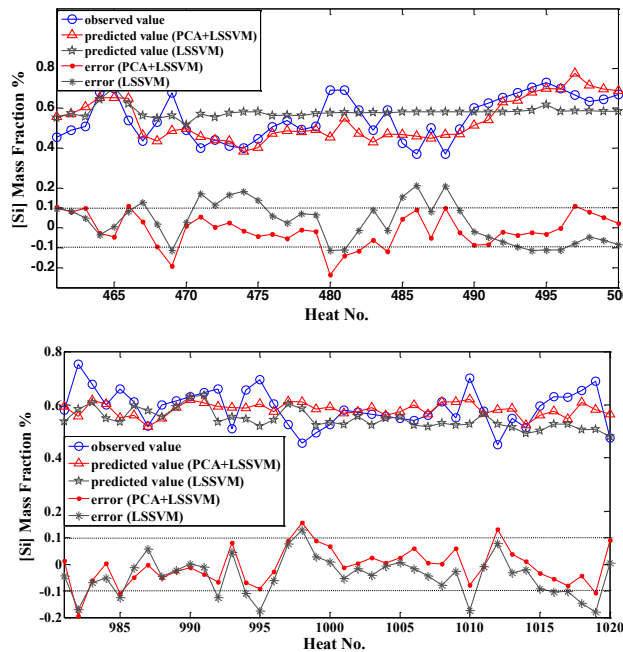


Fig. 2. Comparison of predicted and observed values of two sets of test sets

In addition, the standard of evaluation model can be reflected by the following two indicators:

- (1) RMSE-Root mean square error  
Define the root mean square error

$$RMSE = \left[ \frac{1}{n-1} \sum_{i=1}^n (\hat{y}_i - y_i)^2 \right]^{1/2} \tag{19}$$

Where  $\hat{y}_i$  is the predicted value,  $y_i$  is the observed value,  $n$  is the length of the predicted data set. The range of RMSE is between 0 and 1. If RMSE is closer to 0, this model provide a better prediction result.

- (2) PHT-Percentage of hitting target  
Define the percentage of hitting target

$$PHT_{\varepsilon} = \frac{1}{n} \sum_{i=1}^n H_i \cdot 100\% \tag{20}$$

$H_i$  means hit count, it can be written as:

$$H_i = \begin{cases} 1 \cdots \cdots |\hat{y}_i - y_i| \leq \varepsilon \\ 0 \cdots \cdots otherwise \end{cases} \tag{21}$$

Where  $\hat{y}_i$  is the predicted value,  $y_i$  is the observed value,  $n$  is the length of the predicted data set.  $H_i$  takes only two values: 0 or 1, If  $H_i = 1$ , it conveys that the observed and predicted values of this sample are within the range of error. On the contrary,  $H_i = 0$ , beyond the range of error.  $\varepsilon$  reflects the difference between observed and predicted values. It is a small positive number, Normally,  $\varepsilon$  takes 0.1. The PHT value is between 0 and 1. Generally speaking, when PHT value is greater than 0.8, the model prediction simulation provides a positive result..

Table 3 reflects. It is can be found that the composite model provides more accuracy than single model which applied LSSVM only.

**Table 3.** prediction results of ([Si])

Test set	Basic methods	RMSE	Hits	PHT
461-500	LSSVM	0.103	25	0.625
	PCA+LSSVM	0.082	32	0.8
981-1020	LSSVM	0.084	29	0.725
	PCA+LSSVM	0.070	35	0.875

## 4 Conclusions

In this paper, a new algorithm formed up by combining PCA and LSSVM is used to predict [Si] in Blast Furnace System. The simulation shows that the new method achieves more obvious prediction hit rate than the direct multivariable LSSVM, which indicates that 80% vs 62.5% and 87.5% vs 72.5% respectively. The new algorithm has positive significance. Furthermore, when the continuous sampling is insufficient (learning data segments), the new algorithm has better stability than the direct algorithm.

## 5 Acknowledgements

This research is partially supported by the NSFC (61563018), the NSF of Jiangxi Province 20161ACB20009, 20133BCB23014) and the Foundation of the Office of Education, Jiangxi Province (KJLD13033).

## 6 References

- [1] Ueda, S, Natsui S, Nogami H, Yagi J.I , and A. Tatsuro: Recent progress and future perspective on mathematical modeling of blast furnace [J]. ISIJ Int., 2010, 50(7), pp. 914-923. <https://doi.org/10.2355/isijinternational.50.914>
- [2] X. G. Liu and F. Liu: Optimization and Intelligent Control System of Blast Furnace Iron-making Process, Metallurgical Industry Press, Beijing, (2003).
- [3] X.Y. Liu: Research on [Si] predictive control model of blast furnace temperature based on Bayesian network [D]. Master's degree thesis of Zhejiang University, 2004.
- [4] C.H. Gao, Z.M. Zhou, Z.J. Shao: Chaotic Local-region Linear Prediction of Silicon Content in Hot Metal of Blast Furnace [J]. Acta Metall. Sin., 2005, 41: 433-436.
- [5] M. Zhao, X.G. Liu, S.H. Luo: An Evolutionary Artificial Neural Networks Approach for BF Hot Metal Silicon Content Prediction Based on Chaotic Analysis [J]. LNCS (Lecture Notes in Computer Science), 2005, (3610): 572-575.
- [6] H.W. Guo, L.K. Chen, H.B. Zuo, T.J. Yang: Prediction model of blast furnace temperature based on fuzzy reasoning [J]. Iron and steel research, 2007, 35 (2): 22.
- [7] H.W. Guo, L.K. Chen, H.B. Zuo, T.J. Yang: Application of multidimensional time series fuzzy association rules to prediction of blast furnace temperature [J]. Journal of University of Science and Technology Beijing, 2008, 30 (5): 553.
- [8] Bhattacharya. T, Prediction of silicon content in blast furnace hot metal using partial least squares (PLS). ISIJ International, 2005, 45(12): 1943-1945. <https://doi.org/10.2355/isijinternational.45.1943>
- [9] M.S. Chu, X.Z. Guo, F.M. Shen: Blast furnace mathematical model and solution based on Multi fluid theory [J]. Northeastern University (NATURAL SCIENCE EDITION), 2007, 28 (3): 361.
- [10] L. Jian, C.H. Gao, and Z.Q. Xia, "A sliding-window smooth support vector regression model for nonlinear blast furnace system," Steel. Res. Int., 2011, 82(3), pp. 169-179. <https://doi.org/10.1002/srin.201000082>
- [11] W.L. Zhang, H.M. Lin: Principal component analysis and factor analysis of the similarities and differences and SPSS software [J]. statistical research, 2005, (03): 65-69.

- [12] S.W. Xie: Principal component analysis and factor analysis of mathematical model based on the application of [D]. Shandong University of Technology, 2016.
- [13] P. Wang, Y.W. Deng, Y.P. Tian, F.M. Kuang, F. Yi: Hengyang City, the land ecological security evaluation based on principal component analysis [J]. Economic geography, 2015 (01): 168-172.
- [14] W. Lei: Support vector machine regression prediction model based on principal component analysis [J]. Information technology, 2008, (12): 58-60.
- [15] R.N. Shen, C. Cao, C.J. Fan: Practice and understanding principal component analysis and support vector machine model of Shanghai price prediction research [J]. mathematics, 2013, (23): 11-16.
- [16] V. N. Vapnik, The Nature of Statistical Learning Theory, Springer-Verlag, New York, NY, USA, 2nd edition, 2000. <https://doi.org/10.1007/978-1-4757-3264-1>
- [17] J.H. Zheng: Research on prediction of blast furnace temperature based on support vector machines [D]. Zhejiang University, 2007.
- [18] J. Liu, J. Chen, S. Jiang, and J. Cheng : Online LS-SVM for function estimation and classification [J]. Univ. Sci. Technol. B., vol. 10, no. 5, 2003.
- [19] C. Gao, L. Jian, and S.H. Luo, "Modeling of the thermal state change of blast furnace hearth with support vector machines," IEEE Transactions on Industrial Electronics, vol. 59, no. 2, pp. 1134–1145, 2012. <https://doi.org/10.1109/TIE.2011.2159693>
- [20] L.H. Wang, S. Gao, X. Qu: Prediction of silicon content in molten iron in LSSVM blast furnace based on GA optimization [J]. Journal of Terahertz Science and electronic information, 2013, (04): 641-645.
- [21] Y. Tao: Application and improvement of blast furnace temperature prediction using LSSVM [D]. Shenyang University of Aeronautics and Astronautics, 2014.
- [22] Y.J. Yue, Y.F. Lu, Hui. Z, H.J. Wang: Prediction of blast furnace gas using wavelet analysis ARIMA and LSSVM [J]. Computer measurement and control, 2015, (06): 2128-2131.
- [23] Y.T. Xiong, Z.P. Li, J.Wang, S.W. Feng, Z.W. Li, J.J. Yin, Q.H. Song: Analysis of liquor grains quantitative composition based on least squares support vector machine. Food science, 2016,37 (12): 163-168.
- [24] Y.J. Yue, Y.F. Cheng , H. Zhao, H.J. Wang: Prediction of blast furnace gas based on wavelet analysis ARIMA and LSSVM computer measurement and control [J]. 2015,23 (06): 2128-2131.
- [25] C. Zhao, K.C. Dai: Short term power load forecasting based on adaptive weighted least squares support vector machines [J]. information and control, 2015,44 (05): 634-640.
- [26] R.X. Li: Approximate Lagrange multipliers and KKT conditions for constrained vector optimization [D]. Yunnan University, 2013.

## 7 Authors

**Shihua Luo** is currently an Professor of School of Statistics, Center of Applied Statistics, Jiangxi University of Finance and Economics, Nanchang, 330013, China. He received the B.S. degree and M.S. degree in Mathematics from Jiangxi Normal University, China, in 1998 and 2001, and the Ph.D. degree in Operational Research and Cybernetics from Zhejiang University, China, in 2007. His research interests are in the field of modeling and optimization of complex systems, especially on the extraction of nonlinear characteristics of chemical process and financial process.

**Tianxin Chen** is currently an Postgraduate of School of Statistics, Center of Applied Statistics, Jiangxi University of Finance and Economics, Nanchang, 330013, China. He received the B.S. degree in Statistics from Lanzhou University of Finance and Economics, China, in 2012 and 2016. His research interests are multivariate statistical analysis, data mining and machine learning algorithms.

**Ling Jian** is currently an Associate Professor, where he has been with the School of Science, China University of Petroleum, Qingdao, 266555, China since March 2006. He received the B.S. degree in applied mathematics from Qufu Normal University, Qufu, China, in 2003, the M.Sc. degree in operational research and cybernetics from Zhejiang University, Hangzhou, China, in 2006 and the Ph.D. degree in operational research and cybernetics from the Dalian University of Technology, Dalian, China, in 2012. His scientific interests include machine learning, data mining, kernel methods and optimization.

Article submitted 08 February 2018. Resubmitted 23 February 2018. Final acceptance 31 March 2018. Final version published as submitted by the authors.