

Variable Selection Method for Aluminum Electrolytic Process Based on FNN and RM in KPLS Feature Space

Lizhong Yao^a, Taifu Li^a, Jun Peng^a, Deyong Wu^b, Yingying Su^a, Jun Yi^a

^a School of Electric and Information Engineering, Chongqing University of Science and Technology, Chongqing, 40133, China

Email: yaolizhong225@163.com; litaifuemail@qq.com; pengjun70@126.com; 41965457@qq.com;

^b School of Materials Science and Engineering, Chongqing University of Technology, Chongqing, 400054, China
Email: 279062432@qq.com

Abstract—Selecting the process variables is an important prerequisite for establishing an accurate model of aluminum electrolytic process. A variable selection method is researched and proposed based on the False Nearest Neighbors (FNN) and Randomization Method (RM) (FR) in KPLS(Kernel Partial Least Squares) feature space. Firstly, the KPLS is employed to transform the original space to the PLS feature space; secondly, in the new feature space, the FNN is used to calculate the similarity measure of each variable which first is retained and then reset to zero for evaluating the importance to the dependent variable; then, the RM is utilized to test the significance level of the importance for each variable in turn, so the redundant variables would be excluded; lastly, technical energy consumption model of aluminum electrolytic process is built to verify the presented method. The experimental results show that the method selects out the best process variables of aluminum electrolytic process. Therefore, the research provides a new method of the variable selection for metallurgical industrial processes.

Index Terms—variable selection, aluminum electrolysis, false nearest neighbors(FNN), randomization method(RM), KPLS

I. INTRODUCTION

Aluminum electrolysis process has the multiple characteristics, which gather the multivariate, strong coupling, strong interference and the time-varying parameters in integral whole [1] [2]. When the model of aluminium electrolysis process is established, the complexity of the model will increase exponentially along with decision variables [3]. Thus, the accuracy and efficiency of the model can not be guaranteed. Therefore, how to choose the best combination of process variables is a very important part for establishing an accurate

technical energy consumption model in aluminum electrolytic process.

Traditional methods of variable selection such as forward selection [4], backward elimination [5], stepwise regression [6], ridge regression [7], bootstrap [8] and Partial Least-Squares Regression(PLSR) [9] couldn't apply to selecting variables for nonlinear data, due to the nonlinear distortion of the image data in low-dimensional space [10].

The method of feature extraction based on kernel function is an effective way to eliminate information redundancy of the nonlinear variable [11]. It reduces the dimension of data by eliminating the linear correlation data, meanwhile, solves the problem of dimension disaster of high-dimensional data. The following method is often used, for instance, Kernel Principal Component Regression (KPCR) [12], Kernel Principal Component Analysis (KPCA) and neural networks [13], KPCA and Support Vector Machine (KPCA-SVM) [14], Kernel Fisher Discriminate Analysis (KFDA) [15], Kernel Independent Component Analysis and SVM (KICA-SVM) [16], etc. Unfortunately, the feature extraction just gets the combinations of original variables, and do not meet the purpose of reducing the number of process variables.

For exploring a way to overcome the above defects, we have done some research using the FNN method [10] [17]. Firstly, the initial nonlinear space is converted to linear space based on kernel function; then, the FNN is employed to calculate the similarity measure of the each variable by retaining and removing to determine the importance to the output variable. This approach provides a very important revelation for selecting out the best process variables. However, after further detailed research, we find that although FNN can quantify the important degrees of variables and provides the sorting of important degrees, it doesn't come up with a specific standard to filter out the best variables. Fortunately, randomization method (RM) is a pruning technique, which can test the significant level of statistical indicators based on statistical theory and then exclude no significant indicators [18].

Based on those reasons, A novel variable selection strategy is proposed based on the FNN and RM in Kernel

Manuscript received June 1, 2013; revised July 5, 2013; accepted July 12, 2013.

The work is supported by National Natural Science Foundation of China (No.51075418), Application Development Major Projects of Chongqing (No.cstc2013yykfC0034), Natural Science Foundation Key Project (No.CQ CSTC2013jjB40007), Chongqing Educational Committee Science and Technology Research Project (No.KJ111417 and No.KJ091402), Foundation of Chongqing University of Science and Technology (No. CK2011B04 and CK2011Z01).

Spaces. Taking into account the partial least square (PLS) [19] having many advantages, which not only involves the relationship between independent variables, but also between the independent variable and dependent variable. So, the space of kernel partial least square (KPLS) [20] is used to participate in the process of selecting the best process variables.

II. PRELIMINARIES

A. Kernel Partial Least Squares(KPLS)

KPLS is a nonlinear PLS method combined with the kernel function and PLS. The most commonly kernel function is Gaussian kernel $K(x_i, x_j) = \exp(-\|x_i - x_j\|^2 / c)$, where c is a constant. So the feature subspace can be constructed to infinite dimension. Assuming the main component of the independent variable matrix is represented as t , the main component of the dependent variable matrix as u , the sample data in the original space $\mathbf{R}^{l \times n}$ as \mathbf{X} , the variable dimension as n , the number of samples as l . After the nonlinear transform, it is mapped to high dimensional feature space \mathbf{H} , where its structure is $l \times N$ ($N \gg n$) and Φ is a matrix of $n \times H$, the i -th line is the vector $\varphi(x_i)$. The specific algorithm is presented as follow:

- 1) Initialize vector u randomly;
- 2) Calculate $t = \Phi \Phi^T u$, Normalization $t \leftarrow t / \|t\|$;
- 3) Independent variable spatial weight vector $c = Y^T t$;
- 4) Calculate $u = Yc$, Normalization $u \leftarrow u / \|u\|$;
- 5) Repeat steps 2)—4), until convergence;
- 6) Calculate the residual space of the feature space and the dependent variable space, as in (1).

$$\Phi \Phi^T \leftarrow (\Phi - t t^T \Phi)(\Phi - t t^T \Phi)^T, Y \leftarrow Y - t t^T Y \quad (1)$$

Where: $\Phi \Phi^T$ represents the nuclear matrix K , $K(x_i, x_j) = \varphi(x_i)^T \varphi(x_j)$.

B. False nearest-neighbor method(FNN)

FNN[17] is a method used to determine the input space dimension of autoregressive time series prediction model, belonging to the phase space reconstruction method. Assuming that the phase space is the m dimensions, where a phase point vector is denoted by $X(t) = \{x(t), x(t-\tau), \dots, x(t-(m-1)\tau)\}$, $t = 1, 2, \dots, m$, there must be a nearest neighbor point X_f in a distance, and its distance is $R_m(t) = \|X(t) - X_f(t)\|$. If the dimension of the phase space expand from m to $m+1$, the two - point distance measure will change to:

$$R_{m+1}^2(t) = R_m^2(t) + \|x(t-m\tau) - x_f(t-m\tau)\|^2 \quad (2)$$

Through the above analysis, the importance of decision variables from two models $f_1(x_1, x_2, \dots, x_i, \dots, x_d)$ and

$f_2(x_1, x_2, \dots, x_i, \dots, x_{d+1})$ can be calculated by using FNN algorithm. According to equation (2), we can easily see that characteristics x_{d+1} is or not important, depending on the change $\|x(t-m\tau) - x_f(t-m\tau)\|$ between the $R_m(t)$ and $R_{m+1}(t)$ whether or not significant difference. If X_f is used as the origin of the feature space, that $X_f(t) = (0, 0, \dots, 0)$, the change $\|x(t-m\tau) - x_f(t-m\tau)\|$ between $R_m(t)$ and $R_{m+1}(t)$ has become $\|x_{d+1}\|$. And then according to the pthagorean theorem, calculating $\|x_{d+1}\|$ is changed to calculate the euclidean distance between $\mathbf{A} = (x_1, x_2, \dots, x_d, x_{d+1})$ and $\mathbf{B} = (x_1, x_2, \dots, x_d, 0)$. In actual applications, we use a similarity measure $\cos\theta$ instead of $\|x(t-m\tau) - x_f(t-m\tau)\|$. In order to complete the investigation of all decision variables, specific steps are as follows:

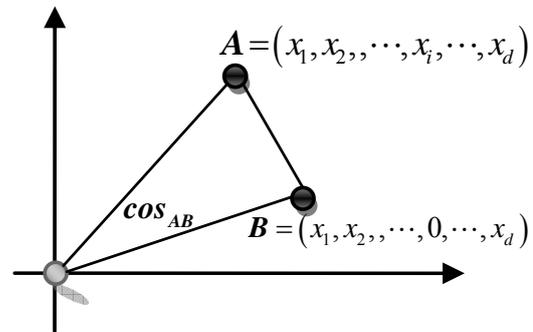


Figure 1. The diagram of FNN

- 1) As shown in Fig.1, d -dimensional data which is composed by all decision variables is viewed as $\mathbf{A} = (x_1, x_2, \dots, x_d)$.

- 2) Remove the original decision variables x_i , we can get a low-dimensional space projection, as in (3).

$$\mathbf{B} = (x_1, x_2, \dots, x_{i-1}, 0, x_{i+1}, \dots, x_d) \quad (3)$$

- 3) Calculate the similarity measure \cos_{AB} between the high-dimensional phase space point \mathbf{A} and its projection \mathbf{B} .

Where:
$$\cos_{AB} = \frac{\mathbf{A} \cdot \mathbf{B}^T}{\|\mathbf{A}\| \cdot \|\mathbf{B}\|} \quad (4)$$

- 4) Examine variables x_1, x_2, \dots, x_d in turn, determine the influence of the input variables to the original data structure by comparing the corresponding change of $\sum_{i=1}^l \cos_{AB}$, and sort according to the change from big to small in the last, where l is the number of samples.

The greater the similarity measure \cos_{AB} , and the smaller the effect to the original data structure when removing the variable; however, the smaller the similarity measure \cos_{AB} , the larger the to the original data structure.

C. Randomization method(RM)

RM [18] is a pruning technique essentially, which do count test significantly of statistical indicators based on

the statistical theory, excluding no significant indicators, and then retaining significant indicators, so achieving the purpose of excluding irrelevant variables.

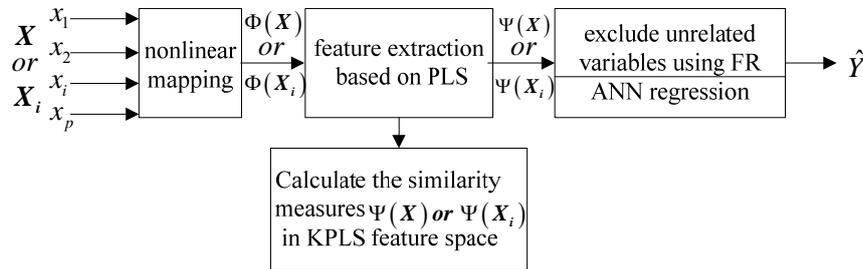


Figure 2. The diagram of decision variable selection based on FR in KPLS feature sapce

In order to make variable importance obtained by FNN more statistical significance, using RM to test the significant level of the variable similarity measures and exclud irrelative variables is an effective way.

Therefor, under the revelation of the RM and FNN, similarity measure of input variables based on FNN as shown in the formula (4) is used to do a random test.Finally,the decision variables can be selected according to the high and low of the significant level. Following these steps:

- 1) According to a large number of data samples normalized, calculate the similarity measure of each decision variable using FNN, and refer to as the standard value.
- 2) Randomly change the order of the sample output set.
- 3) Used the new samples and FNN, recalculate the similarity measure of each decision variable and record as the random value.
- 4) Repeat steps 2) and 3) a large number of times, using count to record the number of repetitions(such as 999), according to the step 3)to record every random value.
- 5) Calculate the significant degree P of similarity measure for each decision variable .
 - a)If the standard value is greater than 0, $P = (N + 1) / (COUNT + 1)$, N is the number of the random value greater than or equal to the standard value.
 - b)If the standard value less than 0, $P = (M + 1) / (COUNT + 1)$, M is he number of the randomization value less than or equal to the standard value.
- 6) If P is less than or equal to the significant level α (such as 0.05), reserve the decision variable, otherwise exclude.

III. THE PROPOSED METHOD

In this section, we describe the proposed variable selection algorithm based on FR in KPLS feature space.

A. The Basic Principles of FR

The basic principles of FR: any variable x_i in the original variable group $\mathbf{X} = (x_1, x_2, \dots, x_i, \dots, x_n)$ is set to zero using FNN, and get a new variables set

$\tilde{\mathbf{X}}_i = (x_1, x_2, \dots, 0, \dots, x_n)$. Then the two points are mapped to feature space by the kernel function and the similarity measure $\cos\theta$ between two points can be calculated in the new space. All the $\cos\theta_i (i = 1, 2, \dots, n)$ are arranged from small to large order, and the original variable which has greater similarity measure indicates it has little contributions to dependent variable. Finally, the RM is used to test the significant level of similarity measure from each variable. If a variable has not significant meaning, then it can be deleted from variables set. The diagram is showed in Fig. 2.

B. The Variable Selection Algorithm

Based on the above analysis, the variable selection algorithm based on FR in KPLS feature space can be described as follow:

1) The KPLS algorithm

Input: data $\mathbf{X} = \{x_1, \dots, x_i\}$, dimension n ;

$\mathbf{Y} = \{y_1, \dots, y_i\}$, dimension m ;

- a) Initialize vector u randomly;
- b) $t = \Phi\Phi^T u, t \leftarrow t / \|t\|$;
- c) $c = Y^T t$;
- d) $u = Yc, u \leftarrow u / \|u\|$;
- e) Repeat steps b)—d), until convergence;
- f) Calculate the residual space of the feature space and the dependent variable space

$$\Phi\Phi^T \leftarrow (\Phi - t t^T \Phi)(\Phi - t t^T \Phi)^T, Y \leftarrow Y - t t^T Y \quad (5)$$

Where $K = \Phi\Phi^T, K(x_i, x_j) = \varphi(x_i)^T \varphi(x_j)$.

Output: Transformed data $\hat{\mathbf{X}} = \{\hat{x}_1, \dots, \hat{x}_i\}$.

2) FR algorithm

a) for $i = 1 : n$

for $j = 1 : l$

Calculate the similarity measure of i -th variable in column j before and after set to zero:

$$\cos\theta_{ij} = \frac{A(:, j) \cdot B(:, j)^T}{\|A(:, j)\| \cdot \|B(:, j)\|} \quad (6)$$

end

Calculate the average of the similarity measure from

TABLE I.
DAILY DATA FROM 11 KINDS OF PROCESS PARAMETERS

z	x_1	x_2	x_3	x_4	x_5	x_6	x_7	...	x_{11}	Y
1	1751	2.55	16.6	17	945	1236.4	19	...	1250	12245
2	1750	2.55	16.3	18	944	1240	20	...	1250	12320
3	1744	2.55	17	15	946	1236	17	...	1240	12268
4	1745	2.59	18	16	947	1231.5	18	...	1220	12558
...
130	1682	2.44	18	14	951	1236.5	15.4	...	1210	11905

TABLE II.
SIX PRINCIPAL COMPONENT SCORES EXTRACTED BY KPLS

the data sample from aluminum electrolytic process

C	l_1	l_2	l_3	l_4	l_5	l_6	l_7	l_8	...	l_{130}
C_1	1.012	0.058	0.129	-0.046	0.233	0.342	-0.237	0.239	...	0.001
C_2	0.149	0.351	-0.829	0.377	-0.345	0.339	0.480	-0.029	...	-0.065
C_3	-0.008	0.577	0.187	0.577	0.546	0.558	0.677	0.1568	...	0.078
C_4	0.512	-0.34	0.135	-0.34	0.435	0.114	-0.042	0.154	...	0.146
C_5	0.289	0.671	-0.569	0.897	-0.446	0.219	0.202	-0.933	...	-0.214
C_6	0.438	0.472	0.452	0.384	0.721	-0.35	0.523	0.677	...	0.084

i -th variable :

$$\cos \theta_i = \frac{1}{l} \sum_{j=1}^n \cos \theta_{ij} \tag{7}$$

end

Record the average of the similarity measure as the standard value (before randomized).

b) Randomly change the order of the sample output set;

c) Repeat steps a) and b), and using *COUNT* to record the number of repetitions (such as 999), according to the step a) to record every random value;

d) Calculate the significant degree P of similarity measure for each decision variable;

① If the standard value is greater than 0,

$$P = (N + 1) / (COUNT + 1) \tag{8}$$

then N as the randomization value greater than or equal to the number of standard values;

② If the standard value less than 0,

$$P = (M + 1) / (COUNT + 1) \tag{9}$$

M is the number of the randomization value less than or equal to the standard value.

e) If P is less than or equal to the significant level α (such as 0.05), reserve the decision variable, otherwise exclude.

According to significant indicators of n variables, the decision variable whose significant level $\alpha \leq 0.05$ will be excluded from the variables set. The original variable group is $\mathbf{X} = (x_1, x_2, \dots, x_n)^T$, and the new set of independent variables after excluding redundant variables is $\hat{\mathbf{X}} = (\hat{x}_1, \hat{x}_2, \dots, \hat{x}_k)$ which reduces $n - k$ dimensions.

IV. SIMULATION EXPERIMENTS AND RESULTS ANALYSIS

A. Variable Selection of Aluminum Electrolytic Process

The nonlinear data set used in the experiment research is collected from the actual production site of aluminum reduction cell. Parameters are respectively in detail line current (A) represented x_1 , molecular ratio (1) x_2 , aluminum level (cm) x_3 , electrolyte levels (cm) x_4 , cell temperature ($^{\circ}C$) x_5 , aluminum tapping volume (kg) x_6 , daily dosage of fluoride salt (kg) x_7 , blanking interval (s) x_8 , cell voltage (mv) x_9 , blanking times (1) x_{10} , aluminum indicator (kg) x_{11} , aluminum electricity consume per ton (KW · h/t - Al) Y . By sampling the electrolyzer No. 224 which is 170KA series from an aluminum factory, we've collected the 130 group daily data from 11 kinds of process parameters, as shown in Table I.

By applying the method proposed in combination with parameter matrixes from Table I, six principal component scores $C_1, C_2, C_3, C_4, C_5, C_6$ can be easily calculated and extracted by KPLS, as shown in Table II.

Then, the x_1, x_2, \dots, x_{11} are projected into a new space built by $C_1, C_2, C_3, C_4, C_5, C_6, Y$. The FR is used to calculate the similarity measure of each process variables in the new space. For instance, x_1 from in the $a = (x_1, x_2, \dots, x_{11})$ is set to zero, so the new

TABLE III.
THE AVERAGE OF SIMILARITY MEASURE AND SIGNIFICANT INDICATORS BASED ON FR IN KPLS FEATURE SPACE

x_i	x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8	x_9	x_{10}	x_{11}
asm	0.693	0.656	0.697	0.846	0.723	0.766	0.742	0.536	0.426	0.942	0.878
α	0.043	0.024	0.032	0.026	0.05	0.042	0.032	0.014	0.004	0.572	0.736

vector $b = (0, x_2, \dots, x_{11})$ can be gotten. After that, the a and b are both mapped to the new space, so the $a^* = (C_{1a}, C_{2a}, \dots, C_{11a}, Y)$ and $b^* = (C_{1b}, C_{2b}, \dots, C_{11b}, Y)$ can be obtained. Then the similarity

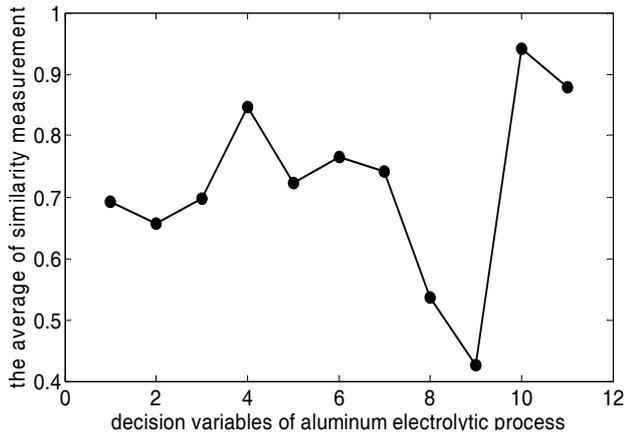


Figure 3. The average similarity measure of process variables in aluminum electrolytic process

measure $\cos a^* b^*$ can be calculated of a^* and b^* . In accordance with the above procedures, the $\cos a^* b^*$ of process variables can be gotten respectively. And then the average of similarity measure(asm) and the significant indicator ($\alpha = 0.05$) of each variable can be calculated and obtained, as shown in Table III. The Fig 3 shows the x_{10}, x_{11} have the highest similarity measure. Bold italics in the table indicate the corresponding variable has a significant effect, which must be retained, and the rest removed. So the variables x_{10}, x_{11} will be removed from the variables set. Finally

$X = (x_1, x_2, x_3, x_4, x_5, x_6, x_7, x_8, x_9, x_{10}, x_{11})$ has become $X = (x_1, x_2, x_3, x_4, x_5, x_6, x_7, x_7, x_9)$.

B. Model Analysis Based on ANN

In order to detailed analysis the precision and effect of the model before and after the variables removed, this study establishes two different models based on significant indicator α .

Model 1: exclude (x_{10}, x_{11}) , establish the mathematical model based on $(x_1, x_2, x_3, x_4, x_5, x_6, x_7, x_8, x_9)$.

Model 2: establish the mathematical model of aluminum electrolytic process based on all variables $(x_1, x_2, x_3, x_4, x_5, x_6, x_7, x_8, x_9, x_{10}, x_{11})$.

The 130 group daily samples of aluminum reduction cell is divided into 100 group training set samples and 30 group testing set samples. Then, the BP Neural Network[21] is used to build a 3-layer feedforward network, that the input of the decision-making parameters is 9 and 11 respectively, the output is aluminum electricity consume per ton, transfer function of the hidden layer and the output layers is sigmoid and purelin respectively. The formula (10) is used to determine the number of hidden nodes. The initial weights and biases of the artificial neural network(ANN) are set to the random number between (-1,1).

$$h = \sqrt{n + m} + k \tag{10}$$

Where: h is the number of hidden layer neurons; n the number of input layer neurons; m the number of output layer neurons; k the constant between 1-10.

After tested based on the formula (10), the model 1 employs 12 hidden layer nodes, and the model 2 employs 13. The model structure is shown in Fig. 4 from model 1 and Fig.5 from model 2.

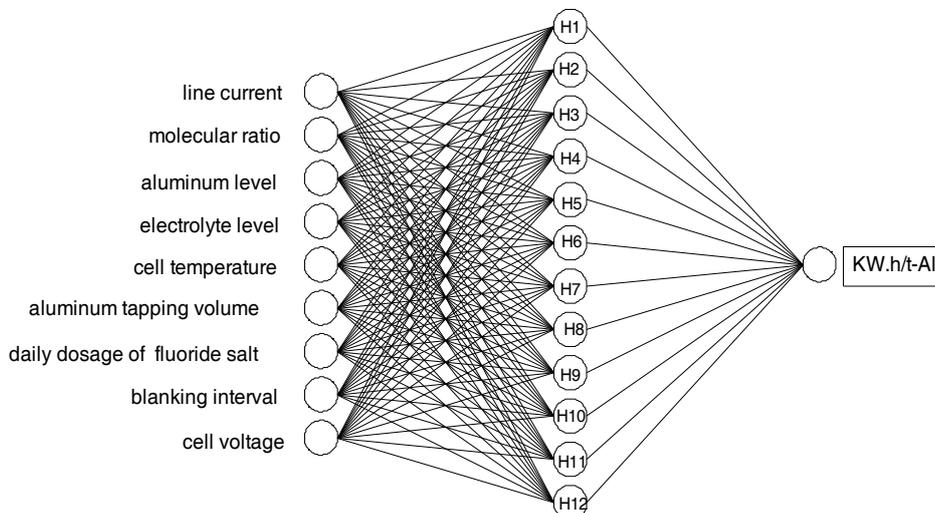


Figure 4. Model 1

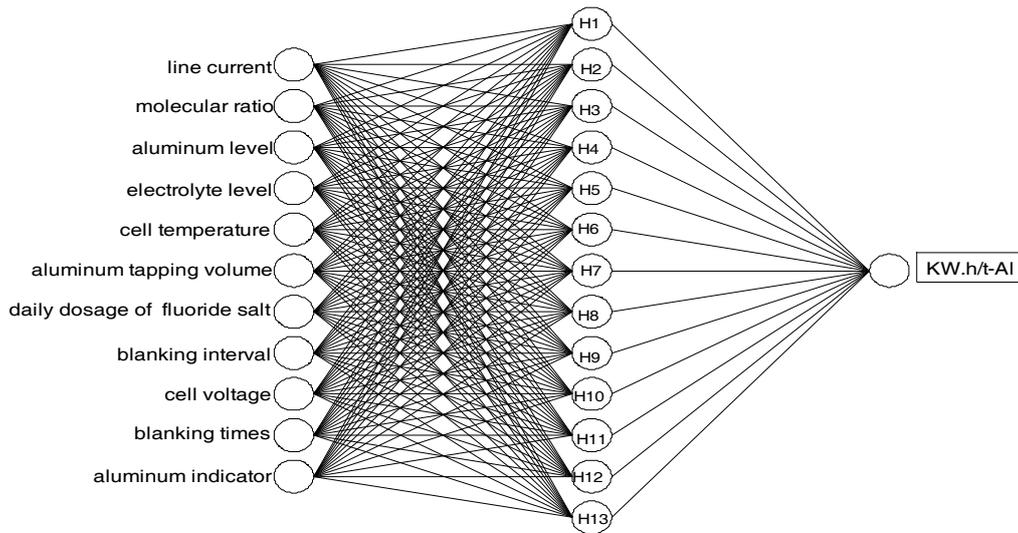


Figure 5. Model 2

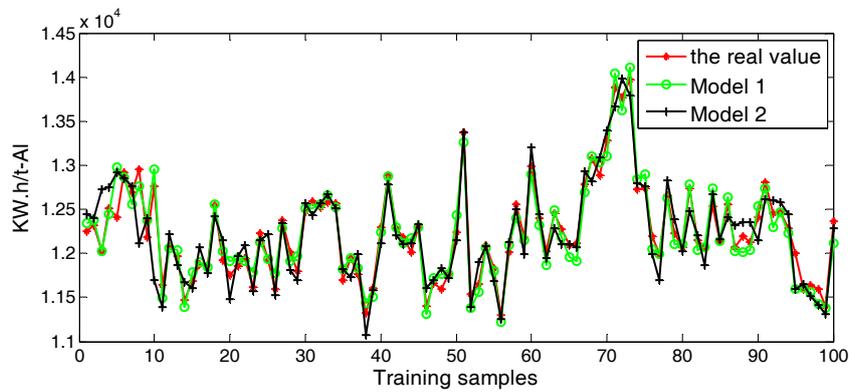


Figure 6. Fitting curve of the two models from the training set

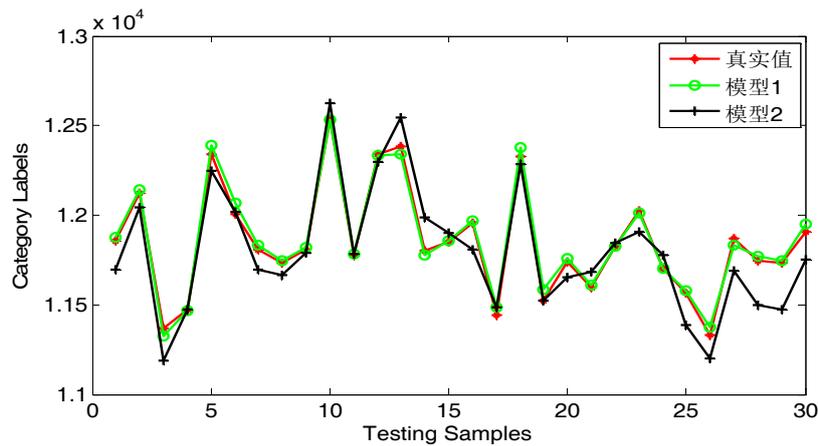


Figure 7. Fitting curve of the two models from the test set

TABLE IV.
Relevant statistical indicators from test set

Model	Max	Min	Average	MSE	running time of model
Model 1	50	1	8	415	1.920
Model 2	200	20	93	12630	2.530

Fig. 6 and Fig. 7 are respectively the fitting curves of the training sets and test sets from the two models. Although both models get a good fitting effect from fitting curve which can be presented in Fig.6, the model 1 has a better fitting curve compared to the model 1, which can be seen clearly in Fig. 7. Meanwhile, according to the Table IV, it's easy to find that the model 1 not only has a higher precision, but also higher efficiency. Therefore, the model 1 which selects out the best process variables achieves a better effect after removing redundant variables.

V. CONCLUSION

In order to select out the best process variables of the energy consumption model in aluminum electrolysis process, a novel method based FR in KPLS feature space is presented. Firstly, the KPLS is employed to transform the original space to the PLS feature space; then, in the new feature space, the FNN is employed to calculate the similarity measure of process variables, so the importance rank of process variables can be obtained; lastly, the RM is used to test the significance level of the importance of each variable in turn, so the unrelated variables can be excluded. The results show that the process energy consumption model in aluminum electrolysis process obtains a high degree of accuracy after the redundant variables excluded.

In the future, the study will continue to explore the selective ability of FR in KPCA, KICA and KCCA feature space for forming a complete set of variable selection techniques and giving the criteria of application of each method.

ACKNOWLEDGEMENTS

The work is supported by National Natural Science Foundation of China (No.51075418), Application Development Major Projects of Chongqing (No.cstc2013yykfC0034), Natural Science Foundation Key Project (No.CQ CSTC2013jjB40007), Chongqing Educational Committee Science and Technology Research Project (No.KJ111417 and No.KJ091402), Foundation of Chongqing University of Science and Technology (No. CK2011B04 and CK2011Z01).

REFERENCES

- [1] G. Huchet, M. Boussuge, V. Maurel, H. Roustan, "Mechanical properties of a spinel-ferrite-based cermet: effects of temperature and oxidation," *Journal of Materials Science*, Springer, vol.48,no.8,pp. 3264-3271, 2013
- [2] C.-Y. Cheung, C. Menictas, J. Bao, M. Skyllas-Kazacos, and B. J. Welch, "Spatial temperature profiles in an aluminum reduction cell under different anode current distributions," *American Institute of Chemical Engineers*, Wiley-Blackwell, vol.59,no.5,pp. 1544-1556, 2013
- [3] M. Abbasi, M. A. Abdul, B. Omidvar, A. Baghvand, "Forecasting municipal solid waste generation by hybrid support vector machine and partial least square model," *International Journal of Environmental Research*, Univ Tehran,vol.7, no.1,pp. 27-38,2013.
- [4] R. D. Pereira, J. Sousa, S. Vieira, S. Reti, S. Finkelstein, "Modified Sequential Forward Selection Applied to Predicting Septic Shock Outcome in the Intensive Care Unit," in *Soft Methods in Probability and Statistics 2013*, 2013 6th International Conference on,vol.190,2013,pp. 469-477
- [5] K. Y. Chan, C. K. Kwong, T. S. Dillon, Y. C. Tsim, "Reducing overfitting in manufacturing process modeling using a backward elimination based genetic programming," *Applied soft computing*, Elsevier, vol.11, no.2, pp. 1648-1656, 2011.
- [6] N. Zhou, J. W. Pierre, D. Trudnowski, "A Stepwise Regression Method for Estimating Dominant Electromechanical Modes," *Power Systems*, IEEE Transactions on, vol.27, no.2, pp. 1051-1059, 2012.
- [7] F. Zhdanov, Y. Kalniskhan, "An identity for kernel ridge regression," *Theoretical Computer Science*, Elsevier, vol.473, no. SI, pp. 157-178, 2013.
- [8] C. X. Wu, J. K. Zhu, D. Cai, C. Chen, J. J. Bu, "Semi-Supervised Nonlinear Hashing Using Bootstrap Sequential Projection Learning," *Knowledge and Data Engineering*, IEEE Transactions on, vol.25, no.6, pp. 1380-1393, 2013.
- [9] J. C. Boulet, D. Bertrand, G. Mazerolles, R. Sabatier, "A family of regression methods derived from standard PLSR," *Chemometrics and Intelligent Laboratory Systems*, Elsevier, vol.120, pp. 116-125, 2013.
- [10] J. Yi, T. F. Li, Y. Y. Su, W. J. Hu, T. Gao, "A variable selection method based on KPCA and FNN for nonlinear system modeling," in *BMEI, 2011 Proceedings IEEE*, May 2011, pp. 832-835.
- [11] J. B. Li, H. J. Gao, "Sparse data-dependent kernel principal component analysis based on least squares support vector machine for feature extraction and recognition," *Neural Computing & Applications*, Springer, vol.21, no.8, pp. 1971-1980, 2013.
- [12] R. C. Zhang, L. H. Wang, S. H. Zhang, "Windows memory analysis based on KPCR," in *Information Assurance and Security 2009*, 2009 5th International Conference on, vol.2, 2009, pp. 677-680
- [13] J. S. Wang, Q. P. Guo, "D-FNN based soft-sensor modeling and migration reconfiguration of polymerizing process," *Applied Soft Computing*, Elsevier, vol.13, no.4, pp. 1892-1901, 2013.
- [14] W. Suphamitmongko, G. L. Nie, R. Liu, "An alternative approach for the classification of orange varieties based on near infrared spectroscopy," *Computers and Electronics in Agriculture*, Elsevier, vol.91, no.2, pp. 87-93, 2013.
- [15] J. H. Li, P. L. Cui, "Improved kernel fisher discriminant analysis for fault diagnosis," *Expert Systems with Applications*, Elsevier, vol.36, no.2, pp. 1423-1432, 2009.
- [16] Z. X. Li, X. P. Yan, C. Q. Yuan, "Intelligent fault diagnosis method for marine diesel engines using instantaneous angular speed," *Journal of Mechanical Science and Technology*, Korean Soc Mechanical Engineers, vol.26, no.8, pp. 2413-2423, 2012.
- [17] H. D. Abarbane, R. Brown, J. J. Sidorowich, "The Analysis of Observed Chaotic Data in Physical Systems," *Rev. Mod. Phys.*, American Physical Soc, vol.65, no.4, pp. 1331-1392, 1993.
- [18] J. D. Olden, D. A. Jackson, "Illuminating the "black box": a randomization approach for understanding variable contributions in artificial neural networks," *Ecological Modelling*, Elsevier, vol.154, no.1-2, pp. 135-1502, 2002.
- [19] P. Susanne, K. Daniel, H. Thomas, "Significant amino acids in aroma compound profiling during yeast fermentation analyzed by PLS regression," *Lwt-Food*

Science and Technology, Elsevier, vol.51, no.2, pp. 423-4322, 2013.

- [20] N. Hadi, F. Abbas, N. Mehrab, "Application of GA-KPLS and L-M ANN calculations for the prediction of the capacity factor of hazardous psychoactive designer drugs," *Medicinal Chemistry Research*, Springer, vol.21, no.9, pp. 2680-2688, 2012.
- [21] J. J. Wang, B. L. Chen, Y. C. Yang. Mehrab, "Approximation of algebraic and trigonometric polynomials by feedforward neural networks," *Neural Computing & Applications*, Springer, vol.21, no.1, pp. 73-80, 2012.

Lizhong Yao received his B.S. degree in automation from Chongqing University of Science and Technology, China in June 2009 and his M.S. degree in Detection Technology and Automatic Equipment from Xi'an Shiyou University, China in June 2013. His current research interest includes modeling and optimization of complex systems.

Taifu Li received his B.Sc, M.Sc and Ph.D all from Chongqing University in 1996, 2000 and 2004 respectively. Now he is a professor and supervisor for M.Sc candidate in Chongqing University of Science and Technology. His main areas of interest and research are intelligent control and soft-sensing.

Jun Peng received his Ph. D. degree from Chongqing University, China in 2003. Now he is a professor and president in Chongqing University of Science and Technology. His main areas of interests are application of chaos theory and digital picture processing.

Deyong Wu received his B.Sc, M.Sc from Wuhan University of Science and Technology, China in 2011. Now he is a Master degree candidate in Chongqing University of Technology. His main areas of interest is materials science.

Jun Yi received his Ph. D. degree from Chongqing University, China in 2010. Now he is an associate professor at the Department of Electrical and Information Engineering, Chongqing University of Science and Technology. He His research interest includes intelligent control and wireless sensor network.

Yingying Su received her B.Sc and M.Sc from Heilongjiang University and Chongqing University of Technology, China in 2005 and 2008 respectively. She started to work on her Ph.D from Chongqing University, China in 2011. Now she is a lecturer in Chongqing University of Science and Technology. Her research interest is computational intelligence.