Soft Sensor Modeling Based on Rough Set and Least Squares Support Vector Machines

LI CHUAN[1,2], WANG SHILONG[2], ZHANG XIANMING[1] and XU JUN[1]
1. Institute of Separation Machinery
Chongqing Technology and Business University
Chongqing 400067
2. College of Mechanical Engineering
Chongqing University
Chongqing 400044
CHINA

Abstract: Soft sensor is an effective tool to estimate industrial process variables which are hard to be measured online for the technical or economical reasons. The modeling methods of the sensor are related to the approximating precision and speed. A soft sensor model with rough set and Least Squares Support Vector Machines (LSSVM) is presented in the paper. The rough set is employed to compress the data for preprocessing, which can get rid of the multicollinearity and reduce the dimension of input variables for the model. To solve the nonlinear and multiple input characteristics of industrial process, the LSSVM is delivered for model regression. The model is applied for moisture content soft sensing of vacuum oil purifier. The result shows that the proposed method features high speed and precise approximation ability, which has better performance of generalization for tracking the trend of the moisture content variety during oil purification.

Key-Words: Soft Sensor, Least Squares Support Vector Machines, Rough Set Theory, Vacuum Oil Purifier, Modeling, Moisture Content

1 Introduction

For the technical or economical reasons, some essential variables in production process are hard to be measured online [1]. There are time lags between laboratory measurement and product line requirement, which makes the production quality is difficult to be controlled accurately [2].

Based on the optimization control rules, a set of easy-to-measure variables concerning primary variables are selected as secondary variables. By means of modeling the relations between the primary variables and secondary variables, soft sensor technology can estimate the primary variables from the model in time [3]. The main modeling methods at present include mechanism based, data-driven based, hybrid and other methods [4].

Support Vector Machine (SVM) is a powerful methodology for solving problems in nonlinear classification, function estimation and density estimation which has also led to many other recent developments in kernel based learning methods in general [5]. Recent years, Least Squares SVM (LSSVM) was presented by Sukens. It is closely related to regularization networks and Gaussian processes, but additionally emphasizes and exploits primal-dual interpretations [6]. Taking into account the noises in the actual processes and multicollinearity between secondary variables, rough set is firstly employed to compress the data for preprocessing [7]. Then LSSVM is delivered to train the extracted knowledge to model the soft sensor.

The presented model is applied for moisture content soft sensing of vacuum oil purifier. There are many complex factors affecting moisture content in oil in the working process of vacuum oil purifier. According to multiple easy-to-measure variables in different sampling point of the plant, such as pressure, flow rate, degree of vacuum, temperature and so on, the moisture content of purified oil is computed by the soft sensor. The proposed model can simulate the work mechanism of vacuum oil purifier and perform real-time measurement on the effect of operation parameters to the process. In the experiment, 150 samples are employed for model training and other 150 samples for testing. Comparing LSSVM with BPNN and RBFNN soft sensor models, the result shows that the proposed method features high speed and precise approximation ability, which has better performance of generalization for tracking the trend of the moisture content variety during oil purification.

2 Soft Sensor Modeling Method

2.1 Rough Set Theory

As a methodology of artificial intelligence domain, rough set theory is an effective tool for representing and reasoning about imprecise or uncertain information. It was introduced by Pawlak in the early 1980s [8]. Based on the notion of indiscernibility and the inability to distinguish between objects, the rough set deals with the approximation of sets or concepts by means of binary relations, typically constructed from empirical data. Such approximations can be said to form models of our target concepts, and hence in its typical use falls in under the bottom-up approach to model construction.

An information system can be defined as $S=(U,R,V,f)$. Where, $U=\{x_1,x_2,...,x_n\}$, is the sample set, i.e. universe. $R=C\bigcup D$, is the attribute set. Its subset A is the set of condition attributes and D is the set of decision attributes. $V=\bigcup V_s$, is the range set of attributes $s\in A\bigcup D$, and $V_s=[l_s,r_s]$, is the range. $f:U\times R\to V$, is the function mapping samples into attributes. At the most conditions, condition attributes are not all the same important in *S*. The redundant attributes can not only waste memory resource, but also interfere right decision making [9]. So the reduction is necessary for decision making in rough set theory. Suppose there is a subset $B\subseteq R$, the indiscernibility of *B* at *U* is

$$IND(B)=\{(x,x')\in U^2 \mid \forall a\in B, a(x)=a(x')\} \quad (1)$$

Indiscernibility is an equivalence property. Let *U/IND(B)* represents all of the equivalent classes of *U*. For arbitrary element $x\in U$, its B equivalent classes are

$$[x]_B=\{y \mid (x,y)\in IND(B)\} \quad (2)$$

For arbitrary object set of the system $X\subseteq U$ and attribute subset $B\subseteq R$, define the lower approximation of *B* to *X* is

$$\underline{BX}=\{x \mid [x]_B\subseteq X\} \quad (3)$$

Define the upper approximation of *B* to *X* is

$$\overline{BX}=\{x \mid [x]_B\bigcap X\neq\Phi\} \quad (4)$$

So in *S*, the arbitrary least subset $B\subseteq R$ with *IND(A)=IND(B)* is a reduct marked *RED(A)*. Following typical steps show what the reduction in rough set theory is.

(1) All of the condition attributes and decision attributes in universe form the decision table.

(2) Discretize the real value attributes with discretization algorithms.

(3) Reduce the discretized decision table and generate the reduced decision rules.

2.2 LSSVM

SVM is developed with small samples machine learning theory [10]. Suppose that there is a sample set $(x_1,y_1),(x_2,y_2),...,(x_k,y_k),...,(x_l,y_l)\in R_n\times R$ whose length is *l*. Where, x_k is the input variable set and y_k is the output variable set. Through a nonlinear function $\Phi(\bullet)$, the input vector is mapped from the original space *x* to a high dimensional space *z* and to construct an optimization hyperplane in accordance with structural risk minimization. Suppose the optimization decision function is

$$f(x)=\omega^T\varphi(x_k)+b \quad (5)$$

The structural risk minimization principle is used to compute ω^T and *b*. Where, for arbitrary input x_k except sample set, $|y-\omega^T\varphi(x_k)-b|\leq\varepsilon$. In this way, the ω^T and *b* computation problem is transformed to be the following optimization problem [11].

$$\min J=\frac{1}{2}\|\omega\|^2+cR_{emp} \quad (6)$$

Where, $\|\omega\|^2$ is the value range controlling the complexity of the model, c>0 is the error penalty function and R_{emp} is the empirical risk, i.e. ε loss function. Different loss functions can form different SVMs. For standard SVM, with the increase of the sample number, the more complicated algorithm makes the computing speed slower. Fortunately, LSSVM is proposed to solve the problem. It differs from the standard SVM in the loss function and equation constraints. The loss function of the standard SVM optimization is the error ξ_i . However, another type is employed by LSSVM to make the optimization problem to be

$$\min J=\frac{1}{2}\omega^T\omega+\frac{1}{2}c\sum_{i=1}^{l}\xi_i^2 \quad (7)$$

$$S.t. \quad y_i=\omega^T\varphi(x_i)+b+\xi_i, \quad i=1,2,...,l$$

According to the upper equation, following Lagarangian function is defined to solve the optimum.

$$L(\omega,b,\xi,\alpha)=\frac{1}{2}\omega^T\omega+\frac{1}{2}c\sum_{i=1}^{l}\xi_i^2-\sum_{i=1}^{l}\alpha_i[\omega^T\varphi(x_i)-y_i+b+\xi_i] \quad (8)$$

The optimization conditions are

$$\frac{\partial L}{\partial\omega}=0,\frac{\partial L}{\partial b}=0,\frac{\partial L}{\partial\xi}=0,\frac{\partial L}{\partial\alpha}=0 \quad (9)$$

Computing from equations above, following equations can be obtained.

$$\begin{cases} \omega - \sum_{i=1}^{l} \alpha_i \varphi(x_i) = 0 \\ \sum_{i=1}^{l} \alpha_i = 0 \\ \alpha_i = c\xi_i \\ \omega^T \varphi(x_i) + b + \xi_i - y_i = 0 \end{cases} \qquad (10)$$

A kernel function $K(x, x_i) = \Phi(x)\Phi(x_i)$ is defined to replace the nonlinear mapping function. Equation (10) is transformed to be the following linear equation.

$$\begin{bmatrix} & \mathbf{I}^T \\ \mathbf{I} & K(x_i, x_j) + \frac{\mathbf{I}}{c} \end{bmatrix} \begin{bmatrix} b \\ \boldsymbol{\alpha} \end{bmatrix} = \begin{bmatrix} 0 \\ \mathbf{y} \end{bmatrix} \qquad (11)$$

Where, $\mathbf{y} = [\hat{y}_1, \hat{y}_2, ..., \hat{y}_l]^T$, $\mathbf{I} = [1,1,...,1]^T$, $\boldsymbol{\alpha} = [\alpha_1, \alpha_2, ..., \alpha_l]^T$. So the regression estimation model of LSSVM is

$$f(x) = \sum_{i=1}^{l} \alpha_i K(x, x_i) + b \qquad (12)$$

2.3 Modeling Steps

As the simulation of abstract thinking, rough set theory represents and reasons about imprecise or uncertain information by means of data set reduction. So the knowledge access process of rough set is just a case-based reasoning. However, its generalization ability is not strong enough. And it can only deal with the discretized data.

With the machine learning algorithm, LSSVM express the implicit relations between input and output data. Instead of minimizing training error, it minimizes generalization risk, which features better generalization performance. However, it is a black-box model. So it just likes imagery thinking.

The general thinking of mankind combines abstract thinking with imagery thinking. Integrating rough set and LSSVM, the rough set is employed as preprocessor. Then the preprocessed data is used for LSSVM modeling. In this way, the attribute dimension and number are all reduced so as to improve the computing speed and estimation precision. Following method shows the main steps of soft sensor modeling based on rough set and LSSVM.

Step 1: Data preprocessing. The sample data form the information system *S*. So the quality of samples is one of the most important factors affecting on the information value. The better the quality is, the more effective the decision making is. The classical rough set theory is based on the integrality and consistency of the data. However, the actual samples knowledge is usually incomplete, inconsistent or duplicate. So after the data preprocessing, a complete, consistent decision table can be obtained.

Step 2: Discretize the real value attributes. As rough set theory can only process the discretized data, such as the high, middle, low oil level, other real value attributes (including condition attributes and decision attributes) should be discretized firstly. For example, moisture content, flow rate, temperature and so on.

Step 3: Reduce decision table. Set R is a equivalent set, and $r \in R$. If *IND(R)=IND(R-{r})*, *r* can be reduced from *R*. The decision table reduction consists of attributes and values reduction.

Step 4: Select LSSVM kernel function and parameters. In the actual applications, different kernel functions can be selected to form different LSSVMs. There are some kernel functions in common use: multinomial kernel function $K(x, x_i) = (x.x_i + 1)^d$, RBF kernel function $K(x, x_i) = \exp(-\| x - x_i \| / 2\sigma^2)$, Sigmoid kernel function $K(x, x_i) = \tanh(\beta x_i + b)$, B-spline kernel function $K(x, x_i) = B_{2n+1}(x - x_i)$ and so on.

Step 5: Soft sensor training. Extract reducts from the optimization decision table. Train LSSVM using samples from the optimization decision table. Examine the model using the testing samples.

Step 6: If the precision of model is not as it should be, select more samples and goto step 1. Otherwise output the model. End.

3 Soft Sensor Model of Vacuum Oil Purifier

Vacuum oil purifier is designed with the vacuum drying theory for moisture abstraction. As the vacuum oil purification is a complicated system process with mass, heat, momentum transfer and phase transition, a mechanism model explaining its process completely has not been constructed yet. Moisture content in purified oil, one of the most important indexes of vacuum oil purification, can be computed from a data-driven soft sensor model.

The experimental platform is a ZL-30S vacuum oil purifier developed by us. The plant has the functions of particles filtration, moisture abstraction, gas removal and dielectrical strength improvement so as to be used in air conditioner compressor product line to purify N56 refrigerator oil. As is shown in Fig. 1, the refrigerator oil is pumped into vacuum separation tower for purification. According to the local data of the plant, following variables are taken into consideration: moisture content at oil inlet C_1, vacuum degree at vacuum separation tower C_2, temperature at oil inlet C_3, temperature at heater outlet C_4, temperature at bottom of separation tower C_5, flow rate at oil inlet C_6, flow rate at circulation loop C_7, flow rate at oil outlet C_8 and oil level of separation tower C_9. The primary variable of soft sensor is the moisture content at oil outlet D_1.

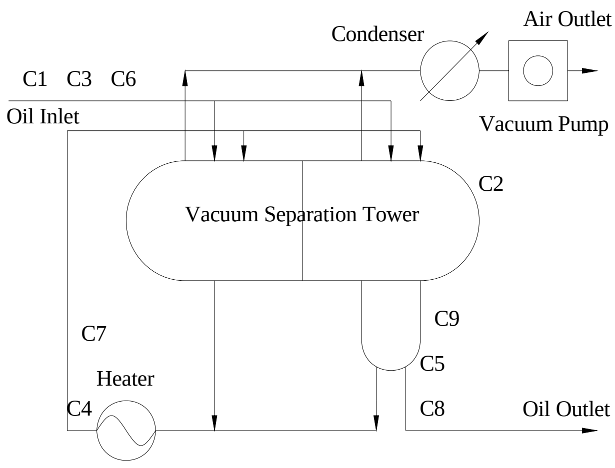

Fig. 1. Flow chart of moisture abstraction

The control system of ZL-30S vacuum oil purifier consists of master-slaver structure. A PLC is employed to be the slaver of the control system for all of the secondary variables acquirement. Through master-slaver communication, secondary variables are delivered to the master, industrial computer, for soft sensor modeling.

300 samples are random chosen from the working process of the vacuum oil purifier. In the experiment, 150 samples are used for model training and other 150 samples for testing. The real values of testing samples are shown in Fig.2.

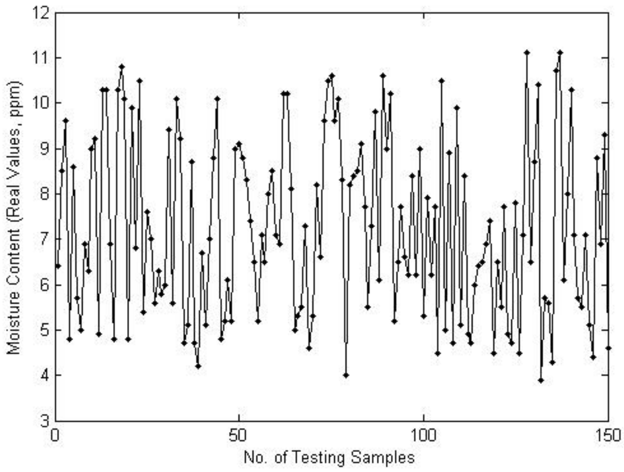

Fig. 2. Real values of testing samples of moisture content

According to the proposed steps, 150 samples form a decision table. After reduce the attributes and attribute values, the input vector dimension is reduced from 9 to 6 and the effective samples are compressed from 150 to 58. Take them as the input variables of LSSVM and let RBF as the kernel function. By cross validation, let c=182.0739 and $\sigma^2 = 11.59043$ to train the soft sensor model.

Then the paper estimate 150 testing samples with trained soft sensor model. Fig. 3 shows the estimation values of testing samples. Comparing with the real values (Fig. 2), Fig. 4 shows the estimation error between real values and estimation values for each testing sample.

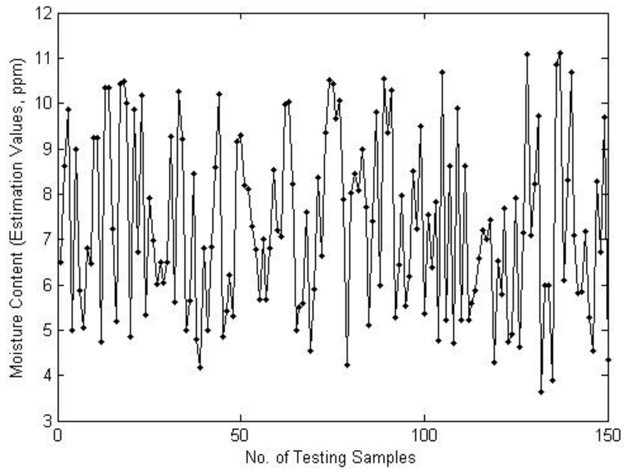

Fig. 3. Estimation values of testing samples (the real values are shown in Fig.2) using soft sensor model based on rough set and LSSVM

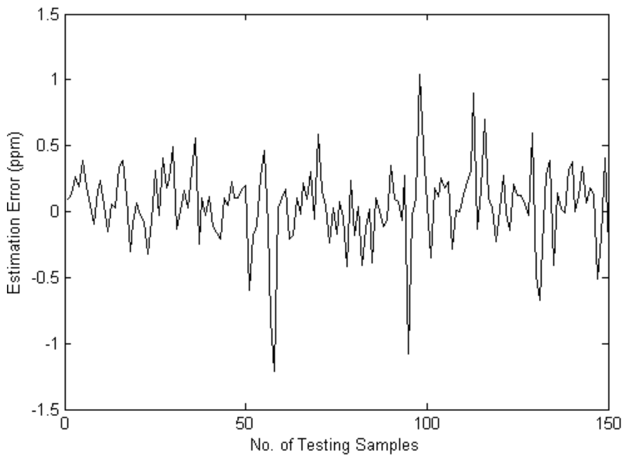

Fig. 4. Error curves of soft sensor model

To validate the effectiveness of the model, the same samples are employed for BPNN and RBFNN to model soft sensor [12]. In the comparative experiments, let BPNN consist of 3-layer networks and the number of hidden layer neurons is 10. The width of RBF function of RBFNN is 0.01. The same 300 samples are used for training and testing. Define Root Mean Square Error (*RMSE*) and Maximal Absolute Error (*MAXAE*) are error weight parameters:

$$\begin{cases} RMSE = \sqrt{\dfrac{\sum_{i=1}^{l}\left(\dfrac{y_i - \hat{y}_i}{y_i}\right)^2}{l}} \\ MAXAE = \max_{i=1}^{l}\left|y_i - \hat{y}_i\right| \end{cases} \tag{13}$$

Table 1 shows that both *RMSE* and *MAXAE* of LSSVM method are all the least. So the proposed method features precise approximation ability, which has better performance of generalization for tracking the trend of the moisture content variety during oil purification.

Table 1. Estimation results of several soft sensor modeling methods

Method	*RMSE*	*MAXAE*
Proposed Method	0.0472	1.2068
BPNN	0.1161	1.8992
RBFNN	0.1841	2.3592

4 Conclusions

A soft sensor model with rough set and Least Squares Support Vector Machines (LSSVM) is presented. The rough set are employed to compress the data for preprocessing, which can get rid of the multicollinearity and reduce the dimension of input variables for the model. To solve the nonlinear and multiple input characteristics of industrial process, the LSSVM is delivered for model regression. The model is applied for moisture content soft sensing in the process of vacuum oil purifier. In the experiment, 150 samples are employed for model training and other 150 samples for testing. Comparing LSSVM with BPNN and RBFNN soft sensor models, the result shows that the proposed method features high speed and precise approximation ability, which has better performance of generalization for tracking the trend of the moisture content variety during oil purification.

References:

[1] Park S. and Han C.A., A nonlinear soft sensor based on multivariate smoothing procedure for quality estimation in distillation columns, *Computers and Chemical Engineering*, Vol.24, 2000, pp. 871-877.

[2] Luo J.X., Shao H.H., Soft sensing modeling using neurofuzzy system based on rough set theory, In: *Proceedings of the 2002 American Control Conference*, Alaska, 2002, pp. 543-548.

[3] Luo J.X., Shao H.H., Selecting secondary measurements for soft sensor modeling using rough sets theory, In: *Proceedings of the 4th World Congress on Intelligent Control and Automation*, Shanghai, 2002, pp. 415-419.

[4] Densai K., BarHe Y., Tambe S.S., etc., Soft-sensor development for fed-batch bioreactors using support vector regression, *Biochemical Engineering Journal*, Vol.27, No.3, 2006, pp. 225-239.

[5] Vapnik V.N., An overview of statistical learning theory *IEEE Trans. Neural Networks*, Vol.10, No.5, 1999, pp. 988-999.

[6] Suykens J.A.K., Vandewalle J., Least squares support vector machine classifiers, *Neural processing Letters*, Vol.9, No.3, 1999, pp. 293-300.

[7] Xu L., Zhang W.J., Comparison of different methods for variable selection, *Analytica Chimica Acta*, Vol.446, 2001, pp. 477-483.

[8] Pawlak Z., Rough sets, *International Journal of Information and Computer Science*, Vol.11, No.5, 1982, pp. 341-356.

[9] Zhai L.Y., Khoo L.P., Fok S.C., Feature extraction using rough set theory and genetic algorithms-An application for the simplification of product quality evaluation, *Computers and Industrial Engineering*, Vol.43, No.4, 2002, pp. 661-676.

[10] David V., Sanchez A., Advanced support vector machines and kernel methods. *Neurocomputing*, Vol.55, No.1-2, 2003, pp. 5-20.

[11] Lin C.J., Formulations of support vector machines: A note from an optimization point of view, *Neural Computation*, Vol.13, No.2, 2001, pp. 307-317.

[12] Marcantonio C., Ada F., Fault diagnosis of electronic analog circuits using a radial basis function network classifier, *Measurement*, Vol.28, 2000, pp. 147-158.