Drifting Modeling Method Using Weighted Support Vector Machines with Application to Soft Sensor[1)]

FENG Rui[1] ZHANG Yue-Jie[1] ZHANG Yan-Zhu[2] SHAO Hui-He[3]

[1] (*Department of Computer Science and Engineer, Fudan University, Shanghai* 200433)

[2] (*Department of Information Engineering, Shenyang Institute of Technology, Shenyang* 110015)

[3] (*Department of Automation, Shanghai Jiaotong University, Shanghai* 200030)

(E-mail: frui@sh163. net)

Abstract The kernel problem in soft sensor of industrial processes is how to build the soft sensor model. However, there exist some questions to some extent in soft sensor model with conventional modeling methods such as global single model and multiple models. Using the high generalization ability of support vector machines (SVMs) and the idea of locally weighted learning (LWL) algorithm, this paper proposes a novel learning algorithm named weighted support vector machines (W_SVMs) which is suitable for local learning. We also present a drifting modeling method based on this algorithm. The proposed modeling method is applied to the estimation of Box-Jenkins gas furnace and FCCU and the simulation results show that the proposed approach is superior to the traditional modeling methods.

Key words Support vector machines, weighted support vector machines, locally weighted learning, modeling

1 Introduction

Traditionally, most of the conventional modeling and identification methods for industrial process are global modeling methods and multiple modeling methods. However, many complex industrial processes are characterized by multivariable, nonlinear, wide operation range, *etc.*, which are too complex to be described by a single model. In order to overcome the aforementioned difficulties, some researchers advocate representing complex processes by multiple models(MM)[1]. However, there also exist some disadvantages such as optimization of model set, switch strategy of sub-models, *etc.* Locally weighted learning(LWL)[2] is a form of instance-based algorithm for learning continuous non-linear mappings. Because the algorithm is particularly appropriate for learning complex and highly non-linear functions, it not only can avoid the single model's poor generalization ability and tremendous computation complexity, but also can overcome the delay between the estimation and actual output brought by sub-models' switch of MM[2]. Support vector machines (SVMs), originally developed by Vapnik[3], is a new learning machine based on the statistical learning theory. For it not only has solid theory, but also can solve many practical problems such as small samples, over learning, high dimension and local minima, SVMs have become a new focuss artificial neural network(ANN). SVMs has shown attractive potential and promising performance in a wide range of fields and applications[4,5]. However, its applications and researches in industrial process modeling have been received little attention.

Combining the idea of LWL algorithm with SVMs, this paper develops an improved SVMs algorithm, named weighted support vector machines (W_SVMs), which is suitable for local learning. A drifting modeling (DM) method which is used to estimate the outputs

1) Supported by National High Technology Research and Development "863" Program of P. R. China (001 AA413130)
Received April 16, 2003; in revised form October 13, 2003
收稿日期 2003-04-16 ; 收修改稿日期 2003-10-13

of complex industrial process is proposed based on W_SVMs. The proposed modeling method is applied Box-Jenkins gas furnace and fluidized catalytic cracking unit (FCCU) and the simulation results show that the property of this proposed approach is superior to those of MM methods[6] and single model based standard SVMs method.

2 Weighted support vector machines

Given the samples $\{\boldsymbol{X}_i, y_i\}_{i=1}^{l}$, where l is the size of samples, $\boldsymbol{X}_i \in R^n$ is the input vector, and $y_i \in R$ is the corresponding desired response, the Vapink SVMs algorithm[3] transforms the data regression problem into a convex optimization problem as (1).

Note that all the training points are treated uniformly in the standard SVMs algorithm. However, in many real-world applications, the effects of the training points are different. It is often that some training points are more important than others in the regression problem. [9] proposed an improved SVMs algorithm named fuzzy support vector machines (F_SVMs) and applied it to a data classification problem successfully. We here propose another improved SVMs algorithm, W_SVMs.

$$\min_{\boldsymbol{\omega},b,\xi,\xi^*} J = \frac{1}{2}\boldsymbol{\omega}^{\mathrm{T}}\boldsymbol{\omega} + C\sum_{1}^{l}(\xi_i + \xi_i^*)$$

$$\text{s. t.} \quad \begin{cases} y_i - \boldsymbol{\omega}^{\mathrm{T}}\varphi(\boldsymbol{x}_i) - b \leqslant \varepsilon + \xi_i \\ \boldsymbol{\omega}^{\mathrm{T}}\varphi(\boldsymbol{x}_i) + b - y_i \leqslant \varepsilon + \xi_i^* \quad i = 1,2,\cdots,l \\ \xi_i, \xi_i^* \geqslant 0 \end{cases} \tag{1}$$

The sample data $\{\boldsymbol{X}_i, y_i\}_{i=1}^{l}$ are pretreated by introducing weights $p_i \in R$ such that the sample data have the form: $\{\boldsymbol{X}_i, y_i, p_i\}_{i=1}^{l}$, where $\delta \leqslant p_i \leqslant 1$ is the weight with a sufficient small $\delta > 0$. Since the weight p_i is the attitude of the corresponding point $\boldsymbol{X}_i$ toward the process output y_i and the parameters ξ, ξ^* are a measure of error in the SVMs, the term $p_i(\xi_i + \xi_i^*)$ is a measure of error with different weights. The regression problem is then regarded as the solution to

$$\min_{\boldsymbol{\omega},b,\xi,\xi^*} J = \frac{1}{2}\boldsymbol{\omega}^{\mathrm{T}}\boldsymbol{\omega} + C\sum_{1}^{l} p_i(\xi_i + \xi_i^*) \tag{2}$$

The constraints in (2) are the same as in (1). It is noted that a smaller p_i reduces the effect of the parameters ξ, ξ^* in problem (2), such that the corresponding point $\boldsymbol{X}_i$ is treated as less important. Similar to Vapnik SVMs algorithm, the optimization problem (2) can be solved in its dual formulation. According to the primal objective function and corresponding constraints stated in (2), A Lagrange function L is presented by introducing a set of dual variables as follows

$$\begin{cases} \max J = \max\limits_{\alpha,\alpha^*} \min\limits_{\boldsymbol{\omega},b,\xi,\xi^*} L = -\frac{1}{2}\sum\limits_{i,j=1}^{l}(\alpha_i - \alpha_i^*)(\alpha_j - \alpha_j^*)\boldsymbol{\varphi}(\boldsymbol{x}_i)^{\mathrm{T}}\boldsymbol{\varphi}(\boldsymbol{x}_j) \\ \qquad\qquad - \varepsilon\sum\limits_{i=1}^{l}(\alpha_i + \alpha_i^*) + \sum\limits_{i=1}^{l} y_i(\alpha_i - \alpha_i^*) \end{cases}$$

$$\text{s. t.} \quad \begin{cases} \sum\limits_{i=1}^{l}(\alpha_i - \alpha_i^*) = 0 \\ \alpha_i, \alpha_i^* \in [0, p_i C] \end{cases} \quad i = 1,2,\cdots,l \tag{3}$$

The $\boldsymbol{\omega}$ and the nonlinear approximate function can be written as follows.

$$\begin{cases} \boldsymbol{\omega} = \sum\limits_{i=1}^{l}(\alpha_i - \alpha_i^*)\boldsymbol{\varphi}(\boldsymbol{x}_i) \\ f(\boldsymbol{x}) = \sum\limits_{i=1}^{l}(\alpha_i - \alpha_i^*)k(\boldsymbol{x}, \boldsymbol{x}_i) + b \end{cases} \tag{4}$$

In comparison with Vapnik SVMs, the difference is that weight p_i is added in (2), which shows that the same $(\alpha_i - \alpha_i^*)$ may denote the different SVs, and then the corresponding input vectors have the different effect on the output of model. Therefore, we can control the precision of model indirectly with the control of weight p_i. The advantage not only helps to improve the estimation precision of model, but also helps to eliminate the sample data with less weight in order to reduce the algorithm complexity. If all p_i are set 1, the W_SVMs algorithm is the same as Vapnik SVMs algorithm, that is, the standard SVMs algorithm is the special case of W_SVMs.

3 Generating the weights

Defining the Euclidian distance of input vector $\boldsymbol{X}_k$ and sample data $\boldsymbol{X}_i$ as $s_{ik} = \| \boldsymbol{X}_i - \boldsymbol{X}_k \|$, weight p_i is the function of s_{ik}:

$$p_i = f(s_{ik}) \quad s_{\min} \leqslant s_{ik} = \| \boldsymbol{X}_i - \boldsymbol{X}_k \| = \sqrt{(\boldsymbol{X}_i - \boldsymbol{X}_k)^{\mathrm{T}}(\boldsymbol{X}_i - \boldsymbol{X}_k)} \leqslant s_{\max} \tag{5}$$

According to LWL method[7], the input vectors corresponding to $s_{ik} = s_{\min}$, would be most important, that is, $p_{\max} = f(s_{\min}) = 1$, and those corresponding to $s_{ik} = s_{\max}$ would be least important, that is, $p_{\min} = f(s_{\max}) = \delta$. If the Euclidian distance is between $(s_{\min}, s_{\max})$, the corresponding input vectors′ weights can be determined by (5).

We can select the weight function as an exponential function of s_{ik}:

$$p_i = f(s_{ik}) = a\left(\frac{1}{s_{ik}} - b\right)^2 + c \tag{6}$$

By applying the boundary conditions, we can get

$$p_i = f(s_{ik}) = (1-\delta)\left(\frac{\dfrac{1}{s_{ik}} - \dfrac{1}{s_{\max}}}{\dfrac{1}{s_{\min}} - \dfrac{1}{s_{\max}}}\right)^2 + \delta \tag{7}$$

4 Drifting modeling method

The idea of the DM method is from LWL algorithm [7]. In order to obtain the sample data which are near the input vectors, the sample database is created with the sample data which are obtained from industrial field. Therefore, the local estimation model can be achieved with the nearby points. The outputs of model are sent to the control equipment to participate in feedback control. The difference between the DM method and MM method is that the training samples are not predetermined, but obtained on-line according to the input vectors. The fundamental frame is illustrated in Fig. 1.

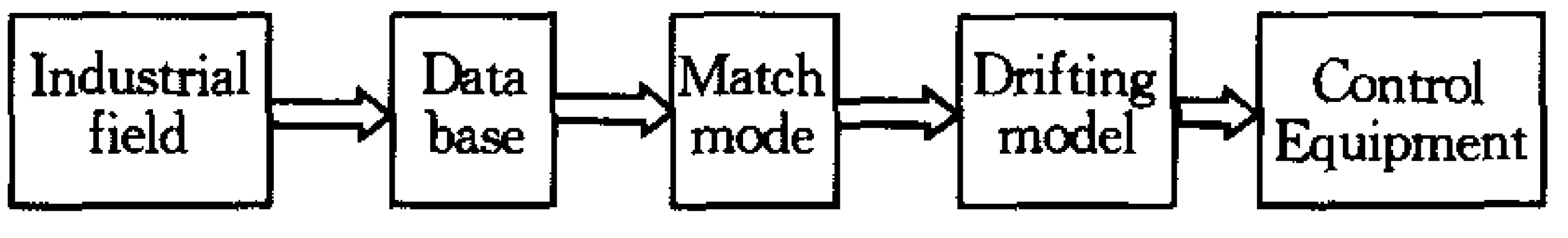


Fig. 1 Frame of the drifting modeling method based on the W_SVMs

The practical drifting modeling method consists of the following steps.

Step 1. Create the sample database. Data preprocessing is needed, which includes eliminating data with obvious errors, filtrating of random errors, eliminating redundancy data and conciliating data. All modes of process should be included in the database;

Step 2. Obtain the nearest sample data of the k_{th} input vector $\boldsymbol{X}_k$ and corresponding weight p_i;

Step 3. Solve the estimation model f_k which is based on the W_SVMs algorithm, and $\hat{y}_k$, the estimation of $\boldsymbol{X}_k$, can be achieved;

Step 4. $k = k+1$, go to Step 2.

5 Simulation and results

In this section, we applied the DM method to modeling Box-Jenkins gas furnace and heavy oil fluidized catalytic cracking unit (FCCU). Some numeric results and discussion are presented to see the benefits of DM using W_SVMs.

5.1 Soft sensor case

The first example is based on absorption stabilization system (ASS) of Shi Jiazhuang refinery's heavy oil fluidized catalytic cracking unit (FCCU)[10]. DM method and Vapnik SVMs are applied to soft sensor modeling. The sample data with 100 samples are actual data obtained from Shi Jiazhuang refinery in May, 1998[10]. RBF kernel function is used in both methods. The parameters are $\mu=0.15$, $c=10$ and $\varepsilon=0.01$. The weight p_i can be determined with (7), where $s_{max}=0.25$ and $\delta=0.001$. The results are given in Fig. 2 and Table 1. Fig. 2 shows respectively the curves of estimated values of drifting model and the corresponding actual values. Table 1 shows the comparison of the two methods. From Table 1, we can see that the RMSE and GMSE of soft sensor drifting model using W_SVMs are less than using Vapnik SVMs method, which shows that the generalization ability of the drifting model is better. Furthermore, the MAXE of the proposed method is less than Vapnik SVMs, which shows that the smoothness of the model's output is better. It also indicates that the drifting model using W_SVMs is suitable for the soft sensor modeling of complex industrial process.

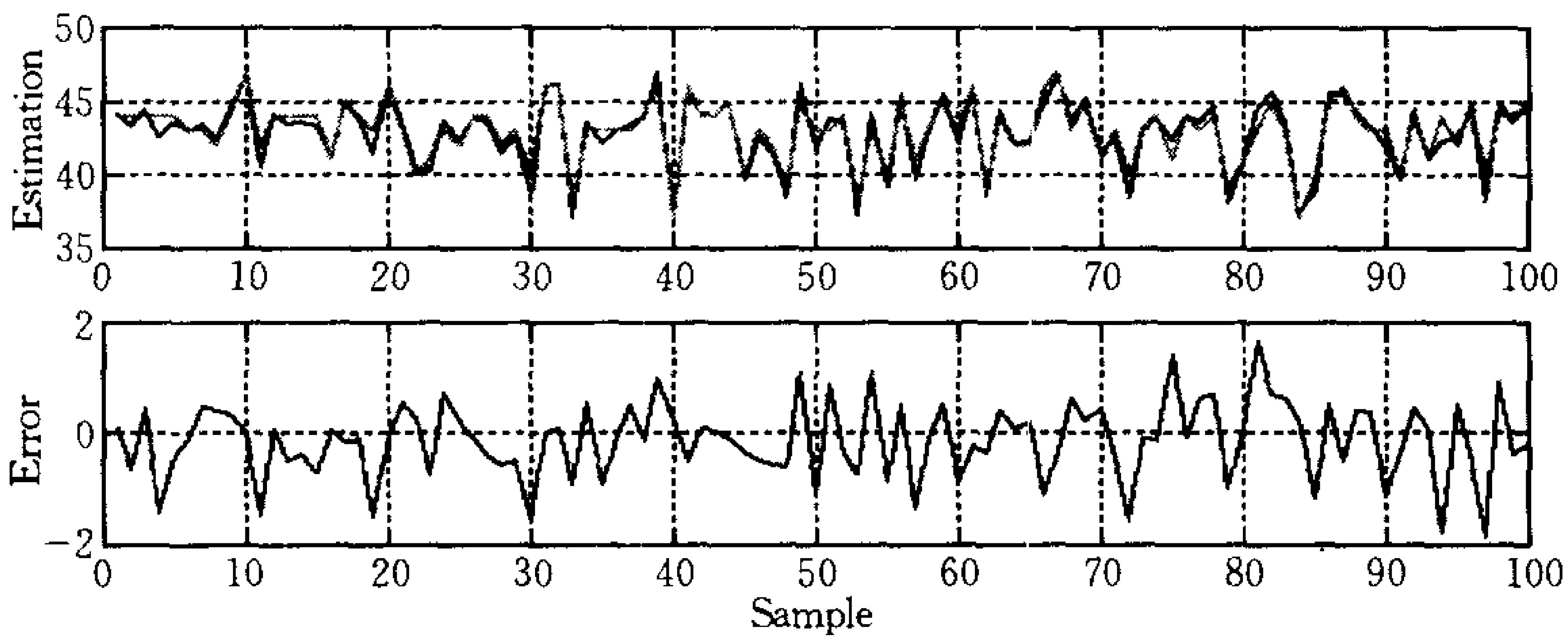


Fig. 2 The curves of W_SVMs drifting model output and actual output ($\mu=0.15, c=10, \varepsilon=0.01$)

Table 1 Performance analysis of W_SVMs and Vapnik SVMs

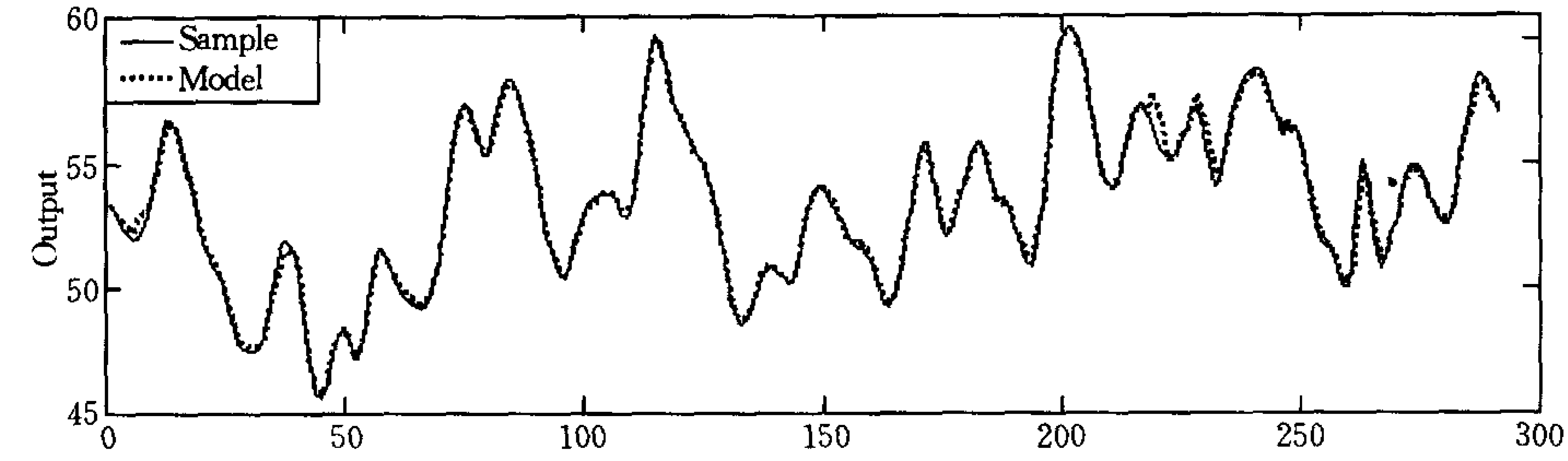
	W_SVMs	SVMs
RMSE	0.7223	0.9644
GMSE	0.0722	0.1275
MAXE	1.8771	2.2445

5.2 Box-jenkins gas furnace

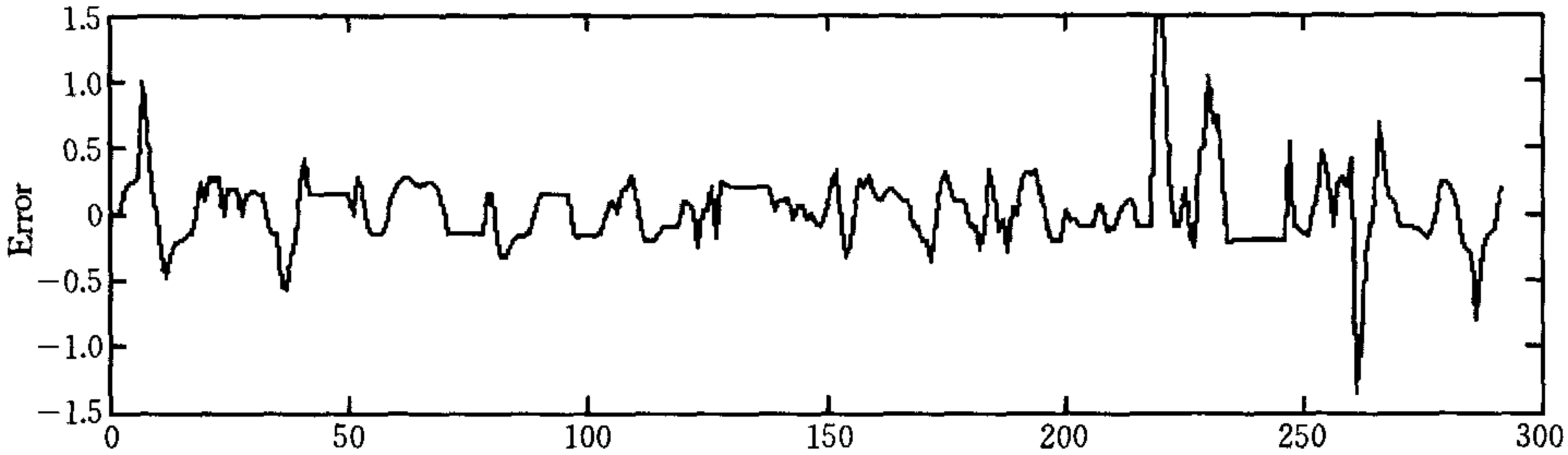
The Box-Jenkins's gas furnace is a famous example of system identification[8], many researchers have proposed different multiple modeling methods to estimate the output of the system[6]. RBF kernel function is used in W_SVMs. Its parameters are $\mu=0.15, \varepsilon=0.01$, $c=10, s_{max}=0.25, \delta=0.001$, and the weight p_i can be determined with (7). The results are given in Fig. 3(a), 3(b) and Table 2. Fig. 3(a), 3(b) show the curves of estimated values with DM method and errors. In Table 2 we present the comparison of our method with several multiple models proposed in the literature, showing the number of variables chosen in each case and the performance of the models.

Table 2 Comparison of box-jenkins gas furnace modeling methods[6]

Model	Tong'77	Pedrycz'84	Xu'87	Peng'88	Sugeno'91	Sugeno'93	Wang'96	EST1/EST2/EST3	W_SVMs
RMSE	0.684	0.565	0.572	0.548	0.261	0.435	0.397	0.400/0.401/0.396	0.2696
No. inputs	2	2	2	2	6	3	2	2	2



(a) The curves of W_SVMs drifting model output and actual output ($\mu=0.15, c=10, \varepsilon=0.01$)



(b) The errors curves of W_SVMs drifting model output and actual output ($\mu=0.15, c=10, \varepsilon=0.01$)

Fig. 3 W_SVMs drifting model

We obtain a superior performance to other "classical" approaches: the RMSE is 32% less than EST3 model[6] and 60% less than Tong'77 model[6]. The MAXE (1.6) is only 3.3% of the smallest output, which indicates that the smoothness of the model is good. The EST1, EST2 and EST3 models also exhibit a good performance[6]. However, the methods EST1 and EST2 model the system with a greater error, and the method EST3 involves a higher computational complexity[6]. The Sugeno' 91 model[6] provides the smallest RMSE, however, the model is based on more variables (six input variables) and more computational complexity. Therefore, our DM approach is suitable for complex industrial process modeling with serious nonlinear and multiple work modes.

6 Conclusion

In this paper, a novel soft sensor drifting modeling (DM) method is proposed to model complex industrial process. The motivation relies on the observation that a single model is usually incapable of modeling these industrial process properties, and multiple modeling methods would lead to a delay between the estimation and real system output. Hence the drifting modeling method based on locally weighted learning (LWL) algorithm is proposed. Because Vapnik SVMs is not suitable for local learning, we introduce sample weight concept and propose an improved SVMs algorithm: weighted support vector machines (W_SVMs). This proposed algorithm can not only improve the estimation precision of model, but also reduce the compute complex by eliminating some learning sample data set with least weight. Supported by comparison with multiple modeling method and single modeling method, the new modeling approach is shown to be suitable for modeling complex industrial process with multiple work modes. Meanwhile, this approach sheds some light on the great potential application in industrial field.

References

1 Takagi T, Sugeno M. Fuzzy identification of systems and its applications to modeling and control. *IEEE Transactions on Systems, Man and Cybernetics*, 1985, (15): 116~132

2 Bottou L, Vapnik V. Local learning algorithms. *Neural Computation*, 1992, 4(6): 888~900

3 Vapnik V. Nature of Statistical Learning Theory. New York: Springer Verlag, 1995
4 Zhang Xue-Gong. Introduction to statistical learning theory and support vector machines. *Acta Automatica Sinica*, 2000, **26**(1): 32~42(in Chinese)
5 Hsu Chih-Wei, Lin Chih-Jen. A comparison of methods for multiclass support vector machines. *IEEE Transactions on Neural Networks*, 2002, **13**(2): 415~425
6 Gomez-Skarmeta A F, Delgado M, Vila M A. About the use of fuzzy clustering techniques for fuzzy model identification. *Fuzzy Sets and Systems*, 1999, **106**(2): 179~188
7 Atkeson C G, Moore A W, Schaal S. Locally weighted learning. *AI Review*, **11**(1): 11~73
8 Box G E P, Jenkins G M. Time Series Analysis Forecasting and Control. San Francisco, CA: Holden-Day, 1970
9 Lin Chun-Fu, Sheng-De Wang. Fuzzy Support Vector Machines. *IEEE Transactions on Neural Networks*, 2002, **13**(2): 464~471
10 Wang Yong-Sheng. Studies on soft sensor theories, methods, software package, and industrial application[Ph. D. dissertation]. Shanghai Jiaotong University, 2000(in Chinese)

FENG Rui Received his master degree and Ph. D. degrees from Northeastern University and Shanghai Jiaotong University in 1997 and 2003, respectively. Now, He works in the Department of Computer Science and Engineering of Fudan University. His research interests include statistical learning theory, support vector machines, and soft sensor technique.

ZHANG Yue-Jie Recived her master degree and Ph. D. degrees from Northeastern University in 1997 and 1999, respectively. She is now an associate professor at Fudan University. Her research interests include data mining and statistical learning theory.

ZHANG Yan-Zhu Recived her master degree from Northeastern University in 1998. She is now an instructor at the Shenyang Institute of Technology. Her research interests include statistical learning theory and the control of high performance computer network.

SHAO Hui-He Professor at the Shanghai Jiaotong University, P. R. China. His research interests include the artificial intelligence control and industrial process control.

基于加权支持向量机的移动建模方法及其在软测量中的应用

冯 瑞[1] 张玥杰[1] 张艳珠[2] 邵惠鹤[3]

[1](复旦大学计算机科学与工程系 上海 200433)

[2](沈阳工业学院信息工程系 沈阳 110015)

[3](上海交通大学自动化系 上海 200030)

(E-mail: frui@sh163.net)

摘 要 工业过程软测量技术的核心问题是建立软测量模型,然而,利用传统全局建模方法与多模型建模方法进行复杂工业过程软测量建模时,在不同程度上存在一些问题.本文利用支持向量机(SVMs)泛化能力强的特点,结合局部加权学习(LWL)算法思想,提出一种适于局部学习的加权支持向量机(W_SVMs)学习算法和基于这种算法的移动建模方法.利用这种建模方法对Box-Jenkins煤气炉和重油催化裂化(FCCU)装置进行分析建模,并与其它不同建模方法进行比较,显示了该方法的优点和有效性.

关键词 支持向量机,加权支持向量机,局部加权学习,建模

中图分类号 TP273