Journal of Control Theory and Applications 4 (2004) 365 - 370

Temperature prediction control based on least squares support vector machines

Bin LIU[1,2], Hongye SU[1], Weihua HUANG[2], Jian CHU[1]

(1. National Laboratory of Industrial Control Technology, Institute of Advanced Process Control, Zhejiang University, Yuquan Campus, Hangzhou Zhejiang 310027, China;

2. Department of Automation, College of Information Science and Engineering, Wuhan University of Science and Technology, Wuhan Hubei 430081, China)

Abstract: A prediction control algorithm is presented based on least squares support vector machines (LS-SVM) model for a class of complex systems with strong nonlinearity. The nonlinear off-line model of the controlled plant is built by LS-SVM with radial basis function (RBF) kernel. In the process of system running, the off-line model is linearized at each sampling instant, and the generalized prediction control (GPC) algorithm is employed to implement the prediction control for the controlled plant. The obtained algorithm is applied to a boiler temperature control system with complicated nonlinearity and large time delay. The results of the experiment verify the effectiveness and merit of the algorithm.

Keywords: Predictive control; Least squares support vector machines; RBF kernel function; Generalized prediction control

1 Introduction

Model prediction control (MPC), based on prediction model and receding horizon optimization, has become an attractive feedback control strategy, because it has found successful applications, especially in the process industry[1]. For this kind of control strategy, the prediction model is a crucial component because the essence of MPC is to optimize the forecast of process behavior[2], and the forecast is accomplished with the prediction model. If the controlled plant is linear or slightly nonlinear, it can be fit by linear prediction model effectively. But, if the plant possesses highly nonlinear characteristics and operates over a large region in variable space, a nonlinear prediction model must be used to approximate the system dynamics. So far linear MPC (LMPC) using linear prediction model is quite satisfying, but nonlinear MPC (NMPC) based on nonlinear prediction model is still an open topic.

The learning algorithms for traditional nonlinear modeling approaches, including classical neural networks and fuzzy modeling, etc., are almost based on the expectation risk minimization principle. These kinds of algorithm often cause the problem of overfitting[3]. Simply speaking, for a given learning task with a given finite amount of training data, the smaller training error may result in the poorer generalization performance. Based on the statistical learning and structural risk minimization principle, Vapnik presented the support vector machines (SVM), which emphasize both the expectation risk and the generalization performance and can be used to approximate nonlinear functions [4]. Furthermore, Suykens [5] presented the LS-SVM method, in which equality constraints are used and the quadratic programming problem existing in the standard SVM need not be solved.

In the modeling process of the complex nonlinear system, an open-loop dynamical data set, which includes sufficient information about the system, is usually necessary for training the system model. If the training data set do not include sufficient information, the generalization capacity of the system model must be improved to ensure the model performance. The objection function of the LS-SVM includes two parts, one of which is responsible for improving the generalization performance and the other is in charge of abating the expectation risk. It is simple for LS-SVM to switch model performance between generalization capability and model precision.

Generally, it is difficult to obtain the process data including the whole bunch of the system information. Therefore, it is very important to obtain a system model with better generalization performance. Compared with other modeling methods, the LS-SVM and even SVM methods possess a remarkable merit that they can easily

Received 28 April 2004; Revised 22 September 2004.

This work has been supported by the National Outstanding Youth Science Foundation of China (No.60025308) and the Teach and Research Award Program for Outstanding Young Teachers in Higher Education Institutions of MOE, China.

improve the generalization performance of the model. In this paper, the LS-SVM with radial basis function (RBF) kernel is employed to build the off-line model of the controlled system, and the pruning procedure[6] is used to impose sparseness property on the off-line model.

2 LS-SVM

SVM is one of the methods by which the statistical learning theory can be applied to practice. It has its own advantages in settling pattern recognition problem with small samples, nonlinearity, and higher dimension. SVM can be easily applied to learning problem, such as function estimation.

On the basis of classical SVM, Suykens[5] presented the LS-SVM approach, in which the following function is used to approximate the unknown function,

$$y(x) = w^{\mathrm{T}}\varphi(x) + b, \tag{1}$$

where, $x \in \mathbb{R}^n, y \in \mathbb{R}, \varphi(\cdot): \mathbb{R}^n \to \mathbb{R}^{n_h}$ is a nonlinear function which maps the input space into a higher dimension feature space, w is the weight vector and b denotes a bias term.

Given training data set $\{x_k, y_k\}_{k=1}^{N}$ with input data $x_k \in \mathbb{R}^n$ and corresponding output $y_k \in \mathbb{R}$, LS-SVM defines an optimization problem as follows:

$$\min_{w,b,e} J(w,e) = \frac{1}{2}w^{\mathrm{T}}w + \gamma\frac{1}{2}\sum_{k=1}^{N} e_k^2, \ \gamma > 0, \tag{2}$$

subject to the equality constraints

$$y_k = w^{\mathrm{T}}\varphi(x_k) + b + e_k, \ k = 1,\cdots,N. \tag{3}$$

To solve this optimization problem, one can define the following Lagrange function,

$$L(w,b,e;\alpha) = J(w,e) - \sum_{k=1}^{N}\alpha_k\{w^{\mathrm{T}}\varphi(x_k) + b + e_k - y_k\}, \tag{4}$$

where α is the Lagrange multiplier. Calculating the partial derivatives of $L(w,b,e;\alpha)$ with respect to w,b,e,α, respectively, one gets the optimal condition for Eq.(4),

$$\begin{cases} \dfrac{\partial L}{\partial w} = 0 \to w = \sum_{k=1}^{N}\alpha_k\varphi(x_k), \\ \dfrac{\partial L}{\partial b} = 0 \to \sum_{k=1}^{N}\alpha_k = 0, \\ \dfrac{\partial L}{\partial e_k} = 0 \to \alpha_k = \gamma e_k, \\ \dfrac{\partial L}{\partial \alpha_k} = 0 \to w^{\mathrm{T}}\varphi(x_k) + b + e_k - y_k = 0, \\ k = 1,\cdots,N. \end{cases} \tag{5}$$

Expressing e_k and w with α_k and b, one can transform the above equality into

$$\begin{bmatrix} 0 & \vec{I}^{\mathrm{T}} \\ \vec{I} & \Omega + \gamma^{-1}I \end{bmatrix}\begin{bmatrix} b \\ a \end{bmatrix} = \begin{bmatrix} 0 \\ y \end{bmatrix}, \tag{6}$$

where $y = [y_1,\cdots,y_N]^{\mathrm{T}}, \vec{1} = [1,\cdots,1]^{\mathrm{T}}, \alpha = [\alpha_1,\cdots,\alpha_N]^{\mathrm{T}}$, and Ω is a square matrix in which the element located on k^{th} column and l^{th} row is

$$\Omega_{kl} = \varphi(x_k)^{\mathrm{T}}\varphi(x_l) = K(x_k,x_l), \ k,l = 1,\cdots,N.$$

Choosing $\gamma > 0$, ensure that the matrix

$$\Phi = \begin{bmatrix} 0 & \vec{I}^{\mathrm{T}} \\ \vec{I} & \Omega + \gamma^{-1}I \end{bmatrix}$$

is invertible. Then we have the analytical solution of a and b

$$\begin{bmatrix} b \\ a \end{bmatrix} = \Phi^{-1}\begin{bmatrix} 0 \\ y \end{bmatrix}. \tag{7}$$

Substituting Eq.(7) into Eq.(5), we get

$$y(x) = \sum_{k=1}^{N}\alpha_k K(x,x_k) + b, \tag{8}$$

where $K(x,x_k)$ is the Kernel function, and it can be any symmetric function satisfying Mercer's condition[7]. In this paper, the RBF is used as the Kernel function,

$$K(x,x_k) = \exp\{-\|x - x_k\|_2^2/\sigma^2\}. \tag{9}$$

Note that in the case of RBF kernel function, LS-SVM has only two undetermined parameters: γ in Eq.(2) and σ in Eq.(9).

Standard SVM possesses a sparseness property in the sense that many α_k values are equal to zeros, but this is not the case in LS-SVM due to the fact that $\alpha_k = \gamma e_k$ in Eq.(5). From Eq.(3) to Eq.(5), we can see that any element in α is not equal to zero. That means that all the data vectors in training set are support vectors. So, LS-SVM loses the sparseness property. To get a sparse support vector set, we use the pruning procedure presented by Suykens[6]. In this procedure, the importance of the k^{th} support vector is determined by the corresponding $|\alpha_k|$ (absolute value of α_k). The bigger $|\alpha_k|$ is, the more important the support vector is. On the premise that model performance index does not degrade seriously, some unimportant support vectors can be removed from the support vector set step by step. In the pruning process, the value of γ and σ can be adjusted to improve the model performance.

The following input-output model

$$y = f(x)$$

is employed to denote the controlled system characteristic. Where, $x = [x(1),x(2),\cdots,x(nu + ny)]$ denotes the regression vector including the past input-output data of the system. $f(\cdot)$, a nonlinear function, is used to fit the system characteristic. nu and ny are input and output orders of the system respectively. Input-output data of the system are collected and constitute the training data set $\{x_k, y_k\}_{k=1}^N$. Here, x_k is the regression vector in different sampling instants and y_k is the system output corresponding to x_k. Using the above LS-SVM with RBF kernel, given selected γ and σ, we get the off-line nonlinear model of the controlled system as follows:

$$y(x) = \sum_{k=1}^{N} \alpha_k \exp\{-\|x - x_k\|_2^2/\sigma^2\} + b, \quad (10)$$

where N is the number of support vectors and is determined by the result of pruning procedure.

3 Model linearization and GPC

3.1 Real-time linearization of model

To avoid solving the nonlinear programming problem, the off-line nonlinear model of the controlled plant (Eq.(10)) is linearized at each sampling period of system running.

Let the current instant be the k^{th} sampling instant and the current sampling point be $x(k)$. For the sake of simple expression, let $x_0 = x(k)$. Linearizing Eq.(10) at x_0, we have

$$\begin{aligned} y(x) &= y(x)\,|_{x=x_0} + \left.\frac{\partial y}{\partial x(1)}\right|_{x=x_0}[x(1) - x_0(1)] + \cdots \\ &\quad + \left.\frac{\partial y}{\partial x(nu+ny)}\right|_{x=x_0}[x(nu+ny) - x_0(nu+ny)] \\ &= y(x)\,|_{x=x_0} - \left.\frac{\partial y}{\partial x(1)}\right|_{x=x_0} x_0(1) - \cdots \\ &\quad - \left.\frac{\partial y}{\partial x(nu+ny)}\right|_{x=x_0} x_0(nu+ny) \\ &\quad + \left.\frac{\partial y}{\partial x(1)}\right|_{x=x_0} x(1) + \cdots \\ &\quad + \left.\frac{\partial y}{\partial x(nu+ny)}\right|_{x=x_0} x(nu+ny) \\ &= p + b_1 * x(1) + \cdots + b_{nu} * x(nu) \\ &\quad - a_1 * x(nu+1) - \cdots - a_{ny} * x(nu+ny), \end{aligned}$$

where p is constant in the current sampling period and depends only on the current sampling point.

Given the following regression vector

$$\begin{aligned} x(k) &= [x(1), \cdots, x(nu+ny)] \\ &= [u(k-1), \cdots, u(k-nu), \\ &\qquad y(k-1), \cdots, y(k-ny)], \end{aligned}$$

the input-output difference equation model is obtained as follows,

$$\begin{aligned} y(k) &= p + b_1 u(k-1) + \cdots + b_{nu} u(k-nu) \\ &\quad - a_1 y(k-1) - \cdots - a_{ny} y(k-ny). \quad (11) \end{aligned}$$

Namely,

$$A(z^{-1})y(k) = B(z^{-1})u(k-1) + p, \quad (12)$$

where

$$A(z^{-1}) = 1 + a_1 z^{-1} + \cdots + a_{ny} z^{-ny},$$
$$B(z^{-1}) = b_1 + b_2 z^{-1} + \cdots + b_{nu} z^{-nu+1}.$$

3.2 GPC of the controlled system

According to the input-output linear model at current sampling time (Eq.(12)), we can characterize the controlled system as a discrete difference equation (Eq.(13)) in the neighbor of the current sampling point.

$$A(z^{-1})y(t) = B(z^{-1})u(t-1) + v(t), \quad (13)$$

where $v(t)$ is the error and disturbance resulting from one fitting current system characterized by Eq.(13). We decompose $v(t)$ as follows

$$v(t) = v_{\text{dc}} + v_{\text{ac}}(t),$$

where, v_{dc}, including the constant p emerging in the linearization process, is the direct-current component independent of time. The amplitude of v_{dc} is equal to the mean of $v(t)$. v_{ac} is the AC component whose mean is zero. Modeling v_{ac} with $w(t)/\Delta$, we can transform the input-output equation at the current time into

$$A(z^{-1})y(t) = B(z^{-1})u(t-nd-1) + v_{\text{dc}} + w(t)/\Delta, \quad (14)$$

where, $w(t)$ is the disturbance with zero mean, $\Delta = 1 - z^{-1}$ is the difference operator. In order to forecast the future system output based on the past input-output data and the future system input, we introduce the following Diophantine equations,

$$1 = E_j(z^{-1})A(z^{-1})\Delta + z^{-j}F_j(z^{-1}), \quad (15)$$

$$E_j(z^{-1})B(z^{-1}) = G_j(z^{-1}) + z^{-j}H_j(z^{-1}). \quad (16)$$

Multiplying both the sides of Eq.(14) by $E_j(z^{-1})\Delta z^j$, we have

$$\begin{aligned} &E_j(z^{-1})A(z^{-1})\Delta y(t+j) \\ &\quad = E_j(z^{-1})B(z^{-1})\Delta u(t+j-1) \\ &\qquad + E_j(z^{-1})(\Delta v_{\text{dc}} + w(t+j)). \quad (17) \end{aligned}$$

v_{dc} is independent of time, so $\Delta v_{\text{dc}} = 0$. Then the above equation is simplified as

$$\begin{aligned} &E_j(z^{-1})A(z^{-1})\Delta y(t+j) \\ &\quad = E_j(z^{-1})B(z^{-1})\Delta u(t+j-1) \\ &\qquad + E_j(z^{-1})w(t+j). \quad (18) \end{aligned}$$

Accordingly, the well-known method[8] is used to obtain the multistep prediction of $y(t)$, which can be denoted as the following expression in the form of vector

$$y^o = Gu + Fy(t) + H\Delta u(t-1). \quad (19)$$

Finally, the control law is constructed as follows:

$$u = (G^{\mathrm{T}}G + \lambda I)^{-1}G^{\mathrm{T}}[y_r - Fy(t) - H\Delta u(t-1)]. \quad (20)$$

All the polynomials, vectors and matrices in the Eqs. (15) ~ (20) have the same definition as in Clarke's paper[8].

3.3 Predictive control based on LS-SVM

Based on the given off-line nonlinear model of the controlled plant, the online linear model is obtained by real-time linearization, and then the GPC algorithm is employed to compute control increment at each sampling instant. The block diagram of the prediction control system based on LS-SVM model is given in Fig. 1.

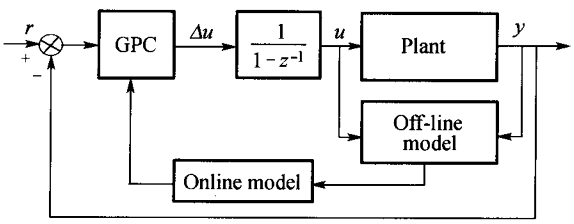

Fig. 1 Structure of the control system.

The algorithm of the prediction control based on LS-SVM is described as follows:

1) Given the dynamical input-output data set of the controlled plant, the off-line model is obtained by the LS-SVM method.

2) The off-line model is linearized according to the input-output data of current instant at each sampling time.

3) Given the linear online model of current sampling period, the GPC strategy is used to compute the control increment.

4) Go to 2), unless the control process is over.

4 Experiment

The EFPT process-control experiment system is a set of integrative experiment equipment. It can provide several kinds of controlled objects such as temperature, pressure and liquid level. In this section, the presented algorithm is used to control the inner barrel temperature of the boiler in the EFPT system.

We construct a computer control system, which is composed of the boiler in the EFPT, the temperature transducer, a data acquisition card PCL818L, a D/A output card PCL726 and an industrial computer. The boiler comprises two-part inner barrel and interlayer. The temperature of both the parts influences each other and the temperature of the interlayer is immesurable, which results in the complicated nonlinear characteristic of the controlled object. The sketch map of the computer control system is shown in Fig. 2.

In Fig. 2, the output signal is collected by the data acquisition card PCL818L and treated by the industrial computer. The control signal is calculated according to the presented control algorithm and then implemented by D/A card PCL726 and SCR control module.

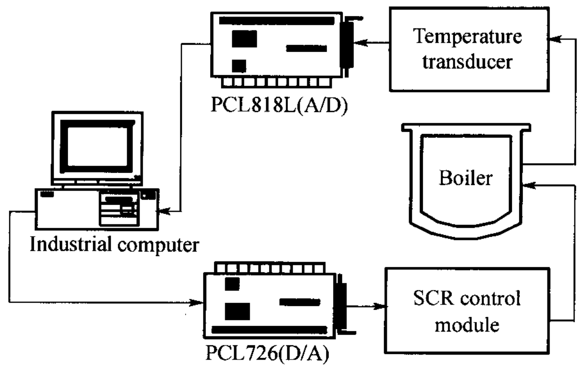

Fig. 2 Sketch map of the computer control system.

Using the LS-SVM method with RBF kernel function introduced in Section 2, we can build the off-line model of the controlled object. The white noise is used as the input signal of the system to produce training data set $\{x_k, y_k\}_{k=1}^{863}$. The boiler's temperature control system is a slow system with nonlinearity and time delay. Considering the specific characteristic of the controlled object, we choose the sampling time as $Ts = 20$s, and choose the order of the input and output variable as $nu = 6$ and $ny = 5$ correspondingly. In fact, the value of nu and ny is influenced by the value of Ts, and the product of nu and Ts must be larger than the dead time. For the normalized process data, a rational span of γ in Eq. (2) should be $5 \leqslant \gamma \leqslant 100$, the smaller γ has better generalization capacity and the bigger γ is of the higher model precision. Here, the parameter γ is chosen as 10. In the process of executing the LS-SVM and pruning algorithm, predictive models with different amount of support vectors can be obtained and the parameter σ in the kernel function can be determined synchronously.

In order to test the approximation performance of the off-line model, a signal composed of four kinds of sine wave with different frequency is employed as test input and the following performance index is used to test the approximation performance

$$VAF = 100 * (1 - \mathrm{var}(YY - YM)/\mathrm{var}(YY)).$$

In the above index, corresponding to the test input, YY is the output of the actual system and YM is the output of the system model. Apparently, the bigger index value represents the better approximation performance.

Performance comparison of four prediction models with different amount of support vectors is given in Table 1, from which we can draw some conclusions. Firstly, in order to solve the sparseness problem (decrease the number of the support vectors), the pruning algorithm can be executed repeatedly until the model performance degrades considerably. Secondly, even if the pruning procedure is not executed and the initial prediction model with 863 support vectors is adopted, the average calculation time of the control variable in each sampling period can be accepted by most practical systems. These results show that the prediction control algorithm presented in this paper is feasible in real-time control.

Table 1 Performance comparison of prediction models with different amount of support vectors.

Number of support vectors	VAF	σ	Average calculation time/s
863	99.8722	6	0.0999
716	99.8729	6	0.0847
552	99.8721	6	0.0828
425	99.8675	6	0.0826

The prediction model with 425 support vectors is used as the off-line model. Given the test input, Fig.3 compares the actual output of the controlled system with the prediction output of the off-line model. In Fig.3, 'Y' is the actual output of the controlled system and 'Y_m' is the prediction output of the off-line model.

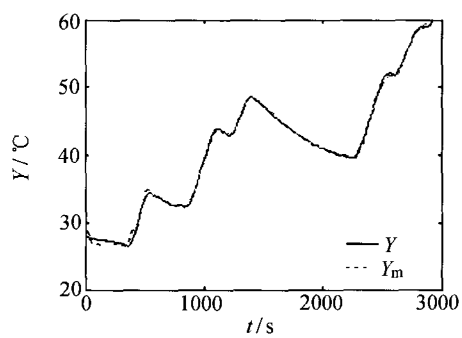

Fig. 3 Comparison of prediction and actual output.

The off-line model is linearized online and the GPC strategy is used to compute the control action at each sampling point. Let prediction horizon be $P = 10$, control horizon be $M = 4$, and make sure that the sampling time is $Ts = 20$s. We get the tracking curve of the closed-loop system in Fig.4, in which 'Y_r' is the reference trajectory and 'Y' is the output of practical system. In Fig.4, the steady state error is less than 0.3℃, the maximum overshoot is 2℃. From Fig.4, we know that the presented prediction algorithm based on LS-SVM can control the given complicated nonlinear temperature system quickly and stably.

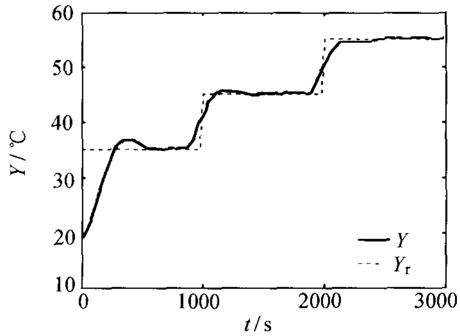

Fig. 4 Tracking curve of boiler temperature.

5 Conclusion

A prediction control algorithm based on LS-SVM is proposed in this paper. LS-SVM is a modeling approach based on the structural risk minimization principle. It can pay attention to both the expectation risk and the generalization performance. Its modeling process has analytic solution formula and less indeterminated parameters. The LS-SVM method is used to build the nonlinear off-line model for the controlled system, and then use the prunning procedure to obtain the LS-SVM model with sparse support vectors on condition that the approximation capacity of the model should not be spoiled. To avoid the nonlinear programming problem to be resolved in each sampling period, we linearize the off-line model online at each sampling point and employ GPC strategy to calculate the control variable. The results of the experiment show the effectiveness of the presented algorithm.

References

[1] S.J. Qin, T.A. Badgwell. A survey of industrial model prediction control technology [J]. *Control Engineering Practice*, 2003, 11(7): 733 – 764.

[2] J. B. Rawlings. Tutorial overview of model prediction control [J]. *IEEE Control Systems Magazines*, 2000, 20(3): 38 – 52.

[3] X. Zhang. Introduction to statistical learning theory and support vector machines [J]. *Acta Automatica Sinica*, 2000, 26(1): 32 – 42.

[4] V. Vapnik. *Statistical Learning Theory* [M]. New York: John Wiley, 1998.

[5] J. A. K. Suykens, J. Vandewalle. Least squares support vector machine classifiers [J]. *Neural Processing Letters*, 1999, 9(3): 293 – 300.

[6] J. A. K. Suykens, L. Lukas, J. Vandewalle. Sparse approximation using least squares support vector machine [C]// *IEEE Int. Symposium on Circuits and Systems*, Geneva, Switzerland, 2000: 757 – 760.

[7] A. J. Smola. *Regression estimation with support vector learning machines* [D]. Munich: Muniche University of Technology, 1996.

[8] D. W. Clarke, C. Mohtadi, P. S. Tuffs. Generalized prediction control-part I: the basic algorithm [J]. *Automatica*, 1987, 23(2): 137 – 148.

Bin LIU was born in 1972. She received a first class B. S. degree from Wuhan University of Science and Technology in 1999. Now she is studying in Zhejiang University as a Ph. D. candidate and is also an associate professor in the automation department of Wuhan University of Science and Technology. Her recent research interests include nonlinear modeling and prediction control. E-mail: bliu@iipc.zju.edu.cn.

Hongye SU was born in 1969. He received a first class B. S. degree from Nanjin University of Chemical Technology in 1990, and received M. S. degree and Ph. D. degree from Zhejiang University in 1993 and 1995, respectively. Now he is a professor in Institute of Advanced Process Control, Zhejiang University. His recent research interests include robust control, time-delay systems, nonlinear systems, DEDS, advanced process control theory and application. E-mail: hysu@iipc.zju.edu.cn.

Weihua HUANG was born in 1976. She received the B. S. degree from Wuhan University of Science and Technology in 1999. Now she is a lectuer in the automation department of Wuhan University of Science and Technology. Her recent research interests include fuzzy control and process control.

Jian CHU was born in 1963. He received the B. E. degree in chemical engineering from Zhejiang University in 1982 and Ph. D. degree from Kyoto University in 1989. Since 1993, he was employed as professor in Industrial Process Control. Now he is associate director of National Engineering Research Center of Industrial Automation, and director of National Lab. of Industrial Control Technology, and he is also one of the professors specially employed as "Changjiang Scholor". His interests are time-delay system, nonlinear control, robust control and application.