2005 American Control Conference
June 8-10, 2005. Portland, OR, USA

Adaptive Inverse Disturbance Canceling Control System Based on Least Square Support Vector Machines

Xiaojing Liu, Jianqiang Yi, Dongbin Zhao
(Laboratory of Complex Systems and Intelligence Science, Institute of Automation
Chinese Academy of Sciences, Beijing 10008)

Abstract-- **Adaptive inverse disturbance canceling control uses some adaptive filters. The neural network methods of training these filters have been fully researched. However, the problems of local minimum, curse of dimensionality and overfitting limit the application of neural networks. Comparatively, Support Vector Machines effectively overcome these limitations. A kind of adaptive inverse disturbance canceling control system based on least squares support vector machines (LS-SVM) is proposed. The approach of modeling and inverse modeling using LS-SVM is presented. A parameter selecting method within the Bayesian evidence framework is given for SVM regression with Gaussian kernel. Simulation results show that the approach is effective.**

Index Terms-- **support vector machine; Bayesian; evidence framework; adaptive inverse; disturbance canceling control**

I. INTRODUCTION

In control theory, it is most common to control plant response and plant disturbance in one process. With adaptive inverse control, however, it is convenient to treat these problems independently. In this way, the dynamic control process is not compromised by the need to reduce plant disturbance. Furthermore, the plant disturbance reduction process is not compromised by the needs of dynamic control. B Widrow and E Walach[1] introduced the functioning of the adaptive plant disturbance canceler and gave the proof of optimality. Disturbance can be canceled even when the plant is nonlinear[2].

Support Vector Machine (SVM) is a new powerful machine learning method developed on Statistical Learning Theory [3]and has many successful applications in function regression, pattern recognition, signal processing[4][5]. Least squares support vector machines (LS-SVM) is an extension of SVM and has good performance in function estimation and regression[6]. However, to obtain a high level of performance, some parameters in SVM still have to be tuned. These include regularization parameter C and kernel parameter σ . These parameters are sometimes just hand-picked by the user. A more disciplined approach is to use a validation set[7], or by data-resampling techniques such as cross-validation and bootstrapping. Mackay[8] proposed a kind of Bayesian evidence framework. Kwok[9] apply the evidence framework to SVM in order to solve classification problems.

In this paper, LS-SVM is used to make the plant model and inverse plant model, and the plant output disturbance is used to drive the inverse plant model to generate filter disturbance for subtraction from the plant input. The ultimate effect is to cancel disturbance at the plant output. A parameter selecting method within the Bayesian evidence framework for SVM regression with Gauss kernel is presented. The approach of modeling and inverse modeling using LS-SVM is discussed. A kind of adaptive inverse disturbance canceling control system based on LS-SVM is proposed. Simulation results are given in the end.

II. LEAST SQUARES SUPPORT VECTOR MACHINES (LS-SVM)

The basic idea of SVM is firstly to map the input data into a high dimensional feature space via a nonlinear map. In the high dimensional feature space, linear decision function is constructed. Then SVM nonlinearly maps inner product of feature space to the original space via a kernel. LS-SVM function regression algorithm is as follow. Consider a training set containing N data point $\{x_k, y_k\}$, $k=1,\cdots,N$, with input $x_k \in R^n$ and output $y_k \in R$, the following regression model is used.

$$y(x) = w \cdot \varphi(x) + b \,. \qquad (1)$$

where $\varphi(x)$ maps the input data into a higher dimensional feature space, w is the coefficient, b is the bias. In LS-SVM for function estimation, the object function is defined as

Manuscript received September 14, 2004. This work was partly supported by NSFC Project (Grant No. 60475030), and International Cooperation Key Project (Grant No. 2004DFB02100) of Ministry of Science and Technology, China. X.J.Liu is with lab of Complex Systems and Intelligence Science, Institute of Automation, Chinese Academy of Sciences, P.O.Box 2728, Beijing 100080,China. (corresponding author to provide phone: 86-10-82615422; e-mail: xiaojing.liu@mail.ia.ac.cn).

Jianqiang Yi and Dongbin Zhao are with lab of Complex Systems and Intelligence Science, Institute of Automation, Chinese Academy of Sciences.

follows:

$$\min J(w,e)=\frac{1}{2}w^{\mathrm{T}}w+\frac{1}{2}C\sum_{k=1}^{N}e_k^2 \qquad (2)$$

subject to the equality constraints

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

where C is regularization parameter and e_k represents the error at the instant k. The corresponding Lagrangian is given by

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

where $\alpha_k\,(k=1,\cdots N)$ are Lagrange multipliers. Considering the optimization conditions

$$\frac{\partial L}{\partial w}=0,\ \frac{\partial L}{\partial b}=0,\ \frac{\partial L}{\partial e_k}=0,\ \frac{\partial L}{\partial \alpha_k}=0 \qquad (5)$$

Optimization problem can be rewritten as

$$\begin{bmatrix} 0 & 1_v \\ 1_v^{\mathrm{T}} & \Omega+\frac{1}{c}I \end{bmatrix}\begin{bmatrix} b \\ \alpha^{\mathrm{T}} \end{bmatrix}=\begin{bmatrix} 0 \\ y^{\mathrm{T}} \end{bmatrix}, \qquad (6)$$

where $y=[y_1,\cdots,y_N]$, $1_v=[1,\cdots,1]$, $\alpha=[\alpha_1,\cdots,\alpha_N]$. From application of the Mercer condition, one obtains

$$\Omega_{kl}=K(x_k,x_l)=\varphi(x_k)^{\mathrm{T}}\varphi(x_l) \qquad (7)$$

The resulting LS-SVM model for function estimation becomes

$$y(x)=\sum_{i=1}^{N}\alpha_i K(x,x_i)+b, \qquad (8)$$

where α_i and b are the solutions to equation (6). The kernel function $K(x,x_i)=\exp\{-|x-x_i|^2/2\sigma^2\}$ is Gaussian kernel in this paper.

III. Parameter Selecting Method within the Bayesian Evidence Framework

Regularization parameter C and kernel parameter σ are very important parameters for SVM. The parameter selecting method for SVM within the Bayesian evidence framework is discussed in literature [10]. The evidence framework is divided into three levels of inference[8]. The first level of inference infers α_i and b. The second level of inference infers regularization parameter C. The third level of inference infers kernel parameter σ.

A. The First Level Inference

Let $\lambda=1/C$. Assuming that D is data space and H is model space, the first level of inference infers the posterior distribution of w by the Bayes rule:

$$p(w/D,\lambda,H)\propto p(D/w,\lambda,H)p(w/\lambda,H) \qquad (9)$$

Assuming that the training data is independently identically distributed and $p(w/\lambda,H)$ follows Gaussian distribution, then

$$p(D/w,\lambda,H)=\prod_{i=1}^{N}p(y_i/x_i,w,\lambda,H)p(x_i/w,\lambda,H) \qquad (10)$$

$$p(w/\lambda,H)=\left(\frac{\lambda}{2\pi}\right)^{\frac{k}{2}}\exp\left(-\frac{\lambda}{2}w^{\mathrm{T}}w\right) \qquad (11)$$

where $p(x_i/w,\lambda,H)$ can be considered as a constant. Suppose $L(y_i,f(x_i))$ is loss function, then

$$p(y_i/x_i,w,\lambda,H)\propto\exp(-L(y_i,f(x_i))) \qquad (12)$$

Substituting (10)~(12) into (9), we get

$$p(w/D,\lambda,H)\propto\exp\left(-\frac{\lambda}{2}w^{\mathrm{T}}w-\sum_{i=1}^{N}L(y_i,f(x_i))\right) \qquad (13)$$

From (13), we known that minimize formulate (1) is just the same as finding the maximum of the posteriori estimate $p(w/D,\lambda,H)$ of w.

B. The Second Level Inference

The second level inference determines the value of λ by maximizing a posteriori estimate $p(\lambda/D,H)$ of λ.

$$p(\lambda/D,H)\propto p(D/\lambda,H)p(\lambda/H)\propto p(D/\lambda,H)\propto$$
$$\int p(D/w,\lambda,H)p(w/\lambda,H)dw\propto$$
$$(\frac{\lambda}{2\pi})^{\frac{k}{2}}\int\exp(-\frac{\lambda}{2}w^{\mathrm{T}}w-\sum_{i=1}^{N}L(y_i,f(x_i)))dw \qquad (14)$$

Let $E_w=\frac{1}{2}w^{\mathrm{T}}w$, $E_d=\sum_{i=1}^{N}L(y_i,f(x_i))$, then

$$p(\lambda/D,H)\propto$$
$$\lambda^{\frac{k}{2}}\int\exp(-\lambda E_W^{MP}-E_D^{MP}-\frac{1}{2}(w-w_{MP})A(w-w_{MP}))dw$$

$$=\lambda^{\frac{k}{2}}\exp(-\lambda E_W^{MP}-E_D^{MP})(2\pi)^{\frac{k}{2}}\det^{-\frac{1}{2}}A \qquad (15)$$

where w_{MP} is an optimal value of w, and

$$A=\frac{\partial^2(\lambda E_w+E_d)}{\partial w}=\nabla^2(\lambda E_w+\sum_{i=1}^{N}L(y_i,f(x_i))).$$

Taking the log function on both sides of (15), we get

$$\ln p(\lambda/D,H)=$$
$$-\lambda E_W^{MP}-E_D^{MP}+\frac{k}{2}\ln\lambda-\frac{1}{2}\ln(\det A)+constant \qquad (16)$$

Maximizing $\ln p(\lambda/D,H)$ is the same as maximizing λ. Then λ_{MP} can be obtained

$$2\lambda_{MP}E_W^{MP}=\gamma \qquad (17)$$

where $\gamma=k-\lambda traceA^{-1}$ is called the effective number of parameters[8].

In LS-SVM function regression algorithm, loss function is

$$L(y_i,f(x_i))=e_i^2=(y_i-w\varphi(x_i)-b)^2 \qquad (18)$$

$$A = \nabla^2(\lambda E_w + \sum_{i=1}^{N} L(y_i, f(x_i))) = \lambda I + B \tag{19}$$

$$B = \sum_{i=1}^{N} 2\varphi(x_i)\varphi^{\mathrm{T}}(x_i) \tag{20}$$

Denoting the eigenvalues of B by ρ_l, we get the effective number of parameters γ [9]

$$\gamma = k - \lambda trace A^{-1} = \sum_{i=1}^{L} \frac{\rho_i}{\lambda + \rho_i} \tag{21}$$

where $L(L \le N)$ denote the number of nonzero eigenvalues.

C. The Third Level Inference

The third level of inference determines the optimal kernel parameter σ by maximizing the posteriori estimate $p(H/D) \propto p(D/H)p(H)$. Assuming a flat prior $p(H)$, we get

$$p(H/D) \propto p(D/H) \propto \int p(D/\lambda, H)p(\lambda/H)d\lambda \propto p(D/\lambda_{MP}, H)/\sqrt{\gamma} \tag{21}$$

$$\ln p(H/D) = -\lambda_{MP}E_W^{MP} - E_D^{MP} + \frac{k}{2}\ln\lambda_{MP} - \frac{1}{2}\ln(\det A) - \frac{1}{2}\ln(k - \lambda_{MP} trace A^{-1}) + constant \tag{22}$$

By maximizing $\ln p(H/D)$, we get the optimal kernel parameter σ.

$$\partial \ln p(H/D)/\partial\sigma = 0\,. \tag{23}$$

For LS-SVM

$$\frac{\partial(\lambda_{MP}E_W^{MP})}{\partial\sigma} = -\lambda_{MP}a_i a_j \frac{\partial K}{\partial\sigma} = -\lambda_{MP}\sum_{i,j=1}^{N} a_i a_j \cdot \exp(-\frac{(x_i - x_j)^2}{2\sigma^2})(x_i - x_j)^2\sigma^{-3} \tag{24}$$

$$\frac{\partial \ln(\det A)}{\partial\sigma} = trace(A^{-1}(\frac{\partial A}{\partial\sigma})) = trace(A^{-1}(\frac{\partial \overline{K}}{\partial\sigma})) \tag{25}$$

$$\frac{\partial \ln(k - \lambda_{MP} trace A^{-1})}{\partial\sigma} = \frac{\lambda_{MP}}{k - \lambda_{MP} trace A^{-1}} trace(A^{-2}(\frac{\partial \overline{K}}{\partial\sigma})) \tag{26}$$

Substituting (22), (24)~(26) into (23), we get

$$\sigma = \left[\frac{2\lambda_{MP}\sum_{i,j=1}^{N} a_i a_j \times \exp(-\frac{(x_i - x_j)^2}{2\sigma^2})(x_i - x_j)^2}{\frac{\lambda_{MP}}{k - \lambda_{MP} trace A^{-1}} trace(A^{-2}(\frac{\partial K}{\partial\sigma})) + trace(A^{-1}(\frac{\partial K}{\partial\sigma}))}\right]^{\frac{1}{3}} \tag{27}$$

IV. Approach of Plant Modeling and Plant Inverse Modeling based on LS-SVM

A plant model and a plant inverse model are needed in adaptive inverse control approach. The plant model has the same transfer function as the plant, while the plant inverse model has a transfer function which is the reciprocal of that of the plant.

Plant modeling based on LS-SVM is illustrated in fig.1. A discrete single input single output nonlinear model

$$y(k+1) = f[y(k), y(k-1), \cdots y(k-n); u(k), u(k-1), \cdots, u(k-m)] \tag{28}$$

where $u(k)$ and $y(k)$ denote the input and output of system at the instant k. Given a series of input signals $u(k-m)$, $u(k-m+1)\cdots, u(k)$ and output signals $y(k-n)$, $y(k-n+1), \cdots, y(k)$, the corresponding output is $y(k+1)$. Let

$$X(i) = (y(i), y(i-1), \cdots, y(i-n), u(i), u(i-1), \cdots, u(i-m)) \quad (i = 1, 2, \cdots, N) \tag{29}$$

then

$$y(i+1) = f(X(i)) \tag{30}$$

Constructing a data set $(X(i), y(i+1))$ and using LS-SVM regression algorithm, we get the estimator $\hat{y}(k+1)$ of the plant output $y(k+1)$.

$$\hat{y}(k+1) = \sum_{i=1}^{N} \alpha_i K(X(i), X(k)) + b \tag{31}$$

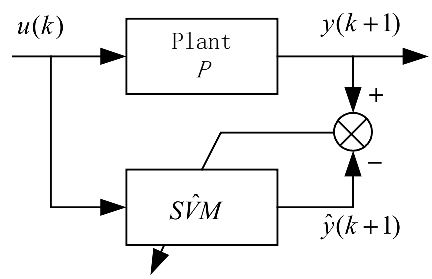

Fig. 1. Plant modeling based on LS-SVM

Plant inverse modeling based on LS-SVM is illustrated in fig.2. The idea is the same as that of plant modeling. Synthetic noise $n(k)$ whose statistical character would be the same as that of the original plant disturbance. Let

$$Y(i) = (\hat{y}(i), \hat{y}(i-1)\cdots\hat{y}(i-n), n(i-1), n(i-2), \cdots n(i-m)) \tag{32}$$

then

$$n(i) = f(Y(i))\,. \tag{33}$$

So we can construct a data set $(Y(i), n(i))$. With LS-SVM regression algorithm, we get the estimator $\hat{n}(k)$ of disturbance $n(k)$.

$$\hat{n}(k) = \sum_{i=1}^{N} \alpha_i^* K(Y(i), Y(k)) + b^* \tag{34}$$

Through learning the samples, we can obtain α_i, b in (31) and α_i^*, b^* in (34) respectively.

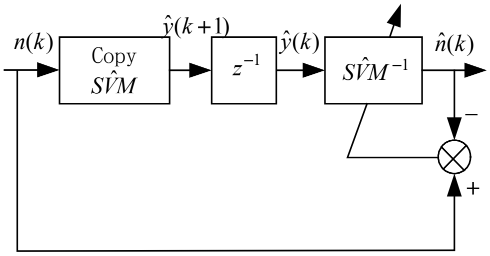

Fig.2 Plant inverse modeling based on LS-SVM

The steps of plant modeling and plant inverse modeling based on LS-SVM within the Bayesian evidence framework are as follow:

Step 1: setting up the data set about input and output;

Step 2: setting original parameter value and training the LS-SVM, obtain the coefficients a_i and b;

Step 3: getting regularization parameter C by level 2;

Step 4: getting kernel parameter σ by level 3;

Step 5: training the LS-SVM again with the new parameter C and σ

Step 6: going back to Step 2 repeatedly until getting the optimal plant model and plant inverse model.

V. Adaptive Inverse Disturbance Canceling Control System

Adaptive inverse disturbance control system based on LS-SVM is diagrammed in fig.3. Disturbance cancelation is accomplished by this system in the following manner. A copy of $S\hat{V}M$, a very close, disturbance-free match to $P(z)$, is fed the same input as the plant $P(z)$. The difference between the disturbed output of the plant and the disturbance-free output of $S\hat{V}M$ is a very close approximation to the plant output disturbance $n(k)$. The approximate $n(k)$ is then input to the filter $S\hat{V}M^{-1}$. The output of $S\hat{V}M^{-1}$ is subtracted from the plant input to effect cancellation of the plant disturbance. Unit delay Δ^{-1} is placed in front of $S\hat{V}M^{-1}$ because we consider that digital feedback links must have at least one unit of delay around each loop. Thus, the current value of the plant disturbance $n(k)$ can be used only for the cancelation of future values of plant disturbance and cannot be used for instantaneous self-cancelation. The effects of these unit delays are small when the system is operated with a high sampling rate, however.

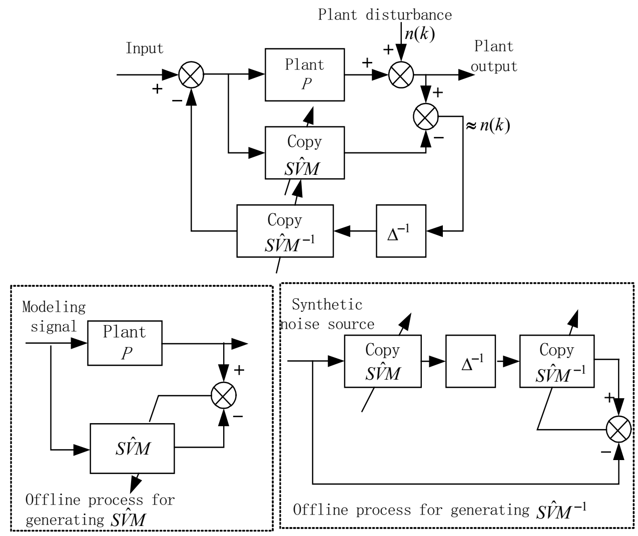

Fig.3 Adaptive inverse disturbance canceling control system based on LS-SVM

VI. Simulation

Given a nonlinear plant

$$y(k+1) = y(k)/(1+y^2(k)) + u^3(k) + n(k) \quad (35)$$

The plant disturbance $n(k)$ is a band-limited white noise signal whose noise power is 0.1. The plant input $u(k)$ is impulse signal whose amplitude is 1. Select 200 sample points to construct a data set. We get the optimal regularization parameter C=270 and kernel parameter σ =70.748 within the Bayesian evidence framework. The simulation result of offline modeling is shown in fig.4. Assuming that the synthetic noise signal for inverse modeling is the same as the plant disturbance $n(k)$, we also get the optimal regularization parameter C=40 and kernel parameter σ =0.211. The simulation result of offline inverse modeling is shown in fig.5. The output of plant without adaptive inverse disturbance control is shown in fig.6. The output of plant with adaptive inverse disturbance control is shown in fig.7. The simulation results show that the adaptive inverse disturbance canceling system based on LS-SVM has a better effect of canceling disturbance.

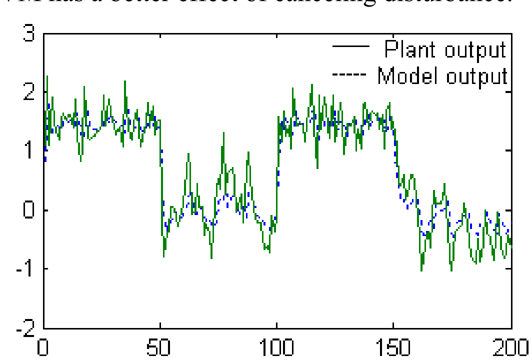

Fig.4 Offline modeling based on LS-SVM

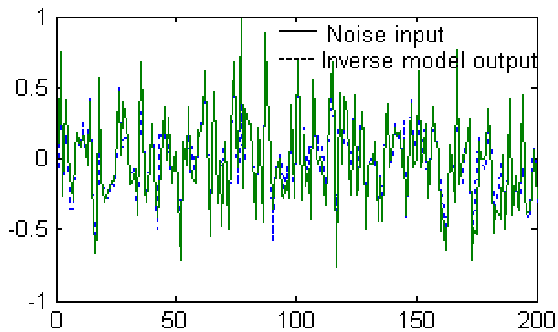

Fig.5 Offline inverse modeling based on LS-SVM

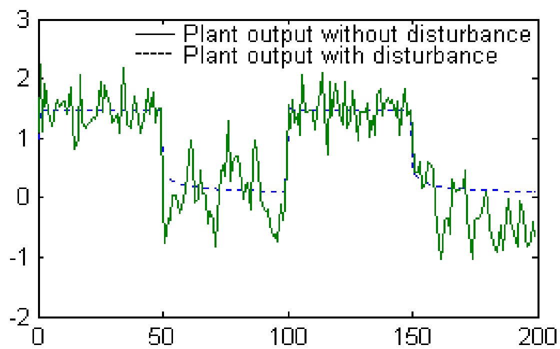

Fig.6 Plant output without adaptive inverse disturbance canceling control

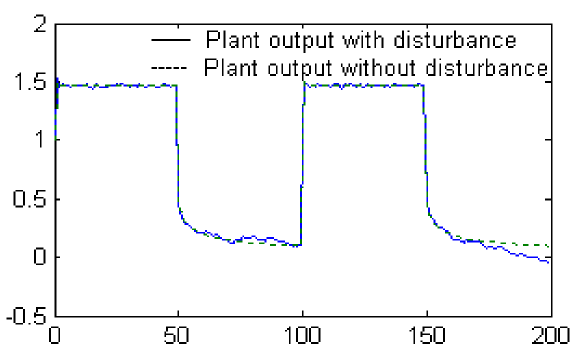

Fig.7 Plant output of the adaptive inverse disturbance canceling control system

VII. Conclusion

In this paper, an approach of modeling and inverse modeling based on LS-SVM is presented. Then, a parameter selecting method within the Bayesian evidence framework is given for SVM regression with Gauss kernel. Finally, a kind of adaptive inverse disturbance canceling control system based on LS-SVM is proposed. Simulation result shows that the approach is effective.

References

[1] B Widrow, E Walach. *Adaptive Inverse Control*. Prentice Hall, 1996.

[2] Lu Zhigang, Wu Shichang and Lu Linghui. *Nonlinear adaptive inverse control and its applicaition*. Publishing company of defence industry. 2004.

[3] V.N. Vapnik, *The nature of statistical learning learning theory*, Springer-Verlag, New York, 1995.

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

[5] J.A.K. Suykens and and J. Vandewalle, "Leasut squares support vector machines classifiers", Neural Network Letters, 1999, Vol.9, No.3, pp: 293-300.

[6] J.A.K. Suykens, "Nonlinear modeling and support vector machines", in Proc. of the 18th IEEE Conference on Instrumentation and Measurement Technology, 2001, pp: 287-294.

[7] K.R. Muller, A.J. Smola, G. Ratsch, B. Scholkopf, J. Kohlmorgen and V. Vapnik, "Predicting time series with support vector machines", in Proc. Int. Conf.Artificial Neural Networks, 1997, pp.999-1004.

[8] D.C. Mackay, "Probable network and plausible predictions-A review of practical Bayesian methods for supervised neural networks", Network Computation in Neural Systems, 1995,6: 469-505.

[9] J.T. Kowk, "The evidence framework applied to support vector machines", IEEE Trans on Neural Network, 2000, Vol.10, No.5, pp:1162-1173.

[10] W.W. Yan, J.L. Chang, H.H. Shao, "Modeling method based on support vector machines within the Bayesian evidence framework ", Control and Decision, 2004,Vol.19, No.5, pp: 525-528.