Proceedings of the Third International Conference on Machine Learning and Cybernetics, Shanghai, 26-29 August 2004

NONLINEAR DYNAMIC SYSTEM IDENTIFICATION USING LEAST SQUARES SUPPORT VECTOR MACHINE REGRESSION

XIAO-DONG WANG[1], MEI-YING YE[2]

[1]College of Information Science and Engineering, Zhejiang Normal University, Jinhua 321004, China
[2]College of Mathematics and Physics, Zhejiang Normal University, Jinhua 321004, China
E-MAIL: wxd@mail.zjnu.net.cn, ymy@mail.zjnu.net.cn

Abstract:

The least squares support vector machine (LS-SVM) regression is presented for the purpose of nonlinear dynamic system identification. The LS-SVM achieves higher generalization performance than the multilayer perceptron (MLP) and radial basis function (RBF) neural networks and no number of hidden units has to be defined. Another key property is that unlike MLP' training that requires nonlinear optimization with the danger of getting stuck into local minima. A difference with the RBF neural networks is that no center parameter vectors of the Gaussians have to be specified. The identification procedure is illustrated using simulated examples. The results indicate that this approach is effective even in the case of additive noise to the system. The LS-SVM can be used as an important alternative to MLP and RBF neural networks in nonlinear dynamic system identification.

Keywords:

Nonlinear systems; system identification; least squares support vector machine

1. Introduction

Controlling nonlinear system has been an active area in recent years. For affine nonlinear systems, when the plant nonlinearities are known, many results have been obtained using feedback linearization methods. However, a common assumption made is that either all or parts of the system dynamics are known. The control problem of a completely unknown plant cannot be solved under such assumption. A common approach to deal with this problem is to utilize the identification technique in the control scheme. The process of identification includes selecting a proper model and adjusting its parameters so that the output of the model approaches to the output of the actual plant under the same input. The nonlinear system identification process has turned out to be one of central parts in constructing tracking controllers.

Recently artificial neural networks (ANNs) have emerged as a powerful learning technique to perform complex tasks in highly nonlinear dynamic environments. Some of the prime advantages of using ANN models are: their ability to learn based on optimization of an appropriate error function and their excellent performance for approximation of nonlinear functions. Until now, most of the ANN-based system identification techniques are based on multilayer feedforward networks such as multilayer perceptron (MLP) trained with backpropagation or more efficient variation of this algorithm [1]–[4]. This is due to the fact that these networks are robust and effective in identification and control of complex dynamic plants. Narendra and Parthasarathy [1] have proposed effective identification and control of dynamic systems using MLP networks. Although the ANNs are developed in system identification, some inherent drawbacks, e.g., the multiple local minima problem, the choice of the number of hidden units and the danger of over fitting, etc., would make it difficult to put the networks into practice. As an alternative to the MLP, there has been considerable interest in radial basis function (RBF) neural networks [5]–[8], primarily because of its simpler structure. The RBF neural networks can learn functions with local variations and discontinuities effectively and also possess universal approximation capability [8]. This network represents a function of interest by using members of a family of compactly or locally supported basis functions, among which radially-symmetric Gaussian functions are found to be quite popular. A RBF neural network has been proposed for effective identification of nonlinear dynamic systems [9,10]. In these networks, however, choosing an appropriate set of RBF centers and the number of hidden units for effective learning still remains as a problem.

The present study focuses on the problem of nonlinear system identification using least squares support vector machine (LS-SVM) regression [11,12]. The LS-SVM is established based on the structural risk minimization principle rather than minimize the empirical error commonly implemented in the neural networks, LS-SVM achieves higher generalization performance than the MLP and RBF neural networks in solving these machine learning problems. Another key property is that unlike MLP' training that requires non-linear optimization with the danger of getting stuck into local minima, training LS-SVM is equivalent to solving a set of linear equations. Consequently, the solution of LS-SVM is always unique and globally optimal. A difference with the RBF neural networks is that no center parameter vectors of the Gaussians have to be specified and no number of hidden units has to be defined because of Mercer's condition.

2. Problem description

A wide class of nonlinear systems with an input u and an output y can be described in discrete time by the NARX (nonlinear autoregressive with exogenous input) input–output model:

$$y(k+1) = f(\mathbf{x}(k)), \tag{1}$$

where $f(\cdot)$ is some nonlinear function, $y(k+1)$ denotes the output predicted at the future time instant $k+1$ and $\mathbf{x}(k)$ is the regressor vector, consisting of a finite number of past inputs and outputs:

$$\mathbf{x}(k) = \begin{bmatrix} y(k) \\ \vdots \\ y(k-n_y+1) \\ u(k) \\ \vdots \\ u(k-n_u+1) \end{bmatrix}. \tag{2}$$

The dynamic order of the system is represented by the number of lags n_u and n_y.

The task of system identification is essentially to find suitable mappings, which can approximate the mappings implied in a dynamic system. The aim of identification is only to obtain an accurate identifier for y. In this work, we use LS-SVM regression mentioned above for identification of nonlinear dynamic systems.

3. Least squares support vector machine regression for identification

In the following, we briefly introduce LS-SVM regression, which can be used for nonlinear system identification.

Consider a given training set of N data points $\{x_k, y_k\}_{k=1}^{N}$ with input data $x_k \in R^n$ and output $y_k \in R$. In feature space SVM models take the form:

$$y(x) = w^T \varphi(x) + b, \tag{3}$$

where the nonlinear mapping $\varphi(\cdot)$ maps the input data into a higher dimensional feature space. Note that the dimensional of w is not specified (it can be infinite dimensional). In LS-SVM for function estimation the following optimization problem is formulated

$$\min \ J(w,e) = \frac{1}{2} w^T w + \gamma \frac{1}{2} \sum_{k=1}^{N} e_k^2 \tag{4}$$

subject to the equality constrains

$$y(x) = w^T \varphi(x_k) + b + e_k, \quad k = 1, \cdots, N. \tag{5}$$

Important differences with standard SVM [13,14] are the equality constrains and the squared error term, which greatly simplifies the problem.

The solution is obtained after constructing the Lagrangian

$$L(w,b,e,\alpha) = J(w,e) - \sum_{k=1}^{N} \alpha_k \{ w^T \varphi(x_k) + b + e_k - y_k \} \tag{6}$$

with Lagrange multipliers α_k. The conditions for optimality are

$$\begin{cases} \dfrac{\partial L}{\partial w} = 0 \ \rightarrow \ w = \sum_{k=1}^{N} \alpha_k \varphi(x_k) \\ \dfrac{\partial L}{\partial b} = 0 \ \rightarrow \ \sum_{k=1}^{N} \alpha_k = 0 \\ \dfrac{\partial L}{\partial e_k} = 0 \ \rightarrow \ \alpha_k = \gamma e_k \\ \dfrac{\partial L}{\partial \alpha_k} = 0 \ \rightarrow \ w^T \varphi(x_k) + b + e_k - y_k = 0 \end{cases} \tag{7}$$

for $k = 1, \cdots, N$. After elimination of e_k, w the solution is given by the following set of linear equations

$$\begin{bmatrix} 0 & \vec{1}^T \\ \vec{1} & \varphi(x_k)^T \varphi(x_l) + \gamma^{-1} I \end{bmatrix} \begin{bmatrix} b \\ \alpha \end{bmatrix} = \begin{bmatrix} 0 \\ y \end{bmatrix}, \tag{8}$$

where $y = [y_1; \cdots; y_N]$, $\vec{1} = [1; \cdots; 1]$, $\alpha = [\alpha_1; \cdots; \alpha_N]$ and the Mercer's condition

$$\Psi(x_k, x_l) = \varphi(x_k)^T \varphi(x_l), \quad k, l = 1, \cdots, N \tag{9}$$

has been applied. This finally results into the following LS-SVM model for function estimation

$$y(x) = \sum_{k=1}^{N} \alpha_k \Psi(x, x_k) + b, \tag{10}$$

where α_k, b are the solution to the linear system, $\Psi(\cdot,\cdot)$ represents the high dimensional feature spaces that is nonlinearly mapped from the input space x. The LS-SVM approximates the function using the Eq.(10).

The choice of the kernel function $\Psi(\cdot,\cdot)$ has several possibilities. In this work, the radial basis function (RBF) function is used as the kernel function of the LS-SVM because RBF kernels tend to give good performance under general smoothness assumptions. Consequently, they are especially useful if no additional knowledge of the data is available. For RBF kernels one has

$$\Psi(x_k, x_l) = \exp(-\|x - x_k\|_2^2 / \sigma^2), \qquad (11)$$

where σ is a positive real constant. Note that in the case of RBF kernels, one has only two additional tuning parameters σ in Eq.(11) and γ in Eq.(4), which is less than for standard SVM.

The method for performing nonlinear system identification is to induce the function $f(\cdot)$ in Eq.(1) using LS-SVM regression, which is trained with a set of finite number of past inputs and outputs as the target value of the LS-SVM regression. This method is often called the sliding window technique as the inputs and outputs slide over the full training set.

4. Simulation studies

In this section, the two simulated examples are used as evaluation of the identification power of LS-SVM.

The results are presented in terms of the accuracy of the identification using the root-mean-square error (RMSE) metric. The RMSE of the validation set is calculated as follows:

$$\mathrm{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \dot{y}_i)^2}\ , \qquad (12)$$

where n denotes the total number of data points in the validation set. y represents the output of the original system and $\dot{y}$ represents the output of the LS-SVM

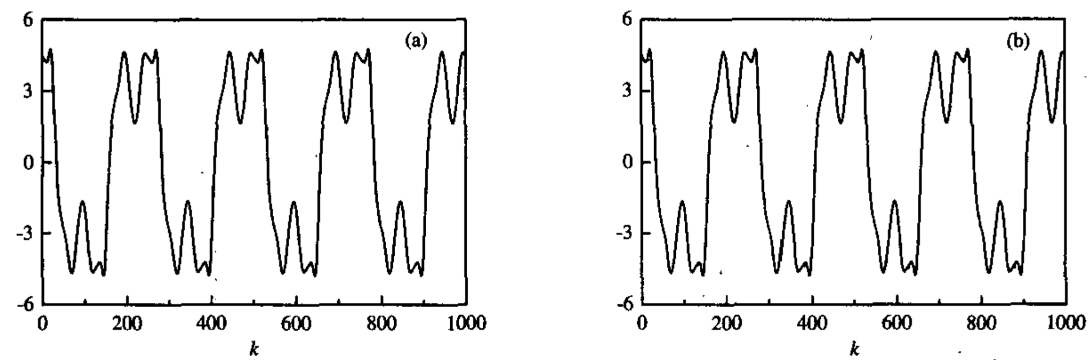

Figure 1. The plant output (solid) and the LS-SVM identifier output (dashed) of example 1 for (a) noise-free and (b) 0.2 noise level.

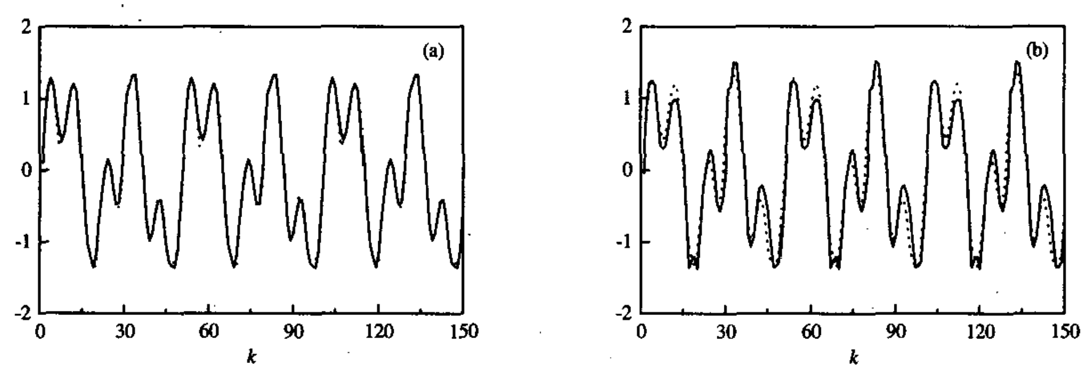

Figure 2. The plant output (solid) and the LS-SVM identifier output (dashed) of example 2 for (a) noise-free and (b) 0.2 noise level.

model.

In real system identification problems, data are usually corrupted by noise. Uncertainty can arise from measurement instruments, system noise, or unmodelled dynamics. In practice, we should take into account the effect of noise. Hence, we have shown that this method works for various noise levels in both examples. To do this we needed some way of measuring the level of the additive noise signal. The noisy training data are derived by adding the normal noise to the noise-free data. We divide the noisy training data, which we call y', in two terms:

$$y' = y + l_n \sigma_s r_n . \quad (13)$$

Here, the first term y is the original noise-free data, the second term $l_n \sigma_s r_n$ is additive noise, where l_n is defined as the noise level which is the noise standard deviation σ_n divided by the standard deviation σ_s of the noise-free data, r_n is chosen from the normal distribution numbers with mean zero, variance one and standard deviation one. We used noise levels of 0.1 and 0.2 to investigate the quality of proposed identification method in comparison with the results of the noise-free identification mentioned above.

Example 1. The plant of the nonlinear system is given by the following difference equation:

$$\begin{aligned} x(k) &= 0.3y(k-1) + 0.6y(k-2) \\ &\quad + 0.6\sin(\pi u(k-1)) + 0.3\sin(3\pi u(k-1)) \quad (14) \\ &\quad + 0.1\sin(5\pi u(k-1)) \end{aligned}$$

where the output at time t is a linear function of past output at times $k-1$ and $k-2$ plus a nonlinear function of the input at time $k-1$. The reference input $u(k-1)$ to the system is selected as

$$u(k-1) = \sin(2\pi(k-1)/250) . \quad (15)$$

Example 2. The plant of the nonlinear system is given by

$$y(k) = \frac{0.2y(k-1) + 0.6y(k-2)}{1 + y(k-1)^2} + \sin(u(k-1)) \quad (16)$$

The reference input is selected as

$$u(k-1) = \sin(2\pi(k-1)/10) + \sin(2\pi u(k-1)/25) \quad (17)$$

Figs. 1 and 2 depict the identification results using the proposed LS-SVM to examples 1 and 2, respectively. In the Figs. 1(b) and 2(b), the noise added to the systems is level of 0.2.

As can be seen from Figs. 1(a) and 2(a), which demonstrate that the identified and actual points of the system are largely indistinguishable if the training data are noise-free, almost perfect identification are achieved. In this situation, the RMSE is 0.0026 for example 1 and 0.062 for example 2.

From the Figs. 1(b) and 2(b), if the training data are corrupted with noise of 0.2 levels, it is verified that additive noise has an influence on the nonlinear system identification. In this situation, the RMSE is 0.100 for example 1 and 0.624 for example 2. It may be seen that the identification of the plant is satisfactory even when a training data are contaminated with additive noise.

5. Conclusions

In this paper, we investigate the problem of nonlinear system dynamic identification using the LS-SVM regression. We also consider the influence of noise on the identification. The results indicate that nonlinear system identification using LS-SVM regression is an effective approach. This approach is found to be superior to that of the neural networks for the complex task of nonlinear system identification even in the case of additive noise to the system. The LS-SVM can be used as an important alternative to MLP and RBF neural networks in nonlinear system identification.

Acknowledgements

The Project Supported by Zhejiang Provincial Natural Science Foundation of China (602145) and Key Construction Subject of Zhejiang Normal University.

References

[1] K. S. Narendra and K. Parthasarathy, "Identification and control of dynamical systems using neural networks," IEEE Trans. Neural Networks, vol. 1, pp. 4–26, Jan. 1990.

[2] D. H. Nguyen and B.Widrow, "Neural networks for self-learning control system," Int. J. Contr., vol. 54, no. 6, pp. 1439–1451, 1991.

[3] G. Cembrano, G. Wells, J. Sarda, and A. Ruggeri, "Dynamic control of a robot arm based on neural networks," Contr. Eng. Practice, vol. 5, no. 4, pp. 485–492, 1997.

[4] S. Lu and T. Basar, "Robust nonlinear system identification using neural network models," IEEE Trans. Neural Networks, vol. 9, pp. 407–429, May. 1998.

[5] T. Poggio and F. Girosi, "Networks for approximation and learning," Proc. IEEE, vol. 78, pp. 1481–1497, Sep. 1990.

[6] J. Moody and C. J. Darken, "Fast learning in networks of locally-tuned processing units," Neural Comput., vol. 1, pp. 281–294, 1989.
[7] J. Park and I. W. Sandberg, "Universal approximation using radial basis function networks," Neural Comput., vol. 3, pp. 246–257, 1991.
[8] E. J. Hartman, J. D. Keeler, and J. M. Kowalski, "Layered neural networks with Gaussian hidden units as universal approximation," Neural Comput., vol. 2, pp. 210–215, 1990.
[9] S. Chen, S. A. Billings, and P. M. Grant, "Recursive hybrid algorithm for nonlinear system identification using radial basis function networks," Int. J. Contr., vol. 55, no. 5, pp. 1051–1070, 1992.
[10] S. V. T. Elanayar and Y. C. Shin, "Radial basis function neural network for approximation and estimation of nonlinear stochastic dynamic systems," IEEE Trans. Neural Networks, vol. 5, pp. 594–603, July. 1994.
[11] J. A. K. Suykens and J. Vandewalle, "Least squares support vector machine classifiers," Neural Processing Letters, vol. 9, pp. 293-300, Jun. 1999.
[12] J. A. K. Suykens, "Least squares support vector machines for classification and nonlinear modeling," Neural Network World, vol. 10, pp. 29-48, Jan. 2000.
[13] V. N. Vapnik, The Nature of Statistical Learning Theory. New York: Springer, 2000.
[14] V. N. Vapnik, Statistical Learning Theory. New York: Wiley, 1998.