Support Vector Regression for Initialization of Radial Basis Function Networks for a Multi-input Multi-output System

Pi-Yun Chen[1], Chia-Ju Wu[2], Yu-Yi Fu[3], Chia-Nan Ko[4], and Jin-Tsong Jeng[5]
[1]Graduate School of Engineering Science and Technology, National Yunlin University of Science and Technology
Douliou, Yunlin 640, Taiwan
[2]Department of Electrical Engineering, National Yunlin University of Science and Technology
Douliou, Yunlin 640, Taiwan
[3,4]Department of Automation Engineering, Nan-Kai University of Technology
Tasotun, Nantou 542, Taiwan
[5]Department of Computer Science and Information Engineering, National Formosa University
Huwei, Yunlin 632, Taiwan
(E-mail: [1]g9310820@yuntech.edu.tw; [2] wucj@yuntech.edu.tw ; [3] t098@nkut.edu.tw ; [4] t105@nkut.edu.tw;
[5]tsong@nfu.edu.tw)

Abstract: When a radial basis function network (RBFN) is used for identification of a nonlinear multi-input multi-output (MIMO) system, the number of hidden layer nodes, the initial parameters of the kernel, and the initial weights of the network must be determined first. For this purpose, a systematic way that integrates the support vector regression (SVR) and the least squares regression (LSR) is proposed to construct the initial structure of the RBFN. First, determine the number of hidden layer nodes and the initial parameters of the kernel by the SVR method. Then the weights of the RBFN are determined by solving a minimization problem based on the concept of LSR. After initialization, an annealing robust learning algorithm (ARLA) is then applied to train the RBFN. With the proposed initialization approach, one can find that the designed RBFN has a fast convergent speed. To show the feasibility and superiority of the annealing robust radial basis function networks (ARRBFNs) for identification of MIMO system, one example is included.

Keywords: Radial basis function networks, Support vector regression, Multi-input multi-output, Least square regression

I. INTRODUCTION

In industry, multi-input multi-output (MIMO) system is applied widely. Structural identification and parameter estimation of nonlinear MIMO system are very important but rather difficult issues in system identification. Radial basis function networks (RBFNs) are widely used for system modeling recently since they have only one hidden layer and have fast convergence speed. RBFNs are often referred to as model-free estimators since they can be used to approximate functions without requiring a mathematical description of how the outputs functionally depend on the inputs. This means that they can build systems from input–output patterns directly, or they learn from examples without any knowledge of the model type.

When RBFNs are used for system identification, the number of hidden nodes, the initial parameters of the kernel, and the initial weights of the network must be determined first. In the past few years, several initializations for RBFNs have been proposed. First, the number of hidden nodes is fixed, and then all kinds of algorithms (such as genetic algorithms and gradient descent method) is used to optimize all the parameters, namely the initial parameters of the kernel, and the initial weights of the network. However, a systematic way to determine the initial structure of a RBFNs for a MIMO system is still not established. Therefore, based on support vector regression (SVR), a least square regression method is proposed to solve this problem in this paper. Given the sample set that describes the input-output relation of a function, the first step of the proposed method is to solve an SVR problem for one input-output of an MIMO system. Then the weights of RBFNs for another input-output can be obtained by using least square regression (LSR). From the SVR results, the initial structure of an RBFN can be determined. After initialization, annealing learning algorithms can then be applied to train the RBFNs. With the proposed initialization method, the MIMO system has fewer neurons and the designed RBFNs have fast convergence speed. To show the feasibility and superiority of the proposed method, simulation results are included for illustration.

II. RBFNS FOR IDENTIFICATION OF NONLINEAR MIMO SYSTEMS

In general, the input-output relation of a nonlinear MIMO system can be expressed as

$$\begin{aligned}\mathbf{y}(t+1) = \mathbf{f}(&\mathbf{y}(t), \mathbf{y}(t-1), \cdots, \mathbf{y}(t-n_y), \\ &\mathbf{x}(t), \mathbf{x}(t-1), \cdots, \mathbf{x}(t-n_u)),\end{aligned} \tag{1}$$

where $\mathbf{f}(t) = [f_1(t), f_2(t), \cdots, f_p(t)]^{\mathrm{T}}$ denotes the nonlinear relation to be estimated. $\mathbf{x}(t) = [x_1(t), x_2(t), \cdots, x_m(t)]^{\mathrm{T}}$ is the input vector, $\boldsymbol{y}(t) = [y_1(t), y_2(t), \cdots, y_p(t)]^{\mathrm{T}}$ is the output vector, n_u and n_y are the maximal lags in the input and output, respectively.

One can use a neural network to estimate the input-output relation of a nonlinear MIMO system. In this paper, an RBFN will be adopted since it has a simple structure. When the Gaussian function is chosen as the radial basis function, an RBFN can be expressed in the form

$$\hat{y}_j(t+1) = \sum_{i=1}^{L} G_i w_{ij} = \sum_{i=1}^{L} w_{ij} \exp\left(-\frac{\left\|\hat{\mathbf{x}} - \mathbf{m}_i\right\|^2}{2\sigma_i^2}\right), \quad \text{for } j = 1, \cdots, p \tag{2}$$

where $\hat{\mathbf{x}}(t) = [\hat{x}_1(t), \hat{x}_2(t), \cdots, \hat{x}_m(t)]^{\mathrm{T}}$ is the input vector, $\hat{\mathbf{y}}(t) = [\hat{y}_1(t), \hat{y}_2(t), \cdots, \hat{y}_p(t)]^{\mathrm{T}}$ is the output vector, w_{ij} is the synaptic weight, G_i is the Gaussian function, $\mathbf{m}_i$ and σ_i are the center and width of G_i, respectively, and L is the number of the Gaussian functions, which is also equal to the number of hidden layer nodes.

Given a set of training input-output pairs $(\mathbf{x}^{(k)}, \mathbf{y}^{(k)})$, $k = 1, 2, \cdots, N$, where the identification problem of the nonlinear MIMO system is to determine the values of L, w_{ij}, $\mathbf{m}_i$, and σ_i, to minimize the following performance index

$$J = \sum_{k=1}^{N} \left\| \mathbf{y}^{(k)} - \hat{\mathbf{y}}^{(k)} \right\|^2, \tag{3}$$

where $\hat{\mathbf{y}}^{(k)}$ is the corresponding output of the RBFN when the input $\hat{\mathbf{x}}$ to the network is equal to $\mathbf{x}^{(k)}$.

In usual cases, the initial values of L, w_{ij}, $\mathbf{m}_i$, and σ_i are chosen first. Then a training algorithm is applied to the RBFN to search for the optimal combination of these values in an iterative manner. However, as mentioned above, there is no way to choose the initial values of L, w_{ij}, $\mathbf{m}_i$, and σ_i systematically. Therefore, in the following section, an SVR approach will be proposed to serve for this purpose.

III. INITIAL STRUCTURE OF RBFNS

1. SVR-based method to determine L, $\mathbf{m}_i$, and σ_i

The proposed SVR-based method can approximate an unknown function. Without loss of generality, an output of the RBFN, say y_1, and its corresponding training pairs, $(\mathbf{x}^{(k)}, y_1^{(k)})$, $k = 1, 2, \cdots, N$, will be used for demonstration. Meanwhile, assume that a set of basis functions, $g_l(\mathbf{x})$, $l = 1, 2, \cdots, M$, is given. Then the problem of function approximation is transformed into finding the parameters of the following basis linear expansion

$$f(\mathbf{x}, \boldsymbol{\theta}) = \sum_{l=1}^{M} \theta_l g_l(\mathbf{x}) + b, \tag{4}$$

where $\boldsymbol{\theta} = (\theta_1, \theta_2, \cdots, \theta_M)$ is a parameter vector to be identified and b is a constant to be determined.

From Vapnik[1], one can find that the solution is to find $f(\mathbf{x}, \boldsymbol{\theta})$ that minimizes

$$R(\boldsymbol{\theta}) = \frac{1}{N} \sum_{i=1}^{N} L_\varepsilon(y_i - f(\mathbf{x}_i, \boldsymbol{\theta})), \tag{5}$$

subject to the constraint

$$\|\boldsymbol{\theta}\|^2 < C, \tag{6}$$

where $L_\varepsilon(\cdot)$ is the ε -insensitive loss function defined as

$$L_\varepsilon(e) = \begin{cases} 0 & \text{for } |e| \le \varepsilon \\ |e| - \varepsilon & \text{otherwise} \end{cases}, \ \varepsilon > 0. \tag{7}$$

By using the Lagrange multiplier method, it was shown that the minimization of (5) leads a dual optimization problem[1].

In this paper, since the Gaussian function is used as the kernel function, (12) can be rewritten as

$$f(\mathbf{x}, \lambda) = \sum_{l=1}^{\#SV} \lambda_l \exp\left(-\frac{\left\|\mathbf{x} - \mathbf{x}_l\right\|^2}{2\sigma_l^2}\right) + b, \tag{8}$$

where $\mathbf{x}_l$ denotes *support vectors* (SVs), $\#SV$ is the number of SVs. Comparing (8) with (2), $\#SV$, l, λ_l, and $\mathbf{x}_l$ in (8) can be regarded as the L, i, w_{i1}, and $\mathbf{m}_i$ in (2), respectively. From the above derivation, the number of hidden layer nodes L, the initial parameters w_{i1}, $\mathbf{m}_i$, and σ_i of the RBFNs can be determined.

2. LSR-based method to determine the synaptic weights

In the above section, the initial values of $w_{11}, w_{21}, \cdots, w_{L1}$ are determined. However, one still needs to determine the initial values of w_{ij}, $1 \le i \le L$, $2 \le j \le p$. Based on the concept of LSR[2], these values can be determined by solving the following problem:

Given L, $w_{11}, w_{21}, \cdots, w_{L1}$, $\mathbf{m}_i$, $1 \le i \le L$, and σ_i, $1 \le i \le L$, determine w_{ij}, $1 \le i \le L$, $2 \le j \le p$, to minimize the following performance index

$$J_{\mathrm{LSR}} = \sum_{k=1}^{N} \left\| \mathbf{y}^{(k)} - \hat{\mathbf{y}}^{(k)} \right\|^2 - \sum_{k=1}^{N} \left(y_1^{(k)} - \hat{y}_1^{(k)} \right)^2, \qquad (9)$$

At first glance, this problem is very similar to the one in (3). However, since the values of L, $w_{11}, w_{21}, \cdots, w_{L1}$, $\mathbf{m}_i$, $1 \le i \le L$, and σ_i, $1 \le i \le L$, are already given, this problem will be very easy to be solved based on the concept of LSR.

IV. ANNEALING ROBUST LEARNING ALGORITHM FOR RBFNS

In the training procedure of the proposed RBFN, the annealing concept[3] in the cost function of robust back-propagation learning algorithm[4] was adopted to overcome the existing problems in robust back-propagation learning algorithm. A cost function for the ARLA is defined here as

$$J_j(h) = \frac{1}{N} \sum_{k=1}^{N} \rho\left[e_j^{(k)}(h); \beta(h) \right], \text{ for } j = 1, 2, \cdots, p \qquad (10)$$

where

$$e_j^{(k)}(h) = y_j^{(k)} - \sum_{i=1}^{L} w_{ij} \exp\left(-\frac{\left\| \mathbf{x}^{(k)} - \mathbf{m}_i \right\|^2}{2\sigma_i^2}\right), \qquad (11)$$

h is the epoch number, $e_j^{(k)}(h)$ is the error between the kth desired output and the kth output of the ARRBFN at epoch h for the jth input-output training data in an MIMO system, $\beta(h)$ is a deterministic annealing schedule acting like the cut-off point, and $\rho(\cdot)$ is a logistic loss function defined as

$$\rho[e_j^{(k)}; \beta] = \frac{\beta}{2} \ln\left[1 + \frac{(e_j^{(k)})^2}{\beta} \right], \quad \text{for} \quad j = 1, 2, \cdots, p \qquad (12)$$

Based on the gradient-descent kind of learning algorithms, the synaptic weights w_{ij}, the centers $\mathbf{m}_i$, and the widths σ_i of Gaussian functions are updated as

$$\Delta w_{ij} = -\eta \frac{\partial J_j}{\partial w_{ij}} = -\eta \sum_{k=1}^{N} \varphi_j(e_j^{(k)}; \beta) \frac{\partial e_j^{(k)}}{\partial w_{ij}}, \qquad (13)$$

$$\Delta \mathbf{m}_i = -\eta \frac{\partial J_j}{\partial \mathbf{m}_i} = -\eta \sum_{j=1}^{p} \sum_{k=1}^{N} \varphi_j(e_j^{(k)}; \beta) \frac{\partial e_j^{(k)}}{\partial \mathbf{m}_i}, \qquad (14)$$

$$\Delta \sigma_i = -\eta \frac{\partial J_j}{\partial \sigma_i} = -\eta \sum_{j=1}^{p} \sum_{k=1}^{N} \varphi_j(e_j^{(k)}; \beta) \frac{\partial e_j^{(k)}}{\partial \sigma_i}, \qquad (15)$$

$$\varphi_j(e_j^{(k)}; \beta) = \frac{\partial \rho(e_j^{(k)}; \beta)}{\partial e_j^{(k)}} = \frac{e_j^{(k)}}{1 + (e_j^{(k)})^2 \big/ \beta(h)}, \qquad (16)$$

where η is a learning constant, $\varphi(\cdot)$ is usually called the influence function. In the ARLA, the annealing schedule $\beta(h)$ has the convergent properties [4].

V. SIMULATION RESULTS

The identification scheme of a nonlinear MIMO system is depicted in Fig. 1. In this scheme, the training input-output data are obtained by feeding a signal $\mathbf{x}(k)$ to the MIMO system and measure its corresponding output $\mathbf{y}(k+1)$. Then subject to the same input signal, the objective of identification is to construct a suitable network model, which produces an output $\hat{\mathbf{y}}(k+1)$ to approximate $\mathbf{y}(k+1)$ as closely as possible.

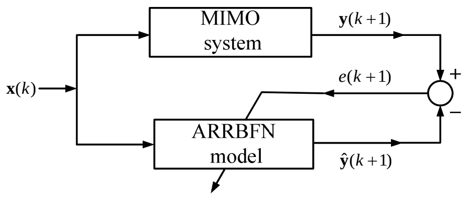

Fig. 1 The proposed identification scheme for an MIMO system

In this section, a two-input two-output nonlinear MIMO systems are used to verify the feasibility of the proposed approach. The root mean squares error (*RMSE*) of the testing data is used to measure the performance of the learning network and is defined as

$$RMSE = \sqrt{\frac{\sum_{k=1}^{N} \left(y_j^{(k)} - \hat{y}_j^{(k)} \right)^2}{N}} \quad \text{for} \quad j = 1, 2 \qquad (17)$$

where $y_j^{(k)}$ is the desired output and $\hat{y}_j^{(k)}$ is the output of the ARRBFN.

Example :

In this example, the nonlinear MIMO system to be identified is described as[5]

$$\begin{bmatrix} y_1(k+1) \\ y_2(k+1) \end{bmatrix} = \begin{bmatrix} \dfrac{y_2(k)}{1 + y_1^2(k)} \\ \dfrac{y_1(k)}{1 + y_1^2(k)} \end{bmatrix} + \begin{bmatrix} x_1(k) \\ x_2(k) \end{bmatrix} \qquad (18)$$

where

$$(x_1(k), x_2(k)) = (\cos(\frac{2\pi k}{100}), \sin(\frac{2\pi k}{100})) \text{ for } 0 \le k \le 300$$

and the 301 training input-output data are generated by substituting into (18) sequentially.

With the training data and following the procedure of the proposed method, the value of L is found to be 15. Meanwhile, the initial values of w_{ij}, $\mathbf{m}_i$, and σ_i can also be determined. When applying the proposed SVR method, the parameters in (6) and (7) are chosen as $C = 1$ and $\varepsilon = 0.2$, respectively. After initialization, the ARLA is then applied to train the RBFN. After 1000 epochs, the *RMSE* values of y_1 and y_2 are found to be 0.0102 and 0.0144, respectively. The details of the simulation results are shown in Fig. 2 through Fig. 7, respectively.

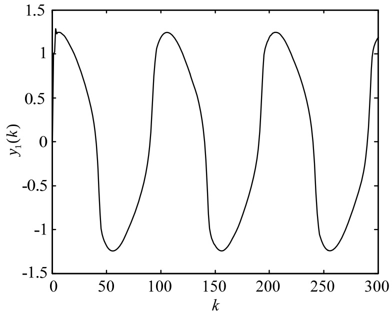

Fig.2 The desired output for y_1

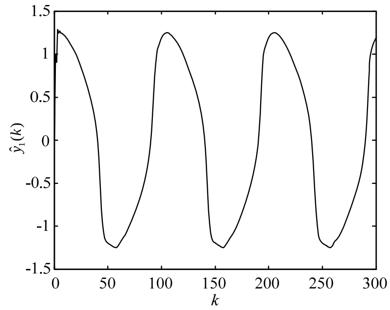

Fig. 3 The final output for y_1 after 1000 epochs

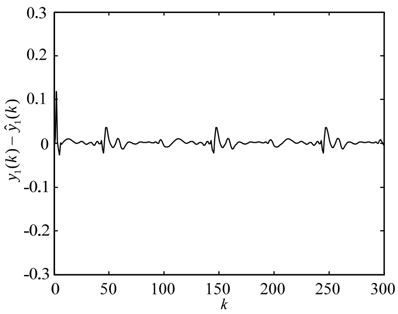

Fig. 4 The plot of $y_1(k) - \hat{y}_1(k)$ after 1000 epochs

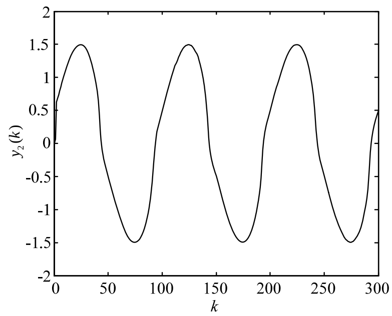

Fig. 5 The desired output for y_2

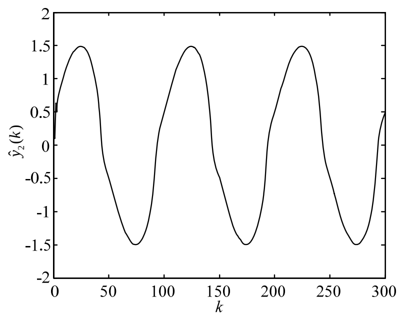

Fig. 6 The final output for y_2 after 1000 epochs

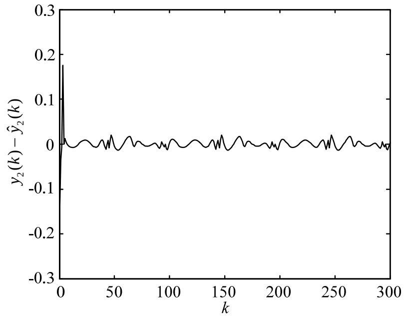

Fig. 7 The plot of $y_2(k) - \hat{y}_2(k)$ after 1000 epochs

7. Conclusion

With the integration of SVR, LSR, and the annealing robust algorithm, an RBFN is used for identification of an MIMO system. The proposed SVR approach has good performance in determining the number of hidden layer nodes and the initial parameters of the kernel. Then based on the values obtained by the SVR method, the synaptic weights can also be determined by using the technique of the LSR. After initialization, the annealing robust learning algorithm is adopted to adjust the parameters of the RBFN to approximate the MIMO system as closely as possible. The simulation results indicated that the proposed method can be used as a reliable technique for identification of nonlinear MIMO systems.

Acknowledgment

This work was supported in part by the National Science Council, Taiwan, R.O.C., under grants NSC96-2628-E-224-007-MY2.

References

[1] Vapnik VN (1995), The Nature of statistical learning Theory. Springer, New York

[2] Larson R, Edwards BH, Falvo DC (2004), Elementary linear algebra, fifth edition. Houghton Mifflin Company, New York

[3] Chuang CC, Jeng JT, Lin PT (2004), Annealing robust radial basis function networks for function approximation with outliers. Neurocomputing 56:123-139

[4] Chuang CC, Su SF, Hsiao CC (2000), The annealing robust backpropagation (BP) learning algorithm. IEEE Trans on Neural Networks 11(5):1067-1077

[5] Purwar S, Kar IN, Jha AN (2007), On-line system identification of complex systems using Chebyshev neural networks. Applied Soft Computing 7:364-372