STRUCTURE DETECTION OF RBF NETWORKS USING DRR ALGORITHM

M. Y. Mashor

School of Electrical and Electronic,
University Science Malaysia,
31750 Tronoh, Perak,
MALAYSIA
E-mail: yusof@eng.usm.my

Abstract

The performance of a neural network strongly depends on the network structure. Therefore, an algorithm that can automatically select the network structure will be very helpful. Various algorithms based on orthogonal estimation have been suggested to perform this task. However, these algorithms require very heavy computational load. This paper presents an algorithm called distance reduction ratio algorithm as an alternative to those algorithms. The proposed algorithm only involves simple calculation steps. Thus, computational load is tremendously reduced without sacrificing the network performance.

1. Introduction

Most of the neural network architectures or training algorithms do no provide a specific function to determine the optimum network structure to describe the identified system. However, the performance of RBF network depends heavily on the network structure especially the input and hidden nodes. Incorrect input nodes or poorly located RBF centres will induce bias to the fitted network model. Mashor (1995) showed that both underfitting and overfitting in input nodes or hidden nodes are undesirable. So without efficient structure detection method or prior knowledge about the identified system the resulting network model will have very little practical value.

Korenberg et al. (1988) introduced an orthogonal estimation algorithm that was based on the orthogonal least squares algorithm to select the significant terms and estimate the parameters of the NARMAX model. The algorithm when coupled with some model validity tests such as correlation tests developed by Billings and Voon (1986) provides good results in practice to detect the structure of the NARMAX models. However, the resulting models are affected by the order in which the terms are processed by the algorithm. The problem has been mentioned in Korenberg et al. (1988).

A more reliable version of the algorithm called orthogonal-forward-regression (OFR) algorithm was presented in Billings et al. (1989). Chen et al. (1991), Chen and Billings (1992) and Shertinsky and Picard (1996) used the OFR algorithm to select the centres of RBF network models from the input and output data of the identified system. However, these algorithms that are based on orthogonal least squares require heavy computation and normally take a few hours to complete the selection process using a PC with Pentium CPU.

In the present study, a new algorithm called *distance reduction ratio algorithm* (DRR) is proposed as an alternative to the orthogonal estimation algorithm for selecting the significant centres. The new algorithm can tremendously reduce computational time (just a few minutes) compared to the method that is based on the orthogonal estimation algorithm without sacrificing the performance of the RBF network.

2. RBF Network with Linear Input Connections

A RBF network with m outputs and n_h hidden nodes can be expressed as:

$$y_i(t) = w_{i0} + \sum_{j=1}^{n_h} w_{ij} f\left(\left\|v(t) - c_j(t)\right\|\right);$$
$$i = 1,\ldots, m \tag{1}$$

where w_{ij}, w_{i0} and $c_j(t)$ are the connection weights, bias connection weights and RBF centres respectively, $v(t)$ is the input vector to the RBF network composed of lagged input, lagged output and lagged prediction error and $\phi(\bullet)$ is a non-linear basis function. $\|\bullet\|$ denotes a distance measure that is normally taken to be the Euclidean norm.

Since neural networks are highly non-linear, even a linear system has to be approximated using the non-linear neural network model. However, modelling a linear system using a non-linear model can never be better than using a linear model. Considering this argument, the RBF network with additional linear input connections is used. The proposed network allows the network inputs to be connected directly to the output node via weighted connections to form a linear model in parallel with the non-linear standard RBF model as shown in Figure 1.

The new RBF network with m outputs, n inputs, n_h hidden nodes and n_l linear input connections can be expressed as:

$$y_i(t) = w_{i0} + \sum_{j=1}^{n_l} l_{ij} vl(t) + \sum_{j=1}^{n_h} w_{ij} f\left(\left\|v(t) - c_j(t)\right\|\right);$$
$$i = 1, 2, \Lambda\ , m \tag{2}$$

where the λ's and vl's are the weights and the input vector for the linear connections respectively. The input vector for the linear connections may consist of past inputs, outputs and noise lags. Since λ's appear to be linear within the network, the λ's can be estimated using the same algorithm as for the w's. As the additional linear connections only introduce a linear model, no significant computational load is added to the standard RBF network training. Furthermore, the number of required linear connections is normally much smaller than the number of hidden nodes in the RBF networks. In the present study, Givens least squares algorithm with additional linear input connection features is used to estimate w's and λ's. Refer to Chen et al. (1992) for implementation of Givens least squares algorithm.

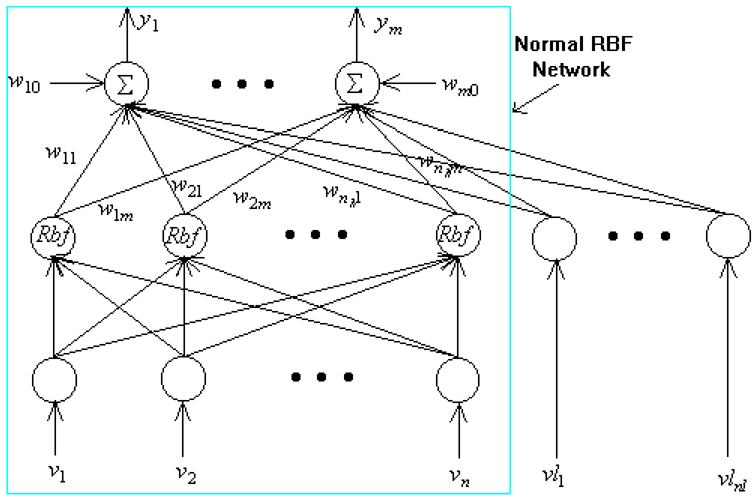

Figure 1. The RBF network with linear input connections

3. Centre Detection Using DRR Algorithm

Improper hidden node or centre selection may induce bias in the network model. The quality of the RBF centres positioned using a clustering algorithm such as *k*-means clustering will normally be affected by the initial centres. Some of the centres may be placed at inappropriate positions and will not give any significant contribution to the network model. In some cases, the over specified centres will cause the network model to become overfitted which will slow down the training process and may induce bias.

In the present study, *distance reduction ratio algorithm* is proposed to select the centres for RBF network from the training data set. The algorithm is an unsupervised algorithm that selects the centres such that the mean squared distance between the centres and the training data defined in equation (3) will be minimised. In contrast, the error reduction methods that use orthogonal least squares (Korenberg et al., 1988; Billings et al., 1989) are supervised algorithms that select the centres such that the mean squared prediction errors are minimised.

In general, the centres of the RBF network should be selected in such a way that the identified data can be properly represented. In other words, the total distance between the centres and the training data should be minimised. The mean squared distance (MSD) between the centres and the training data can be calculated by:

$$MSD = \frac{1}{N}\sum_{t=1}^{N} M_{jt}\left(\left\| v(t) - c_j \right\|\right)^2 ; \quad j = 1,2,\Lambda\ n_h \tag{3}$$

where the n_h, N, $v(t)$ and c_j are the number of centres, number of training data, input vector and centres respectively. M_{jt} is a membership function which means that the data $v(t)$ belongs to centre c_j and the data are assigned to the nearest centres.

The mean square distance in equation (3) is maximum when no centre is included in the network model to give

$$D_{\max} = \frac{1}{N}\sum_{t=1}^{N}\left(v(t)\right)^2 \qquad (4)$$

By including the centre c_j in the network model as the first significant centre, the total squared distance will be reduced by a distance reduction ratio, *DRR* defined as

$$DRR_j = \frac{D_{\max} - \frac{1}{N}\sum_{t=1}^{N} M_{jt}\left(\left\|v(t)-c_j\right\|\right)^2}{D_{\max}},$$

$$j = 1, 2, \ldots, N \qquad (5)$$

Including the centre c_j in the network model as the *k*-th significant centre gives the distance reduction ratio as follows

$$DRR_j = \frac{D_{\max} - \frac{1}{N}\sum_{t=1}^{N} M_{pt}\left(\left\|v(t)-c_p\right\|\right)^2}{D_{\max}};$$

$$j = k, k+1, \ldots, N;\ p = 1, \ldots, k-1, k \qquad (6)$$

and $c_p = [c_l, c_j]$ where c_l $(1 \le l \le k\text{-}1)$ are the previously selected centres and c_j is the current candidate centres.

The value of *DRR* varies with the suitability of the centre locations, the larger the value of DRR_j the more significant is the centre c_j to the process of minimising the mean squared distance. Due to the strong correlation between the mean squared distance and the mean squared prediction error of the RBF network model, *DRR* defined in equation (5) and (6) may be used to indicate the significance of each candidate centre to the overall RBF network performance.

The algorithm to select the significant centres using the DRR method can be implemented as follows:

1) Assign the input vector to be used for the RBF network model and use all the training data as the candidate centres; and choose the required number of centres, n_h.

2) Calculate the *DRR* value for each candidate centre according to equation (5) and select the centre that is associated with the largest *DRR* as the first significant centre.

3) Exclude the selected centre from the candidate centres and use the remaining centres as the candidates for selecting the next significant centre.

4) Calculate the *DRR* value for each candidate centre, c_j ($j \ge k$), according to equation (6) and select the centre that is associated with the largest *DRR* as the *k*-th significant centre

5) Repeat steps 3 and 4 until the number of the selected centres is equal to n_h.

After the centres have been selected, the weights of the RBF network are estimated using Givens least squares algorithm. Because the selection procedure involves only a few simple calculation steps, the centre selection process requires much less computational time compared to the selecting procedure that is based on the orthogonal estimation algorithm Korenberg et al. (1988) and Billings et al. (1989).

4. Results

The efficiency of the proposed DRR algorithm to detect the significant hidden nodes was tested using a simulated data set and two real data sets. In this simulations, all the network models have the following specifications, ρ = 1000.0, β_0 = 0.99, β(0) = 0.95 and $\phi(\bullet)$ was selected to be the thin-plate-spline (refer to Chen et al. (1992) for the definition of these parameters). During the calculation of the mean squared error (MSE), the noise model was excluded from the model since the noise model will normally cause the mean squared error to become unstable in the early stage of training. This is because if the noise model is included, the MSE will consist of model predicted output of the predicted noise, $\hat{\varepsilon}(t)$, and one step ahead predicted output of the process model. Since the process model is very bad at the early training stage, $\hat{\varepsilon}(t)$ will be very large. So model predicted output of the noise model based on $\hat{\varepsilon}(t)$ may become unstable.

Example 1

System S1 was described by the following difference equation:

$$y(t)=1.0+0.3y(t-1)-0.4y(t-3)+$$
$$0.1y(t-5)y(t-1)+0.2y(t-1)u(t-3)+$$
$$0.4u(t-1)0.6u(t-2)-0.4u(t-5)+e(t)+$$
$$0.2e(t-2)-0.8e(t-4)+0.5e(t-6)$$

where $u(t)$ was a uniformly distributed random sequence [-1,+1] and $e(t)$ was a Gaussian white noise sequence with zero mean and variance 0.005. System S1 was used to generate 1000 pairs of data input and output. The first 500 data were used to train the network and the other 500 data were used for testing the fitted network model.

The network was assigned to have the following input vectors:

$$v(t)=[u(t-1)\Lambda\ u(t-3)u(t-5)y(t-1)y(t-3)y(t-5)]$$

and a bias input

$$vl(t)=[e(t-2)e(t-4)e(t-6)]$$

The distance reduction method was used to select the 40 most significant centres from the training data. One step ahead prediction (OSA), correlation tests and MSE of the network model using the selected centres are shown in Figures 2, 3 and 4 respectively. The resulting network gives good correlation tests and predicts quite well, hence the network model is considered as an adequate representation of the system S1.

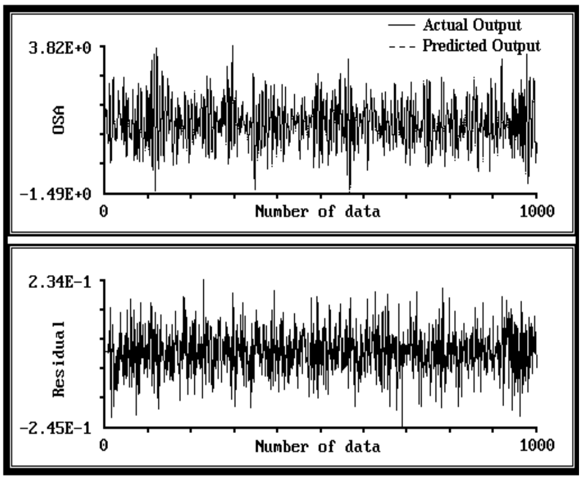

Figure 2:- OSA for system S1

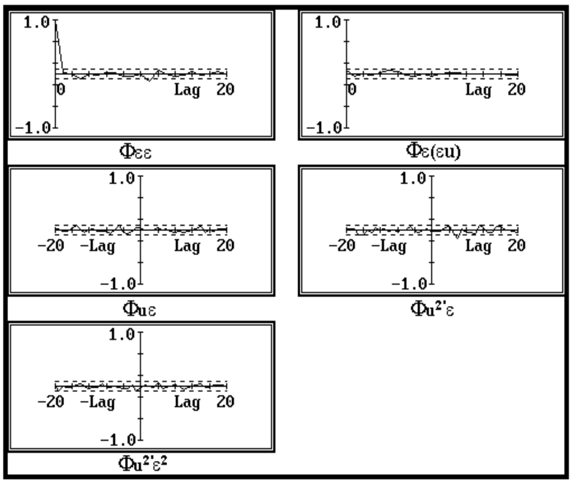

Figure 3: Correlation tests for system S1

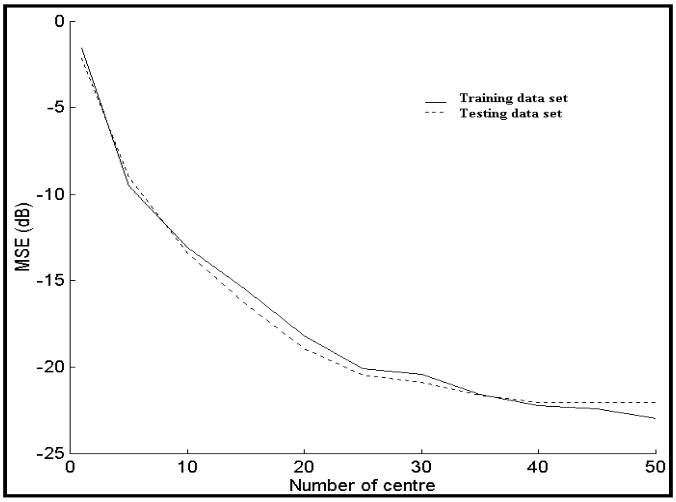

Figure 4: MSE test for system S1

Example 2

A data set of 1000 input-output samples were taken from a tension leg platform. The first 600 data were used for training and the rest of the data were used for testing the fitted network model. The network was assigned to have the following input vectors:

$$v(t) = [u(t-1) \Lambda\ u(t-10) y(t-1) y(t-3) y(t-4)]$$
$$vl(t) = [y(t-2) y(t-3) y(t-4) y(t-5) e(t-1) e(t-2)]$$

and a bias *input*

where *u*'s, *y*'s and *e*'s are the past input, output and prediction error respectively. The distance reduction method was used to select the 45 most significant centres from the possible candidates of 600 training data. OSA test, correlation tests and MSE of the network using the selected centres are shown in Figures 5, 6 and 7 respectively. The resulting network passes all the correlation tests and predicts very well hence the network is considered as an adequate representation of the system.

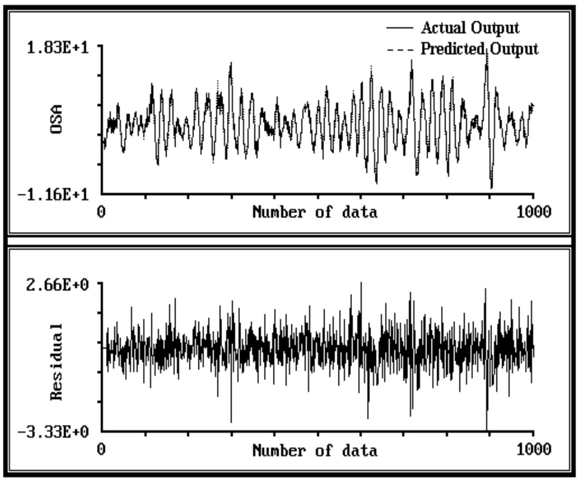

Figure 5. OSA for system S2

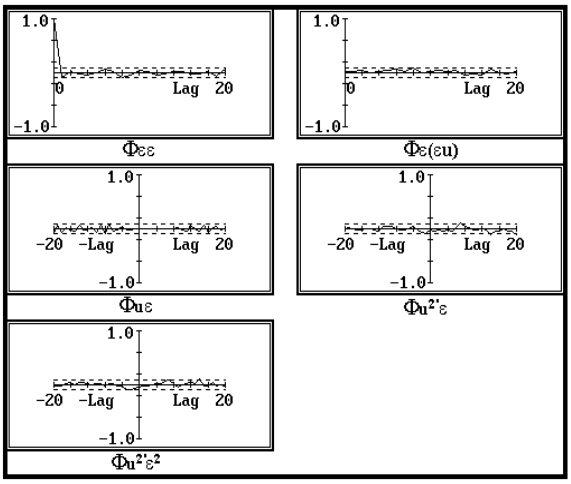

Figure 6. Correlation tests for system S2

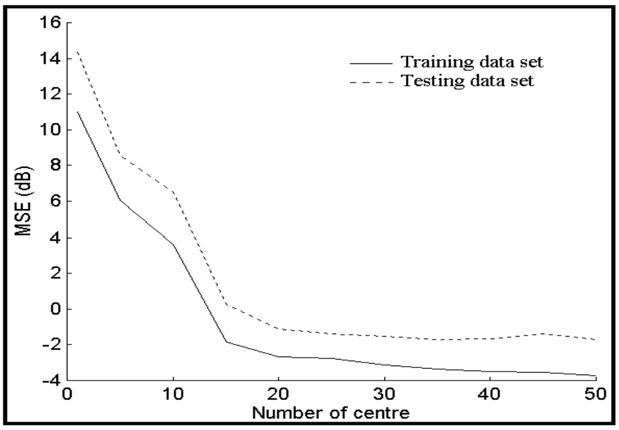

Figure 7. MSE test for system S2

Example 3

The second data set (S3) was taken from a heat exchanger system and consists of 1000 samples. The first 500 data were used to train the network and the remaining 500 data were used to test the fitted network model. The network has been trained using the following specification:

$$v(t) = \begin{bmatrix} u(t-1) & u(t-2) & y(t-1) & y(t-4) \\ e(t-3) & e(t-4) & & \end{bmatrix}$$

$$vl(t) = \begin{bmatrix} u(t-1) & u(t-2) & y(t-1) & y(t-4) \\ e(t-3) & e(t-4) & & \end{bmatrix}$$

with bias input

The distance reduction method was used to select the 35 most significant centres from the possible candidates of 600 training data. OSA test, correlation tests and MSE of the network using the selected centres are shown in Figures 8, 9 and 10 respectively. The resulting network passes all the correlation tests and predicts very well hence the network is considered as an adequate representation of the system.

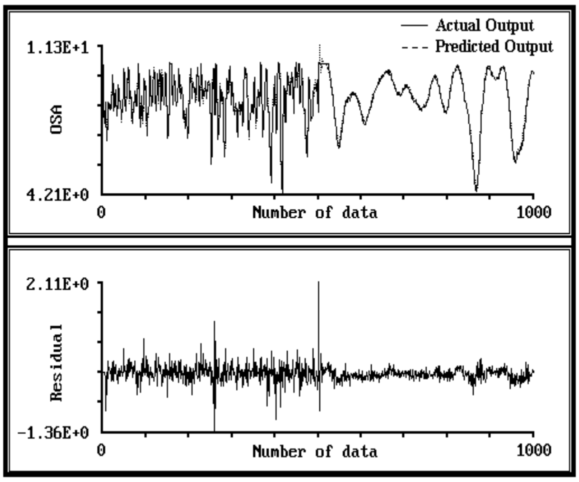

Figure 8. OSA for system S3

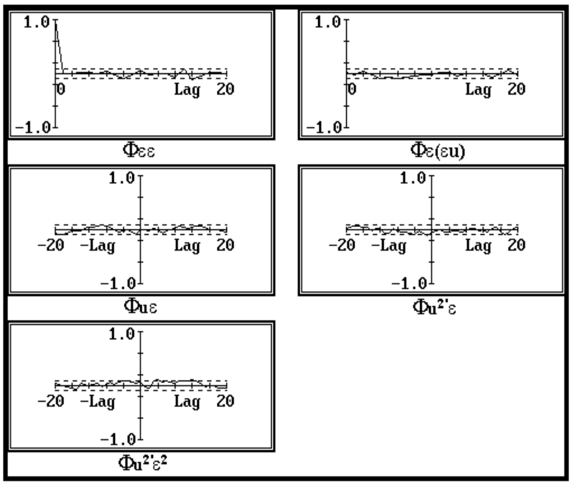

Figure 9. Correlation tests for system S3

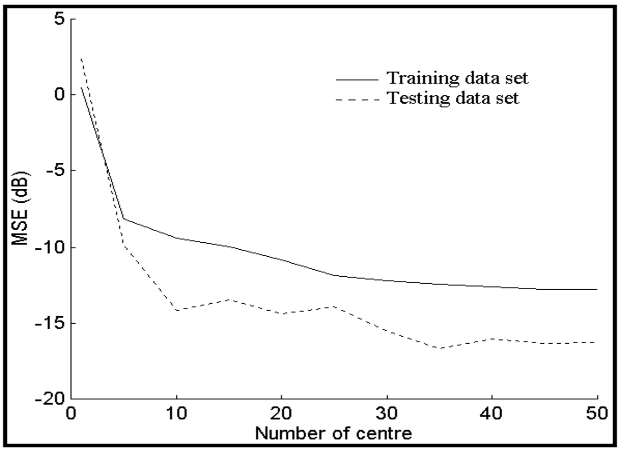

Figure 10. MSE test for system S3

These examples illustrate that the distance reduction method has successfully selected the significant centres from the training data. In all examples the resulting RBF network models provided good predictions and correlation tests. Hence, the method may be considered as an alternative to the orthogonal-forward-regression algorithm used by Chen et al. (1991) and Chen and Billings (1992). This method offers the extra advantage of much simpler calculations and hence shorter detection time.

5. Conclusion

The DRR algorithm was introduced as an alternative to the selection procedures that are based on orthogonal estimation algorithm to detect the hidden nodes for RBF networks. This algorithm is an unsupervised method that selects the centres such that the distance between the selected centres and the training data is minimised. The results in section 4 show that the proposed centre detection algorithm is adequate to select significant centres. Due to simple calculation steps, the DRR algorithm can perform centre selection in a much less computational load compared to the selection procedures based on orthogonal estimation without sacrificing RBF network performance.

______

References

1. Billings, S.A. and Voon, W.S.F., 1986, "Structure detection and model validity tests in the identification of non-linear systems", *Proc. IEE, Part D*, Vol. 127, pp. 272-285.

2. Billings, S.A., Chen, S. and Korenberg, M.J., 1989, "Identification of MIMO non-linear systems using a forward-regression orthogonal estimator", *Int. J. Control*, Vol. 49, pp. 2157-2189.

3. Billings, S.A., Jamaluddin, H.B. and Chen, S., 1992, "Properties of neural networks with applications to modelling non-linear dynamical systems", *Int. J. of Control*, Vol. 55, pp. 193-224.

4. Chen, S., Cowan, C.F.N. and Grant, P.M., 1991, "Orthogonal least squares learning algorithm for radial basis function networks", *IEEE Trans. on Neural Networks*, Vol. 2, pp. 302-309.

5. Chen, S. and Billings, S.A., 1992, "Neural networks for non-linear dynamic system modelling and identification", *Int. J. of Control*, Vol. 56 (2), pp. 319-346.

6. Chen, S., Billings, S.A. and Grant, P.M., 1992, "Recursive hybrid algorithm for non-linear system identification using radial basis function networks", *Int. J. of Control*, Vol. 55, pp. 1051-1070.

7. Korenberg, M.J., Billings, S.A, Liu Y.P. and McIlroy, P.J., 1988, "Orthogonal parameters estimation algorithm for non-linear stochastic systems". *Int. J. Control*, Vol. 48 (1), pp. 193-210.

8. Mashor, M.Y., 1995, System identification using radial basis function network, *PhD thesis*, University of Sheffield, United Kingdom.

9. Shertinsky A. and Picard, R.W., 1996, "On the efficiency of the orthogonal least squares training method for radial basis function networks". *IEEE Trans. on Neural Networks*, Vol. 7, pp. 195 - 200.