Comparison of Multilayer Perceptron and Radial Basis Function networks as tools for flood forecasting

A. W. JAYAWARDENA, D. A. K. FERNANDO & M. C. ZHO
Department of Civil and Structural Engineering, The University of Hong Kong, Pokfulam Road, Hong Kong

Abstract This paper presents a comparison between two Artificial Neural Network (ANN) approaches, namely, Multilayer Perceptron (MLP) and Radial Basis Function (RBF) networks, in flood forecasting. The basic difference between the two methods is that the parameters of the former network are nonlinear and those of the latter are linear. The optimum model parameters are therefore guaranteed in the latter, whereas it is not so in the more popularly adopted former approach. The two methods are applied to predict water levels at stations in an experimental drainage basin and in a major river in China during storm periods. The RBF network based models give predictions comparable in accuracy to those from the MLP based models. It is also observed that the RBF approach requires less time for model development since no repetition is required to reach the optimum model parameters.

INTRODUCTION

Reliable forecasts form the basis of any warning system used as a non-structural means of flood disaster mitigation. Many of the techniques used in the past for flood forecasting are based on some assumed relationship between the rainfall and the corresponding runoff. More recently however, the Artificial Neural Network (ANN) approach which can learn from training data sets and does not require a prior knowledge of an explicit relationship between the rainfall and runoff has been applied to several situations. Most of such applications have used the Multi Layer Perceptron (MLP) type ANN models coupled with the error Back Propagation (BP) algorithm (e.g. Isobe et al., 1994; Jayawardena & Fernando, 1995 a,b; Liong & Chan, 1993; Raman & Sunilkumar, 1995; and Smith & Eli, 1995). MLP however is highly nonlinear in its parameters. The BP algorithm which uses the method of steepest descent does not guarantee convergence to globally optimum set of parameters. Several attempts of trial and error are therefore required to choose the “best” from a set of locally optimum parameters.

An alternative to the MLP is the Radial Basis Function (RBF) network (Bianchini et al., 1995; Chen et al., 1991) which has linear parameters and has found applications in other areas such as electrical and electronic engineering. Park & Sandberg (1991) proved theoretically that the RBF type ANNs are capable of universal approximations and learning without local minima, thereby guaranteeing convergence to globally optimum parameters. For hypothetical situations, Moody & Darken (1989) demonstrated that the RBF type networks learn faster than MLP networks. This study attempts to apply the RBF approach to real situations of flood water level predictions and to compare the model performances with those of MLP network models.
RADIAL BASIS FUNCTION (RBF) TYPE NETWORK

Network architecture and composition

An RBF network is a two-layer feed-forward type network in which the input is transformed by the basis functions at the hidden layer. At the output layer, linear combinations of the hidden layer node responses are added to form the output. The name RBF comes from the fact that the basis functions in the hidden layer nodes are radially symmetric. Chen et al. (1991) report that the choice of the basis function is not crucial to the performance of the network. The most common choice however, is the Gaussian function which can be defined by a mean and a standard deviation. Figure 1 shows a schematic diagram of an RBF network with N, L and M respectively of input, hidden and output layer nodes for the general transformation of ND points of \(X(X_1, \ldots, X_i, \ldots, X_{ND})\) in the input space to points \(Y(Y_1, \ldots, Y_i, \ldots, Y_{ND})\) in the output space.

The parameters of an RBF type neural network are the centres \((U_j)\) and the spreads \((\sigma_j)\) of the basis functions at the hidden layer nodes, and the synaptic weights \((w_k)\) of the output layer nodes. The RBF centres are also points in the input space. The basis function response for an input depends on the distance between the point representing the input \((X)\) and the RBF centre \((U_j)\). The RBFs at the hidden layer nodes produce non-zero responses only when the input falls within a small localized region of the basis functions' centre.

Network functioning

In RBF networks, the connections between the input and the hidden layers are not weighted. The inputs therefore reach the hidden layer nodes unchanged. For an input \(X\), the \(j\)th hidden node produces a response \(h_j\) given by,

\[
h_j = \exp\left[-\frac{\|X - U_j\|^2}{2\sigma_j^2}\right]
\]

(1)

---

Fig. 1 Schematic diagram of a Radial Basis Function Network.
where \( \| X' - U_j \| \) is the distance between the point representing the input \( X' \) and the centre of the \( j \)th hidden node as measured by some norm. In this study the Euclidean norm is used. The output \( y_{ik} \) of the network at the output node is given by,

\[
y_{ik} = \sum_{j=1}^{L} h_j w_{kj}
\]

In the special case where the number of hidden layer nodes is equal to the number of data in the training set \( L = ND \) and the RBF centres coincide with the inputs \( U_j = X_i \), where \( i = j = 1, 2, \ldots, ND \), the hidden layer response according to equation (1) becomes unity for \( j = i \). If the basis functions are truly localized, the response of the other hidden layer nodes will be near zero (i.e. \( \sigma_j \) for \( j \neq i \) are such that \( h_j \equiv 0 \) for \( j \neq i \)). It can also be seen that equation (2) gives the exact output when the output layer weight is equal to the output (the contribution to the weighted summation from \( j = i \) is \( y_{ik} \) and that from all \( j \neq i \) is nearly zero). In the ideal case therefore, RBF network can be made to map points in \( N \)-dimensional input space exactly on to points in \( M \)-dimensional output space. This however, is not practical when \( ND \) is large in which case a few input points are chosen to represent the entire input data set.

**Training or determination of model parameters**

Training of RBF networks is carried out using a hybrid procedure consisting of both supervised and unsupervised learning methods. The output layer is trained by a supervised learning method, similar to that used in the BP algorithm in which the synaptic weights are updated in proportion to the difference between the network output and the target output. Training of the hidden layer on the other hand involves the determination of the radial basis functions by specifying appropriate \( U_j \) and \( \sigma_j \) for each node. These parameters are dependent only on the inputs and are independent of the outputs, making this part of the learning process an unsupervised one.

The original RBF method requires that there be as many RBF centres as there are distinct data points in the input space. This however, is not possible in practice because of the large numbers of input points found in most real situations. Moreover, the inputs usually occur in clusters making overlapping of receptive fields inevitable. Choosing all points as RBF centres will therefore lead to an unnecessarily large network involving long training and computation times. An effective way of reducing the number of nodes in the hidden layer is by “clustering” the input points such that each point falls into one of the hyperspheres which collectively span the entire input space (Wasserman, 1993). Each of the RBF centres \( (U_j) \) will then be located at the centres of each cluster. Definition of the “centre” of the cluster depends on the type of clustering technique adopted. In this study \( k \)-means clustering algorithm (Moody & Darken, 1989) is adopted where the centre is defined as the centroid (in this context, the centroid is the centre of gravity of unit masses located at each point belonging to the cluster). The value of \( \sigma_j \) is computed as the mean distance from the centre of the cluster to other points that form the cluster. The number of hidden nodes is equal to the number of clusters \( k \).
APPLICATIONS

Prediction of water levels at Shang Qiao station in the Shang Qiao drainage basin

The MLP and RBF network type models described above were first applied to predict flood water levels at a gauging station in the Shang Qiao experimental drainage basin (inset 1 in Fig. 2, area 110 km²) which lies to the west of the Pearl River in southern China (~112°30'E, ~22°40'N). The reason for using water levels instead of discharges is because they are more practical indicators of the level of flooding. Hourly rainfalls at Ban Chun and Shang Qiao and hourly water levels at Shang Qiao hydrological station for six storm periods extracted from hydrological data year books compiled by the Guangdong Provincial General Hydrological Station, were used in the study. A summary of the data used is given in Table 1.

In general, the lead time of a forecasting model can be increased at the expense of the accuracy of predictions. In this case a lead time of 2 h was found to be a

Inset 1: Experimental drainage basin

Inset 2: Liu Xie River

Fig. 2 Study areas of the Pearl River delta.
Table 1 Summary of data used for the applications.

<table>
<thead>
<tr>
<th>Application</th>
<th>Storm event no.</th>
<th>Year</th>
<th>Event duration (number of hours)</th>
<th>Observed mean water level (m)</th>
<th>Data used for</th>
</tr>
</thead>
<tbody>
<tr>
<td>Prediction of water levels at Shang Qiao station</td>
<td>1</td>
<td>1987</td>
<td>4 April–9 April (144)</td>
<td>2.93</td>
<td>Training</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>1987</td>
<td>22 July–4 August (336)</td>
<td>3.15</td>
<td>Testing</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>1988</td>
<td>13 May–20 May (192)</td>
<td>3.16</td>
<td>Testing</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>1988</td>
<td>26 June–2 July (168)</td>
<td>3.01</td>
<td>Testing</td>
</tr>
<tr>
<td></td>
<td>5</td>
<td>1988</td>
<td>28 October–2 November (168)</td>
<td>3.08</td>
<td>Testing</td>
</tr>
<tr>
<td></td>
<td>6</td>
<td>1988</td>
<td>14 November–20 November (168)</td>
<td>3.15</td>
<td>Testing</td>
</tr>
<tr>
<td>Prediction of water levels at Tai Bin Chong town station</td>
<td>1</td>
<td>1987</td>
<td>5 April–8 April (65)</td>
<td>16.12</td>
<td>Testing</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>1987</td>
<td>20 May–25 May (110)</td>
<td>17.67</td>
<td>Training</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>1988</td>
<td>16 May–18 May (23)</td>
<td>16.10</td>
<td>Testing</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>1988</td>
<td>23 June–24 June (40)</td>
<td>16.22</td>
<td>Testing</td>
</tr>
<tr>
<td></td>
<td>5</td>
<td>1988</td>
<td>20 July–21 July (18)</td>
<td>16.19</td>
<td>Testing</td>
</tr>
</tbody>
</table>

Reasonable compromise. The 2-h-ahead level prediction models were developed by taking the present water level at Shang Qiao station SWL, to be dependent on its past values SWL, -2, SWL, -3, SWL, -4, past rainfall values at Shang Qiao SR, -2, SR, -3, SR, -4 and past rainfall values at Ban Chun BR, -2, BR, -3, BR, -4. The data for the first storm period were used for training and the remainder was used for testing the performance of the models.

The stopping criterion used in this study was termination upon completion of predefined number of training iterations. The number 200 was chosen after a few initial runs which indicated that appreciable learning of all networks occur very quickly and the reduction in error becomes insignificant well before the 200 limit is reached. In the case of MLP type of networks, the number of nodes in the hidden layer was varied from 4 to 10. Each network was trained 10 times starting with different initial weights. The weights that correspond to the minimum mean squared error during the 10 training cycles were taken as the final model parameters. For the

Table 2 Performance of models in predicting water levels.

<table>
<thead>
<tr>
<th>Application</th>
<th>ANN model</th>
<th>Predicted mean water level (m)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Prediction of water levels</td>
<td></td>
<td>* r.m.s. error as a % of observed mean water level</td>
</tr>
<tr>
<td>Storm event no.</td>
<td>1*</td>
<td>2</td>
</tr>
<tr>
<td>at Shang Qiao station</td>
<td>RBF</td>
<td>2.93</td>
</tr>
<tr>
<td></td>
<td>MLP</td>
<td>5.75</td>
</tr>
<tr>
<td></td>
<td></td>
<td>2.93</td>
</tr>
<tr>
<td></td>
<td></td>
<td>1.46</td>
</tr>
</tbody>
</table>

| Storm event no.      | 1 | 2* | 3 | 4 | 5 |
| at Tai Bin Chong town station | RBF | 16.65 | 17.67 | 15.97 | 16.28 | 16.68 |
|                       | MLP | 2.99 | 1.61 | 2.68 | 2.43 | 3.59 |
|                       |     | 16.18 | 17.65 | 15.95 | 16.09 | 16.20 |
|                       |     | 4.13 | 0.99 | 3.22 | 3.63 | 2.52 |

* Event used for training the networks.
RBF type networks, several network configurations were developed by clustering the training data with $k$ varying from 4 to 10. These RBF networks were then each trained over 200 iterations. Training in the RBF networks however need not be repeated for each network configuration as the weights are linear.

In any simulation exercise the most suitable model is chosen on the basis of the least error produced during both training (calibration) and testing (validation) stages. At the model development stage however, it has to be chosen on the basis of the least error during the training stage alone. On this basis, an MLP network with six nodes and an RBF network with five nodes in their hidden layers were chosen to represent MLP and RBF models. The statistics of the predictions made with the two models are given in Table 2. The predicted mean water levels agree well with the observed mean water levels (in Table 1) for all storms. The prediction errors for each storm period (in italics) indicate that the two models are comparable. The two hour ahead predictions of the water levels obtained by the RBF and MLP models are shown in Fig. 3 along with the corresponding observed values and rainfalls at Shang Qiao (SR) and Ban Chun (BR).

**Prediction of river water levels at Tai Bin Chong town**

The second application was to predict water levels at Tai Bin Chong gauging station which lies in the downstream part of the Liu Xie River, a tributary of the Pearl River (inset 2 in Fig. 2). It is the last gauged station where the water levels are not affected by the tidal variations. The aim is to predict water level at this station using the upstream water levels and rainfall in the region. Data corresponding to five flood events that occurred in 1987 and 1988, extracted from the hydrological data books, are summarized in Table 1.

A 5-h-ahead forecasting model was found to give a reasonable compromise between the lead time and the accuracy of predictions. In order to develop this five hour ahead flood level forecasting model, the present water level at Tai Bin Chong town TWL$_t$ was taken to be a function of TWL$_{(t-5)}$, TWL$_{(t-6)}$, TWL$_{(t-7)}$, past water levels at Wan Chung reservoir WWL$_{(t-11)}$, WWL$_{(t-12)}$, WWL$_{(t-13)}$, WWL$_{(t-14)}$, rainfall at Sha Shi SR$_{(t-10)}$, SR$_{(t-11)}$, SR$_{(t-12)}$, SR$_{(t-13)}$, SR$_{(t-14)}$, SR$_{(t-15)}$, and rainfall at Wan Chung reservoir WR$_{(t-11)}$, WR$_{(t-12)}$, WR$_{(t-13)}$, WR$_{(t-14)}$. It is likely that some of these parameters are correlated with each other. Nevertheless they were retained because the network considered is not too large. For the MLP network the number of nodes in the hidden layer was increased from 8 to 15. A procedure similar to that adopted in the first application revealed that the representative MLP and RBF models are those with 10 and 9 nodes respectively in their hidden layers. These models were then used for predictions in the remaining floods. The statistics of the predictions from the two models are shown in Table 2. While the predicted mean water levels for each event compare well with those observed (in Table 1), the prediction errors are also quite low. It can also be seen that the RBF network is capable of producing predictions to the same order of accuracy as the MLP network. The predicted and observed water levels are plotted in Fig. 4.
Comparison of two artificial neural network approaches as tools for flood forecasting

Fig. 3 Predicted and observed water levels at Shang Qiao station.
Fig. 4(a): Storm period 1

Fig. 4(b): Storm period 2

Fig. 4(c): Storm period 3

Fig. 4(d): Storm period 4

Fig. 4(e): Storm period 5

Fig. 4 Predicted and observed water levels at Tai Bin Chong town station.
CONCLUSIONS

Radial Basis Function type Artificial Neural Network models were developed to predict water levels at stations in an experimental drainage basin and in a major river during storm periods. As far as accuracy of predictions is concerned the performance of the RBF model using the k-means clustering technique compares well with that of the MLP with error back propagation method. The RBF network based models are linear in the parameters and therefore guarantee convergence to their optimum values for a particular network architecture. Development of the RBF network model therefore requires less trial and error and thus, less time and effort, than that needed by the MLP with BP approach.

Acknowledgements Part of the work reported in this paper was carried out under a collaboration project with the Guangdong Provincial Research Institute of Water Conservancy and Hydropower, China. Partial funding for the study was provided by the Hong Kong Research Grants Council under award no. HKU 272/95E.

REFERENCES