Comparison of MLP, CCL and CGT Networks for Prediction of Rainfall

N. Vivekanandan
Assistant Research Officer
Central Water and Power Research Station, Pune 411024, Maharashtra, India
E-mail: anandaan@rediffmail.com

Abstract: Prediction of rainfall for a region is of utmost importance for planning and design of irrigation and drainage systems as also for command area development. Assessment of rainfall can be carried out by different approaches like deterministic, stochastic, conceptual and soft computing. This paper illustrates the use of Artificial Neural Network (ANN) for prediction of rainfall at Fatehabad and Hansi stations. ANN models such as Multi-Layer Perceptron (MLP), Cascade Correlation (CCL) and Conjugate Gradient (CGT) are applied to train the network data. Model performance indicators such as correlation coefficient, model efficiency and root mean square error are used to evaluate the performance of the ANN models. The study showed that the CCL network is better suited amongst three networks studied for prediction of rainfall at Fatehabad and MLP for Hansi.

Keywords - Correlation, Mean Square Error, Neural Network, Rainfall, Model Efficiency

I. INTRODUCTION

Prediction of rainfall for a region is of utmost importance for planning and design of irrigation and drainage systems as also for command area development. Since the distribution of rainfall varies over space and time, it is required to analyze the data covering long periods and observed at various locations to arrive at reliable information for decision support. Further, such data need to be analyzed in different ways, depending on the issue under consideration. For example, analysis of consecutive days of rainfall is more relevant for drainage design of agricultural lands, whereas analysis of weekly rainfall data is relevant for planning of cropping pattern. Likewise, analysis of monthly, seasonal and annual data is more useful for water management practices [1]. Approaches such as deterministic, conceptual, stochastic and Artificial Neural Network (ANN) are commonly used for prediction of rainfall. Past research experience shows that there is an abundance of literature on development of deterministic, conceptual and stochastic models [2]. In this context, ANN is considered an effective tool for prediction of meteorological variables such as rainfall, temperature and wind speed, etc; and hence used in the present study.

ANN modelling procedures adapt to complexity of input-output patterns and accuracy goes on increasing as more and more data become available. Figure 1 shows the architecture of ANN that consists of input layer, hidden layer, and output layer. In turn, these layers have a certain number of neurons or units, so the units are also called input units, hidden units and output units. From ANN structure, it can be easily understood that input units receive data from external sources to the network and send them to the hidden units, in turn, the hidden units send and receive data only from other units in the network, and output units receive and produce data generated by the network, which goes out of the system. In this process, a typical problem is to estimate the output as a function of the input. This unknown function may be approximated by a superposition of certain activation functions such as tangent, sigmoid, polynomial, and sinusoid in ANN. A common threshold function used in ANN is the sigmoid function (f(S)) expressed by Eq. (1), which provides an output in the range of 0<f(S)<1 [3].

\[ f(S) = \left[1 + \exp\left(-S_i\right)\right]^{-1} \]

where \( S_i \) is the characteristic function of \( i^{th} \) layer, \( I_i \) is the input unit of \( i^{th} \) layer, \( O_i \) is the output unit of \( i^{th} \) layer, \( W_{ij} \) is the synaptic weights between input and hidden layers, \( N \) is the number of observations and \( M \) is the number of neurons of hidden layer. The sigmoid function is chosen for mathematical convenience because it resembles a hard-limiting step function for extremely large positive and negative values of the incoming signal and also gives sufficient information about the response of the processing unit to inputs that are close to the threshold value.

Figure 1: Architecture of ANN

Number of networks such as Multi Layer Perceptron (MLP), Cascade Correlation (CCL), Conjugate Gradient (CGT), Radial basis function, Bayesian, etc are commonly used for training the network data [4]. The objective in training the network is to reduce the global error between the predicted and targeted outputs. From the research reports on ANN, it is understood that number of researchers has applied different networks for prediction of rainfall for various regions [5-12]. But there is no general agreement in applying particular network for rainfall prediction for a region though different networks are available for training the network data. In the present study, an attempt has been made to train the network data with MLP, CCL and CGT networks for prediction of rainfall at Fatehabad and Hansi. Model performance indicators
(MPIs) such as Correlation Coefficient (CC), Model Efficiency (MEF) and Root Mean Square Error (RMSE) are used to evaluate the performance of the models with a specific objective to identify the most suitable network for rainfall prediction. The procedures adopted in training the networks with MLP, CCL and CGT networks, and computation of MPIs is briefly described in the ensuing sections.

II. METHODOLOGY

Multi-Layer Perceptron Network (MLPN)

MLPN is the most widely used for rainfall prediction and its architecture with single hidden layer is shown in Figure 1. Gradient descent is the most commonly used supervised training algorithm in MLPN [13]. Each input unit of the training data set is passed through the network from the input layer to output layer. The network output is compared with the desired target output and output error (E) is computed using Eq. (2). This error is propagated backward through the network to each neuron, and the connection weights are adjusted based on Eq. (3).

\[ E = \frac{1}{2} \sum_{i=1}^{N} (P_i - P_i^*)^2 \]  \hspace{1cm} (2)

where \( P_i \) is the observed rainfall for \( i \)th year and \( P_i^* \) is the predicted rainfall for \( i \)th year.

\[ \Delta W_{ij}(M) = -\epsilon \frac{\partial E}{\partial W_{ij}} + \alpha \Delta W_{ij}(M-1) \]  \hspace{1cm} (3)

where \( W_{ij} \) is the synaptic weights between input and hidden layers, \( \Delta W_{ij}(M) \) is the weight increments between \( i \)th and \( j \)th units during \( M \) neurons (units) and \( \Delta W_{ij}(M-1) \) is the weight increments between \( i \)th and \( j \)th units during \( M-1 \) neurons. In MLPN, momentum factor (\( \alpha \)) is used to speed up training in very flat regions of the error surface to prevent oscillations in the weights and learning rate (\( \epsilon \)) is used to increase the chance of avoiding the training process being trapped in local minima instead of global minima [14].

Cascade Correlation Network (CCLN)

In MLPN, all weights change simultaneously while responding to a new training pattern. This process results in an unnecessary motion of the network leading to more efforts and time. This can be avoided in the CCLN by training only one layer of weights and keeping the rest of the weights unchanged [15]. In CCLN, hidden nodes are added one by one starting from zero during the training until the training termination criterion is reached. The CCLN does not involve learning by descending down the error gradient, but by maximizing the effect (or correlation) of the new hidden unit output on the residual error. It also does not involve transmission of the error as detailed in MLPN [16]. The training procedure of CCLN is as follows:

i) Consider only the input and output units.

ii) Train direct the input-output weights over the entire training set using delta rule. This process does not require back propagation through the hidden units.

iii) Add one hidden unit by following the separate procedure. Then freeze its input weights and train all output weights once again by using the delta rule.

(a) Consider the new hidden unit; (b) Connect it to all input units as well as to all other existing hidden units; and (c) Take all training data sets one by one and adjust the input weights of this new hidden unit after each training set. This adjustment is done so as to maximize the overall correlation between the new hidden unit value and the residual error.

iv) Repeat step (iii) until minimum error is reached or a specified maximum number of iterations are over.

Conjugate Gradient Network (CGTN)

CGTN differs from MLPN is gradient calculations and subsequent corrections to weights and bias. In CGTN, a search direction \( d_k \) is computed at each training iteration ‘k’ and the error function \( f(S) \) is minimized along it with the use of a line search. The gradient descent does not move down the error gradient as in the foregoing back propagation method but along a direction that is conjugate to the previous step. The change is gradient is thus taken as orthogonal to the previous step with the advantage that the function minimization carried out in each step, is fully preserved because of the lack of any interference from the subsequent steps [17][18]. The five-step iteration process is as follows:

i) Initialize weight vector \( P \) by using uniform random numbers from the interval (-0.5, 0.5). Calculate error gradient \( \nabla E \) by using back propagation at this point.

Select initial search direction \( d_0 = -\nabla E \).

ii) For each iteration ‘j’, determine constant \( \alpha_j \), which minimizes the error function \( f(P_j + \alpha_j d_j) \) by line search. Here \( d_j \) is the search direction at iteration ‘j’.

Update the weight factor \( P_{j+1} \) to \( P_j \) using \( P_{j+1} = P_j + \alpha_j d_j \).

iii) If the error at this iteration, ‘j+1’ is acceptable or if a specified number of computations of the function and gradients is reached, terminates the algorithm.

iv) Otherwise, get the new director vector \( d_{j+1} \) by using the error gradient \( \nabla E_{j+1} \) at the iteration ‘j+1’ such that \( d_{j+1} = -\nabla E_{j+1} \). If ‘j+1’ is an integral multiple of \( N \) (where \( N \) is the dimension of \( P \)); otherwise, \( d_{j+1} = -\nabla E_{j+1} + \beta_j d_j \) with \( \beta_j = (\frac{1}{\|\nabla E\|^2} \frac{d_j^T \nabla E}{d_j^T \nabla E_{j+1}}) \).

Here, \( \nabla E_{j+1} \) is the error gradient at iteration ‘j’.

v) Repeat the above steps for the next iteration.

Normalization of Data

By considering the nature of sigmoid function adopted in ANN, the training data set values are normalized between 0 and 1 by Eq. (4) and passed into the network [19]. After the completion of training, the output values are denormalized to provide the results in original domain.

\[ \text{NOR}(P_i) = \frac{P_i - \text{Min}(P)}{\text{Max}(P) - \text{Min}(P)} \]  \hspace{1cm} (4)

where \( \text{NOR}(P_i) \) is the normalized value of \( P_i \), \( \text{Min}(P) \) is the minimum value of \( P_i \) and \( \text{Max}(P) \) is the maximum value of \( P_i \).
Model Performance

The performance of predicted rainfalls using MLP, CCL and CGT networks are analyzed by CC, MEF and RMSE, and are:

\[
CC = \frac{\sum_{i=1}^{N}(P_i - \overline{P})(\hat{P}_i - \overline{\hat{P}})}{\left(\sum_{i=1}^{N}(P_i - \overline{P})^2 \sum_{i=1}^{N}((\hat{P}_i - \overline{\hat{P}})^2}\right)^{1/2}} \quad \text{… (5)}
\]

\[
MEF (\%) = \left(1 - \frac{\sum_{i=1}^{N}(P_i - \hat{P}_i)^2}{\sum_{i=1}^{N}(P_i - \bar{P})^2}\right) \times 100 \quad \text{… (6)}
\]

\[
RMSE = \left(\frac{1}{N}\sum_{i=1}^{N}(P_i - \hat{P}_i)^2\right)^{1/2} \quad \text{… (7)}
\]

where \(\overline{P}\) is the average observed rainfall and \(\overline{\hat{P}}\) is the average predicted rainfall [20].

III. APPLICATION

An attempt has been made to predict the rainfall at Fatehabad and Hansi using MLP, CCL and CGT networks. The normal annual rainfalls observed at Fatehabad and Hansi are 400 mm and 460 mm respectively. Fatehabad station is located between the latitude 29° 25' N and longitude 75° 38' E in Fatehabad district of Haryana. Similarly, Hansi station is located between the latitude 29° 06' N and longitude 75° 58' E in Hisar district of Haryana. Annual rainfall data observed at Fatehabad and Hansi rain-gauge stations for the period 1954-2005 are used. The data for the period 1954-1985 are used for training (TRG), data for the period 1986-1995 for validation (VAL) and data for the period 1996-2005 for cross-validation (CVL).

IV. RESULTS AND DISCUSSIONS

Statistical software, namely, SPSS Neural Connection was used to train the network data with different combinations of parameters to determine optimum network architecture of MLP, CCL and CGT networks for prediction of rainfall at Fatehabad and Hansi.

Prediction of Rainfall using ANN

For Fatehabad station, the parameters of \(\alpha=0.6\) and \(\varepsilon=0.07\) were used in optimising the network architecture of MLP. Similarly, the factors of \(\alpha=0.8\) and \(\varepsilon=0.10\) were used in optimising the MLP network architecture of Hansi. The optimum network architectures with model parameters were used to predict the rainfall at Fatehabad and Hansi. The model performance of MLP, CCL and CGT networks were evaluated by MPIs and given in Tables 1 and 2 for the stations under study.

Table 1: Comparison of MPIs using MLP, CCL and CGT Networks for Fatehabad

<table>
<thead>
<tr>
<th>Model Performance Indicators</th>
<th>MLP (1-18-1)</th>
<th>CCL (1-15-1)</th>
<th>CGT (1-25-1)</th>
</tr>
</thead>
<tbody>
<tr>
<td>TRG</td>
<td>VAL</td>
<td>CVL</td>
<td>TRG</td>
</tr>
<tr>
<td>CC</td>
<td>0.978</td>
<td>0.989</td>
<td>0.995</td>
</tr>
<tr>
<td>MEF (%)</td>
<td>95.0</td>
<td>99.0</td>
<td>98.6</td>
</tr>
<tr>
<td>RMSE (mm)</td>
<td>31.1</td>
<td>24.0</td>
<td>28.1</td>
</tr>
</tbody>
</table>

Table 2: Comparison of MPIs using MLP, CCL and CGT Networks for Hansi

<table>
<thead>
<tr>
<th>Model Performance Indicators</th>
<th>MLP (1-25-1)</th>
<th>CCL (1-20-1)</th>
<th>CGT (1-30-1)</th>
</tr>
</thead>
<tbody>
<tr>
<td>TRG</td>
<td>VAL</td>
<td>CVL</td>
<td>TRG</td>
</tr>
<tr>
<td>CC</td>
<td>0.978</td>
<td>0.988</td>
<td>0.991</td>
</tr>
<tr>
<td>MEF (%)</td>
<td>95.5</td>
<td>99.2</td>
<td>99.4</td>
</tr>
<tr>
<td>RMSE (mm)</td>
<td>30.3</td>
<td>21.6</td>
<td>19.1</td>
</tr>
</tbody>
</table>

From Table 1, it may be noted that: (i) RMSE on the predicted rainfall using CCL network is less than the corresponding values of MLP and CGT networks and the architecture of CCL is better suited to train the network data; (ii) RMSE on the predicted rainfall using CCL at VAL and CVL periods are about 23 mm; (iii) There is generally a good correlation between the observed and predicted rainfall using MLP, CCL and CGT networks, with CC values varying from 0.978 to 0.995 and (iv) MEF on rainfall prediction using CCL at TRG, VAL and CVL periods are varied from about 95% to 99%.

Similarly, from Table 2, it may be noted that (i) RMSE on the predicted rainfall using MLP network is less than the corresponding values of CCL and CGT networks and the architecture of MLP is better suited to train the network data; (ii) RMSE on the predicted rainfall using MLP at VAL and CVL periods are about 22 mm and 19 mm respectively; (iii) There is generally a good correlation between the observed and predicted rainfall using MLP, CCL and CGT networks, with CC values varying from 0.978 to 0.991 and (iv) MEF on rainfall prediction using MLP at TRG, VAL and CVL periods are varied from about 96% to 99%. Figures 2 and 3 give the plots of observed and predicted rainfalls (using MLP, CCL and CGT networks) for Fatehabad and Hansi respectively.
Analysis Based on Statistical Parameters

The statistical parameters such as mean, standard deviation (SD), skewness and kurtosis for the observed and predicted rainfalls using CCL network for Fatehabad and MLP network for Hansi, were computed and given in Table 3.

Table 3: Statistical parameters of observed and predicted rainfalls using CCL network for Fatehabad and MLP network for Hansi

<table>
<thead>
<tr>
<th>Statistical parameters</th>
<th>Fatehabad</th>
<th>Predicted rainfall using CCL network</th>
<th>Hansi</th>
<th>Predicted rainfall using MLP network</th>
</tr>
</thead>
<tbody>
<tr>
<td>Observed rainfall</td>
<td>TRG: 373.2, VAL: 338.8, CVL: 342.5</td>
<td>TRG: 366.6, VAL: 317.6, CVL: 344.2</td>
<td>TRG: 342.6, VAL: 284.6, CVL: 275.8</td>
<td>TRG: 338.6, VAL: 285.5, CVL: 275.9</td>
</tr>
<tr>
<td>Predicted rainfall</td>
<td>TRG: 338.8, VAL: 284.6, CVL: 275.8</td>
<td>TRG: 366.6, VAL: 317.6, CVL: 344.2</td>
<td>TRG: 342.6, VAL: 284.6, CVL: 275.8</td>
<td>TRG: 338.6, VAL: 285.5, CVL: 275.9</td>
</tr>
<tr>
<td>SD (mm)</td>
<td>226.9</td>
<td>317.6</td>
<td>225.0</td>
<td>317.5</td>
</tr>
<tr>
<td>Skewness</td>
<td>-0.246</td>
<td>-0.058</td>
<td>0.379</td>
<td>0.291</td>
</tr>
<tr>
<td>Kurtosis</td>
<td>-0.836</td>
<td>-1.498</td>
<td>-1.244</td>
<td>-0.323</td>
</tr>
</tbody>
</table>

From Table 3, it may be noted that the percentage of variation between the average predicted rainfall, with reference to average observed rainfall, at TRG, VAL and CVL periods, is about 0.5% to 2.0% for Fatehabad when
CCL is applied; and 0% to 1% for Hansi when MLP is applied. From the analysis of the results based on MPIs and statistical parameters, it is suggested that CCL network could be used for prediction of rainfall at Fatehabad and MLP network for Hansi.

V. CONCLUSIONS

The paper described the procedures involved in prediction of rainfall for using MLP, CCL and CGT networks for Fatehabad and Hansi. Performance analysis based on MPIs showed that CCL network with network architecture of 1-15-1 and MLP network with 1-25-1 are better suited for training the network data. The MPIs obtained from VAL and CVL periods indicated that CCL network is comparatively better than MLP and CGT for Fatehabad. The performance analysis also showed that MLP gives better results for Hansi. The results indicated that the RMSE on rainfall prediction vary from about 23 mm to 30 mm for Fatehabad when CCL is applied; and 19 mm to 30 mm when MLP is applied. The study showed that the MEF in rainfall prediction at TRG, VAL and CVL periods vary from 95% to 99% for the stations under study. The study showed that the CC given by CCL network at TRG, VAL and CVL periods are in the range of 0.978 to 0.995 for Fatehabad; and 0.978 to 0.991 for Hansi when MLP is applied. The paper presented that the percentages of variation on the average predicted rainfall, with reference to the average observed rainfall, are in the range of 0% to 2% for the stations under study. Based on performance analysis, the study suggested that CCL could be used for rainfall prediction for Fatehabad and MLP for Hansi. The plot of observed and predicted rainfalls using MLP, CCL and CGT networks at Fatehabad and Hansi are developed and presented in the paper.

ACKNOWLEDGMENTS

The author is grateful to the Director, Central Water and Power Research Station, Pune, for providing the research facilities to carry out the study. The author is thankful to India Meteorological Department, Pune, for making available the rainfall data used in the analysis.

REFERENCES