Comparison of Artificial Neural Networks ANN and Statistica in Daily Flow Forecasting

Mahdi Moharrampour, Mohammad Kherad Ranjbar, Nazi Abachi, Mehrnush Zoghi, Mohammad Reza Asadi Asad Abad

Young Researchers Club, Islamic Azad University, Buinzahra Branch, Buinzahra, Iran.
Same Technical and Vocational Training College, Karaj Branch, Islamic Azad University, Karaj, Iran.
Department of Mathematical, Buinzahra Branch, Islamic Azad University, Buinzahra, Iran.
Department of Architecture, Buinzahra Branch, Islamic Azad University, Buinzahra, Iran.

Mahdi Moharrampour, Mohammad Kherad Ranjbar, Nazi Abachi, Mehrnush Zoghi, Mohammad Reza Asadi Asad Abad; Comparison of artificial neural networks ANN and Statistica in daily flow forecasting

ABSTRACT

Two decades with the advent of methods based on artificial intelligence and genetics. Algorithm directly based upon different parameters to predict water engineering is highly developed. Accurate prediction of flow in rivers, always as one of the most important factors in safe and economic design of facilities and structures related to river water has been considered by researchers. In this study, the method of artificial neural network and statistica model, are used to forecast daily river flow in north of Iran and the results of these models are compared with Observed daily values. The observed data that are used in this study start from 1992 to 2010 in 18 year’s period (6550 days). The study of two sigmoid functions and tangent hyperbolic MLP model was used to train and eventually compared with the results of STATISTICA software, the results showed that the ability of the neural network model output better than the statistical software

Key words: water resources management, flow forecasting, artificial neural network (ANN), the statistical software, Gharasoo River.

Introduction

The foundation of Support Vector Machines (SVM) was given by Vapnik, a Russian mathematician in the early 1960s, based on the Structural Risk Minimisation principle from statistical learning theory and gained popularity due to its many attractive features and promising empirical performance. SVM has been proved to be effective in classification by many researchers in many different fields such as electric and electrical engineering, civil engineering, mechanical engineering, medical, financial and others. Recently, it has been extended to the domain of regression problems. In the river flow modelling field, Liong & Sivapragasam compared SVM with Artificial Neural Networks (ANN) and concluded that SVM’s inherent properties give it an edge in overcoming some of the major problems in the application of ANN in addition to Due to the complexity of the methods like artificial intelligence (ANN) and Support vector machine (SVM), simpler methods with much more efficiency can be used in some initial studies. In this study, statistica software was used too for the first time in order to predict the daily discharge. This paper compares two expert models in daily flow forecasting. The support vector machine (SVM) and statistica model, are used to forecast daily river flow in north of Iran and the results of these models are compared with Observed daily values Gharasoo River is the case study and Gharasoo river data is used for this article.

Case study area and data:

Gharasoo River basin is in Golestan province, northeast of Iran. This basin is located (54) to (54-45) east latitude and (36-36) to (36-59) north longitude. Basin area is 1678.1 km2. Maximum height of this basin is 3200 meters from sea level and the length of Main River is 108.005 km. Fig. 1 shows the natural plan and location of Gharasoo River.

Corresponding Author

Mahdi Moharrampour, Young Researchers Club, Islamic Azad University, Buinzahra Branch, Buinzahra, Iran.
E-mail: m62.mahdi@yahoo.com
Fig. 1: Natural plan and location of GHRASOO River.

More than 4 rain measurement stations are existed over this river, but because of lack of statistics for all stations, in this research 4 stations are used. Gharasoo station as exit discharge of this basin and Ziarat, Shastkalateh and Kordkooy as input of this basin in three different locations.(Table 1).

Table 1: Specification of Gharasoo basin stations.

<table>
<thead>
<tr>
<th>Province</th>
<th>Code</th>
<th>Location</th>
<th>River</th>
<th>Longitude</th>
<th>Latitude</th>
</tr>
</thead>
<tbody>
<tr>
<td>Golestan</td>
<td>12-050</td>
<td>Gharasoo</td>
<td>Gharasoo</td>
<td>54-03-00</td>
<td>36-50-00</td>
</tr>
<tr>
<td>Golestan</td>
<td>12-043</td>
<td>Naharkhoran</td>
<td>Ziarat</td>
<td>54-28-00</td>
<td>36-46-00</td>
</tr>
<tr>
<td>Golestan</td>
<td>12-045</td>
<td>Shastkalate</td>
<td>Shastkalate</td>
<td>54-20-00</td>
<td>36-45-00</td>
</tr>
<tr>
<td>Golestan</td>
<td>12-049</td>
<td>Ghaz mahalle(pole jadde)</td>
<td>Kordkooy</td>
<td>54-05-00</td>
<td>36-47-00</td>
</tr>
</tbody>
</table>

Preprocessing data:

Preprocessing of data includes selection of effective variable, selection of training and test patterns and normalizing the patterns. The goal of normalizing is that all values in one pattern are in a range. Pattern normalizing exchanges all values to a specified interval such as [0-1] or [-1-1].

After normalizing all patterns, period of case study was selected between 1989 to 2007 (18 years). For this period, there are 6550 daily patterns for every station. 75% of these data are used for support vector machine training and 25% of these data are used for test. Fig. 2 shows daily flow hydrograph of Gharasoo River for training period and Fig. 3 shows daily flow hydrograph of Gharasoo River for test period.

Fig. 2: Daily flow hydrograph of Gharasoo River for training period.
Designed and developed simulation models using artificial neural network:

To determine the neural network architecture also consider having the software defaults Qnet2000, a process of trial and error in determining the number of hidden layer neurons and the threshold is also used.

Moreover, avoiding the phenomenon over fitness number of training cycles is determined by trial and error with over fitness could be two reasons Number of training cycles required and the other one too bulky structure of neural network As was stated in both cases to avoid trial and error method is used .In this study, the neural network training function of the t sigmoid functions and tangent hyperbolic in the MLP, the acceptable application performance in similar processes Is used and the results are compared To determine the best number of hidden layer neurons is required The root mean square error of network output Count the number of neurons for each hidden layer neurons is selected in a drawing graphs Root mean square error is the lowest number of neurons as the number of neurons in the hidden layer is selected To determine the number of hidden layers in a network similar to this should be done. It is worth mentioning that to find the rate and momentum coefficient of tried-and-error method is used In the methods of teaching - oriented, the number and type of model input parameters is important Since the structure of the neural network input, there is a constant and uniform The results presented in the articles will help Accordingly, the following five input pattern has been studied:

\[ Q(t) = f\{P_g(t), P_n(t), P_{ab}(t), P_p(t), \]
\[ 1) P_g(t-1), P_n(t-1), P_{ab}(t-1), \]
\[ 2) Q(t-1), Q(t-2) \}

\[ Q(t) = f\{Q_n(t), Q_n(t-1), Q_p(t), \]
\[ 2) Q(t-1), Q(t-2) \}

\[ Q(t) = f\{P_g(t), P_n(t), P_{ab}(t), P_p(t), \]
\[ 3) P_p(t), P_g(t-1), P_n(t-1), P_{ab}(t-1), \]
\[ P_p(t-1), Q(t-1), Q(t-2) \}

\[ Q(t) = f\{Q(t-1), Q(t-2) \}

\[ Q(t) = f\{Q(t-1), Q(t-2) \}

\[ Q(t) = f\{P_g(t), P_n(t), P_{ab}(t), P_p(t), \]
\[ 4) P_p(t-1), P_p(t-1), P_{ab}(t-1), \]
\[ P_p(t-1) \}

\[ Q(t) = f\{Q(t-1), Q(t-2) \}

\[ Q(t) = f\{Q(t-1), Q(t-2) \}

In these equations:

Q: Daily average discharge of Gharasoo station
Q_n: Daily average discharge of Naharkhoran station
Q_p: Daily average discharge of Polejadde station
P_n: Daily average rainfall of Naharkhoran station
P_{ab}: Daily average rainfall of Shastkalateh station
P_p: Daily average rainfall of Polejaddeh station
P_g: Daily average rainfall of Gharasoo station

The other patterns were built near each of their results with one of the top models have not gained, and here the expression patterns .Here is the question raised is which of the five models in the best performance will determine the daily discharge To answer this question with the two input patterns, sigmoid functions and tangent hyperbolic is evaluated .In the present study to compare the performance of models of the correlation coefficient R2 is used.
Table 2: Review of ANN performance results (sigmoid).

<table>
<thead>
<tr>
<th>Input pattern</th>
<th>Pattern1</th>
<th>Pattern1</th>
<th>Pattern1</th>
<th>Pattern1</th>
<th>Pattern1</th>
</tr>
</thead>
<tbody>
<tr>
<td>neuron number in input layer</td>
<td>14</td>
<td>6</td>
<td>10</td>
<td>2</td>
<td>8</td>
</tr>
<tr>
<td>neuron number in hidden layer</td>
<td>5</td>
<td>4</td>
<td>4</td>
<td>2</td>
<td>7</td>
</tr>
<tr>
<td>neuron number in output layer</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>R²</td>
<td>0.9314</td>
<td>0.9152</td>
<td>0.9246</td>
<td>0.9074</td>
<td>0.3049</td>
</tr>
<tr>
<td>RMSE</td>
<td>0.02423</td>
<td>0.02646</td>
<td>0.02492</td>
<td>0.02749</td>
<td>0.06227</td>
</tr>
</tbody>
</table>

Table 3: Review of ANN performance results (tangent hyperbolic).

<table>
<thead>
<tr>
<th>Input pattern</th>
<th>Pattern1</th>
<th>Pattern1</th>
<th>Pattern1</th>
<th>Pattern1</th>
<th>Pattern1</th>
</tr>
</thead>
<tbody>
<tr>
<td>neuron number in input layer</td>
<td>14</td>
<td>6</td>
<td>10</td>
<td>2</td>
<td>8</td>
</tr>
<tr>
<td>neuron number in hidden layer</td>
<td>5</td>
<td>4</td>
<td>4</td>
<td>2</td>
<td>7</td>
</tr>
<tr>
<td>neuron number in output layer</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>R²</td>
<td>0.9287</td>
<td>0.9148</td>
<td>0.9247</td>
<td>0.9091</td>
<td>0.3045</td>
</tr>
<tr>
<td>RMSE</td>
<td>0.02466</td>
<td>0.02651</td>
<td>0.02489</td>
<td>0.02725</td>
<td>0.06233</td>
</tr>
</tbody>
</table>

What the tables (2) and (3) can be concluded that the model trained with the first input pattern with 14 provinces sigmoid for the input parameters (1-5-14) means 14 input neurons, 5 hidden neurons and Way that the results of the model trained with a threshold function tangent hyperbolic Has created a better and less error, the predicted peak discharge operation is successful.

Predicting the course of the river by using software statistica:

By using the data related to the structure of discharge proposed exit point with the software and statistica had predicted the results with the results of the structure proposed by comparison, svm and just as in the table (2) is determined by the structure proposed by statistica software performance as a result.

Figure 3 and 4 shows the observed discharge and predicted discharge with statistica software. Figure 5 shows the results of both observed and predicted discharge. It shows that predicted maximum discharge is lower than observed discharge[2].

Fig. 4: Mean square error curve with respect to the number of hidden neurons.

Fig. 5: Comparison of model outputs in the training and testing (first model).
Fig. 6: Comparison of model outputs in the training and test data with best-fit line (first model).

Table 2: Results of statistical function.

<table>
<thead>
<tr>
<th>Ways</th>
<th>RMSE</th>
</tr>
</thead>
<tbody>
<tr>
<td>statistica</td>
<td>0.095025</td>
</tr>
</tbody>
</table>

**Fig. 3:** Daily discharge hydrograph of Gharasoo Station (Observed).

**Fig. 4:** Daily discharge hydrograph of Gharasoo Station (Predicted).
Fig. 5: Daily discharge hydrograph of Gharasoo Station (Observed and Predicted).

Conclusion:

1. five input pattern is introduced, the two patterns in both the sigmoid functions and tangent hyperbolic, provided acceptable results.
   The first pattern in both the sigmoid functions and tangent hyperbolic less error than the other four patterns of established and most successful model was among the five models.
2. Sigmoid the results of the tangent hyperbolic with higher correlation coefficient and the RMSE is less.
3. In the prediction of river flow and rainfall the day before and two days before the day Gharah Station plays a fundamental role in the model, so the results can be seen as.
4. If the minimum error in the network model and if we consider the minimum error and minimum parameters, model 4 is the best model.
5. The rate of learning and distance education had an important effect on network performance.
6. The structure of the proposed structure, use patterns, the MLP network with 14 input parameters as Sigmoid provinces (1-5-14) means 14 input neurons, 5 hidden neurons and output neurons, compared with a much better software STATISTICA.

According to what mentioned above, we can say that the Sigmoid function as network transfer function is a reasonable thing, because it predicts that tangent hyperbolic as well as the function is successful software STATISTICA.

References