FLOOD STAGE FORECASTING USING CLASS SEGREGATION METHOD BASED ON NEURAL NETWORKS MODELS

SUNGWON KIM

Ph.D./P.E., Associate Professor, Department of Railroad and Civil Engineering, Dongyang University, 1 Kyochon-dong, Yeongju, 750-711, Republic of Korea
(phone: +82-54-630-1241, fax: +82-54-637-8027, e-mail: swkim1968@dyu.ac.kr)

The new methodology which combines Kohonen self-organizing map neural networks model (KSOM-NNM) and the conventional neural networks models such as feedforward neural networks model (FFNNM) and generalized neural networks model (GRNNM) is introduced to forecast flood stage in the Nakdong River, Republic of Korea. It is possible to train without output data in KSOM-NNM. KSOM-NNM is used to classify the input data before it combines with the conventional neural networks model. Four types of models such as SOM-FFNNM-BP, SOM-GRNNM-GA, FFNNM-BP and GRNNM-GA are used to train and test performances respectively. From the statistical analysis for training and testing performances, SOM-GRNNM-GA shows the best results compared with the other models such as SOM-FFNNM-BP, FFNNM-BP and GRNNM-GA. Furthermore, FFNNM-BP shows vice-versa. From this research, the new methodology can be suggested to forecast flood stage and construct flood warning system in river basin.

Keywords: flood stage forecasting, KSOM-NNM, GRNNM, FFNNM.

INTRODUCTION

Flood forecasting and warning is one of the nonstructural measures receiving increased and sustained attention from hydrologists and the public. Flood forecasting can provide important information for the water resources managements such as forewarning and proper reservoir regulation during the low flows (Thirumalaiah and Deo [17]). Flood stage can be forecasted continuously by direct or indirect method. Direct methods use statistical correlation techniques (Mutreja et al. [11]), and indirect methods widely use the physical-based rainfall-runoff models for the flood stage forecasting. The forecasted runoff is later converted to a flood stage using a rating curve. For this purpose, an accurate rating curve is essential (Kim and Salas [8]). The physical-based rainfall-runoff models rely on mathematical descriptions of the physical processes that take place in a watershed and river basin. These models employ some form of the continuity and momentum equations in differential forms that are applied and integrated throughout the different components of the models. Furthermore, the advent of advanced computer technology, geographical information systems, and digital elevation models has made it possible to apply the physical-based rainfall-runoff models efficiently in a distributed context (Rosso et al. [13]).
In this study, the specific operation method, Kohonen self-organizing map (KSOM) and the conventional neural networks model is coupled together, is developed to forecast flood stage at Jindong station of Nakdong River, Republic of Korea. Flood stage data at Koryeonggyo and Jeokpogyo stations are used for the input nodes of neural networks models. The optimal parameters of neural networks models can be determined using training performance, and the validation of the neural networks models can be carried out using test performance, respectively. From this study, the new methodology can be suggested to forecast flood stage and construct flood warning system in river basin.

NEURAL NETWORKS MODELS AND GENETIC ALGORITHM

Kohonen self-organization map neural networks model
The Kohonen self-organizing map neural networks model (KSOM-NNM) performs a mapping from a continuous input space to a discrete output space, preserving the topological properties of the input nodes. This means that points close to each other in the input space are mapped to the same or neighboring processing elements (PEs) in the output space. The idea of soft competition is crucial to understanding the function of KSOM-NNM. The connection weights from the input layer to the output/Kohonen layer perform association between connection weights and input nodes. The PE whose the connection weights vector is closest the present input node wins the competition. In KSOM-NNM, not only the winner of competition but also its neighborhood has their connection weights updated according to the competitive rule. To simplify the computation of the computation of the algorithm, the lateral inhibition networks is assumed to produce Gaussian distribution centered at the winning PE. Thus, instead of recursively computing the activity of each PE, we simply find the winning PE and assume the other PE have an activity proportional to the Gaussian function evaluated at each PE’s distance from the winner (Kohonen [9] & [10]; Tsoukalas and Uhrig [18]; Islam and Kothari [4]; Bowden et al. [2]). Figure 1 shows the developed KSOM-NNM structure having two input nodes in the input layer and 4-by-4 PEs in the Kohonen layer.

Figure 1. The developed KSOM-NNM structure

Figure 2. A schematic diagram of the watershed
Generalized regression neural networks model
The generalized regression neural networks model (GRNNM), which is applied to this study, is the modified forms of radial basis function neural networks model (RBFNNM). GRNNM is composed of four layers, that is, input layer, hidden layer, summation layer, and output layer. And GRNNM is the neural networks model based on the nonlinear regression theory. Input layer, hidden layer, and summation layer nodes are connected completely, whereas output layer node is connected with only some of summation layer nodes. Each output layer node is connected with the summation node and division node of summation layer, and the connection weight is not composed between summation layer and output layer (Kim and Kim [6]; Kim et al. [7]; Specht [15]; Tsoukalas and Uhrig [18]; Wasserman [19]).

Feedforward neural networks model
The feedforward neural networks model (FFNNM) has an input layer, an output layer, and one or more hidden layers between the input and output layers. Each of the nodes in a layer is connected to all the nodes of the next layer, and the nodes in one layer are connected only to the nodes of the immediate next layer. The strength of the signal passing from one node to the other depends on the connection weights of the interconnections. The hidden layers enhance the network’s ability to model complex functions (Simpson [14]; Haykin [3]).

Genetic algorithm
In this study, GA is used to determine not only the proper individual smoothing factor but also the overall smoothing factor for each input layer node. The training performance of GRNNM using the GA progresses largely in two parts. The first part uses training data to train GRNNM. The second part uses the developed GRNNM in the first part to test the entire range of the smoothing factor for the optimal operation over the testing data. And, GA can produce GRNNM which can be operated best over the testing data. When the neural networks model is trained using GA, however, it requires much time than the BP as search method (Neuroshell 2 [12]).

RESEARCH DATA AND SCOPE
To apply KSOM-NNM, GRNNM and FFNNM, Jindong station is selected for the output node, and Koryeonggyo and Jeokpogyo stations of the Nakdong River are selected for the input nodes. Flood stage data for 4 credible flood stage events that occurred between 1990 and 2005 are collected from the Internet homepage of water management information system (www.wamis.go.kr). A schematic diagram of the watershed, which has an area 23,656.3 km², is shown in Figure 2. Flood stage data are prepared from the measurements made at 1-hour intervals at Koryeonggyo, Jeokpogyo and Jindong stations. A summary of the flood stage data used in this study is shown in Table 1. Flood stage event 1 ~ 3 are used for the training performance to parameterize the neural networks
models. Flood stage event 4 is used for the test performance to validate the neural networks models.

Table 1. A summary of the flood stage data

<table>
<thead>
<tr>
<th>Station</th>
<th>Date</th>
<th>Total Hour (hr)</th>
<th>Hourly Maximum Stage (m)</th>
<th>Remarks</th>
</tr>
</thead>
<tbody>
<tr>
<td>Koryeonggyo</td>
<td>08/22/1991 – 08/31/1991</td>
<td>240</td>
<td>8.35</td>
<td>Training</td>
</tr>
<tr>
<td></td>
<td>07/11/1993 – 07/20/1993</td>
<td>240</td>
<td>8.21</td>
<td>Training</td>
</tr>
<tr>
<td></td>
<td>08/07/1993 – 08/16/1993</td>
<td>240</td>
<td>9.05</td>
<td>Training</td>
</tr>
<tr>
<td>Jeokpogyo</td>
<td>08/22/1991 – 08/31/1991</td>
<td>240</td>
<td>8.34</td>
<td>Training</td>
</tr>
<tr>
<td></td>
<td>07/11/1993 – 07/20/1993</td>
<td>240</td>
<td>7.90</td>
<td>Training</td>
</tr>
<tr>
<td></td>
<td>08/07/1993 – 08/16/1993</td>
<td>240</td>
<td>8.97</td>
<td>Training</td>
</tr>
<tr>
<td></td>
<td>09/11/2003 – 09/20/2003</td>
<td>240</td>
<td>11.82</td>
<td>Testing (Typhoon MAEMI)</td>
</tr>
<tr>
<td></td>
<td>07/11/1993 – 07/20/1993</td>
<td>240</td>
<td>8.01</td>
<td>Training</td>
</tr>
<tr>
<td></td>
<td>08/07/1993 – 08/16/1993</td>
<td>240</td>
<td>9.01</td>
<td>Training</td>
</tr>
<tr>
<td></td>
<td>09/11/2003 – 09/20/2003</td>
<td>240</td>
<td>11.70</td>
<td>Testing (Typhoon MAEMI)</td>
</tr>
</tbody>
</table>

APPLICATION OF NEURAL NETWORKS MODELS

Training performance

In this study, three types of methodology for neural networks models development are selected. First, KSOM-NNM pre-processing and the observed data are used for the pre-processing method of training performances, respectively. Second, GRNNM and FFNNM are used for the training performances of conventional neural networks models. Third, backpropagation (BP) and genetic algorithm (GA) are used for the training performances. Therefore, the four types of neural networks models are classified into SOM-FFNNM-BP, SOM-GRNNM-GA, FFNNM-BP and GRNNM-GA, respectively. This study set up, as the training data of SOM-FFNNM-BP, SOM-GRNNM-GA, FFNNM-BP and GRNNM-GA, flood stage event 1 ~ 3 data among the hourly data from 1990 to 2005. So, the total number of data used for the training performance is composed of 720 time series data
The statistical index for examining the performance of SOM-FFNNM-BP, SOM-GRNNM-GA, FFNNM-BP and GRNNM-GA are correlation coefficient (CC), root mean square error (RMSE), Nash-Sutcliffe coefficient (E) and mean absolute error (MAE). And, learning rate is fixed as 0.5, the initial connection weight is fixed as 0.3, the maximum iteration number is fixed as 10,000, and the neighborhood size is fixed as 3 for training performance of KSOM-NNM pre-processing method. And, learning rate and momentum is fixed as 0.1, the initial connection weight is fixed as 0.3, the maximum iteration number is fixed as 5,000, and the training threshold is fixed as 0.001 for training performance of FFNNM-BP.

**Testing performance**

This study set up, as the testing data of SOM-FFNNM-BP, SOM-GRNNM-GA, FFNNM-BP and GRNNM-GA, flood stage event 4 data among the hourly data from 1990 to 2005. So, the total number of data used for the testing performance is composed of 240 time series data. In this study, the testing performance applied the cross-validation method to overcome the overfitting problem of SOM-FFNNM-BP, SOM-GRNNM-GA, FFNNM-BP and GRNNM-GA. The cross-validation method (Stone [16]) is not to train all the training data until SOM-FFNNM-BP, SOM-GRNNM-GA, FFNNM-BP and GRNNM-GA reach the minimum RMSE, but is to cross-validate with the testing data at the end of each training stage. Generally, the maximum 40% of the total training data are used as the testing data. If the over-fitting problem occurs, the convergence process over the mean square error of testing data will not decrease but will increase as the training data are still trained (Bishop [1]; Kim and Kim [5]). Figure 3 (a)-(b) shows the comparison of the results of training and testing performances for flood stage forecasting at Jindong station. In case of Figure 3 (b), we can consider that there is much error in the peak flow forecasting even if the best SOM-GRNNM-GA. And, Table 2 shows the statistical analysis of training and testing performances for flood stage forecasting at Jindong station.

![Figure 3. (a) Comparison of flood stage hydrograph (Jindong, Training data) ; (b) Comparison of flood stage hydrograph (Jindong, Testing data)]
Table 2. The statistical analysis of training and testing performances for flood stage forecasting (Jindong station)

<table>
<thead>
<tr>
<th>Station</th>
<th>Statistical Index</th>
<th>Training</th>
<th>Testing</th>
<th>Training</th>
<th>Testing</th>
<th>Training</th>
<th>Testing</th>
<th>Training</th>
<th>Testing</th>
</tr>
</thead>
<tbody>
<tr>
<td>Jindong</td>
<td>CC</td>
<td>0.991</td>
<td>0.980</td>
<td>0.997</td>
<td>0.976</td>
<td>0.974</td>
<td>0.977</td>
<td>0.996</td>
<td>0.977</td>
</tr>
<tr>
<td></td>
<td>RMSE (m)</td>
<td>0.312</td>
<td>0.969</td>
<td>0.172</td>
<td>1.034</td>
<td>0.625</td>
<td>1.007</td>
<td>0.190</td>
<td>1.005</td>
</tr>
<tr>
<td></td>
<td>E</td>
<td>0.979</td>
<td>0.908</td>
<td>0.994</td>
<td>0.895</td>
<td>0.916</td>
<td>0.901</td>
<td>0.992</td>
<td>0.901</td>
</tr>
<tr>
<td></td>
<td>MAE (m)</td>
<td>0.227</td>
<td>0.649</td>
<td>0.107</td>
<td>0.659</td>
<td>0.449</td>
<td>0.786</td>
<td>0.146</td>
<td>0.799</td>
</tr>
</tbody>
</table>

CONCLUSIONS AND FUTURE RESEARCH TASKS

In this study, the specific neural networks models are developed and applied to forecast flood stage. For the pre-processing method of training performance, KSOM-NNM is applied to classify the flood stage data. From this study, SOM-GRNNM-GA shows the best results compared with the other models such as SOM-FFNNM-BP, FFNNM-BP and GRNNM-GA. Furthermore, FFNNM-BP shows vice-versa. However, there is some weakness to have fair results in the scientific evaluation. The lag-time distribution cannot be considered and the only two stations are applied for the input nodes of neural networks models for flood stage forecasting. Furthermore, the only 4 flood stage events are used to train and test the neural networks models. If lag-time distribution, many input stations, and many flood stage events can be considered, the results of this study can be suggested the different ones. The training performance of KSOM-NNM pre-processing method, however, can be derived the best results for flood stage forecasting, and the new methodology can be considered for flood stage forecasting.

REFERENCES


