Lecture Notes in Artificial Intelligence, Vol. 4031, 2006, pp. 111-116

Evaluation of Several Algorithms in Forecasting Flood

C.L. Wu[1] and K.W. Chau[1]

[1]Department of Civil and Structural Engineering, Hong Kong Polytechnic University,
Hunghom, Kowloon, Hong Kong, People's Republic of China
cekwchau@polyu.edu.hk

Abstract. Precise flood forecasting is desirable so as to have more lead time for taking appropriate prevention measures as well as evacuation actions. Although conceptual prediction models have apparent advantages in assisting physical understandings of the hydrological process, the spatial and temporal variability of characteristics of watershed and the number of variables involved in the modeling of the physical processes render them difficult to be manipulated other than by specialists. In this study, two hybrid models, namely, based on genetic algorithm-based artificial neural network and adaptive-network-based fuzzy inference system algorithms, are employed for flood forecasting in a channel reach of the Yangtze River. The new contributions made by this paper are the application of these two algorithms on flood forecasting problems in real prototype cases and the comparison of their performances with a benchmarking linear regression model in this field. It is found that these hybrid algorithms with a "black-box" approach are worthy tools since they not only explore a new solution approach but also demonstrate good accuracy performance.

1 Introduction

Numerical models for flood propagation in a channel reach can broadly be classified into two main categories: conceptual models [1-5]; and, empirical models based on system analysis or "black-box" approach. Huge amount of data are usually required for calibration of these conceptual models. In many cases, a simple "black-box" model may be preferred in identifying a direct mapping between inputs and outputs. During the past decade, several nonlinear approaches, including artificial neural network (ANN), genetic algorithm (GA), and fuzzy logic, have been employed to solve flood forecasting problems. Smith and Eli [6] applied a back-propagation ANN model to predict discharge and time to peak over a hypothetical watershed. Tokar and Johnson [7] compared ANN models with regression and simple conceptual models. Liong *et al.* [8] employed an ANN approach for river stage forecasting in Bangladesh. Cheng and Chau [9] employed fuzzy iteration methodology for reservoir flood control operation. Chau and Cheng [10] performed a real-time prediction of water stage with ANN approach using an improved back propagation algorithm. Chau [11] calibrated flow and water quality modeling using GA. Cheng *et al.* [12] combined a fuzzy optimal model with a genetic algorithm to solve multiobjective rainfall-runoff model calibration. Chau [13-14] performed river stage forecasting and rainfall-runoff correlation with particle swarm optimization technique. Cheng *et al.* [15] carried out long-term prediction of discharges in Manwan Reservoir using ANN models.

In this paper, two hybrid algorithms, namely, genetic algorithm-based artificial neural network (ANN-GA) and adaptive-network-based fuzzy inference system (ANFIS), are applied for flood forecasting in a channel reach of the Yangtze River. To the knowledge of the authors, these types of algorithms have never been applied to hydrological and water resources problems. The new contributions made by this paper are the application of these two algorithms on flood forecasting problems in real prototype cases and the comparison of their performances with a benchmarking linear regression (LR) model in this field.

2 Genetic Algorithm-Based Artificial Neural Network (ANN-GA)

A hybrid integration of ANN and GA, taking advantages of the characteristics of both schemes, may be able to increase solution stability and improve performance of an ANN model. A genetic algorithm-based artificial neural network (ANN-GA) model is developed here wherein a GA [16] is used to optimize initial parameters of ANN before trained by conventional ANN. In the GA sub-model, the objective function used for initializing weights and biases is represented as follows:

$$\min J(W,\theta) = \sum_{i=1}^{p} \left| Y_i - f(X_i, W, \theta) \right| \tag{1}$$

where W is the weight, θ is the bias or threshold value, i is the data sequence, p is the total number of training data pairs, X_i is the ith input data, Y_i is the ith measured data, and $f(X_i, W, \theta)$ represents simulated output. The main objective of the sub-model is to determine optimal parameters with minimal accumulative errors between the measured data and simulated data.

3 Adaptive-Network-Based Fuzzy Inference System (ANFIS)

In this study, the output of each rule is taken as a linear combination of input variable together with a constant term. The final output is the weighted averaged of each rule's output. The fuzzy rule base comprises the combinations of all categories of variables. As an illustration, the following shows a case with three input variables and a single output variable. Each input variable (x, y, and z) is divided into three categories. Equally spaced triangular membership functions are assigned. The categories are assigned: "low," "medium," and "high." The number of rules in a fuzzy rule base is c^n, where c is the number of categories per variable and n the number of variables. The optimal number of categories is obtained through trials and performance comparison. The format of the rule set contains an output $o_{i,j,k}$ for a combination of category i of input variable x, category j of input y, and category k of input variable z, respectively.

If a rule is triggered, the corresponding memberships of x, y, and z will be computed. The weight $w_{i,j,k}$ to be assigned to the corresponding output $o_{i,j,k}$ will be furnished by the result of a specific T-norm operation. Multiplication operation is adopted here. A single weighted average output will then be acquired by combining the outputs from all triggered rules as follows:

$$o = \frac{\sum w_{i,j,k} \cdot o_{i,j,k}}{\sum w_{i,j,k}} \tag{2}$$

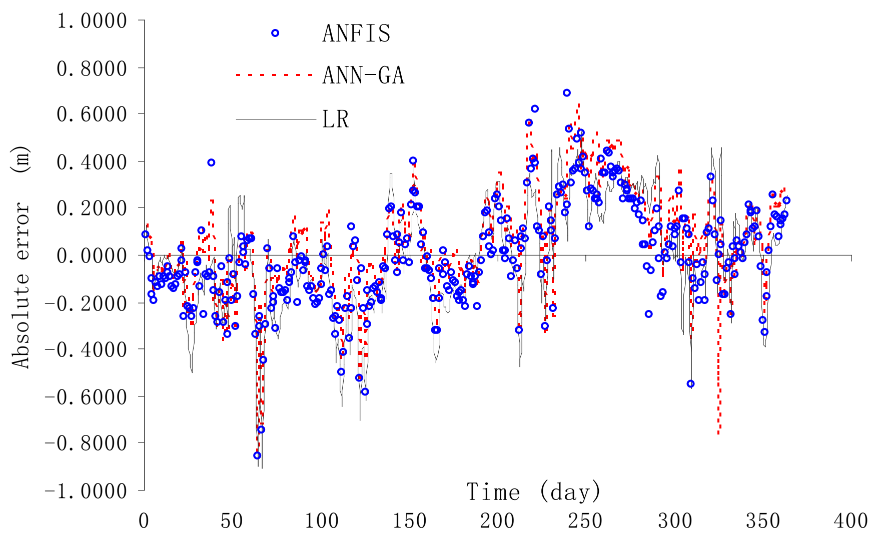

Fig. 1. Performance comparison in terms of absolute errors for different algorithms

For this flood forecasting model, some parameters, including each triangular membership function and the consequence part of each rule, have to be obtained through learning by ANN. The algorithm is able to enhance the intelligence when working in uncertain, imprecise, and noisy environments and to accomplish faster convergence. It possesses the characteristics of both the neural networks, including learning abilities, optimization abilities, and connectionist structures, and the fuzzy control systems, including human like "if-then" rule thinking and ease of incorporating expert knowledge, etc. In this system, the parameters defining the shape of the membership functions and the consequent parameters for each rule are determined by the back-propagation learning algorithm and the least-squares method, respectively.

4 Application Case

The studied channel reach from Luo-Shan to Han-Kou is located at the middle of the Yangtze River. The water elevation at Luo-Shan station ranges from 17.3m during the non-flooding period to 31.0m during the flooding period whilst the mean levels are 20.8m and 27.1m during the non-flooding and flooding periods, respectively. The key objective of this study is to forecast water stages of the downstream station, Han-Kou, on the basis of its counterparts at the upstream station, Luo-Shan.

For the ANN-GA model, a three-layer network is adopted with three input nodes and one output node. As an initial data preprocessing, the input and output data are normalized to be ranging between 0 and 1, corresponding to the minimum and the maximum water stages, respectively. ANN-GA models are trained with different number of nodes in the hidden layer so as to determine the optimal network geometry for these data sets. A testing set is incorporated so as to avoid the overfitting problem. Training is stopped when the error learning curve of the testing set starts to increase whilst that of the training set is still decreasing. It is found that, amongst them, the architecture with 3 nodes in the hidden layer is the optimal.

For an ANFIS model, more number of categories will furnish higher accuracy, but at the same time will have the disadvantages of larger rule bases and higher computation cost. Trial and error procedure is performed with a view to selecting the appropriate number of variable categories. Careful treatment is also made to avoid overfitting, though it is anticipated that more subspaces for the ANFIS model might result in better performance. An optimal number of categories of 3 is adopted, after having taken into consideration of the computational time, root mean square error in training (RMSE_tra), and root mean square error in validation (RMSE_vali).

Table 1. Performance comparison for different models in flood prediction

Models	RMSE_tra (m)	RMSE_vali (m)	Training time (s)	Number of parameter
LR	0.238	0.237	Nil	4
ANN-GA	0.213	0.226	135	16
ANFIS	0.204	0.214	49	135

5 Results and Analysis

The performance comparison of the LR, ANN-GA, and ANFIS models in forecasting 1-day lead time water levels at Han-Kou on the basis of the upstream water levels at Luo-Shan station during the past three days is shown in Figure 1. The fluctuation of absolute error is the largest for the LR model and is smallest for the ANFIS model. Table 1 shows the performance comparison using RMSE_tra, RMSE_vali, training time, and number of parameters. The ANFIS model is able to attain the highest accuracy, yet requires less training time than ANN-GA model. However, it should be noted that the ANFIS model involves more number of parameters than the other two models.

Their differences in performance can be explained somewhat by the fact that the LR model can only fit a linear function to input-output data pairs whilst both the ANN-GA and ANFIS models can contort themselves into complex forms in order to handle non-linear problems. It is justifiable that an ANN-GA model with 16 parameters is more flexible than LR model with 4 parameters since the coupling of ANN and GA can take advantage of the local optimization of ANN and the global optimization of GA. The results indicate that the local approximation approach of the ANFIS model has better performance in mapping the connectivity of input-output data pairs than the global approximation approach of the ANN-GA model. More importantly, the ANN-GA model entails more training time than the ANFIS model due to the time consuming searching nature of GA. Nevertheless, with the recent rate of development of computer technology, it will not be a major constraint. As such, it is trusted that hybrid algorithms, including ANN-GA and ANFIS, will have significant potentials as alternatives to conventional models in solving hydrological problems.

6 Conclusions

In this paper, two hybrid "black-box" models are applied for real flood forecasting. Both ANN-GA and ANFIS models are able to produce accurate flood predictions of the channel reach between Luo-Shan and Han-Kou stations in the Yangtze River. Amongst them, the ANFIS model, having the characteristics of both ANN and FIS, is the optimal in terms of the simulation performance, yet requires a larger amount of parameters in comparison with the benchmarking LR model. The ANN-GA model adequately combines the advantage of ANN with the advantage of GA, yet consumes most computation cost. Both ANN-GA and ANFIS models could be considered as feasible alternatives to conventional models. The new contributions made by this paper are the application of these two algorithms on flood forecasting problems in real prototype cases and the comparison of their performances with a benchmarking model in this field.

7 Acknowledgement

This research was supported by the Internal Competitive Research Grant of Hong Kong Polytechnic University (A-PE26).

References

1. Chau, K.W., Jiang, Y.W.: 3D Numerical Model for Pearl River Estuary. Journal of Hydraulic Engineering ASCE **127(1)** (2001) 72-82
2. Chau, K.W., Jin, H.S.: Numerical Solution of Two-Layer, Two-Dimensional Tidal Flow in a Boundary Fitted Orthogonal Curvilinear Coordinate System. International Journal for Numerical Methods in Fluids **21(11)** (1995) 1087-1107

3. Chau, K.W., Jin, H.S., Sin, Y.S.: A Finite Difference Model of Two-Dimensional Tidal Flow in Tolo Harbor, Hong Kong. Applied Mathematical Modelling **20(4)** (1996) 321-328
4. Chau, K.W., Lee, J.H.W.: Mathematical Modelling of Shing Mun River Network. Advances in Water Resources **14(3)** (1991) 101-124
5. Chau, K.W., Lee, J.H.W.: A Microcomputer Model for Flood Prediction with Application. Microcomputers in Civil Engineering **6(2)** (1991) 109-121
6. Smith, J., Eli, R.N.: Neural-Network Models of Rainfall-Runoff Process. Journal of Water Resources Planning and Management, ASCE **121(6)** (1995) 499-508
7. Tokar, A.S., Johnson, P.A.: Rainfall-Runoff Modeling using Artificial Neural Networks. Journal of Hydrologic Engineering, ASCE **4(3)** (1999) 232-239
8. Liong, S.Y., Lim, W.H., Paudyal, G.N.: River Stage Forecasting in Bangladesh: Neural Network Approach. Journal of Computing in Civil Engineering, ASCE **14(1)** (2000) 1-8
9. Cheng, C.T., Chau, K.W.: Fuzzy Iteration Methodology for Reservoir Flood Control Operation. Journal of the American Water Resources Association **37(5)** (2001) 1381-1388
10. Chau, K.W., Cheng, C.T.: Real-time Prediction of Water Stage with Artificial Neural Network Approach. Lecture Notes in Artificial Intelligence **2557** (2002) 715-715
11. Chau, K.W.: Calibration of Flow and Water Quality Modeling using Genetic Algorithm. Lecture Notes in Artificial Intelligence **2557** (2002) 720-720
12. Cheng, C.T., Ou, C.P., Chau, K.W.: Combining a Fuzzy Optimal Model with a Genetic Algorithm to solve Multiobjective Rainfall-Runoff Model Calibration. Journal of Hydrology **268(1-4)** (2002) 72-86
13. Chau, K.W.: River Stage Forecasting with Particle Swarm Optimization. Lecture Notes in Computer Science **3029** (2004) 1166-1173
14. Chau, K.W.: Rainfall-Runoff Correlation with Particle Swarm Optimization Algorithm. Lecture Notes in Computer Science **3174** (2004) 970-975
15. Cheng, C.T., Chau, K.W., Sun, Y.G., Lin, J.Y.: Long-Term Prediction of Discharges in Manwan Reservoir using Artificial Neural Network Models. Lecture Notes in Computer Science **3498** (2005) 1040-1045
16. Goldberg, D.E., Kuo, C.H.: Genetic Algorithms in Pipeline Optimization. Journal of Computing in Civil Engineering ASCE **1(2)** (1987) 128-141