Computers & Geosciences 70 (2014) 63–72

Contents lists available at ScienceDirect

Computers & Geosciences

journal homepage: www.elsevier.com/locate/cageo

Linear genetic programming application for successive-station monthly streamflow prediction

Ali Danandeh Mehr [a,*], Ercan Kahya [b], Cahit Yerdelen [c]

[a] Istanbul Technical University, Civil Engineering Department, Hydraulics Division, 34469 Maslak, Istanbul, Turkey
[b] Istanbul Technical University, Civil Engineering Department, Hydraulics Division, Istanbul, Turkey
[c] Ege University, Civil Engineering Department, Hydraulics Division, Izmir, Turkey

ARTICLE INFO

Article history:
Received 9 November 2013
Received in revised form
22 March 2014
Accepted 29 April 2014
Available online 8 May 2014

Keywords:
Artificial neural networks
Linear genetic programming
Streamflow prediction
Successive stations

ABSTRACT

In recent decades, artificial intelligence (AI) techniques have been pronounced as a branch of computer science to model wide range of hydrological phenomena. A number of researches have been still comparing these techniques in order to find more effective approaches in terms of accuracy and applicability. In this study, we examined the ability of linear genetic programming (LGP) technique to model successive-station monthly streamflow process, as an applied alternative for streamflow prediction. A comparative efficiency study between LGP and three different artificial neural network algorithms, namely feed forward back propagation (FFBP), generalized regression neural networks (GRNN), and radial basis function (RBF), has also been presented in this study. For this aim, firstly, we put forward six different successive-station monthly streamflow prediction scenarios subjected to training by LGP and FFBP using the field data recorded at two gauging stations on Çoruh River, Turkey. Based on Nash–Sutcliffe and root mean squared error measures, we then compared the efficiency of these techniques and selected the best prediction scenario. Eventually, GRNN and RBF algorithms were utilized to restructure the selected scenario and to compare with corresponding FFBP and LGP. Our results indicated the promising role of LGP for successive-station monthly streamflow prediction providing more accurate results than those of all the ANN algorithms. We found an explicit LGP-based expression evolved by only the basic arithmetic functions as the best prediction model for the river, which uses the records of the both target and upstream stations.

1. Introduction

Artificial neural networks (ANNs) are from popular artificial intelligence (AI) techniques broadly used in various fields of geoscience. They are capable of using field data directly and modelling the corresponding phenomena without prior knowledge of it. Successful results of ANN application in geoscience particularly in hydrological predictions have been extensively published in recent years (e.g. Minns and Hall, 1996; Nourani et al., 2008; Besaw et al., 2010; Piotrowski et al., 2014). Our review concerning the application of different ANN structures in streamflow forecasting indicated that it has been received tremendous attention of research (e.g. Dawson and Wilby, 1998; Dolling and Varas, 2002; Cannas et al., 2006; Kerh and Lee, 2006; Kisi and Cigizoglu, 2007; Adamowski, 2008; Kişi, 2009; Shiri and Kisi, 2010; Marti et al., 2010; Nourani et al., 2011; Abrahart et al., 2012; Can et al., 2012; Krishna, 2013; Kalteh, 2013; Danandeh Mehr et al., 2013).

Minns and Hall (1996) introduced ANNs as rainfall-runoff models and demonstrated that they are capable of identifying usable relationships between discharges and antecedent rainfalls. Kerh and Lee (2006) applied an ANN-based model using information at stations upstream of Kaoping River to forecast flood discharge at the downstream station which lacks measurements. They found that the back-propagation ANN model performs relatively better than the conventional Muskingum method. Besaw et al. (2010) developed two different ANN models using the time-lagged records of precipitation and temperature in order to forecast streamflow in an ungauged basin in the US. The authors explained that ANNs forecasts daily streamflow in the nearby ungauged basins as accurate as in the basin on which they were trained. Can et al. (2012) used streamflow records of nine gauging stations located in Çoruh River basin to model daily streamflow in Turkey. They compared the performance of their ANN-based models with those of auto regressive moving average (ARMA) models and demonstrated that the ANNs resulted in higher performance than ARMA. A comprehensive review concerning

* Corresponding author. Tel.: +90 553 417 8028; fax: +90 212 285 6587.
E-mail addresses: danandeh@itu.edu.tr, danandehmehr@yahoo.com (A. Danandeh Mehr), kahyae@itu.edu.tr (E. Kahya), cahit.yerdelen@ege.edu.tr (C. Yerdelen).

http://dx.doi.org/10.1016/j.cageo.2014.04.015
0098-3004/

the application of different ANN structures in river flow prediction has been presented by Abrahart et al. (2012).

In spite of providing satisfactory estimation accuracy, all aforementioned ANN-based models are implicit and often are criticized as ‘ultimate black boxes’ that are difficult to interpret (Babovic, 2005). Depending on the number of applied hidden layers, they may produce huge matrix of weights and biases. Consequently, the necessity of additional studying in order to develop not only explicit but also precise models still requires serious attention.

Genetic programming (GP) is a heuristic evolutionary computing technique (Koza, 1992; Babovic, 2005) that has been pronounced as an explicit predictive modelling tool for hydrological studies (Babovic and Abbott, 1997a; Babovic and Keijzer, 2002). The capability of GP to model hydro-meteorological phenomena as well as its degree of accuracy are of the controversial topics in recent hydroinformatic studies (e.g. Ghorbani et al., 2010; Kisi and Shiri, 2012; Yilmaz and Muttil, 2014; Wang et al., 2014).

After Babovic and Abbott (1997b), who pronounced GP as an advanced operational tool to solve wide range of hydrological modelling problems, GP and it's variants/advancements were considered broadly in different hydrological processes such as rainfall-runoff (Babovic and Keijzer, 2002; Khu et al., 2001; Liong et al., 2001; Whigham and Crapper, 2001; Nourani et al., 2012), sediment transport (Babovic, 2000; Aytek and Kisi, 2008; Kisi and Shiri, 2012), sea level fluctuation (Ghorbani et al., 2010), precipitation (Kisi and Shiri, 2011), evaporation (Kisi and Guven, 2010), and others.

GP and its variants have also been received remarkable attention in the most recent comparative studies among different AI techniques (e.g. Ghorbani et al., 2010; Kisi and Guven, 2010; Kisi and Shiri, 2012). In the field of streamflow forecasting, Guven (2009) applied linear genetic programming (LGP), an advancement of GP, and two versions of neural networks for daily flow prediction in Schuylkill River, USA. The author demonstrated that the performance of LGP was moderately better than that of ANNs. Wang et al. (2009) developed and compared several AI techniques comprising ANN, neural-based fuzzy inference system (ANFIS), GP, and support vector machine (SVM) for monthly streamflow forecasting using long-term observations. Their results indicated that the best performance in terms of different evaluation criteria can be obtained by ANFIS, GP and SVM. Londhe and Charhate (2010) used ANN, GP, and model trees (MT) to forecast river flow one-day in advance at two gauging stations in India's Narmada Catchment. The authors concluded that ANN and MT techniques perform almost equally well, but GP performs better than its counterparts. Ni et al. (2010) applied GP to model the impact of climate change on annual streamflow of the West Malian River, China. They compared the results of GP with those of ANN and multiple linear regression models and indicated that GP provides higher accuracy than the others. Yilmaz and Muttil (2014) used GP to predict river flows in different parts of the Euphrates River basin, Turkey. They compared the results of GP with those of ANN and ANFIS and demonstrated that GP are superior to ANN in the middle zone of the basin. Among the most recent comparative studies between different AI techniques, Wang et al. (2014) proposed the singular spectrum analysis (SSA) in order to modify SVM, GP, and seasonal autoregressive (SAR) models. They applied the modified models to predict monthly inflow for three Gorges Reservoirs and indicated that modified GP is slightly superior to modified SVM at peak discharges prediction. Although there are some other comparative studies between GP and different AI techniques, to the best of our knowledge, there is no research examining the performance of LGP for successive-station monthly streamflow prediction in comparison with different ANN structures/algorithms.

The main goals and motivation of our study are (i) to further enhance the available LGP modelling tool to provide an explicit expression for successive-stations streamflow prediction and (ii) for the first time, to compare the efficiency of LGP with three different ANN algorithms for monthly streamflow prediction. In this way, at the first stage, we put forward six different successive-station prediction scenarios structured by commonly used feed-forward back propagation neural network algorithm (FFBP). Then, using LGP technique a new set of explicit expressions has been generated for these scenarios. We performed a comparative performance analysis between the proposed LGP and FFBP models using Nash–Sutcliffe efficiency and root mean square error measures. As a consequence of the first stage of the study, the best scenario was identified and discussed. In the second stage, two other ANN algorithms, namely generalized regression neural networks (GRNN), and radial basis function (RBF) neural networks were utilized to restructure the best prediction scenario. Ultimately, we put forward a discussion about both accuracy and applicability of different ANN and LGP models.

It is observed that some of gauging stations are closed down in all over the world where the stations are no longer required or funding to support continued operation is limited. Using successive-station prediction strategy, in case of developing a plausible model between a pair of upstream–downstream stations, the model can be used as a substitute for the station(s) which is at risk for discontinuation. In addition, since inputs of the successive-station prediction models are only time-lagged streamflow observations, such models are also considered more useful for the catchments with sparse rain gauge stations (Besaw et al., 2010). The successive-station strategy also tends to decrease the lagged prediction effect of commonly proposed single-station runoff-runoff modes which has been mentioned by some researchers (Chang et al., 2007; De Vos and Rientjes, 2005; Muttil and Chau, 2006; Wu et al., 2009a).

2. Overview of FFBP, GRNN, and RBF networks

ANNs are from black-box regression methods which are commonly used to find out the nonlinear systems attitude. FFBP networks are probably the most popular ANNs in hydrological problems (Tahershamsi et al., 2012; Krishna, 2013) which considered as general nonlinear approximations (Hornik et al., 1989). The primary goal of this algorithm is to minimize the estimation error by searching for a set of connection weights, synaptic weights, which cause the network to produce outputs closer to the targets. They are typically composed of three parts: (a) input layer including a number of input nodes, (b) one or more hidden layers and (c) a number of output layer nodes. The number of hidden layers and relevant nodes are two of the design parameters of FFBP networks. A neural network with too many nodes may overfit the data, causing poor generalization on data not used for training, while too few hidden units may underfit the model (Fletcher et al., 1998).The input nodes do not perform any transformation upon the input data sets. They only send their initial weighted values to hidden layer nodes. The hidden layer nodes typically receive the weighted inputs from the input layer or a previous hidden layer, perform their transformations on it, and pass the output to the next adjacent layer which is generally another hidden layer or an output layer. The output layer consists of nodes that receive the hidden layer outputs and send it to the modeller. Initial synapses are progressively corrected during the training process that compares predicted outputs with corresponding observations and back-propagates any errors to minimize them. The design issues, training mechanisms and application of FFBP in hydrological studies have been the subject of different studies (e.g. Abrahart et al., 2012; Nourani et al., 2013). Therefore, to avoid duplication, we only introduced the main