Applied Soft Computing 38 (2016) 329–345

Contents lists available at ScienceDirect

Applied Soft Computing

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

Integrating Support Vector Regression and a geomorphologic Artificial Neural Network for daily rainfall-runoff modeling

Seiyed Mossa Hosseini [a,*], Najmeh Mahjouri [b]

[a] Natural Geography Dept., University of Tehran, P.O. Box 14155-6465, Tehran, Iran
[b] Faculty of Civil Engineering, K.N. Toosi University of Technology, Tehran, Iran

ARTICLE INFO

Article history:
Received 15 June 2013
Received in revised form 11 July 2015
Accepted 12 September 2015
Available online 17 October 2015

Keywords:
Rainfall-runoff modeling
Geomorphologic characteristics
Support Vector Regression (SVR)
Artificial Neural Networks (ANNs)
Genetic algorithm
Fuzzy inference system (FIS)

ABSTRACT

In spite of the efficiency of the Artificial Neural Networks (ANNs) for modeling nonlinear and complicated rainfall-runoff (R-R) process, they suffer from some drawbacks. Support Vector Regression (SVR) model has appeared to be a powerful alternative to reduce some of these drawbacks while retaining many strengths of ANNs. In this paper, to form a new rainfall-runoff model called SVR-GANN, a SVR model is combined with a geomorphologic-based ANN model. The GANN is a three-layer perceptron model, in which the number of hidden neurons is equal to the number of possible flow paths within a watershed and the connection weights between hidden layer and output layer are specified by flow path probabilities which are not updated during the training process. The capabilities of the proposed SVR-GANN model in simulating the daily runoff is investigated in a case study of three sub-basins located in a semi-arid region in Iran. The results of the proposed model are compared with those of ANN-based back propagation algorithm (ANN-BP), traditional SVR, ANN-based genetic algorithm (ANN-GA), adaptive neuro-fuzzy inference system (ANFIS), and GANN from the standpoints of parsimony, equifinality, robustness, reliability, computational time, simulation of hydrograph ordinates (peak flow, time to peak, and runoff volume) and also saving the main statistics of the observed data. The results show that prediction accuracy of the SVR-GANN model is usually better than those of ANN-based models and the proposed model can be applied as a promising, reliable, and robust prediction tool for rainfall-runoff modeling.

1. Introduction

Development and application of rainfall-runoff simulation models have been an appealing subject of research in hydrology. For simulating rainfall-runoff (R-R) relation in a watershed, two main approaches, namely knowledge-based and data-driven approaches, have been developed [1]. Many researches have been conducted based on the use of knowledge-based approaches such as physical and conceptual approaches [3]. These methods in general, mimic complexity of real world runoff behavior and conceptualize runoff processes and catchment properties [4,5]. An alternative to the knowledge-based approach is the data-driven approach which is based on extracting and reusing information implicitly existing in hydrologic data [22]. Sources and descriptions of many watershed models and comparative assessment of their performances are well addressed by researchers [6–8].

Several models have been developed to simulate the watershed response to rainfall by Artificial Neural Networks (ANNs), due to their efficiency in modeling nonlinear and complicated rainfall-runoff relationship [9–12]. Among several topologies of ANNs, the feed forward back propagation type is frequently used in rainfall-runoff simulation and also river flow forecasting due to having less complexity than other architectures [13,14]. In spite of the efficiency of the ANNs in modeling nonlinear and complicated R-R process, they suffer from some drawbacks [15–17]. Efforts to overcome these shortcomings in the ANNs models used to simulate the rainfall-runoff process can be summarized in three categories:

(1) Incorporating the physical and geomorphologic characteristics of a watershed directly in the ANN architecture (here named GANN) especially in the connecting weights of the model [18–21]. Zhang and Govindaraju [22] established a three-layer feed-forward ANN that explicitly accounts for the geomorphologic catchment characteristics in the network architecture to estimate direct runoff hydrographs resulted from several storms over two watersheds in the state of Indiana, USA. The structures of geomorphologic ANN as well as network

* Corresponding author. Tel.: +98 216 1113; fax: +98 216 6413065.
E-mail addresses: smhosseini@ut.ac.ir (S.M. Hosseini), mahjouri@kntu.ac.ir (N. Mahjouri).

http://dx.doi.org/10.1016/j.asoc.2015.09.049
1568-4946/

weights were determined using watershed geomorphology. The authors compared the results of the developed model with those obtained using the Geomorphologic Instantaneous Unit Hydrograph theory (GIUH) based on path probability. Their study showed that the GANN model is a promising tool for estimating direct runoff. They also used the ANN-based geomorphologic model developed by [22] for surface and subsurface flow simulation in Heng-Chi watershed in Taiwan [23]. The results of this model were also compared with ones obtained using GIUH model developed by [24] based on kinematic-wave approximation and Darcy's law. The authors suggested that the ANN model showed better performance than GIUH model in terms of time to peak and peak flow estimation.

(2) Using an evolutionary optimization algorithm in the ANN structure to improve the optimization process of the parameters during the training phase. The most well-known training algorithms are the classical feed-forward algorithm [25], conjugate gradient algorithm [26,27] and methods based on second-order gradients such as the Levenberg–Marquardt (LM) algorithm [28]. These gradient-based algorithms usually are prone to be trapped in locally optimal solutions in multi-modal and multi-dimensional non-linear optimization problems such as rainfall-runoff modeling. The recently proposed nature-inspired evolutionary computation algorithms are able to work not only with non-differentiable functions but also with functions with a considerable number of local minima [29]. Hybridizing the ANNs with such optimization algorithms (e.g., genetic algorithm, particle swarm optimization, etc.) can decrease the chance of over-fitting and provide better accuracy, while increasing the computational time. Wu and Chau [30] and Chau [31] developed particle swarm optimization-based ANN (PSO-ANN) model for prediction of Shing Mun River stage in Hong Kong and showed that the PSO-ANN model has higher efficiency than back propagation-based ANN model. Haykin [32] discussed several data-driven optimization training algorithms, such as Levenberg–Marquardt algorithm and scaled conjugate gradient algorithm. Rogers et al. [33] used the genetic algorithm for optimal field-scale groundwater remediation together with ANN.

(3) Integrating the ANN model with the fuzzy inference system (FIS), entitled ANFIS model. The rainfall-runoff process is a complex non-linear outcome of various hydrologic parameters, i.e., precipitation intensity, evaporation, geomorphology of watershed, infiltration of water into the soil and depression storage as well as interactions between groundwater and surface water flows [34,35]. The deficiencies in data including missing data, noisy data, insufficient data, and also vagueness of these variables can cause difficulties in rainfall-runoff modeling. Previous works of application of the FIS integrated with the ANN model structure, so-called neuro-fuzzy system, showed that it could be a well-organized tool to deal with these challenges in rainfall-runoff modeling [12,36,37].

Recently, Support Vector Regression (SVR) has been successfully applied in the fields of water resources engineering and hydrology for purposes such as runoff prediction [38–40], flood forecasting [41,42], lake water level prediction [43,44], regionalization of contaminants in aquifers [45], etc.

In comparison with ANNs, SVRs are able to learn more effectively when using scarce and incomplete hydrologic data [46,47]. This advantage is because of two outstanding features of SVRs: their excellent capability in generalization of the unseen data (testing phase) and their proficiency for application in large scale problems using only a small number of support vectors [48]. The SVRs are also preferred due to reducing the memory footprint of the learned predictor, as well as low computational cost [49]. Liong and Sivapragasam [41] detailed eight facts about superiority of SVRs over ANNs in simulating and predicting complex phenomena in hydrogeology and water resources, while holding all the strengths of ANN.

The key questions in using different approaches for R-R modeling described above are how these models perform in simulating runoff and how one can incorporate the capabilities of the ANN- and SVR-based models in one model to promote the reliability of the watershed runoff simulation. The main purpose of this paper is developing a new model which combines the traditional SVR model with a geomorphologic ANN model (GANN) to simulate the daily runoff in a watershed. The performance of the proposed model (SVR-GANN) is compared with different ANN-based models including ANN with embedded geomorphologic characteristics (GANN), ANN with genetic algorithm (ANN-GA), ANN adopted with the fuzzy inference system (adapted neuro-fuzzy inference system – ANFIS) and also traditional SVR model from the viewpoints of simplicity (parsimony), equifinality, robustness, reliability, computational time, hydrograph ordinates and saving the main statistics of the observed data.

2. Materials and methods

In this section, the structure of the SVR-GANN model, which is developed by combining the GIUH theory, a three-layer perceptron ANN model, a traditional SVR model and a genetic algorithm-based optimization model is described. At first, a brief description of each individual model is given. Then, the structure and details of the proposed GANN model are provided. Also, the ANN-based models, which are applied for comparing their results with those of the proposed model, are briefly explained and finally, the fitness function and evaluation criteria used for evaluating the results of the models are described.

2.1. GIUH theory

Before explaining the structure of GANN model, a brief review of the GIUH theory is given in this section. The GIUH principle assumes that the watershed can be represented as a linear, time-invariant system, such that the total runoff from surface and subsurface flow in response to input rainfall can be obtained based on the following convolution integral [50]:

$$Q(t) = \int_0^t I(\tau) * u(t-\tau)\, d\tau \tag{1}$$

where $u(t)$ is IUH ordinate at time t, $Q(t)$ is direct flow in watershed outlet, and $I(t)$ is excess rainfall at time t. The unit response hydrograph is formulated in terms of travel path probabilities and travel times over portions of the land surface, subsurface zones and channel links (Fig. 1). The spatial structure of the basin, expressed in terms of the Strahler order of the network, is explicitly considered for formulating the GIUH.

Let x_{o_i} denotes the ith overland flow region, x_i the ith surface flow order, and $i = 1, 2, 3, \ldots, \Omega$ (Ω denotes the watershed order). The total IUH of watershed, $u(t)$, as a linear system of surface flows is expressed as follows:

$$u(t) = \sum_{w_s \in W_s} \left| f_{x_{o_i}}(t) * f_{x_i}(t) * f_{x_j}(t) * \cdots * f_{x_\Omega}(t) \right|_{w_s} \cdot P(w_s),$$
$$W_s = \langle x_{o_i}, x_i, x_j, \ldots, x_\Omega \rangle \tag{2}$$

where $f_{x_i}(t)$ is the travel time probability density function in state x_i, W_s is the surface flow path space which includes $\{x_{o_i}, x_i, x_j, \ldots, x_\Omega\}$ with total number of $N(w_s)$, $P(w_s)$ is the probability of the surface