Available online at www.sciencedirect.com
ScienceDirect
Applied Soft Computing 7 (2007) 585–592
Applied Soft Computing
www.elsevier.com/locate/asoc

Hybrid neural network models for hydrologic time series forecasting

Ashu Jain *, Avadhnam Madhav Kumar

Department of Civil Engineering, Indian Institute of Technology, Kanpur 208016, India

Received 3 May 2005; received in revised form 12 January 2006; accepted 9 March 2006
Available online 9 June 2006

Abstract

The need for increased accuracies in time series forecasting has motivated the researchers to develop innovative models. In this paper, a new hybrid time series neural network model is proposed that is capable of exploiting the strengths of traditional time series approaches and artificial neural networks (ANNs). The proposed approach consists of an overall modelling framework, which is a combination of the conventional and ANN techniques. The steps involved in the time series analysis, e.g. de-trending and de-seasonalisation, can be carried out before gradually presenting the modified time series data to the ANN. The proposed hybrid approach for time series forecasting is tested using the monthly streamflow data at Colorado River at Lees Ferry, USA. Specifically, results from four time series models of auto-regressive (AR) type and four ANN models are presented. The results obtained in this study suggest that the approach of combining the strengths of the conventional and ANN techniques provides a robust modelling framework capable of capturing the non-linear nature of the complex time series and thus producing more accurate forecasts. Although the proposed hybrid neural network models are applied in hydrology in this study, they have tremendous scope for application in a wide range of areas for achieving increased accuracies in time series forecasting.

Keywords: Artificial neural networks; Streamflow forecasting; Time series modelling; Rainfall–runoff process; Hydrology; Hybrid models

1. Introduction

Time series forecasting has received tremendous attention of researchers in the last few decades. This is because the future values of a physical variable, which are measured in time at discrete or continuous basis, are needed in important planning, design and management activities. The time series forecasting methods have found applications in very wide areas including but not limited to finance and business, computer science, all branches of engineering, medicine, physics, chemistry and many interdisciplinary fields. Conventionally, the researchers have employed traditional methods of time series analysis, modelling, and forecasting, e.g. Box–Jenkins methods of auto-regressive (AR), auto-regressive moving average (ARMA), auto-regressive integrated moving average (ARIMA), auto-regressive moving average with exogenous inputs (ARMAX), etc. The conventional time series modelling methods have served the scientific community for a long time; however, they provide only reasonable accuracy and suffer from the assumptions of stationarity and linearity. Artificial neural networks (ANNs) were introduced as efficient tools of modelling and forecasting about two decades ago. One can find numerous ANN applications in a wide range of areas for time series forecasting. In spite of the great deal of time and effort spent by the researchers in both conventional and soft computing techniques for time series forecasting, the need of producing more and more accurate time series forecasts has forced the researchers to develop innovative methods to model time series. One can find many studies that involve developing neural network models for time series forecasting. This paper presents a study aimed at achieving accurate forecasts of a hydrologic time series using a combination of traditional time series and neural network approaches. The paper begins with a brief review of the time series forecasting using neural networks in a wide range of fields.

2. Time series forecasting using neural networks

Neural networks have been applied in many areas for time series forecasting. This section reviews some important studies reported in literatures since early 1990s. Arizmendi et al. [1] made accurate predictions of the airborne pollen concentrations using the time series data and neural networks. The predictions were found to be better than the traditional time series

* Corresponding author. Tel.: +91 512 259 7411 (O); fax: +91 512 259 7395.
E-mail address: ashujain@iitk.ac.in (A. Jain).

1568-4946/$ – see front matter
doi:10.1016/j.asoc.2006.03.002

forecasting made using pollen concentrations and other meteorological parameters. Srinivasan et al. [2] employed a four-layered feed-forward neural network trained using back-propagation (BP) to make hourly predictions of electric load for a power system. Prediction accuracies with 1.07% error on weekdays and 1.80% on weekends were achieved, which was much superior to the predictions accuracies obtained from tradition time series forecasting methods. Kaastra and Boyd [3] presented an eight-step procedure to design neural network forecasting model for financial and economic time series. Ansuj et al. [4] compared ARIMA models with BPANN models for sales forecasting in Brazil and found BPANN models to be much superior. Zhang and Hu [5] employed ANNs for forecasting British pound and US dollar exchange rates. They also evaluated the impact of the number of input and hidden neurons and the size of the data on the ANN model performance. The sensitivity analyses showed that the input neurons affect the performance more than the hidden neurons and improved accuracies can be achieved with a larger sample size. Bezerianos et al. [6] employed radial basis function (RBF) neural networks for the assessment and prediction of the heart rate variability. Li et al. [7] modelled damping in tall buildings using BPANNs and AR models. The models were used to predict the damping values at high values of vibrations, which are difficult to obtain through field measurements. Nguyen and Chan [8] proposed a multiple neural network (MNN) approach for making multistep ahead forecasts of hourly customer demand for gas at a compression station in Sask., Canada. It was found that the MNN model performed better than the single ANN model developed on the same data set.

Most of the studies reported above were simple applications of using traditional time series approaches and ANNs. Many of the real-life time series are extremely complex to be modelled using simple approaches especially when high accuracy is required. Hansen and Nelson [9] used synergic combination of neural networks and traditional time series methods to make economic forecasts using revenue data from UT, USA. They concluded that the economic forecasts resulting from synergic combination were more accurate than the individual forecasts from traditional time series and neural networks. Kalaitzakis et al. [10] investigated several approaches including Gaussian encoding BP, window random activation, RBF networks, real-time recurrent neural networks (RNNs) and their innovative variations for short-term electric load forecasting. The models proposed were able to achieve significantly more accurate forecasts as compared to the simple AR and BPANN models. In addition, they also proposed a parallel processing approach for making 24-h ahead electric load forecasts. Zhang [11] proposed a hybrid ARIMA and ANN model to take advantage of the two techniques and applied the proposed hybrid model to some real data sets. He concluded that the combined model can be an effective way to improving forecasts achieved by either of the models used separately. Gao and Er [12] made non-linear auto-regressive moving average with exogenous inputs (NARMAX) model based time series prediction with fuzzy neural networks (FNNs) using both feed-forward and recurrent methods. They also proposed an efficient algorithm for model structure determination and parameter estimation for producing improved forecasts for NARMAX time series models.

It is clear that neural networks have been applied to a wide range of disciplines for time series forecasting using simple approaches. Some recent studies clearly demonstrate that there is a strong need to exploit the advantages of both traditional time series methods and ANNs in order to achieve increased accuracy in time series forecasting. Although one can find many applications of developing hybrid models in a variety of areas, such attempts have been limited in hydrology and water resources system modelling. The next section is devoted to explaining the problem of hydrologic modelling, its need and the state of art in hydrologic time series modelling.

3. Hydrologic time series modelling

One of the key hydrologic variables is the streamflow at a location in a river in a catchment, which is measured in cubic metres per second. The availability of accurate streamflow forecasts at a location in a river in a catchment is important in many water resources management and design activities such as flood control and management and design of various hydraulic structures such as dams and bridges. Streamflow forecasts can be generated using two types of mathematical models: rainfall–runoff models that use both climatic and hydrologic data and streamflow models that use only the hydrologic data. Usually, the researchers have relied on conventional modelling techniques, either deterministic/conceptual models that consider the physics of the underlying process or systems theoretic/black-box models that do not. Deterministic and black-box models of varying degree of complexity have been employed in the past for modelling rainfall–runoff process with varying degree of success. The streamflow process in a catchment is a complex and non-linear process affected by many and often interrelated physical factors. The factors affecting the streamflow response of a catchment subjected to rainfall input include: (a) storm characteristics, i.e. intensity and duration of rainfall events, (b) catchment characteristics, i.e. size, shape, slope and storage characteristics of the catchment, percentage of the catchment contributing streamflow at the outlet at various time steps during a rainfall event, (c) geomorphologic characteristics of a catchment, i.e. topography, land use patterns, vegetation and soil types that affect the infiltration and (d) climatic characteristics such as temperature, humidity and wind characteristics. The influence of these factors and many of their combinations in generating streamflow is an extremely complex physical process and is not understood clearly [13]. Moreover, many of the deterministic or conceptual rainfall–runoff models need a large amount of data for calibration and validation purposes and are computationally extensive. As a result, the use of deterministic/conceptual models of the rainfall–runoff process has been viewed rather sceptically by the researchers and has not become very popular [14]. ANNs have been proposed as efficient tools for modelling and prediction and are supposed to possess the capability to