Applied Mathematical Modelling 38 (2014) 3512–3522

Contents lists available at ScienceDirect

Applied Mathematical Modelling

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

Multi-step prediction of time series with random missing data

Xuedong Wu [a,*], Yaonan Wang [b], Jianxu Mao [b], Zhaoping Du [a], Chunhua Li [a]

[a] School of Electronics and Information, Jiangsu University of Science and Technology, Zhenjiang, Jiangsu 212003, China
[b] College of Electrical and Information Engineering, Hunan University, Changsha, Hunan 410082, China

ARTICLE INFO

Article history:
Received 30 April 2012
Received in revised form 1 November 2013
Accepted 26 November 2013
Available online 4 January 2014

Keywords:
Multi-step prediction
Random missing data
Multilayer perceptron ANN
Nonlinear filters

ABSTRACT

Multi-step prediction is still an open challenge in time series prediction. Moreover, practical observations are often incomplete because of sensor failure or outliers causing missing data. Therefore, it is very important to carry out research on multi-step prediction of time series with random missing data. Based on nonlinear filters and multilayer perceptron artificial neural networks (ANNs), one novel approach for multi-step prediction of time series with random missing data is proposed in the study. With the basis of original nonlinear filters which do not consider the missing data, first we obtain the generalized nonlinear filters by using a sequence of independent Bernoulli random variables to model random interruptions. Then the multi-step prediction model of time series with random missing data, which can be fit for the online training of generalized nonlinear filters, is established by using the ANN's weights to present the state vector and the ANN's outputs to present the observation equation. The performance between the original nonlinear filters based ANN model for multi-step prediction of time series with missing data and the generalized nonlinear filters based ANN model for multi-step prediction of time series with missing data is compared. Numerical results have demonstrated that the generalized nonlinear filters based ANN are proportionally superior to the original nonlinear filters based ANN for multi-step prediction of time series with missing data.

1. Introduction

Time series prediction is a growing field of interest and it plays an important role in nearly all fields of science and engineering. Unlike one-step prediction, multi-step prediction tasks are more difficult [1], since they have to deal with various additional complications, like accumulation of errors, reduced accuracy, and increased uncertainty [2].

The prediction domain has been influenced by linear statistical methods for a long time. However, in the late 1970s and early 1980s, it becomes increasingly clear that linear models are not adapted to many real applications [3]. In the last two decades, machine learning models have drawn attention and have established themselves as serious contenders to classical statistical models in the prediction community [4]. These models, also called black-box or data-driven models [5], are examples of nonparametric nonlinear models which use only historical data to learn the stochastic dependency between the past and the future. For instance, multilayer perceptron ANNs [6–9], radical basis function ANNs [10,11], recurrent neural network [12,13], time-delay ANNs [14] and least squares support vector regression [15] are the typical different machine learning models for time series prediction.

The aforementioned time series prediction models [6–15] assume that the observations are sampled with the same frequency and the sampled data is complete. However, missing data is a very frequent problem in many scientific fields,

* Corresponding author. Tel.: +86 13921598973.
E-mail address: woolcn@163.com (X. Wu).

http://dx.doi.org/10.1016/j.apm.2013.11.029

usually due to faults in data acquisition [16–28]. In order to carry out a precise time series analysis and obtain reliable prediction results, it is necessary to deal effectively with the missing data. Moreover, in many applications, the time series dimension is large and the model depends on a large number of parameters, so the task of model parameter estimation is challenging in various respects as well [29]. Particularly, state space formulations allow obtaining minimum mean square estimations of the parameters together with the corresponding mean square errors by means of nonlinear filters such as the extended Kalman filtering (EKF) [30], the unscented Kalman filtering (UKF) [31] and the Gaussian particle filtering (GPF) [32]. These methods can be applied to time series prediction and can be implemented in a computationally efficient way [9,33]. However, the model representations are not valid in the presence of missing data. With the basis of EKF, UKF and GPF, Wu and Song [27] developed the corresponding generalized nonlinear filters by using a sequence of independent Bernoulli random variables to model random missing data, then the generalized nonlinear filters combined with the state space presentation of multilayer perceptron ANN were applied to one-step prediction of time series. Wu et al. [28] used the generalized EKF and the generalized UKF for one-step prediction of ground radioactivity time series with missing data based on the radial basis function ANN.

Multilayer perceptron ANNs are quite popular among different types of ANNs [34], and the nonlinear filters have the advantage of implementation in a computationally efficient way when they are applied to time series prediction. In this paper, one novel approach for multi-step prediction of time series with random missing data is presented based on the proposed generalized nonlinear filters in Wu and Song [27] and the state space presentation of multilayer perceptron ANN, and the performance between the original nonlinear filters based ANN model in Wu and Wang [9] for multi-step prediction of time series with missing data and the generalized nonlinear filters based ANN model for multi-step prediction of time series with missing data is compared.

The remainder of the paper is organized as follows. The multi-step prediction framework of time series with random missing data is discussed in detail in the next Section. Section 3 contains a description of the generalized nonlinear filters when missing data is present. The five-step prediction simulation results of Mackey-Glass time series and equipment's temperature time series with missing data are presented in Section 4. Finally, we summarize the paper and propose the future work in Section 5.

2. The multi-step prediction framework of time series with random missing data

The ANN model used in our study is a three-layer perceptron ANN with two hidden layers and one output layer. Its configuration is demonstrated in Fig. 1 and its details are given as follows.

(i) $[A_{i,j}]$: weights from input neuron (i) to jth first hidden layer neuron.
(ii) $[B_{j,k}]$: weights from first hidden layer neuron (j) to kth second hidden layer.
(iii) $[C_k]$: weights from third hidden layer neuron (k) to output neuron.
(iv) b_{1j}, b_{2k} and b_{31} denote the bias of each neuron.

The logistic *sigmoid* function $g(\cdot)$ is:

$$g(\mathbf{s}) = 1/[1+\exp(-\mathbf{s})]. \tag{1}$$

With this configuration, then the output versus the inputs to the ANN can be computed as:

$$O = \sum_l C_l g\left[\sum_j B_{j,l} g\left(\sum_i A_{i,j} I_i + b_{1j}\right) + b_{2k}\right] + b_{31}. \tag{2}$$

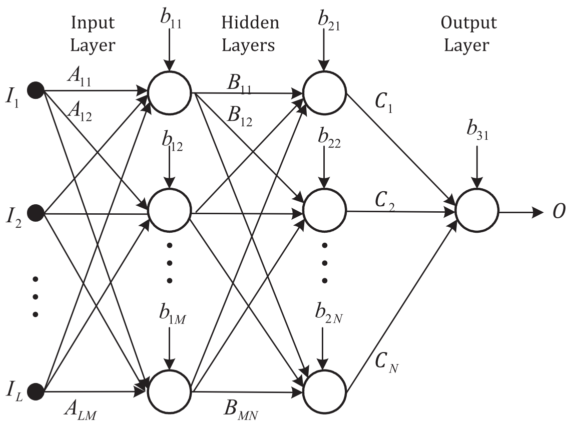

Fig. 1. The configuration of feed-forward neural network.