Application of Nonlinear Filtering Trained RBF Networks to Multi-step Prediction of Time-series with Delayed Observations

Xuedong WU1,†, Yaonan WANG2, Weiting LIU1, Zhiyu ZHU1, Yue TAN3

1 School of Electronics and Information, Jiangsu University of Science and Technology, Zhenjiang Jiangsu 212003, China
2 College of Electrical and Information Engineering, Hunan University, Changsha Hunan 410082, China
3 Department of Physics and Telecom Engineering, Hunan City University, Yiyang, Hunan 413000, China

Abstract

The multi-step prediction problem of chaotic time series with one sampling delay is investigated in this paper. The delay is considered to be random and is modelled by a binary white noise with values of zero or one, and these values indicate that the observation arrives on time or that it is delayed by one sampling time. Based on the original extended Kalman filtering (EKF) and the Unscented Kalman filtering (UKF), we can obtain corresponding additional two kinds of nonlinear filtering methods with one observation sampling delay which are shortened as DEKF and DUKF in this paper. Using the radial basis function (RBF) neural network prototypes and the network weights as state equation, and the output of RBF neural network to present the observation equation, the input vector to the network is composed of predicted chaotic signal with given length, and the multi-step prediction results are represented by the predicted observation value of nonlinear filtering methods. To show the advantage of DEKF and DUKF, we applied them to the five-step prediction of Mackey-Glass time-series with one sampling delay and compared them with the original EKF and UKF. Experimental results have demonstrated that the DEKF and DUKF are proportionally superior to the original EKF and UKF. Moreover, DUKF is a better choice for Mackey-Glass time-series five-step prediction in comparison with DEKF.

Keywords: Nonlinear Filtering Methods with One Sampling Delay; Chaotic Time-series Multi-step Prediction; RBF Neural Network Approximation

1. Introduction

For time-series long-term forecasting problems, there have been several prediction models to data, but the development of a more accurate model is very difficult because of high non-linear and non-stable relations between input and output data. Therefore, an issue on how to construct multi-step predictive models of time series has been a major research topic bearing immediate and significant practical implications. Several techniques have been used to deal with this challenging subject, such as radical basis function neural network [1, 2], multilayer perceptron neural networks [3, 4], evolutionary recurrent neural network [5], wavelet process neural network [6], partial least squares regression [7] and support vector machine [8-10], which possess the abilities to approximate nonlinear systems.

Different time series prediction approaches such as neural networks and support vector regression, have been successfully used to develop forecasting models. A problem that has not yet received proper attention, however, is how to update such forecasting models when new data arrives, i.e. when a new event of the considered time series occurs. Although online prediction has very important significance for real world

† Corresponding author.

Email addresses: woolcn@163.com (Xuedong WU).

1553-9105/ Copyright © 2011 Binary Information Press
March, 2011
applications, almost all the models at hand are not applicable online [10]. Moreover, chaotic signal may suffer from noises, so noisy time series prediction is attractive and challenging since it is essential in many fields, such as forecasting, modeling, signal processing, economic and business planning.

Considering the online prediction and noise problem of time series, another powerful class of prediction is the stochastic dynamic one such as EKF and UKF [11]. The problem is formulated as an estimation problem and is presented in state space form. Kalman filtering and the least absolute value filtering algorithms are examples of such dynamic approaches. Unlike static approaches, where the whole set of data is used to obtain the optimal solution, dynamic filters are recursive algorithms. In recursive filters, the estimations are updated using each new observation. Dynamic filters are well suited to on-line digital processing as data are processed recursively, and they have been used extensively in estimation problems for dynamic systems. Ma et al. [12] and van der Merwe [13] use the EKF and UKF for one-step prediction of chaotic time series, Zhang et al. [14] propose the particle filtering (PF) algorithm to deal with the non-Gaussian noise in one and two-step prediction. However, all these proposed algorithms [12-14] only consider the uncertain observations perturbed by additive noises.

In standard nonlinear filtering problems such as EKF, UKF and PF, they are usually assumed that the observations generated by the system are available in real time. However, in many practical situations the arrival of the observations can be affected by a random delay. This frequently occurs in communication networks with remote sensors in which, for example, the level of congestion within the network affects the time in which the information sent from the sensor arrives at the processor; thus, the observations available to be filtered may not be updated. So it’s necessary to obtain corresponding nonlinear filtering methods with observation sampling delay based on the original nonlinear filtering methods.

RBF neural network is considered as a good candidate for the prediction problem due to its rapid learning capacity and, therefore, it has been applied successfully to nonlinear time series modeling and forecasts. However, the traditional RBF network encounters the primary problem that the network performance is very likely to be affected by additive noises [15], moreover, the arrival of the observations may be affected by a random sampling delay. Based on the combined use of the RBF neural network and the nonlinear filtering methods with one sampling delay, we are concerned with the multi-step prediction of chaotic time series with one sampling delay in this paper. The selection of the RBF prototypes and the network weights can be viewed as a system identification problem, this paper proposes the use of the nonlinear filter methods with one observation sampling delay for the learning procedure, and the multi-step prediction results are represented by the predicted observation value of nonlinear filtering methods.

In the following sections, we describe how the nonlinear filtering methods with one random sampling delay can be applied to RBF optimization. We demonstrate its performance on five-step prediction of Mackey-Glass time-series with one sampling delay. Section 2 provides the kinematic model of chaotic time-series multi-step prediction, and Section 3 contains simulation results and Section 4 draws some conclusion.
2. The Kinematic Model of Chaotic Time-series Multi-step Prediction

2.1. The Radial Basis Function Neural-network Configuration

An RBF consists of the \( m \)-dimensional input \( \mathbf{u} = [I_1 \ I_2 \ \cdots \ I_m] \) being passed directly to a hidden layer. Suppose there are \( c \) neurons in the hidden layer. Each of the \( c \) neurons in the hidden layer applies an activation function which is a function of the Euclidean distance between the input and an \( m \)-dimensional prototype vector. Each hidden neuron contains its own prototype vector as a parameter. The output of each hidden neuron is then weighted and passed to the output layer. The outputs of the network consist of sums of the weighted hidden layer neurons. Fig. 1 shows a schematic of an RBF network. The response of an RBF of the form of Fig. 1, where the hidden layer functions have the form of

\[
g(s) = \exp[-s^2/\beta^2] \quad \text{(where} \ \beta \ \text{is a real constant)},
\]

can be written as follows:

\[
y = [w_{11} \ \cdots \ w_{1c}] [g\|\mathbf{u} - \mathbf{v}_1\|^2 \ \cdots \ g\|\mathbf{u} - \mathbf{v}_c\|^2]^T = \mathbf{W}[g\|\mathbf{u} - \mathbf{v}_1\|^2 \ \cdots \ g\|\mathbf{u} - \mathbf{v}_c\|^2]^T.
\]

Fig. 1 Radial Basis Function Network Architecture

2.2. Neural-network Training Model Based on the Nonlinear Filtering Methods with One Observation Sampling Delay

Inspired by the successful use of the Kalman filter for training neural networks, we present the chaotic predictive nonlinear model with one observation sampling delay in this section. In general, we can view the optimization of the weight matrix \( \mathbf{W} \) and the prototypes \( \mathbf{v}_j \) as a weighted least-squares minimization problem, where the error vector is the difference between the RBF outputs and the target values for those outputs. Consider the RBF network of Fig. 1 with \( m \) inputs, \( c \) prototypes, and \( n \) outputs. In order to cast the optimization problem in a form suitable for the nonlinear filtering methods, we let the elements of the weight matrix \( \mathbf{W} \) and the elements of the prototypes \( \mathbf{v}_j \) constitute the state of a nonlinear system, and we let the output of the RBF network constitute the output of the nonlinear system to which the nonlinear filter is applied. Then the state of the nonlinear system can be represented as

\[
\mathbf{\theta} = [\mathbf{W}^T \ \mathbf{v}_1^T \ \cdots \ \mathbf{v}_c^T]^T.
\]
The vector $\theta$ thus consists of all $(n(c+1)+mc)$ of the RBF parameters arranged in a linear array. In order to execute a stable nonlinear filtering algorithm, we need to add some artificial process noise $w_k$ and observation noise $v_k$ to the system model. In addition, we use a sequence of independent Bernoulli random variables $\gamma_k$ (binary switching sequence taking the values 0 or 1) to denote the observation which can either be delayed by one sampling period with a known probability $p_k$ or updated with probability $1-p_k$. The nonlinear system model to which the nonlinear filter with one observation sampling delay can be applied is

$$\theta_{k+1} = \theta_k + w_k, \quad y_k = \gamma_k h(\theta_k, u_k) + v_k,$$

where the vector $\theta_k$ is the state of the system at time $k$, $w_k$ is the process noise, $y_k$ is the observation vector, $v_k$ is the observation noise, $\gamma_k$ is a sequence of independent Bernoulli random variables, and $h(\bullet)$ are nonlinear vector functions of the state and input vector.

We assume that at time $k=1$ the real observation $\tilde{y}_1$ is always available for the estimation, but at each time $k > 1$ the observation can either be delayed by one sampling period, with a known probability $p_k$, or updated, with probability $1-p_k$; thus, the available observations at time $k > 1$ is given by

$$y_k = \begin{cases} \tilde{y}_{k,1}, & \text{with probability } p_k \\ \tilde{y}_k, & \text{with probability } 1-p_k \end{cases}$$

It is also assumed that a delay in the observation at time $k$ is independent of delays at times $j \neq k$; therefore, if $\gamma_k$ denotes a sequence of independent Bernoulli random variables (binary switching sequence taking the values 0 or 1) with $P[\gamma_k = 1] = p_k$, the available observations to estimate the signal can be described by

$$y_k = (1-\gamma_k)\tilde{y}_k + \gamma_k\tilde{y}_{k,1}$$

It is clear that the Bernoulli variables in formula (5) model the delays in the following sense: if $\gamma_k = 1$, then $y_k = \tilde{y}_{k,1}$ and the observation is delayed by one sampling period; otherwise, $\gamma_k = 0$ implies that $y_k = \tilde{y}_k$ or, equivalently, that the observation is updated.

Due to the values of the variable $\gamma_k$, the density function of $y_k = \gamma_k h(\theta_k, u_k) + w_k$ has the following mixture form

$$g(y_k) = p_k g(y_k | \gamma_k = 1) + (1-p_k) g(y_k | \gamma_k = 0),$$

where $g(y_k | \gamma_k = 1)$ is the density of the vector $h(\theta_k, u_k) + v_k$ and $g(y_k | \gamma_k = 0)$ is the density of $v_k$. Hence,

$$\begin{aligned}
E[y_k] &= p_k \cdot E[h(\theta_k, u_k)] + (1-p_k) \cdot E[v_k], \\
\text{Cov}[y_k, y_{k'}] &= p_k \cdot \text{Cov}[h(\theta_k, u_k), h(\theta_{k'}, u_{k'})], \\
\text{Cov}[y_k] &= p_k \cdot \text{Cov}[h(\theta_k, u_k)] + (1-p_k) \cdot E[h(\theta_k, u_k)] \cdot E[h^T(\theta_k, u_k)] + R_k.
\end{aligned}$$

Therefore, one way to obtain approximations of the above statistics is to approximate $E[h(\theta_k, u_k)], \text{Cov}[h(\theta_k, u_k)]$, and to substitute the approximations in formula (7). So it is easy to derive the DEKF and DUKF based on the literature [16, 17].

2.3. Prediction Methods

The first input vector to neural-network predictor for prediction is as follows:

$$u_1 = \begin{bmatrix} y_{1} & y_{1+S} & \cdots & y_{1+ms} \end{bmatrix}^T,$$
where $S$ is the prediction steps. These predicted values are symbolized with $y_{p1}, y_{p2}, \cdots$ and then are shifted to input vectors of predictor for next prediction. Then after first $m$-step of prediction, the input regression vector to network is the previous predicted values, each predicted data is shifted to input vector for subsequent prediction. The schematic of prediction method is depicted in Fig. 2.

3. Experimental Setup and Experimental Results

The Mackey-Glass equation originally has been proposed as a model of blood cell regulation, and the Mackey-Glass has been used in literature as a benchmark model due to its chaotic characteristics. In this section, we describe and illustrate the use of nonlinear filter with one observation sampling delay training for the RBF network based chaotic time series five-step prediction problem. The Mackey-Glass chaotic time series is generated by the following differential equation [18]:

$$\frac{dx(t)}{dt} = \frac{ax(t-\tau)}{1+x^m(t-\tau)} - bx(t). \quad (9)$$

The simulated data were obtained by using the fourth-order Runge-Kutta method for formula (9) and the initial conditions were taken to be $a = 0.2, b = 0.1, r = 10, \tau = 17$ and $x(0) = 0.1$, individually.

A RBF neural network, which constitutes three inputs, fifteen prototypes and one output, is trained with EKF, UKF, DEKF and DUKF algorithms in an online fashion, respectively. Moreover, we assume that the probability $p_k$ is constant in the simulation. To enable a fair result of the chaotic time-series five-step prediction proposed in this paper, the prediction is averaged across a Monte Carlo simulation consisting of 50 runs. The MSE (mean squared error) is defined as:

$$MSE = \frac{1}{K} \sum_{k=1}^{K} (y_k - y_{pk})^2, \quad (10)$$

where $y_k$ and $y_{pk}$ are observed and predicted values.

A time series samples set with length of 1120 is generated by formula (9), and we take the last predicted 120 data as experimental results. Fig. 3 (where “Idea” is the real values of Mackey-Glass chaotic time-series, and Fig. 5 is the same) and Fig. 4 plotted the five-step predicted results and absolute error with $p_k = 0.9$, and the five-step predicted results and absolute error with $p_k = 0.5$ are shown in Fig. 5 and Fig. 6. The MSE for different nonlinear filtering algorithms with $p_k = 0.9$ and $p_k = 0.5$ is summarized in Table 1.

4. Conclusions

A new challenge in the field of time-series prediction is the online prediction problem of time series with additive noise and random observation sampling delay. This paper develops an online RBF model to solve
the five-step prediction problem of Mackey-Glass time-series in an advancing time-series based on the nonlinear filtering methods with one sampling delay. The simulation experimental comparison between the original nonlinear filtering methods (EKF and UKF) and the nonlinear filtering methods with one observation sampling delay (DEKF and DUKF) demonstrates the effectiveness and efficiency in Mackey-Glass time-series five-step prediction. Experimental results have shown that the DEKF and DUKF are proportionally superior to the original EKF and UKF. Moreover, DUKF is a better choice in comparison with DEKF.

Table 1 The MSE of Mackey-Glass Time-Series Five-Step Prediction Based on EKF, UKF, DEKF and DUKF with \( p=0.5 \) and \( p=0.9 \)

<table>
<thead>
<tr>
<th>Probability</th>
<th>EKF</th>
<th>GEKF</th>
<th>UKF</th>
<th>GUKF</th>
</tr>
</thead>
<tbody>
<tr>
<td>( p=0.5 )</td>
<td>0.3175</td>
<td>0.2112</td>
<td>0.0017</td>
<td>7.2923e-004</td>
</tr>
<tr>
<td>( p=0.9 )</td>
<td>0.0976</td>
<td>0.0527</td>
<td>0.0003</td>
<td>8.0343e-005</td>
</tr>
</tbody>
</table>

Fig. 3 Five-Step Prediction Results of Mackey-Glass Time-Series with \( p=0.9 \)

Acknowledgement

This work is supported by the State Key Program of National Natural Science of China(Grant No.60835004), the Natural Science Foundation of Jiangsu Province of China (Grant No. BK2009727) and Natural Science Foundation of the Jiangsu Higher Education Institutions of China(Grant No. 10KJB510004), National Natural Science Foundation of China (Grant No. 61075028).
References


