Neurocomputing 82 (2012) 186–195

Contents lists available at SciVerse ScienceDirect

Neurocomputing

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

Prediction for noisy nonlinear time series by echo state network based on dual estimation

Chunyang Sheng, Jun Zhao *, Ying Liu, Wei Wang

School of Control Science and Engineering, Dalian University of Technology, China

ARTICLE INFO

Article history:
Received 29 July 2011
Received in revised form
3 November 2011
Accepted 5 November 2011
Communicated by Z. Wang
Available online 3 January 2012

Keywords:
Echo state network
Dual estimation
Kalman filter
Time series
Prediction

ABSTRACT

When using echo state networks (ESNs) to establish a regression model for noisy nonlinear time series, only the output uncertainty was usually concerned in some literature. However, the unconsidered internal states uncertainty is actually important as well. In this study, an improved ESN model with noise addition is proposed, in which the additive noises describe the internal state uncertainty and the output uncertainty. In terms of the parameters determination of this prediction model, a nonlinear/linear dual estimation consisting of a nonlinear Kalman filter and a linear one is proposed to perform the supervised learning. For verifying the effectiveness of the proposed method, the noisy Mackey Glass time series and the generation flow of blast furnace gas (BFG) in steel industry practice are both employed. The experimental results demonstrate that the proposed method is effective and robust for noisy nonlinear time series prediction.

1. Introduction

Theoretically, recurrent neural networks are universal approximators, and as such, have an excellent ability of approximating any nonlinear mapping to any degree of accuracy [1]. Echo state network, a kind of recurrent neural network, exhibits good performance for the prediction of nonlinear or non-Gaussian dynamic system [2]. The dynamic reservoir of ESN, instead of the hidden layer of generic neural network, involves a large number of sparsely connected neurons that shows a sound memory characteristic; furthermore, only the output connections need to be determined during the learning process, which simplifies the establishment of the network. Recently, such network had been successfully applied to time series prediction [3,4], short-term load forecasting [5], signal processing [6,7] and automatic control [8].

On the other hand, ESNs fundamentally suffer from two basic limitations despite its well popularity. First, the ill-conditioned solution associated with linear regression or recursive least squares method could hardly be avoided in the learning process, and such solution might deviate from the real system. As mentioned in [9], the large output weights impaired the generalization capability of the model that means the model input slightly deviated from the training data such that the relatively poor results would occur. In addition, the model with large weights might lead to the lack of stability when ESN features the output feedback. The second issue is the unsatisfactory low performance when the sample uncertainties involved. Since the core of ESN lies in the dynamic reservoir driven by the external input, the update of internal states often accompany with the uncertainty from input. Thus, an inaccurate target value might be obtained. In practice, the uncertainty in real-world dynamic system is fairly prevalent; for example, consider the situation where the various noises are often introduced into an industrial system by sensors. The low accuracy of ESNs in industrial time series prediction was addressed in [10], and the similar demonstrations had been reported in other fields such as sea clutter prediction [11], nonlinear system modeling [12], and nonlinear filtering [13].

To avoid the ill-condition, the eigenvalue spread of the correlation matrix of reservoir activation signals was proposed in [9], where the ill-condition phenomenon was prevented by adding noise to the reservoir during training so as to promote the stability and robustness of the network. However, this method failed to avoid the ill-condition thoroughly and the additive noise might impair the prediction accuracy. Subsequently, the singular value decomposition [10], the biologically motivated learning method [12] and the swarm intelligence optimization [13] were adopted to train the network. All of these methods avoided the ill-condition and improved the accuracy for noiseless time series more or less, but as for the noisy nonlinear series, the quality of these methods had not been proved or demonstrated. In [10] presented, the empirical mode decomposition was used for noise reduction; yet, the noise-reduced sample still comprised of the uncertainties. In [14] reported, the Bayesian method that focused on the distribution of output weights was

* Corresponding author.
E-mail address: zhaoj@dlut.edu.cn (J. Zhao).

0925-2312/$ - see front matter © 2011 Elsevier B.V. All rights reserved.
doi:10.1016/j.neucom.2011.11.021

employed to perform the network training. Although the prediction accuracy with the Bayesian based ESN was improved thanks to the considered output uncertainty, the uncertainty of the internal states resulted from the intrinsic sample was ignored. In addition, a kind of decoupled ESN based on the thought of multiple reservoirs was proposed in [11] and [15], in which those methods were somewhat suitable for complex problems, but the internal states uncertainty was also not mentioned.

In this study, an improved ESN model with noise addition is proposed to predict the noisy nonlinear time series, in which the uncertainties from internal states and outputs are meanwhile considered in accordance with the industrial practice. For the optimal model parameters, a nonlinear/linear dual estimation method is designed, in which a nonlinear Kalman filter serves to estimate the internal states, and a linear one estimates the output weights. The contribution of this paper lies in the following three aspects. First, the ill-conditioned solution could be avoided by using the dual estimation for the parameters determination. Second, the implementation process of the proposed method is much easier than that of the classical dual estimation based RNN. Finally, due to the consideration of internal states uncertainty, the accuracy and the robustness of the model are greatly enhanced. To demonstrate the accuracy and the robustness of the proposed method, the standard and noisy Mackey-Glass time series are first employed as the testing examples. And, the proposed method is further adopted to predict the generation flow of blast furnace gas in steel industry. The experimental results indicate the proposed method presents a good performance for the prediction of noisy industrial time series.

The rest of the paper is organized as follows. Section 2 presents an improved ESN model with additive noise for prediction and elaborates how to determine the parameters of this model. In Section 3, several types of nonlinear Kalman filters are introduced to estimate the internal states of the established model. And, the corresponding output weights estimation is described in Section 4. Section 5 carries out the two classes of simulation experiments to verify the effectiveness of the proposed method. Finally, we summarize the paper and give the future work in Section 6.

2. Improved ESN based on dual estimation

We review the standard form of ESN first in this section and present an improved version considering the noises involvement.

2.1. Improved ESN with noise addition

Echo state network contains the input layer, the reservoir and the output layer. Many sparsely connected neurons in the reservoir guarantee the echo property of the network [2]. The classical recursive formula of ESN reads as

$$\mathbf{x}_k = f(\mathbf{W}^{in}\mathbf{u}_k + \mathbf{W}\mathbf{x}_{k-1} + \mathbf{W}^{back}\mathbf{y}_{k-1}) \tag{1}$$

$$\mathbf{y}_k = f^{out}(\mathbf{W}^{out}(\mathbf{u}_k, \mathbf{x}_k, \mathbf{y}_{k-1})) \tag{2}$$

where $\mathbf{u}_k$ is the exogenous input; the dimension of input is equal to m. $\mathbf{x}_k$ is the internal states, the dimensionality of $\mathbf{x}_k$ is N. $\mathbf{y}_k$ is the output, the dimensionality of $\mathbf{y}_k$ is L. $\mathbf{W}^{in} = (W_{i,j}^{in}) \in \mathbb{R}^{N\times m}$ denotes the input weights; $\mathbf{W} = (W_{i,j}) \in \mathbb{R}^{N\times N}$ denotes the internal weights of the neurons in reservoir. To provide proper memorization capabilities, $\mathbf{W}$ should be sparse whose connectivity level ranges from 1% to 5% and its spectral radius should be less than 1; and $\mathbf{W}^{out} = (W_{i,j}^{out}) \in \mathbb{R}^{L\times(m+N+L)}$ denotes the output weights. f and f^{out} are the activation functions of internal neurons and output neurons, respectively.

When using echo state network to establish a regression model for noisy time series, the output uncertainty used to be considered in the existing literature. In general, the independent Gaussian noise sequences reflecting the difference between the observation and the expected output are introduced into the output formula (2). However, as for as noisy time series, the internal states are still uncertain. In this study, an improved ESN considering the noise addition becomes

$$\mathbf{x}_k = f(\mathbf{W}^{in}\mathbf{u}_k + \mathbf{W}\mathbf{x}_{k-1}) + v_{k-1} \tag{3}$$

$$y_k = \mathbf{W}^{out} \cdot [\mathbf{u}_k, \mathbf{x}_k] + n_k \tag{4}$$

where y_k is a scalar quantity that shows the network is a single-output model, $v_{k-1} \in \mathbb{R}^{N\times 1}$ and n_k are independent white Gaussian noise sequences. And, $\forall k$, $E[v_{k-1}] = E[n_k] = \mathbf{0}$; $\forall i,j$, $E[v_i v_j^T] = \mathbf{R}^v$, $E[n_i n_j^T] = \sigma_n^2$. In (3), $v_{(k-1)}$ reflects the uncertainty of internal states, and n_k reflects the output uncertainty in (4).

2.2. The dual estimation

For the established model above, both the internal states and the output weights are unknown. Hence, it is a tough task for traditional learning methods, such as linear regression or recursive least squares method, to estimate the internal states and output weights

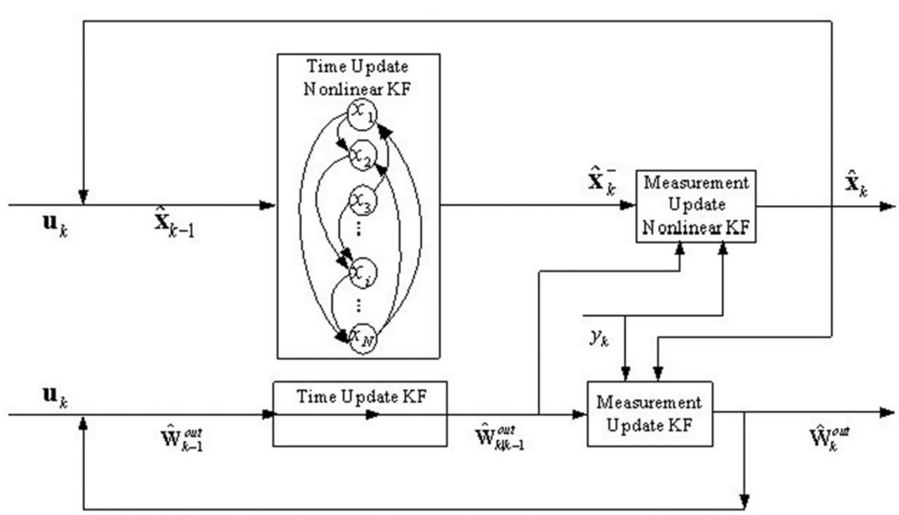

Fig. 1. The proposed dual estimation architecture.