Noisy Chaotic Time Series Prediction Based on Wavelet Echo State Network

Fei Yang, Binxing Fang, Chunlu Wang, Xingquan Zuo, Ruiming Zhong

School of Computer Science, Beijing University of Post and Telecommunications, Key Laboratory of Trustworthy Distributed Computing and Service of the Ministry of Education of China, Beijing 100876, China;
1 Nanjing Panda Electronics Company limited, Nanjing 210002, China

1 yangfei800103@163.com 2 fangbx@bupt.edu.cn 3 wangcl@bupt.edu.cn 4 zuoxq@bupt.edu.cn 5 ruiming_zhong@163.com

Abstract

As a research focus of intelligence algorithm, the prediction of classic noiseless chaotic time series has a great development in recent years. However, the existing prediction models cannot get good performance for real-world chaotic time series because of the interference of noise components. In order to take full advantage of the property of real-world chaotic time series, the paper proposes a novel prediction model based on wavelet transform and echo state network (WESN). The basic idea of WESN is that firstly wavelet decomposition is used to separate the chaotic dynamics component and noise components, then the gotten components can be predicted by echo state network (ESN) independently, and finally the prediction results of time series are obtained by assembling the prediction values of all components. By using real-world sunspot time series for verification, the prediction results show that the proposed model has higher prediction accuracy by comparing with the models of direct echo state network (DESN), and direct echo state network (DSVM).

Keywords: Echo State Network, Noisy Chaotic Time Series, Wavelet Transform, Decomposition Scale

1. Introduction

With development of chaos theory, many of natural phenomena such as sunspots [1], traffic flow [2], weapon equipment [3] and electricity system [4], could be understood afresh. Considered the significant value of these special real-world time series, it is a popular and challenging problem to explore a high accuracy model to predict them recently. As a research hotspot of machine learning, prediction of the classic chaotic time series, which are generated by deterministic systems, has been achieved by many advanced models, such as radial basis function (RBF) neural network [5], support vector machine (SVM) [6], RPNNs [7] and echo state networks (ESN) [8]. Among these prediction models, ESN becomes one of the most interested models for many scholars, and the prediction accuracy of Mackey-Glass chaotic time series in Science (2004) is improved by more than 2000 times over other models [8]. As previous researches shown, these traditional models including ESN are suitable for predicting noiseless chaotic time series, but they cannot get the satisfactory accuracy for predicting real-word time series because of the interference of heavy-noise components [9].

In the pass years, different methods have been implemented for removing the noise to improve the prediction accuracy of real-world time series, such as linear filter and nonlinear filter. However, the noise reduction methods usually used are difficult to upgrade the prediction performance radically, because it will result in the loss of chaotic property of originate time series [10]. Recently, the wavelet theory has been widely studied and applied in signal process. Different from the methods of simple noise reduction, the multi-scale decomposition method based on wavelet transform can effectively separate the signal components and noise components of time series [11], then the every generated component including noisy component can be treated with different prediction models in order to raise the accuracy.

By combining the merits of multi-scale decomposition method and echo state network, a universal model called as WESN is proposed in this paper to predict noisy chaotic time series. Firstly, the feature of wavelet transform is utilized to decompose the main signal component, expressing low-frequency property and minor multiple noise components expressing high frequency property. Subsequently, on
basis of analyzing the property of each component, the adaptive ESN models, which are very effective for predicting noiseless chaotic series with high accuracy, are used to predict each component. Finally, the prediction result of time series is got by assembling the predicted values of all components. To verify effectiveness of the proposed model, it is applied for predicting sunspots time series, which are considered as a classical real-world chaotic series with noisy chaotic property and used usually to examine the model performance, and is compared with the other classical models, including DESN, and DSVM. In addition, the key influence factors are analyzed to reveal the property of WESN model.

The remaining parts of this paper are organized as follows: Section 2 gives an overview of DESN for time series prediction. In the section 3, we present the WESN model. The comparative experiments and the analysis of key influence factors of WESN are given in section 4. In section 5, we give conclusions of this paper.

2. Direct echo state network

2.1. ESN architecture

Without loss of generality, ESN is a new neural network with classical three layers architecture, including $K$ input units, $N$ internal units called reservoir and $L$ output units as shown in Fig.1. The value of the input, internal and output units at the sampling time $n$ are denoted simply by $u(n) = (u_1(n), \ldots, u_K(n))^T$, $x(n) = (x_1(n), \ldots, x_N(n))^T$, $y(n) = (y_1(n), \ldots, y_L(n))^T$, respectively, and $T$ implies transposed inversed-matrix.

The update of the reservoir state and observed output of ESN are denoted as follows:

$$x(n) = f(W^{in}u(n) + Wx(n-1))$$

(1)

$$\hat{y}(n) = f^{out}(W^{out}x(n))$$

(2)

where $f$ is the activation function of reservoir (typically sigmoidal type), and $f^{out}$ is the output activation function of the network (typically sigmoidal or linear type). $W^{in}$, $W$ and $W^{out}$ are the weight matrices for input connections, the reservoir and output connections, respectively [12]. The elements of $W^{in}$ and $W$ are fixed prior to training with random values drawn from a uniform distribution over a symmetric range.

Different from traditional neural networks, the kernel reservoir of ESN is full of neurons with large number and sparse connection, so the special architecture of reservoir makes it have powerful dynamic to approximate the input signal. As shown from Eq. (1), the current state of reservoir is always associated with its previous state. This kind of property is called short-term memory, making ESN be extraordinary suitable for time series prediction. Meanwhile, on a benchmark task of predicting the classical Mackey-Glass chaotic time series, the performance of reservoir is influenced by some important parameters, including reservoir size, spectral radius,
sparseness, and input weight connections, so these parameters need to be set appropriately before ESN is used.

2.2. Determine input and output by phase space reconstruction

According to chaos theory, chaotic time series can be described completely by mathematical method, named phase space reconstruction. The main principle of phase space reconstruction is explained briefly as follow. According to the Taken principal [13], each point of chaotic series is able to be replaced by a multi-dimensional vector which has enough dimensions to include all the property of the point. Therefore, a single dimensional time series can be mapped into a reconstructed multi-dimensional vector to obtain the better dynamic feature.

Under the phase space reconstruction, the chaotic time series \{d(n) | n = 1, 2, \cdots, b\} can be reconstructed as follow:

\[ D(n) = [d(n), d(n - \tau), \cdots, d(n - (m - 1)\tau)]^T \]

where \(D(n)\) is the phase point of \(d(n)\); \(\tau\) and \(m\) are best delay time interval and minimum embedding dimension, respectively. Here, we adopt the \(D(n)\) as the input \(u(n)\) of ESN model in Eq. (1), so the number of input units is equal to \(m\).

In order to perform \(s\)-step prediction of time series, the direct prediction method is adopted [9] and the number of output units \(T\) is settled to 1, i.e. the desired output of ESN model is

\[ y(n) = d(n + s) \]

where \(y(n)\) is \(s\)-step desired prediction values of time series.

2.3. Training and predicting

Due to the invariance of \(W^{in}\) and \(W\) during the training, the \(W^{out}\) is the only weight matrix needed be trained. Considering a supervised task on a pair of time series which can be used to compute the output weight \(W^{out[12]}\), the training set \(n_1\) of time series from the sampling time \(T_0\) to \(T_1\) is expressed by \((u(n_1), y(n_1))_{n_1=1}^{T_1}\), where \(u(n_1) = [d(n_1), \cdots, d(n_1 - (m - 1)\tau)]\) is the training input series and \(y(n_1) = d(n_1 + s)\) is the desired output series. The training risk is to adjust the weights of matrix \(W^{out}\) to minimize the averaged squared error of training set. Therefore, the training output is computed by a simple linear regression method to minimize the squared error

\[ \| W^{out}X - Y \|^2 \]

where \(X = (x(T_0), \cdots, x(T_1))\) and \(Y = (y(T_0), \cdots, y(T_1))\) are the states of reservoir and the desired output series. Therefore, the training weight \(W^{out}\) can be computed by \(M^+\) which is the pseudo-inversion matrix of \(X\) and denoted by

\[ W^{out} = (M^+Y)^T \]

According to Eq. (1) and (2), the output of ESN can be computed by

\[ \hat{y}(n_2) = f^{out}(W^{out}x(n_2)) \]

where \(\hat{y}(n_2)\) is the predicted value of time series, \(n_2 = T_1 + 1, \cdots, b\) is the predicted sampling time.

3. Proposed model

Due to the interference of heavy-noise components, the DESN model is not very suitable to predict noisy chaotic time series. To overcome the shortage of DESN model, in this paper we propose the WESN model by integrating the wavelet theory and ESN to predict chaotic time series with noise. The prediction process includes the following four steps:

**Step 1** Decompose time series \(d(n)\) by wavelet. Based on MRA [14], after \(M\)-scale decomposition, the time series \(d(n)\) can be expressed by a linear combination of scaling function \(\phi(n)\) and wavelet function \(\psi(n)\)

\[ d(n) = \sum_k d_{M,k}\phi_{M,k}(n) + \sum_{j=1}^M \sum_k d_{j,k}\psi_{j,k}(n) \]
where $p_j(t)$ is the approximation coefficients of the $j$-scale, and mainly presents the clear chaotic signal with low-frequency property, $q_j(t)$ is the detail coefficients of the $j$-scale, and mainly presents the noisy signal with high-frequency property, $k$ is the time shift exponent. $\phi(n)$ and $\psi(n)$ can be determined by mother function and denoted by

$$
p_j(t) = \sum_k h(k-2t)p_{j-1}(k)
$$

$$
q_j(t) = \sum_k g(k-2t)p_{j-1}(k)
$$

where $h$ and $g$ are the high-pass and low-pass filters respectively.

**Step 2)** Single branch reconstruction for time series. According to MRA theory, the approximation coefficients of the $j-1$-scale can be reconstructed by

$$
p_{j-1}(t) = \sum_k \tilde{h}(t-2k)p_j(k) + \sum_k \tilde{g}(t-2k)q_j(k)
$$

where $\{\tilde{h}, \tilde{g}\}$ is synthesis filter bank, corresponding to $\{h, g\}$, which is usually called as analysis filter bank. By utilizing the approximation coefficients and detail coefficients of each scale, the time series can be reconstructed approximately (losing very few energy) by a single low-frequency component $a_M(n)$ and multiple high-frequency components $d_j(n)$

$$
d(n) \approx a_M(n) + \sum_{j=1}^M d_j(n)
$$

**Step 3)** Multiple components prediction of ESN. According to private property, each component in step 2 can be predicted independently by the DESN model. Firstly, the input and $s$-step output of each component is established by phase space reconstruction method; then the weight readout $W_{\text{out}}$ of the $i$th component can be determined by the training approach; finally the $s$-step prediction result of the $i$th component is computed and denoted by:

$$
Y_i(n_2) = f_{\text{out}}(W_{\text{out}}x_i(n_2))
$$

where $i = (1, 2, \ldots, M + 1)$ is the number of each component.

**Step 4)** Integrate all components. After the prediction of each component, the final $s$-step prediction result can be obtained by assembling the predicted result of each component

$$
\hat{y}(n_2) = \sum_{i=1}^{M+1} Y_i(n_2)
$$

where $\hat{y}(n_2)$ is the final observed prediction result of WESN.

4. Experiments

4.1. Experimental data and analysis

Real-world sunspots are the phenomenon of continuous changing of solar. It has a great influence on the climate and hydrology of Earth, so it is very significant to research it. Because of its extensive noisy chaotic property, the sunspots prediction problem is usually use to examine the performance of prediction models.

The sunspot number data used in this paper are monthly mean sunspots number. In order to contrast easily, we use sunspots data which are obtained from [7]. Using the same set in [7], a total of 3000 sunspots data are chose for the prediction task in the below experiments. 1400 data is used as training data, and the remaining 1600 data as test data.

Using mutual information method [15] and G-P algorithm [16], the delay time $\tau$ and minimum embedding dimension $m$ are given by 10 and 3 respectively. According to Eq. (3), the input of prediction model is denoted by

$$
D(n) = [d(n), d(n - 10), d(n - 20)]^T
$$

4.2. Evaluation criterion
To evaluate the prediction model, the error of unbiased mean square error (\(\varepsilon_{\text{rmse}}\)) is used in this study and denoted by:

\[
\varepsilon_{\text{rmse}} = \sqrt{\frac{1}{(b - T_1)\sigma^2} \sum_{n_2 = T_1 + 1}^{b} (\hat{y}(n_2) - y(n_2))^2}
\]  \hspace{1cm} (16)

where \(y(n_2)\) and \(\hat{y}(n_2)\) are the desired prediction value from original sunspots series and the observed prediction values respectively; \(\sigma^2\) is the variance of the series \(y(n_2)\) and \(b\) are equal to 3000 and 1400 respectively according to section 4.1.

4.3. Experimental result and analysis

The sunspots data are rescaled in the range (-1,1) before they are input into prediction models. Three experiments are completed in this study. First, we use the WESN model to predict sunspots time series. Then, the prediction results of WESN are compared with other two classical prediction models, i.e. DESN, and DSVM. Finally, we analyze the key influence factors to excavate the property of the WESN.

4.3.1 Sunspots prediction based on WESN model

Under the conditions of wavelet ‘db35’ base and decomposition scale 5, the original sunspots time series are decomposed as shown in Fig.2, where the \(a_j(n)\) and \(d_j(n)\) are the low-frequency and high-frequency components of scale-\(j\) respectively.

According to Eq.(12), the sunspots time series \(d(n)\) can be denoted by:

\[
d(n) \approx a_5(n) + \sum_{i=1}^{5} d_i(n)
\]  \hspace{1cm} (17)

So the prediction of the sunspots time series \(d(n)\) is transformed into predicting each of components, named \(d_1(n), \cdots, d_5(n)\) and \(a_5(n)\).

ESN models for predicting each component can be divided into two kinds of patterns as shown in Table 1. Since little noise exists, the low-frequency component \(a_5(n)\) is predicted by pattern I with larger size of reservoir to achieve a high accurate prediction, while the high-frequency components \(d_1(n), \cdots, d_5(n)\), which contain a plenty of noise, are predicted by pattern II with less size of reservoir to avoid over-fitting. Because of having little effect on the prediction performance, the spectral radius, sparseness and input weight connections are set to 0.75 (smaller than 1 to guarantee the stability of ESN), 10% and uniform over (-0.25,0.25) according to the papers [17] and [12].

<table>
<thead>
<tr>
<th>Table 1. Parameters setting of WESN</th>
</tr>
</thead>
<tbody>
<tr>
<td>Items</td>
</tr>
<tr>
<td>Pattern I</td>
</tr>
<tr>
<td>Pattern II</td>
</tr>
</tbody>
</table>
The twenty-month value (predicted step $s = 20$) is taken as example to be predicted by WESN model. Because of the transient effect of ESN [8,12], the first 100 points of prediction results are discarded, so the remaining 1500 sampling points is used as test data to predict.

Fig. 3 shows the prediction results based on the WESN model, including the contrastive curves and prediction error. From Fig. 3(a), we can see that the difference of desired values and observed values is very small. It means that prediction error is very small and the prediction result is quite accuracy. The Fig. 3(b) gives the actual prediction error of WESN model.

**4.3.2 Performance comparison**

In order to further illustrate the effectiveness of this proposed model, the WESN is compared with two classical models, i.e. DESN and DSVM. DESN model is given by the same parameters as pattern I in Table 1, and DSVM model adopts three loss functions (quadratic, $\varepsilon$-insensitive and Huber) and two
kernel functions (RBF, polynomial). The three models predict the multi-step sunspots time series, including the six-month, ten-month, fifteen-month, and twenty-month values, i.e. predicted step $s = 6, 10, 15, 20$ respectively.

<table>
<thead>
<tr>
<th>Model</th>
<th>DESN</th>
<th>DSVM</th>
<th>WESN</th>
</tr>
</thead>
<tbody>
<tr>
<td>Loss function</td>
<td>- Quadratic</td>
<td>ε-insensitive</td>
<td>Huber</td>
</tr>
<tr>
<td>Kernel function</td>
<td>RBF Poly</td>
<td>RBF Poly</td>
<td>RBF Poly</td>
</tr>
<tr>
<td>$s = 6$</td>
<td>10.871 3.659</td>
<td>3.485</td>
<td>3.335</td>
</tr>
<tr>
<td>$s = 20$</td>
<td>27.091 15.689</td>
<td>16.469</td>
<td>15.15</td>
</tr>
</tbody>
</table>

The prediction error $\varepsilon_{\text{rmse}}$ of the three models are summed in the Table 2. Among the three models, DESN get the worst prediction results performance, while the better prediction results of DSVM is to adopt ε-insensitive loss function and RBF kernel function. The WESN model can get obviously the better prediction results than the other two models.

### 4.3.3 Key parameters analysis of WESN

In this part, the effect of some parameters on WESN model performance is analyzed to reveal the property of WESN model. Due to possess most of energy (>98%) of sunspots time series, the low-frequency component is the key composition to influence the final prediction results. By analyzing carefully, the size of reservoir for predicting low-frequency component and scale for decomposing the sunspots time series are both the most important factors to the low-frequency component, so these two parameters will directly affect the prediction performance of WESN. So the following two experiments are designed to analyze the sensitivity of reservoir size and decomposition scales. The twenty-month value is used as the predicted step again.

#### (1) Reservoir size

Given the fixed parameters, the 5-scale is used in this experiment. We vary the reservoir size of ESN in pattern I of Table 1, from 20 to 120 with the interval 20, and remain other parameters unchanged. The prediction error is shown in Fig.3.

![Fig.3 Prediction error change with reservoir size](image)

It can be seen from Fig.3 that: with the increase of reservoir size, the prediction error $\varepsilon_{\text{rmse}}$ continues to reduce until the reservoir size is larger than 100. In this case, the prediction accuracy cannot be improved even the size of reservoir is increased. This is because that the low-frequency series is approximated completely to get the enough prediction accuracy when the reservoir size is large enough, the prediction error is caused mainly by high-frequency components. In this case, the prediction accuracy is hard to be improved reduce by increasing the reservoir size.
Decomposition scale
Given the fixed parameters, the parameters of WESN for predicting low-frequency and high-frequency components are set as Table 1 in this experiment. We vary the decomposition scale from 1 to 7 with the interval 1 to test the sensitivity of WESN model.

As shown in Fig.4, the error curve illustrates that the decomposition scale has important influence on the prediction performance. With the increase of decomposition scale, the prediction error $\varepsilon_{r.m.s.}$ continues to reduce until the decomposition scale arrives at 5 or 6. Hence the prediction accuracy cannot be improved continuously by enlarging decomposition scale. The reason is that the low-frequency component, which possesses most of energy, has enough noiseless dynamic property to be predicted accurately when the decomposition scale is large enough. We can observe that the prediction error will not change largely although the decomposition scale increases. This is because that the high-frequency components of higher scale have the low-frequency property such as the component $d_5$ in Fig.3 and can be predicted with high accuracy. The prediction error is caused mainly by the high-frequency components of lower scale, so the prediction performance is stable even the decomposition scale increases constantly.

5. Conclusion
In this paper, a new model that combines wavelet transform and echo state network, is proposed to predict chaotic time series with noise. Using the multi-scale property of wavelet to decompose time series, each component can be predicted respectively by ESN model according to private property to get good performance. The method is applied to predict classical real-world noisy chaotic sunspots time series, its effectiveness is verified completely by comparing with prediction results of DESN, DSVM. Finally, the sensitivity of key parameters of WESN is analyzed to discuss the influence of key factors on the performance of WESN. In addition, the proposed model can also better meet the requirements to predict other similar noisy chaotic time series.

6. Acknowledgement
This research was supported by Chinese Universities Scientific Fund under Grant No. 2009RC0208.

7. References


