Vol 17 No 6, June 2008
1674-1056/2008/17(06)/1998-06

Chinese Physics B

© 2008 Chin. Phys. Soc.
and IOP Publishing Ltd

Time series prediction using wavelet process neural network*

Ding Gang(丁 刚)†, Zhong Shi-Sheng(钟诗胜), and Li Yang(李 洋)

School of Mechatronics Engineering, Harbin Institute of Technology, Harbin 150001, *China*

(Received 6 October 2007; revised manuscript received 29 October 2007)

In the real world, the inputs of many complicated systems are time-varying functions or processes. In order to predict the outputs of these systems with high speed and accuracy, this paper proposes a time series prediction model based on the wavelet process neural network, and develops the corresponding learning algorithm based on the expansion of the orthogonal basis functions. The effectiveness of the proposed time series prediction model and its learning algorithm is proved by the Mackey–Glass time series prediction, and the comparative prediction results indicate that the proposed time series prediction model based on the wavelet process neural network seems to perform well and appears suitable for using as a good tool to predict the highly complex nonlinear time series.

Keywords: time series, prediction, wavelet process neural network, learning algorithm
PACC: 0545, 0270

1. Introduction

Time series prediction plays an important role in various fields, ranging from economics through physics to engineering.[1] The previous time series prediction methods mainly include statistical models fitting and traditional artificial neural networks. The statistical models fitting methods like Auto-Regression (AR), Moving Average (MA), Auto-Regression Moving Average (ARMA) are all linear mapping ones. There are only a few nonlinear models such as double linear model, yet they are limited in the degree, which makes it hard to find suitable expression for the varying process in the complex systems. In 1989, Hornik and Funahashi proved respectively that multilayer feedforward neural networks can approximate to any continuous function.[2,3] It seems that artificial neural networks have great potential to high nonlinear and uncertain systems. Nowadays, artificial neural networks were widely adopted to predict the nonlinear time series.[4−7]

However, in real world, the inputs of many complicated systems are time-varying functions or processes. It is difficult for traditional artificial neural networks to predict the outputs of these systems accurately and directly, due to the pact that inputs of the traditional artificial neural networks are all instantaneous discrete values. To overcome this limitation, He and Liang proposed artificial process neuron model in 2000.[8] From a point view of architecture, process neuron model is similar to traditional artificial neuron model. The major characteristics, which distinguish process neuron from traditional artificial neuron, are that inputs, outputs and corresponding connection weights of process neuron can be time-varying functions. Process neural networks (PNN) are comprised of densely interconnected process neurons, and the special characteristics of the process neuron make the PNN very suitable to simulate the aforementioned complicated systems.[9−11]

In order to optimize the performance of the PNN, the architecture of the PNN must be selected. It is well known that wavelets offer an adequate framework for the representation of 'natural' signals and images that are described by piecewise smooth functions, with rather sharp transitions between neighbouring domains. This results in a local correlation structure of signal samples, which decays fast for farther apart ones. A similar structure exists in the frequency domain, corresponding to the band structure of the natural sources spectra. Wavelet representations are adequate building blocks for rather general

*Project supported by the National Natural Science Foundation of China (Grant No 60572174), the Doctoral Fund of Ministry of Education of China (Grant No 20070213072), the 111 Project (Grant No B07018), the China Postdoctoral Science Foundation (Grant No 20070410264), and the Development Program for Outstanding Young Teachers in Harbin Institute of Technology (Grant No HITQNJS.2007.010).
†E-mail: dingganghit@163.com

functions of the mentioned type.[12] Combining the advantages offered by PNN, on one hand, has wavelet representation, on the other, is an attractive idea. According to this analysis, a time series prediction model based on wavelet process neural network (WPNN) is proposed in this paper to predict nonlinear time series.

The plan of this paper is as follows: In Section 2, the wavelet process neuron model is presented, which sets the foundation for the remainder of this paper. In Section 3, a time series prediction model based on the WPNN is proposed. In Section 4, a corresponding learning algorithm for the proposed time series model is developed. To simplify the computational complexity of the learning algorithm, appropriate orthogonal functions are selected as basis functions to expand the input functions and the corresponding weight functions of the WPNN at the same time. In Section 5, the effectiveness of the proposed time series prediction model based on the WPNN and its learning algorithm is proved by the Mackey–Glass time series prediction, and the comparative prediction results highlight the performance of the proposed time series prediction model. Conclusions are given in Section 6.

2. Wavelet process neuron

The wavelet process neuron model is composed of three sections: inputs, an activation unit and output. This is based on the fact that a biological neuron is composed of three basic parts: a dendrite, a soma and an axonal tree. Generally, the inputs and connection weights of the neuron in traditional neural networks are discrete values. However, the inputs and connection weights of the wavelet process neuron are continuous time-varying functions. An aggregation operator on time is added to the wavelet process neuron, which provides the WPNN with the capability of handling two dimensional information of time and space simultaneously. The architecture of the wavelet process neuron is depicted in Fig.1.

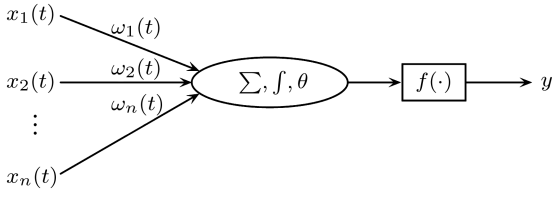


Fig.1. The sketch diagram of wavelet process neuron model.

The output of the wavelet process neuron model can be expressed as

$$y = f\left(\sum_{i=1}^{n}\int_0^T \omega_i(t)x_i(t)\mathrm{d}t - \theta\right). \tag{1}$$

Where $x_i(t) \in C[0,T]$ is the ith input function, $i = 1,2,\ldots,n$, $C[0,T]$ denotes the space of the continuous functions, $\omega_i(t)$ is the ith weight function, θ is the threshold, $f(\cdot)$ is the activation function, which is related to the wavelet transform theory.

Wavelet transform is a classic mathematical theory which is considered as a great breakthrough following Fourier analysis, and has been widely used in various fields. Suppose that $\psi(x) \in L^2(R)$ satisfies the following admissibility condition:

$$C_\psi = \int_R \frac{\left|\hat{\psi}(\omega)\right|^2}{|\omega|}\mathrm{d}\omega < \infty. \tag{2}$$

Where $\hat{\psi}(\omega)$ denotes the Fourier transform. Take $\psi(x)$ as a mother wavelet, then it can generate a family of functions by the operation of dilation and translation.[13] Where $\psi_{a,b}(x)$ can be expressed as

$$\psi_{a,b}(x) = \frac{1}{\sqrt{a}}\psi\left(\frac{x-b}{a}\right). \tag{3}$$

Where the parameters a and b are the dilation and translation parameters expressed in real number R, respectively. The basic idea behind the wavelet transform is to represent arbitrary function $f(x)$ as a superposition of wavelets. The continuous wavelet transform of $f(t)$ is given by

$$W_f(a,b) = (f,\psi_{a,b}) = \frac{1}{\sqrt{|a|}}\int_{-\infty}^{+\infty}\psi_{a,b}(x)f(x)\mathrm{d}x. \tag{4}$$

In terms of wavelet transformation theory, the inverse continuous wavelet transform can be obtained through the following formula

$$\begin{aligned} & f(x) \\ = & \frac{1}{C_\psi}\int_0^{+\infty}\frac{\mathrm{d}a}{a^2}\int_{-\infty}^{+\infty}W_f(a,b)\psi_{a,b}(t)\mathrm{d}b \\ = & \frac{1}{C_\psi}\int_0^{+\infty}\frac{\mathrm{d}a}{a^2}\int_{-\infty}^{+\infty}W_f(a,b)\frac{1}{\sqrt{a}}\psi\left(\frac{t-b}{a}\right)\mathrm{d}b. \end{aligned} \tag{5}$$

3. Time series prediction model based on WPNN

A time series is a sequence of vectors, $x(m)$, $m = 0,1,\ldots$, where m represents elapsed time. For-

mally, the time series prediction based on traditional artificial neural networks can be stated as: find a function $G : R^N \to R$ such as to obtain an estimate of x at time $m+h \quad (h>0)$, from the N time steps back from time m, so that: $x(m+h) = G(x(m), x(m-1), \ldots, x(m-N+1))$. Normally h will be one, so that $G(\cdot)$ will forecast the next value of x. It is obvious that the inputs of the traditional artificial neural networks are discrete values, it is difficult to express how $x(m-1)$ influences $x(m)$ and how $x(m-N+1)$ influences $x(m-N)$at the same time, and so on. This is a shortcoming of the time series prediction method based on traditional artificial neural networks, which makes it have low prediction accuracy. In order to overcome this limitation, a novel time series prediction method based on the WPNN is proposed in this section. The inputs and the corresponding connection weights of the WPNN are continuous time-varying functions. $(x(m), x(m-1), \ldots, x(m-N+1))$ can be used to generate a continuous function $x_m(t)$ by some mathematical method, where $t \in [m-N+1, m]$. Then, the time series prediction based on the WPNN can be stated as: find a functional $F : C([m-N+1, m]) \to R$ such as to obtain an estimate of x at time $m+1$ on $[m-N+1, m]$, so that $x(m+1) = F(x_m(t))$.

The topological architecture of the proposed time series prediction model based on the WPNN is depicted in Fig.2.

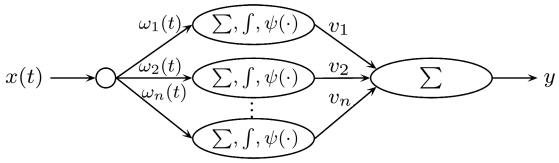


Fig.2. The topological architecture of the time series prediction model based on WPNN.

From Fig.2, it can be seen that the time series prediction model based on the WPNN is comprised of three layers. The first layer is the input layer, which has only one node. The second layer is the hidden layer, which is composed of n wavelet process neurons. The last layer is the output layer, which also has only one node.

Let $x(t) = x_m(t)$, $m = T$ and $N = T+1$, then the estimate of $x(m+1)$ by the time series prediction model based on the WPNN can be expressed as:

$$y = \sum_{i=1}^{n} v_i \psi \left(\frac{\int_0^T \omega_i(t) x(t) \mathrm{d}t - b_i}{a_i} \right). \qquad (6)$$

Where $\omega_i(t)$ is the connection weight function between the i-th wavelet process neuron in the hidden layer and the input unit. v_i is the connection weight between the i-th wavelet process neuron in the hidden layer and the output unit. $\psi(\cdot)$ is the wavelet basis function of the wavelet process neurons in the hidden layer.

4. Learning algorithm

Suppose that $d = x(m+1)$, and y is the corresponding estimate of $x(m+1)$ by the proposed time series prediction model based on the WPNN. Thus, the mean square error of the proposed model can be defined below

$$\begin{aligned} E &= \frac{1}{2}(y-d)^2 \\ &= \frac{1}{2}\left(\sum_{i=1}^{n} v_i \psi \left(\frac{\int_0^T \omega_i(t) x(t) \mathrm{d}t - b_i}{a_i} \right) - d \right)^2 . \end{aligned} \qquad (7)$$

It is obvious that Eq.(7) is very difficult to be implemented by the computer directly. In order to raise the efficiency of the computation and the adaptability to practical problems resolving of the proposed time series prediction model based on the WPNN, a set of appropriate orthogonal basis functions are introduced into the input space of the WPNN.

Let $x(t) = x_m(t)$, by Weierstrass approximation theorem, if any $\varepsilon > 0$ is given, then there exists a polynomial $P(t)$ on $[0, T]$ such that$|x(t) - P(t)| < \varepsilon$ for all $t \in [0, T]$. In short, any $x(t) \in C[0, T]$ can be uniformly approximated on $[0, T]$ by polynomials $P_k(t)$ $(k = 1, 2, \ldots)$ to any degree of accuracy. Therefore, $x(t)$ can be written in the form $x(t) = \sum_{k=1}^{K} a_k P_k(t)$, where the coefficient $a_k \in R$ is easy to find. Suppose that the set of $P_k(t)$ is independent. There is a relationship between orthogonality and independence. It is possible to convert a set of independent functions $P_k(t)$ into a set of orthogonal functions $b_k(t)$ that spans the same space. The standard procedure to accomplish this conversion is called Gram–Schmidt orthogonalization. Therefore, the input function $x(t)$ and the connection weight function $\omega_i(t)$ can be expanded respectively as $x(t) = \sum_{k=1}^{K} c_k b_k(t)$ and $\omega_i(t) = \sum_{k=1}^{K} \omega_{ik} b_k(t)$, $c_k, \omega_{ik} \in R$, at the same time.

According to the property of the orthogonal basis

functions, $\int_0^T b_k(t)b_l(t)\mathrm{d}t = \begin{cases} 1 & l=k \\ 0 & l\neq k \end{cases}$, thus, Eq.(6) can be substituted by

$$y = \sum_{i=1}^{n} v_i \psi \left(\frac{\sum\limits_{k=1}^{K} \omega_{ik} c_k - b_i}{a_i} \right). \tag{8}$$

This allows us to conveniently write out an expression for mean square error of the proposed model as

$$\begin{aligned} E &= \frac{1}{2}(y-d)^2 \\ &= \frac{1}{2}\left(\sum_{i=1}^{n} v_i \psi \left(\frac{\sum\limits_{k=1}^{K} \omega_{ik} c_k - b_i}{a_i} \right) - d \right)^2. \end{aligned} \tag{9}$$

The learning algorithm will adjust v_i, ω_{ik}, a_i and b_i to minimize the mean square error of the proposed time series prediction model based on the WPNN according to Eq.(9).

For the convenience of analysis, Z_i is defined as $Z_i = \dfrac{\sum\limits_{k=1}^{K} \omega_{ik} c_k - b_i}{a_i}$. Applying gradient descent on this estimate of the mean square error, we have

$$\begin{aligned} \Delta v_i &= -\frac{\partial E}{\partial v_i} = -(y-d)\psi(Z_i), \\ \Delta \omega_{ik} &= -\frac{\partial E}{\partial \omega_{ik}} = -(y-d)v_i\psi'(Z_i)\frac{c_k}{a_i}, \\ \Delta a_i &= -\frac{\partial E}{\partial a_i} = -(y-d)v_i\psi'(Z_i)\left(-\frac{Z_i}{a_i}\right), \\ \Delta b_i &= -\frac{\partial E}{\partial b_i} = -(y-d)v_i\psi'(Z_i)\left(-\frac{1}{a_i}\right). \end{aligned} \tag{10}$$

This yields the update rules for each iteration step to be

$$\begin{aligned} v_i(s+1) &= v_i(s) + \alpha \Delta v_i, \\ \omega_{ik}(s+1) &= \omega_{ik}(s) + \beta \Delta \omega_{ik}, \\ a_i(s+1) &= a_i(s) + \gamma \Delta a_i, \\ b_i(s+1) &= b_i(s) + \eta \Delta b_i. \end{aligned} \tag{11}$$

Where α, β, γ and η are learning rate, s is the iteration times.

The usually used basis functions $\psi(\cdot)$ include Gaussian radial basis functions, B-spline basis functions, wavelet basis functions and some neuro-fuzzy basis functions. In this paper, Morlet wavelet function is selected as the activation function of the wavelet process neurons in the hidden layer of the WPNN. The Morlet wavelet is a cosine modulate Gauss wave. It exhibits good locality features in both time and frequency. The Morlet wavelet function can be written as

$$\begin{aligned} &\psi(x) \\ &= \cos\left(1.75\frac{x-b}{a}\right)\exp\left(-0.5\left(\frac{x-b}{a}\right)^2\right). \end{aligned} \tag{12}$$

The complete procedure of the learning algorithm is summarized as follows:

step 1 Select appropriate orthogonal basis functions to expand the input function and the connection weight functions of the time series prediction model based on the WPNN.

step 2 Set output error goal $\varepsilon > 0$; initialize iteration $s = 0$, and set max iteration as S.

step 3 Initialize v_i, ω_{ik}, a_i and b_i.

step 4 Utilize Eq.(9) to calculate the mean square error E, if $E < \varepsilon$ or $s > S$, go to step6; otherwise, go to step5.

step 5 Update v_i, ω_{ik}, a_i and b_i according to Eq.(10) and Eq.(11); $s+1 \to s$, go to step 4.

step 6 Output results; stop.

5. Mackey–Glass time series prediction

In this section, the performance of the proposed time series prediction model based on the WPNN is validated by the Mackey–Glass time series prediction. The time series is generated by the chaotic Mackey–Glass differential delay equation[14] defined as

$$\frac{\mathrm{d}x}{\mathrm{d}t} = \frac{Ax(t-\tau)}{1+x(t-\tau)^{10}} - Bx(t). \tag{13}$$

Where t is a variable, x is a function of t, τ is the time delay. The initial conditions used in this paper are set as $A = 0.2$, $B = 0.1$, $\tau = 17$ and $x(0) = 1.2$.

Fourth order Runge–Kutta method with step size 0.1 is used to generate a time series with 408 points such as $\{x_j\}_{j=1}^{408}$. Eight consecutive points such as $(x_i, x_{i+1}, \ldots, x_{i+7})$ in the time series are given to generate an input function IF_i by nonlinear least-squares method, where $i = 1, 2, \ldots, 400$, and the ninth point x_{i+8} is used as the corresponding output. Thus, it can get 400 couples of samples such as $\{IF_i; x_{i+8}\}_{i=1}^{400}$. The first 200 couples of samples $\{IF_i; x_{i+8}\}_{i=1}^{200}$ are utilized

to train the proposed time series prediction model based on the WPNN. The model used in this section is composed of 3 layers, the first layer is the input layer with one unit, the second layer is the hidden layer, which consists of 10 wavelet process neurons, and the last layer is the output layer with one unit. The orthogonal Legendre basis functions are selected to expand the input functions and the connection weight functions of the WPNN, and the number of the basis functions is set to 6. According to Ref.[15], initialize the dilation and translation parameters of the WPNN. The error goal is set to 0.001, the initial learning rate is set to 0.01, and the max iteration number is set to 2000. After 546 iterations, the time series prediction model based on the WPNN has converged. The last 200 couples of samples $\{IF_i; x_{i+8}\}_{i=201}^{400}$ are selected to validate the time series prediction model based on the WPNN. The validation results are shown in Fig.3.

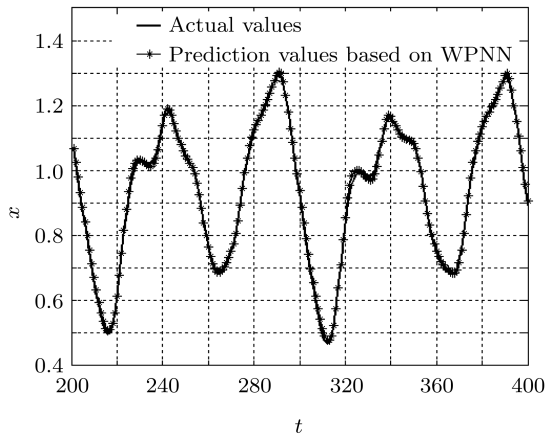


Fig.3. Mackey–Glass time series prediction based on WPNN.

In order to compare the performance of the WPNN with that of the traditional neural networks, a multilayer feedforward neural network (MFNN) with 3 layers are used to predict the same Mackey–Glass time series $\{x_j\}_{j=1}^{408}$. Eight consecutive points such as $(x_i, x_{i+1}, \ldots, x_{i+7})$ in the time series are used as the inputs of the MFNN, and the ninth point x_{i+8} is used as the corresponding output. Thus, it can get 400 couples of samples such as $\{x_i, x_{i+1}, \ldots, x_{i+7}; x_{i+8}\}_{i=1}^{400}$. The first 200 couples of samples $\{x_i, x_{i+1}, \ldots, x_{i+7}; x_{i+8}\}_{i=1}^{200}$ are utilized to train the MFNN. The first layer of the MFNN is the input layer with eight units, the second layer is the hidden layer which has 60 traditional artificial neurons with Sigmoid activation function, and the last layer is the output layer with one unit. The error goal is set to 0.001, the initial learning rate is set to 0.01, and the max iteration number is set to 2000. After 1408 iterations, the MFNN has converged. The last 200 couples of samples $\{x_i, x_{i+1}, \ldots, x_{i+7}; x_{i+8}\}_{i=201}^{400}$ are selected to validate the MFNN. The validation results are shown in Fig.4.

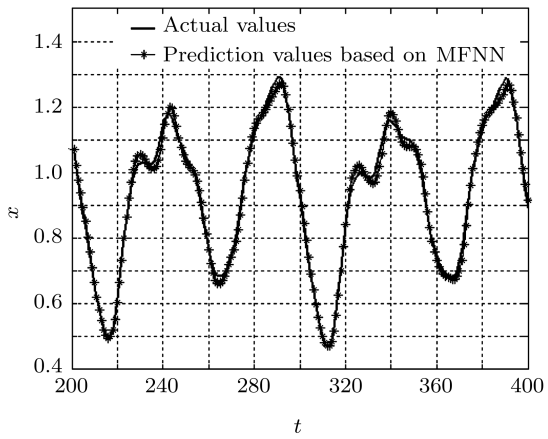


Fig.4. Mackey–Glass time series prediction based on MFNN.

From Figs.3 and 4, the analysis of the comparative prediction results is described in Table 1.

Table 1. The comparative prediction results analysis.

Prediction model	Convergence iterations	Total absolute error/%	Average relative error/%
Based on WPNN	546	1.5768	0.92
Based on MFNN	1408	2.9574	1.57

The comparative prediction results as shown in Fig.3, Fig.4 and Table 1 indicate that the WPNN has a higher prediction accuracy than the MFNN. The WPNN seems to perform well and appears suitable for using as a better tool to predict nonlinear time series.

6. Conclusion

In real world, the inputs of many complicated systems are time-varying functions or processes. How to predict the outputs of these systems with high speed and accuracy has received significant attention. The theoretical foundation of the traditional time series prediction methods is the function approximation theory. In this paper, it presents that the time series prediction problem can be seen as a functional approximation problem. According to this analysis, a novel time series prediction method based on the WPNN is proposed to solve the aforementioned problem, and the corresponding learning algorithm is developed. In order to reduce the computational complexity of the learning algorithm, a group of appropriate orthogonal functions are introduced into the input function space of the WPNN, and then the input functions and the corresponding connection weight functions are expanded by the orthogonal basis functions at the same time. With the orthogonality of the orthogonal basis functions, the aggregation operation of the wavelet process neurons on time is simplified. The effectiveness of the proposed time series prediction model based on the WPNN and its learning algorithm is proved by the Mackey–Glass time series prediction. The comparative prediction results indicate that the WPNN has a faster convergence speed and higher prediction accuracy than the MFNN. The WPNN seems to perform well and appears suitable for using as a better tool to predict the high complex nonlinear time series.

References

[1] Cui W Z, Zhu C C, Bao W X and Liu J H 2005 *Chin. Phys.* **14** 922

[2] Hornik K, Stinchcombe M and White H 1989 *Neural Networks* **2** 359

[3] Funahashi K 1989 *Neural Networks* **2** 183

[4] Zhang J F and Hu S S 2007 *Acta Phys. Sin.* **56** 713 (in Chinese)

[5] Liu H, Liu D and Deng L F 2006 *Chin. Phys.* **15** 1196

[6] Li H C, Zhang J S and Xiao X C 2005 *Chin. Phys.* **14** 2181

[7] Yu Z H and Cai Y L 2006 *Acta Phys. Sin.* **55** 1659 (in Chinese)

[8] He X G and Liang J Z 2000 *Chin. Engin. Sci.* **2** 40 (in Chinese)

[9] Ding G and Zhong S S 2005 *Neural Network World* **15** 137

[10] Ding G and Zhong S S 2006 *Neural Network World* **16** 15

[11] Ding G and Zhong S S 2007 *Acta Phys. Sin.* **56** 1224 (in Chinese)

[12] Cristea P, Tuduce R and Cristea A 2000 *Proc. Neural Network Appl. Elec. Engin.* **1** 5

[13] Daubechies I 1990 *IEEE Trans. Inf. Theory* **136** 961

[14] Mackey M C and Glass L 1977 *Science* **197** 287

[15] Zhao X Z, Zou C H and Chen T J 2003 *J. South China Univ. Tech.* **31** 77 (in Chinese)