Learning and Approximation of Chaotic Time Series Using Wavelet-Networks

V. Alarcon-Aquino, E. S. García-Treviño, R. Rosas-Romero, L.G. Guerrero-Ojeda
Departamento de Ingenieria Electrónica
Universidad de las Américas Puebla,
Sta. Catarina Mártir. Cholula, Puebla. C.P. 72820. MEXICO
E-mail: vicente.alarcon@udlap.mx, erilke@yahoo.com.mx

Abstract

This paper presents a wavelet neural-network for learning and approximation of chaotic time series. Wavelet networks are a class of neural network that take advantage of good localization and approximation properties of multiresolution analysis. This kind of network use wavelets as activation functions in the hidden layer and a non-iterative method is used for its learning. Comparisons are made between a wavelet network, tested with two different wavelets, and the typical feed-forward networks trained with the back-propagation algorithm. The results reported in this paper show that wavelet networks have better approximation properties than its similar back-propagation networks.

Keywords: wavelet networks, wavelets, chaos theory, back-propagation networks, approximation theory, multiresolution analysis.

1. Introduction

Wavelet neural networks are a novel powerful class of neural networks that incorporate the most important advantages of multiresolution analysis introduced by Mallat in 1989 [6]. These networks preserve all the features of common neural networks, but in addition, have rigorous mathematical foundations. In wavelet networks, the activation functions have good localization properties; this important characteristic allows hierarchical and multiresolution learning as well as transparent design of the network.

Wavelets are a new family of basis functions that combine powerful properties such as orthogonality, compact support, localization in time

and frequency, and fast algorithms. Wavelets have generated a tremendous interest in both theoretical and applied areas, especially over the past few years [10]. Wavelet based methods have been used for approximation theory, pattern recognition, compression, numerical analysis, computer science, electrical engineering, physics etc.

Wavelet networks are a class of neural networks that employ wavelets as activation functions. These have been recently researched as an alternative approach to the traditional neural networks with sigmoidal activation functions. Wavelet networks have attracted great interest, because of their advantages over other networks schemes (see e.g., [9]). There are two main research groups that combine wavelet theory and neural networks, both with a comprehensive framework. The first is based on the work of Zhang and Benveniste [8] [11], that introduces a (1+½) layer neural network based on wavelets, but their approach is still using a type of backpropagation algorithm to train de network. The second is based on the work initiated by Bakshi and Stephanopoulos [7], in which wavelets are used in a structure similar to radial basis function networks, but a non-iterative and hierarchical method is used for learning [5]. We based this work on the second approach, following the concept of locally supported basis functions such as radial basis function networks and from wavelet decomposition in signal processing.

Function approximation involves estimating the underlying relationship from a given finite input-output data set, and it has been the fundamental problem for a variety of applications in pattern classification, data mining, signal reconstruction, and system identification [4]. Recently, feed-forward neural networks such as multilayer perceptrons and radial basis function networks

have been widely used as an alternative approach to function approximation since they provide a generic black-box functional representation. Furthermore, these networks have shown to be capable of approximating any continuous function defined on a compact set in $\mathbb{R}^n$, where $\mathbb{R}$ denotes real numbers, with arbitrary accuracy.

In this work we show the powerful and the abilities of this kind of networks, making approximation of time series. Particularly, we present wavelet networks applied to the approximation of chaotic time series. For this purpose comparisons are made between a wavelet network, tested with two different wavelets, and the typical feed-forward networks trained with the backpropagation algorithm [3]. The results reported in this work show clearly that wavelet networks have better approximation properties than its similar backpropagation networks. The reason for that is the firm theoretical foundations of wavelet networks, that combine, the mathematical methods and tools of multiresolution analysis with the neural network framework.

The remainder of this paper is organized as follows. Section 2 briefly reviews wavelet theory. Section 3 describes the wavelet network structure. In Section 4 chaotic time series are presented. In Section 5 comparisons between wavelet networks and back-propagation networks are made and discussed. Finally, Section 5 presents the conclusions of this work.

2. Review of wavelet theory

Wavelet transforms involve representing a general function in terms of simple, fixed building blocks at different scales and positions. These building blocks are generated from a single fixed function called mother wavelet by translation and dilation operations. The continuous wavelet transform considers a family

$$\psi_{a,b}(x) = \frac{1}{\sqrt{|a|}}\psi\left(\frac{x-b}{a}\right), \qquad (1)$$

where $a \in \mathbb{R}^+$, $b \in \mathbb{R}$, with $a \neq 0$, and $\psi(\cdot)$ satisfies the admissibility condition. For discrete wavelets the scale (or dilation) and translation parameters in (1) are chosen such that at level m the wavelet $a_0^m \psi(a_0^{-m}x)$ is a_0^m times the width of $\psi(x)$. That is, the scale parameter $\{a = a_0^m : m \in \mathbb{Z}\}$ and the translation parameter $\{b = kb_0a_0^m : m, k \in \mathbb{Z}\}$. This family of wavelets is thus given by

$$\psi_{m,k}(x) = a_0^{-m/2}\psi(a_0^{-m}x - kb_0), \qquad (2)$$

so the discrete version of wavelet transform is

$$d_{m,k} = \langle g(x), \psi_{m,k}(x)\rangle$$

$$= a_0^{-m/2}\int_{-\infty}^{+\infty} g(x)\psi(a_0^{-m}x - kb_0)dx, \qquad (3)$$

$\langle\,\cdot\,,\,\cdot\,\rangle$ denotes the L^2-inner product. To recover $g(x)$ from the coefficients $\{d_{m,k}\}$, the following stability condition should exist [2],

$$A\|g(x)\|^2 \leq \sum_{m\in\mathbb{Z}}\sum_{k\in\mathbb{Z}}|\langle g(x), \psi_{m,k}(x)\rangle|^2 \leq B\|g(x)\|^2,$$

$$(4)$$

with $A > 0$ and $B < \infty$ for all signals $g(x)$ in $L^2(\mathbb{R})$ denoting the frame bounds. These frame bounds can be computed from a_0, b_0 and $\psi(x)$ [12]. The reconstruction formula is thus given by

$$g(x) \approx \frac{2}{A+B}\sum_{m\in\mathbb{Z}}\sum_{k\in\mathbb{Z}}\langle g(x), \psi_{m,k}(x)\rangle\psi_{m,k}(x).$$

$$(5)$$

Note that the closer A and B, the more accurate the reconstruction. When $A = B = 1$, the family of wavelets then forms an orthonormal basis [12].

2.1 Orthonormal bases and multiresolution analysis

The mother wavelet function $\psi(x)$, scaling a_0 and translation b_0 parameters are specifically chosen such that $\psi_{j,n}(x)$ constitute orthonormal bases for $L^2(\mathbb{R})$ [6], [12]. To form orthonormal bases with good time-frequency localisation properties, the time-scale parameters (b, a) are sampled on a so-called dyadic grid in the time-scale plane, namely, $a_0 = 2$ and $b_0 = 1$, [6], [12]. Thus, from Eq. (2) substituting these values, we have a family of orthonormal bases,

$$\psi_{m,k}(x) = 2^{-m/2}\psi(2^{-m}x - k), \qquad (6)$$

Using Eq. (3), the orthonormal wavelet transform is thus given by

$$\langle g(x), \psi_{m,k}(x)\rangle = $$
$$2^{-m/2}\int_{-\infty}^{+\infty} g(x)\psi_{m,k}(2^{-m}x - k)dx$$

$$(7)$$

and the reconstruction formula is obtained from Eq. (5). A formal approach to constructing orthonormal bases is provided by multiresolution analysis (MRA) [6]. The idea of MRA is to write a function $g(x)$ as a limit of successive approximations, each of which is a smoother version of $g(x)$. The successive approximations thus correspond to different resolutions [6].

2.2 Discrete Wavelet Transform: Decomposition and Reconstruction

Since the idea of multiresolution analysis is to write a signal $g(x)$ as a limit of successive approximations, the differences between two successive smooth approximations at resolution 2^{m-1} and 2 give the detail signal at resolution 2^m. In other words, after choosing an initial resolution L, any signal $g(x) \in L^2(\mathbb{R})$ can be expressed as [6], [12]:

$$g(x) = \sum_{k \in \mathbb{Z}} c_{L,k} \phi_{L,k}(x) + \sum_{m=L}^{\infty} \sum_{k \in \mathbb{Z}} d_{m,k} \psi_{m,k}(x),$$
(8)

where the detail or wavelet coefficients $\{d_{m,k}\}$ are given by Eq. (7), while the approximation or scaling coefficients $\{c_{m,k}\}$ are defined by

$$c_{m,k} = 2^{-m/2} \int_{-\infty}^{+\infty} g(x) \phi_{m,k}(2^{-m}x - k) dx \qquad (9)$$

Equations (7) and (9) express that a signal $g(x)$ is decomposed in details $\{d_{m,k}\}$ and approximations $\{c_{m,k}\}$ to form a multiresolution analysis of the signal [6].

3. Description of wavelet-networks

As stated previously, we based this work on the wavelet network approach introduced by Bakshi and Stephanopoulos [7] in which wavelets are used in a structure similar to radial basis function networks.

Equation (8) represents the basic framework for the construction of wave-nets. A wave-net is defined as a one hidden layer network whose basis functions are drawn from a family of orthogonal wavelets [7]. Multiresolution analysis is the framework for its learning. In fact, by incorporating the basis functions at a particular resolution, namely $m=0$ here, the first approximation to the desired function is obtained [14],

$$g_o(x) = \sum_{k \in \mathbb{Z}} c_{0,k} \phi_{0,k}(x) \qquad (10)$$

Then, by including the wavelets of the same resolution, a finer approximation is provided,

$$g_{-1}(x) = g_o(x) + \sum_{k \in \mathbb{Z}} d_{0,k} \psi_{0,k}(x) \qquad (11)$$

In general, adding wavelets of a higher resolution to a current approximation, a finer approximation of $g(x)$ is obtained,

$$g_{m-1}(x) = g_m(x) + \sum_{k \in \mathbb{Z}} d_{m,k} \psi_{m,k}(x) \qquad (12)$$

This process can be continued until the function $g(x)$ is re-produced as in equation (8). Figure 1 shows a typical wavenet structure. At each stage of learning, the approximation problem may be expressed as

$$\widehat{g}_m(x) = \sum_{k \in \mathbb{Z}} c_{m,k} \theta_{m,k}(x) \qquad (13)$$

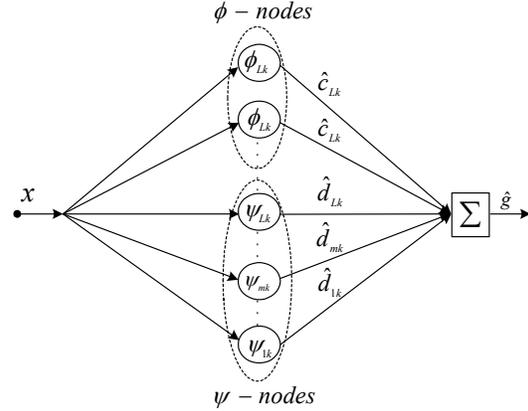where $\qquad \widehat{g}_m(x) = g(x) - g_m(x)$

Figure 1. A typical wavenet structure

Now, if one considers a practical case of finite discrete sequences of inputs and outputs, the equation (13) leads to

$$\widehat{g} = AC \qquad (14)$$

where

$$\widehat{g} = \begin{bmatrix} \widehat{g}(x_1) \\ \widehat{g}(x_2) \\ \vdots \\ \widehat{g}(x_n) \end{bmatrix} \quad A = \begin{bmatrix} \theta_1(x_1) & \dots & \theta_k(x_1) \\ \theta_1(x_2) & \dots & \theta_k(x_2) \\ \vdots & & \vdots \\ \theta_1(x_n) & \dots & \theta_k(x_n) \end{bmatrix}$$

$$C = \begin{bmatrix} c_1 \\ c_2 \\ \vdots \\ c_k \end{bmatrix} \qquad (15)$$

The least squares solution of (14) is given by

$$\begin{aligned} C &= ((A^T A)^{-1} A^T)\widehat{g} \\ &= A^+ \widehat{g} \end{aligned} \qquad (16)$$

where A^+ is the pseudo-inverse of matrix A. If data are regularly sampled and the basis functions are orthonormal, $\sum_k \theta^2(k) = 1$ then $A^T A$ becomes an identity matrix and equation is simplified to:

$$C = A^T \widehat{g} \qquad (17)$$

4. Chaotic time series

Chaos is the mathematical term for the behavior of a system that is inherently unpredictable.

Unpredictable phenomena are readily apparent in all areas of life [1]. Many systems in the natural world are now known to exhibit chaos or non-linear behavior, the complexity of which is so great that they were previously considered random. The unraveling of these systems has been aided by the discovery, mostly in this century, of mathematical expressions that exhibit similar tendencies [2]. In this section, we present briefly the four chaotic time series used in this work (see Figure 2).

4.1 Logistic map

It was explored by ecologist and biologist who used it to model population dynamics. It was popularized by Robert May in 1976 as an example of a very simple nonlinear equation being able to produce very complex dynamics. The mathematical equation that describes the logistic map has the form:

$$x(t+1) = ax(t)(1 - x(t)) \qquad (18)$$

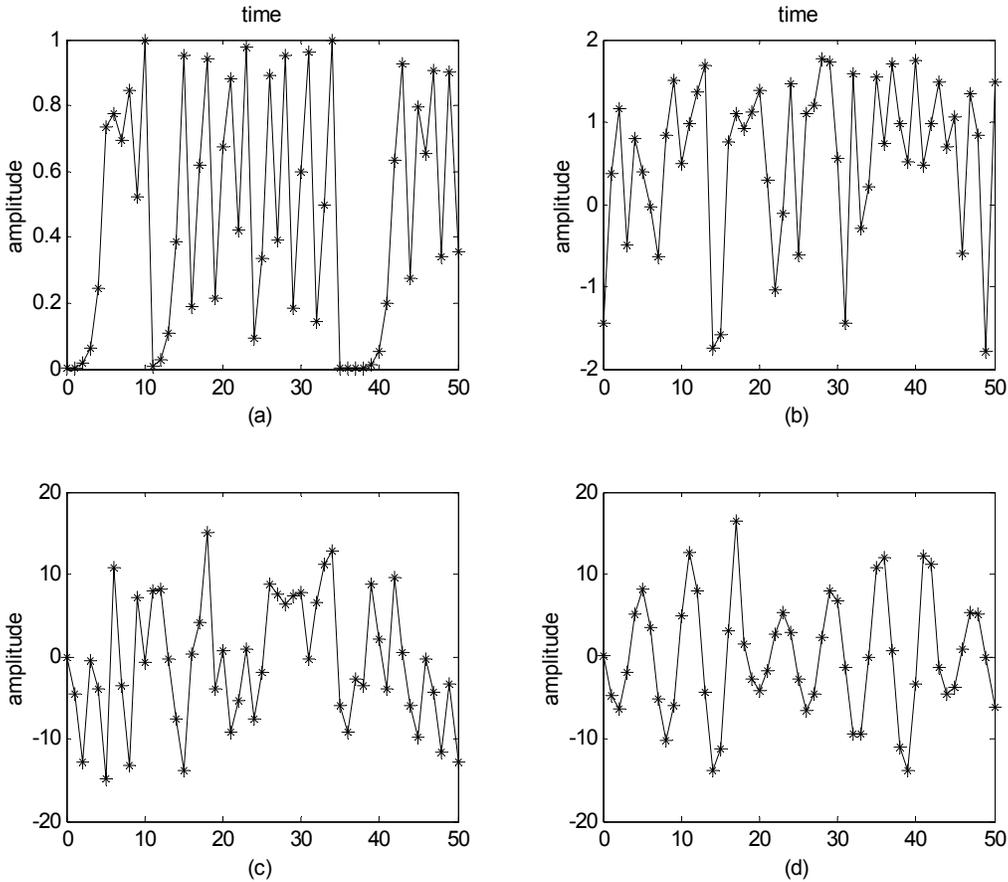
Figure 2. Chaotic series: (a) Logistic map, (b) Henon attractor, (c) Lorenz attractor, (d) Rossler attractor.

where $a = 4$, and the initial condition $x(0) = 0.000104$.

4.2 *Rossler attractor*

It was introduced for Otto Rossler and arose from work in chemical kinetics. The system is described with 3 coupled non-linear differential equations:

$$\frac{dx}{dt} = -y - z \qquad (19)$$
$$\frac{dy}{dt} = x + ay$$
$$\frac{dz}{dt} = b + z(x - c)$$

where $a = 0.15$, $b = 0.20$, $c = 10$.

4.3 Lorenz attractor

It was introduced by E. N. Lorenz, in 1963. It was derived from a simplified model of atmospheric interactions. The system is most commonly expressed as 3 coupled non-linear differential equations:

$$\frac{dx}{dt} = a(y - x) \qquad (20)$$
$$\frac{dy}{dt} = x(b - z) - y$$
$$\frac{dz}{dt} = xy - cz)$$

where $a = 10$, $b = 28$, $c = 8/3$.

4.4 Hénon Attractor

It was introduced by M, Hénon in 1976, and was derived from a study of chaotic functions trajectories.

$$\frac{dx}{dt} = a + by - x^2 \qquad (21)$$
$$\frac{dy}{dt} = x$$

where $a = 1.4$, $b = 0.3$.

The chaotic series described previously are considered for this work because they illustrate how the complex behavior can easily be produced by simple equations with non-linear elements and feedback. Figure 2 shows these four chaotic time series plotted for the first 50 samples.

5. Simulation results

This work present wavelet networks applied to the approximation of four of the most known chaotic time series: Logistic Map, Lorenz Attractor, Hénon Attractor, and Rossler Attractor. Particularly, this section presents comparisons between a wavelet network, tested with two different wavelets, and the typical feed-forward networks trained with the backpropagation algorithm [3].

The last three time series mentioned previously were obtained from [14]. Lorenz attractor was sampled every 0.01 time steps, Rossler and Hénon series every 0.1 time steps. For an appropriate learning, and taking into account the features of the activation functions used by the networks analyzed, the four series were normalized in the range of 0-1. For the basis functions of the wavelet network approach we first test it with orthogonal Battle-Lemarie wavelet (wavenet(1)) and afterwards with Haar wavelet (wavenet(2)). For the back-propagation network, we use tangent-sigmoidal activation functions, and it was trained with a learning rate of 0.1, the momentum term was set at 0.07, and 5000 iterations. The same architecture was used for BPN and for the wavenets, that is, 1-6-1 (one input layer, six hidden units and one output unit). For all series the first 300 samples were used for training and the next 200 samples for testing.

Table 1. Simulations results

Logistic map		
Approach	MSE(training)	MSE (testing)
Wavenet (1)	0.001443	0.001201
Wavenet (2)	0.049707	0.050050
BP	0.121968	0.133198
Hénon attractor		
Approach	MSE(training)	MSE (testing)
Wavenet (1)	0.065681	0.0797552
Wavenet (2)	0.028381	0.031613
BP	0.074312	0.080172
Lorenz attractor		
Approach	MSE(training)	MSE (testing)
Wavenet (1)	0.059465	0.058034
Wavenet (2)	0.060698	0.060266
BP	0.060410	0.062313
Rossler attractor		
Approach	MSE(training)	MSE (testing)
Wavenet (1)	0.039531	0.041005
Wavenet (2)	0.041139	0.042490
BP	0.053366	0.055777

As can be seen in Table 1, wavelet networks outperform BPN when using Battle-Lemarie and Haar Wavelets. This is due to the fact that neural networks based on conventional single-resolution schemes cannot learn difficult time series, with

abrupt changes in its behavior; consequently the training process often converge very slowly and the trained network may not generalize well. It is worth noting that the wavenet(1) performs a better approximation than the wavenet(2). This is due to the fact that the Battle-Lemarie wavelet has higher degree of smoothness than the Haar wavelet. A higher degree of smoothness in a wavelet corresponds to better frequency localization [10].

6. Conclusions

The results reported in this work show clearly that wavelet networks have better approximation properties than its similar backpropagation networks. The reason for that is the firm theoretical foundations of wavelet networks, that combine, the mathematical methods and tools of multiresolution analysis with the neural network framework.

The wavelet selected as basis function for the studied approach play an important role in the approximation process. Generally the selection of the wavelet depends on the kind of function to be approximated. In our particular case, due to the features of time series studied, a wavelet with good high-frequency localization like Battle-Lemarie, showed better approximation results.

7. References:

[1] R. L. Devaney, "Chaotic Explosions in Simple Dynamical Systems", The Ubiquity of Chaos, Edited by Saul Krasner. American Association for the Advancement of Science. Washington D. C., U.S.A. 1990

[2] J. Pritchard, "The Chaos CookBook: A practical programming guide", Part of Reed International Books. Oxford. Great Britain. 1992.

[3] S. Y. Kung, "Digital Neural Networks", Department of Electrical Engineering, Princeton University, Prentice Hall. New Jersey. U.S.A. 1993.

[4] S. Sitharama Iyengar, E.C. Cho, Vir V. Phoha, "Foundations of Wavelet Networks and Applications", Chapman & Hall/CRC. U.S.A. 2002

[5] A. A. Safavi, H. Gunes, J.A. Romagnoli, "On the Learning Algorithms for Wave-Nets", Department of Chemical Engineering, The University of Sydney. Australia. 2002,

[6] S. G. Mallat, "A theory for Multiresolution Signal Decomposition: The Wavelet Representation", IEEE Transactions on Pattern Analysis and Machine Intelligence. Vol II. No 7. July 1989

[7] B. R. Bakshi, G.Sthepanopoulos, "Wavelets as Basis Functions for Localized Learning in a Multiresolution Hierarchy", Laboratory for Intelligent Systems in Process Engineering, Department of Chemical Engineering. Massachusetts Institute of Technology, Cambridge, MA 02139, 1992

[8] Q. Zhang, A. Benveniste, "Wavelet Networks", IEEE Transactions on Neural Networks. Vol 3. No 6. July 1992

[9] E. A. Rying, Griff L. Bilbro, and Jye-Chyi Lu. "Focused Local Learning With Wavelet Neural Networks", IEEE Transactions on Neural Networks, Vol. 13, No. 2, March 2002.

[10] B. Jawerth, W. Sweldens. "An overview of wavelet based multiresolution analyses", SIAM, 1994.

[11] Xeiping Gao, Fen Xiao, Jun Zhang, Chunhong Cao. "Short-term Prediction of Chaotic Time Series by Wavelet Networks", WCICA, Fifth World Congress on Intelligent Control And Automation, 2004.

[12] I, Daubechies. "Ten Lectures on Wavelets", New York. SIAM. 1992.

[13] A. A. Safavi, J. A.Romagnoli "Application of Wave-nets to Modellling and Optimisation of a Chemical Process" Safavi, A.A. Romagnoli, J.A. Proceedings., IEEE International Conference on Neural Networks, 1995.

[14] E, R Weeks. *http://www.physics.emory.edu/~weeks/*