Journal of Engineering and Development, Vol. 12, No. 3, September (2008) ISSN 1813-7822

Genetically Based Wavenets for System Modeling

Asst. Prof. Dr. Mohammed A. Abdala
College of Information Engineering
Nahrain University, Baghdad, Iraq

Abstract

The combination of wavelet theory and neural networks has lead to the development of wavelet networks (wavenets). Wavenets are feed-forward neural networks which used wavelets as activation functions, and the basis used in wavenet has been called wavelons. Wavenets are successfully used in identification problems. The strength of wavenets lies in their capabilities of catching essential features in "frequency-rich" signals. In wavenet, both the translation and the dilation of the wavelets (wavelons) are optimized besides the weights.

The wavenet algorithm consists of two processes: the self-construction of networks and the minimization of errors. In the first process, the network structure is determined by using wavelet analysis. In the second process, the approximation errors are minimized. The wavenets with different types of frame wavelet function are integrated for their simplicity, availability, and capability of constructing unknown nonlinear function. Thus, wavelet can identify the localization of unknown function at any level.

In addition, genetic algorithms (GAs) are used successfully in training wavenets since GAs reaches quickly the region of the optimal solution. Tests show that GA obtains best weight vector and produces a lower sum square error in a short period of time.

الخلاصـــــــة

لقد أدى مزج نظرية المويجات والشبكات العصبية إلى تطوير شبكات المويجات (Wavenet). ان شبكات المويجات هي شبكات عصبية أمامية التغذية تستخدم المويجات كدوال التحفز. أن الأسس المستخدمة في شبكة المويجات سميت بـ (Wavelons). لقد تم استخدام الشبكة المويجية في مشاكل التمييز بنجاح. أن قوة الشبكة المويجية تكمن في قدرتها على امتلاك الخصائص الضرورية في الإشارات الغنية بالترددات. في الشبكة المويجية يتم اختيار الأمثل لكل من التنقلات والسعة للمويجة بالإضافة إلى الأوزان.

تتكون الخوارزمية من عمليتين: التركيب الذاتي للشبكات وتقليل الأخطاء. في العملية الأولى يتم تحديد هيكل الشبكة باستخدام تحليل المويجة. أما في العملية الثانية فيتم تقليل أخطاء التقريب.

لقد تم استخدام الخوارزميات الجينية بنجاح لتدريب الشبكة المويجية وذلك لأن الخوارزميات الجينية تصل بسرعة إلى منطقة الحل الأمثل. لقد أظهرت الاختبارات التي أجريت في هذا البحث بان الخوارزميات الجينية تحصل على مجموعة الأوزان الأفضل وتقدم اقل مجموع لمربعات الأخطاء في فترة زمنية قصيرة.

1. Introduction

The wavelet networks (wavenets) is an approach for system identification in which nonlinear functions are approximated as the superposition of dilated and translated versions of a single function, which is localized both in the time and frequency domains [1,2]. Given a sufficiently large number of network elements (called "wavelets"), any square-integrable function can be approximated to arbitrary precision [3]. The following are the advantages of Wavenet over similar architectures [4]:

(i) Unlike the sigmoidal functions used in neural networks of the multi–layer perceptron type, wavelets are specially localized. As a result, training algorithms for wavelet networks typically converge in a smaller number of iterations when compared to multi–layer perceptrons.

(ii) The magnitude of each coefficient in a wavenet can be related to the local frequency content of the function being approximated. Thus, more parsimonious architectures can be achieved, by preventing the unnecessary assignment of wavelets to regions where the function varies slowly. This property is not shared by networks of conventional radial basis functions (Gaussians).

(iii) It is not difficult to add or delete the hidden units not affecting the performances because the basis function is localized in time-frequency plane.

2. Wavelet

The term "wavelet" as it implies means a little wave. This little wave must have at least a minimum oscillation and a fast decay to zero, in both the positive and negative directions, of its amplitude [5]. Wavelets are a powerful tool for function analysis and synthesis. To begin the discussion, a brief introduction to the continuous wavelet transform and inverse wavelet transform will be given. Following this, a comparison with Fourier analysis will be presented to highlight the strengths of wavelet theory. The environments being considered will result in functions dependent on state (position) as opposed to time. Therefore, the domains of interest will be space-frequency as opposed to time-frequency. This section concludes with a discussion on discrete implementations of the inverse wavelet transform, which is the form used in the wavelet network (wavenet) [6].

2-1 Continuous Wavelet Transform

The continuous wavelet transform (CWT) can be expressed as follows:

$$\mathbf{w_i(\alpha,\beta)=\int_{-\infty}^{\infty} f(x)\psi(\alpha x-\beta)dx} \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \mathbf{(1)}$$

The parameter α controls the width (dilation) of the wavelet, while β controls the position (translation). The variable ψ is a family of functions based on a single analyzing wavelet, along with its translated and dilated versions. The result of this transform is a set of wavelet coefficients, ranging over different widths of wavelets, called levels. The translation parameter is adjusted according to the dilation parameter to ensure that the entire operating range is covered. As an example of what a family of analyzing wavelets would look like, consider the Mexican hat (second derivative of the Gaussian) wavelet in **Fig.(1)**.

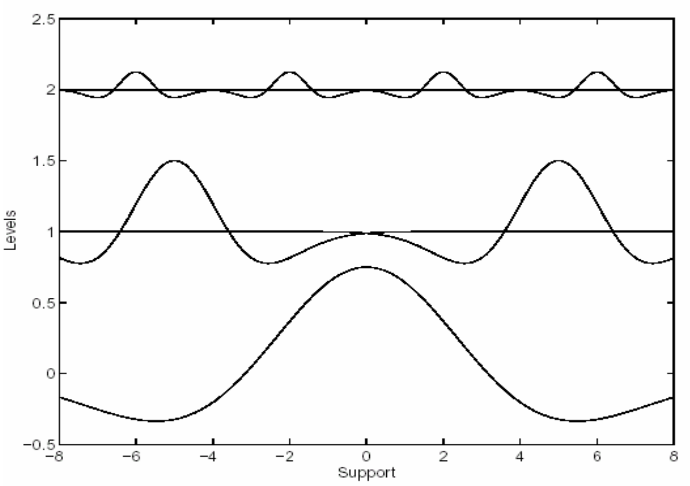


Figure (1) A conceptual view of how the wavelet transform works using a Mexican hat wavelet

The wavelet transform analyzes a function at different wavelet levels. A wavelet level is defined as a set of translated wavelet with constant dilation. Wavelets have been compared to microscopes with the ability to zoom in on short lived high frequency features. This fact can be noted by observing that as the wavelet becomes more contracted, it approaches a discontinuity. Using more levels in the analysis results in the ability to extract more detailed (high frequency) features.

Of greater interest to this work is how to use wavelet theory to reconstruct functions. This is achieved by applying the inverse wavelet transform:

$$f(x) = \int\int w_i(\alpha,\beta)\psi(\alpha x - \beta)d\alpha d\beta \ \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots .. \ (2)$$

Which states that a function can be reconstructed by summing over the entire range of translated and dilated wavelets [6].

2-2 Wavelet and Fourier Analysis

Fourier analysis shows that any periodic function can be represented by the sum of sines and cosines. Similarly, wavelet theory shows that an arbitrary function can be represented by the sum of wavelets. The following points regarding wavelet theory will be highlighted:

- Wavelets are capable of representing functions that are choppy, discontinuous, and potentially non-periodic.

- The wavelet transform represents the function in both space and scale (frequency), while the Fourier transforms addresses frequency alone.

An advance in Fourier analysis provides a method to represent a function in both space and frequency. The concept is referred to as the short time Fourier transform (STFT). The basic approach is to fix a window size of constant length, and change the frequency within the window using a finite section of a sine or cosine function. Once the window size has been defined it remains constant throughout the entire space-frequency plane. Although the approach is sensible, it suffers from the issue of defining window size. If the window is large then low frequency content is captured but it becomes difficult to pinpoint the location of high-frequency features. If the window size is small then high frequency features are captured with good accuracy, however the low-frequency content is missed. Based on the above notion it would be nice if both large and small windows could be used. This is precisely what wavelet theory provides. Wavelets automatically adjust the size of the window, through the dilation parameter, to compensate for high and low frequency content.

The above comparison was a brief introduction into the differences between Fourier and wavelet analysis. The intent was to introduce some characteristics inherent to wavelet analysis that make it a powerful tool when applied to nonlinear functions. For a full and detailed comparison between Fourier and wavelet theory, the reader is referred to [7].

2-3 Discretization Methods

Since the functions of interest to this work will always consist of data points sampled at a constant time step, it is appropriate to obtain a discretized version of the inverse wavelet transform (Eqn. (2)). There are various methods for obtaining a discretized version of Eq.(2). The general structure for the inverse discrete wavelet transform is as follows:

$$f(x) = \sum_{i} w_i \psi(\alpha_i x - \beta_i) \quad \text{..} \quad (3)$$

For stable reconstruction of the function, the above form must meet certain criteria. One common way is to construct a family of wavelets that constitute an orthonormal basis. Another method is to create a family of wavelets that constitute a frame [6].

2-4 Wavelet Frame

Two categories of wavelet functions, namely, orthogonal wavelets and wavelet frames, were developed separately by different groups. Orthogonal wavelet decomposition is usually associated to the theory of multiresolution analysis [8]. The fact that orthogonal wavelets cannot be expressed in closed form is a serious drawback for their application to function approximation and process modeling. Conversely, wavelet frames are constructed by simple operations of translation and dilation of a single fixed function called the mother wavelet, which must satisfy conditions that are less stringent than orthogonality conditions [9].

To constitute a frame, the discrete wavelet transform generated by sampling the wavelet parameters (time/frequency) on a grid or lattice, which is not equally spaced (unlike the short-time Fourier transform, STFT), must satisfy the admissibility condition and the lattice points must be sufficiently close to satisfy basic information-theoretic needs [5].

A daughter wavelet $\psi_i(x)$ is derived from its mother wavelet $\psi(x)$ by the relation:

$$\boldsymbol{\psi_i(x) = \psi(\alpha_i x - \beta_i)} \quad \text{...} \textbf{(4)}$$

where,

β_i *: represent translation factor and*

α_i *: dilation factor* [9].

The family of functions generated by (ψ) can be defined as:

$$\mathbf{W} = \boldsymbol{\alpha_i^{-1/2}\psi(\alpha_i x - \beta_i)} \quad \text{...} \textbf{(5)}$$

A family of wavelets is called a frame if the energy of the wavelet coefficients (sum of their squared moduli), relative to the signal, lies between two positive constants A and B ($A > 0$ and $B < \infty$), called frame bounds.

$$\mathbf{A.E_f} \le \boldsymbol{|w_i(\alpha,\beta)|^2} \le \mathbf{B.E_f} \quad \text{...} \textbf{(6)}$$

Where $(E_f = \sum_i |x_i|^2)$ is defined to be the energy of the function to be identified. If a family of functions satisfies the above expression then it automatically passes the admissibility condition [10].

3. Wavelet Network (wavenet)

The similarity between the inverse wavelet transform and neural network motivates the use of a single layer feed forward network [6].

A wavenet, as shown in **Fig.(2)**, is a neural network with two layers whose output nodes form a linear combination of wavelet basis functions that are calculated in the hidden layer of the network. The basis used in wavenet has been given the name “wavelons”. These wavelons produce a localized response for an input impulse. That is, they produce a non-zero output when the input lies within a small area of the input space [11].

The hidden neuron activation functions (wavelons) are wavelet functions [12]. There are two commonly used types of activation functions: global and local. Global activation functions are active over a large range of input values and provide a global approximation to the empirical data. Local activation functions are active only in the immediate vicinity of the given input value [13].

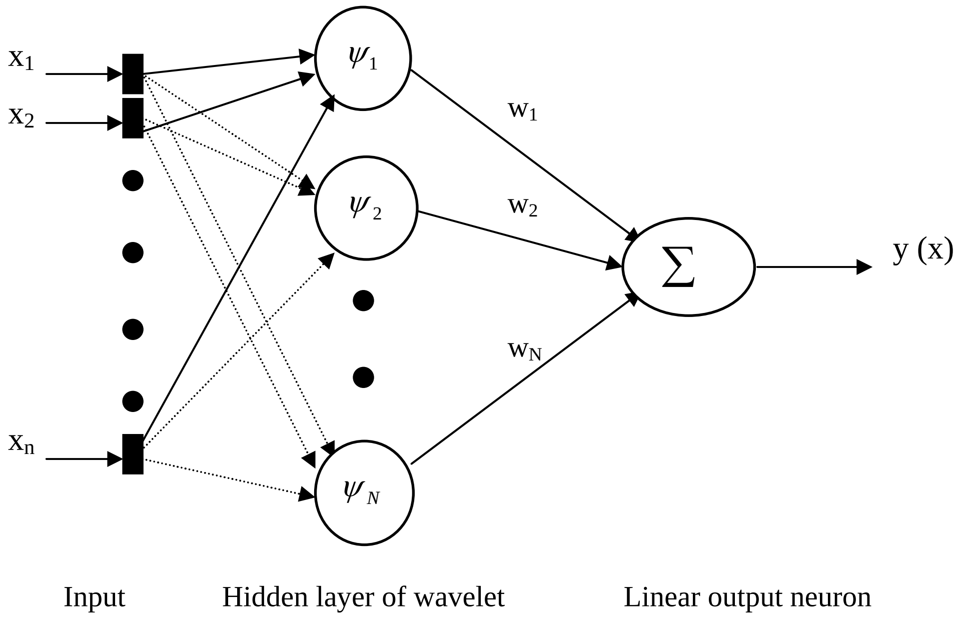


Figure (2) Feed forward wavelet network (wavenet)

From the above model, the wavenet has the following form:

$$\mathbf{y(x)=\sum_{i=1}^{N} w_i \psi(\alpha_i x-\beta_i)} \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \mathbf{(7)}$$

where,

- *The parameter (w_i), is a weight parameter between the hidden layer and the output unit (wavelet coefficient).*
- *The parameter (α_i) and (β_i), are the dilation and translation parameter respectively.*
- *The parameter (N) is the number of wavelons.*
- *Where (n) is the number of input samples.*

3-1 Final Model

The identification process will be applied to a class of environments that typically exhibit linear regions along with the nonlinearities. In order to produce a result that captures both of these aspects **[6]**, additional connections are made directly from input to the output neurons, as shown in **Fig.(3)**. The final model takes the form **[12]**:

$$\mathbf{y(x)=\sum_{i=1}^{N} w_i \psi(\alpha_i x-\beta_i)+cx+b} \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \mathbf{(8)}$$

Where parameter (c) and (b) are the linear coefficients and bias term respectively and (N) is the number of wavelons in the network.

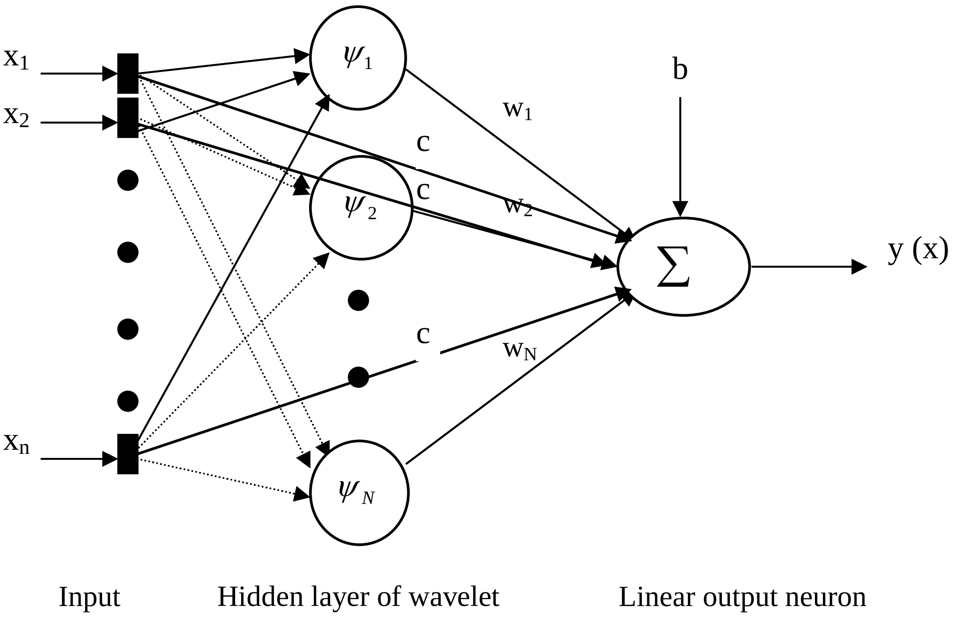


Figure (3) Feed forward wavelet network (wavenet) with direct connections

The wavelet family, which constitutes a frame, consists of the mother wavelet (Mexican hat wavelet) located at the origin along with its translated and dilated versions.

$$\mathbf{y(x)=\sum_{i=1}^{N} w_i\left((\alpha_i(x-\beta_i))^2-1\right).e^{-0.5(\alpha_i(x-\beta_i))^2}+cx+b} \ \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \ \mathbf{(9)}$$

The following matrix of size $n\times(N+2)$ will be defined for use in the function (9).

$$\mathbf{y(x)=}\begin{bmatrix} \left((\alpha_1(x_1-\beta_1))^2-1\right)e^{-0.5(\alpha_1(x_1-\beta_1))^2} & \ldots\ldots & \left((\alpha_N(x_1-\beta_N))^2-1\right)e^{-0.5(\alpha_N(x_1-\beta_{N1}))^2} & x_1 & 1 \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ \left((\alpha_1(x_n-\beta_1))^2-1\right)e^{-0.5(\alpha_1(x_n-\beta_1))^2} & \ldots\ldots & \left((\alpha_N(x_n-\beta_N))^2-1\right)e^{-0.5(\alpha_N(x_n-\beta_N))^2} & x_n & 1 \end{bmatrix}\begin{bmatrix} w_1 \\ \cdot \\ \cdot \\ \cdot \\ w_N \\ c \\ b \end{bmatrix} \ \ldots \ \mathbf{(10)}$$

3-2 Initializing Wavenets

Due to the fact that wavelets are rapidly vanishing functions [9], the following should be noted:

- A wavelet may be too local if its dilation parameter is too small.
- It may sit out of the domain of interest if the translation parameter is not chosen appropriately.

The initial set of wavelets is determined from the input data and the desired resolution. The desired resolution corresponds to the number of wavelet levels which are defined on a dyadic grid [6].

The input vector is pre-processed component-wise using the following expression:

$$\overline{\mathbf{X}} = \frac{\mathbf{X} - \mathbf{x}_{\mathbf{cent}}}{\mathbf{x}_{\mathbf{half}}} \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \mathbf{(11)}$$

where,

$$\mathbf{x}_{\mathbf{cent}} = \frac{1}{2}\left(\mathbf{x}_{\mathbf{min}} + \mathbf{x}_{\mathbf{max}}\right) \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \mathbf{(12)}$$

$$\mathbf{x}_{\mathbf{half}} = \frac{1}{2}\left(\mathbf{x}_{\mathbf{max}} - \mathbf{x}_{\mathbf{min}}\right) \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \mathbf{(13)}$$

For the duration of the section a bar above a variable denotes pre-processed data. The idea behind choosing wavelets for the frame is to construct a pyramid that grows upward with increasing dilation. Based on the range of input data and specified number of levels the wavelets are scanned to obtain translation and dilation parameters.

$$\mathbf{translation = round}\left(\overline{\mathbf{X}} \times \mathbf{2^{level-1}} + \mathbf{0.5}\right) \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots.. \mathbf{(14)}$$

The variable level corresponds to wavelet level, that in practice usually four or five levels is sufficient. Using a dyadic grid for the wavelet levels, the first dilation parameter is, $(dilation = 2^0 = 1)$. The mother wavelet, centered at the origin $(\beta = 0)$ with dilation parameter $(\alpha = 1)$, is always included in the truncated frame. Therefore, the truncated frame consists of the following wavelets:

$$\overline{\alpha} = \mathbf{[1\ 2\ \ldots.2^{level}]} \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots.. \mathbf{(15)}$$

To add the translation parameter, its value is modified to ensure that it covers the entire operating range. This is accomplished using the following relationship:

$$\overline{\beta} = \frac{\mathbf{(translation - 0.5)}}{\mathbf{2^{level-2}}} \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots.. \mathbf{(16)}$$

The process continues through (m) wavelet levels with a final dilation value of $(dilation = 2^m)$. As the dilation parameter gets large, the wavelet becomes more contracted and will only be included if the input data exhibits high frequency characteristics.

To complete the process, the translation and dilation parameters are post-processed using the following equations to obtain the final set of wavelet parameters:

$$\alpha = \frac{\bar{\alpha}}{x_{half}} \quad \ldots\ldots (17)$$

$$\beta = (\bar{\beta} \times x_{half}) + X_{cent} \quad \ldots\ldots (18)$$

4. Design Algorithm

In this section, an algorithm to design wavenets using genetic algorithms is proposed. The rough flow of this algorithm is as follows.

The proposed GA for training wavenet can be summarized as follows:

Step (1): Initialize translation and dilation parameters of all hidden wavelon and create an initial population of chromosomes and set current population to this initial population. Initialize crossover rate (Pc), mutation rate (Pm), and population size (N). Chromosome contains weights (w), linear coefficients (c) and bias (b).

Step (2): Approximates any desired signal $\hat{y}(x)$ by:

$$\hat{y}(x) = \sum_{i=1}^{N} w_i \psi(\alpha_i x - \beta_i) + cx + b \quad \ldots\ldots (19)$$

Step (3): Compute the sum square error (SSE) for all chromosmes, SSE is equal to the differences between target class number and the calculated class number. Thus by denoting

$$e(x) = y(x) - \hat{y}(x) \ldots\ldots (20)$$

By a time-varying error function at time (x), where $y(x)$ is the desired response. The energy function is defined by:

$$SSE = E = \frac{1}{2}\sum_{x=1}^{n} e(x)^2 \quad \ldots\ldots (21)$$

Step (4): Evaluation, where each chromosome is evaluated according to its value of the fitness function.

$$F = \frac{1}{E + \varepsilon} \quad \ldots\ldots (22)$$

Step (5): Save best population (beginning of elitism technique).

Step (6): Selection in which parents for reproduction based on their fitness are selected.

Step (7): Recombination, which generates two offspring from the two parent's chromosomes through the exchange of real string (crossover).

Step (8): Mutation, which apply a random mutation to each offspring.

Step (9): Replace old population with new population.

Step (10): Copy best population to the new population in the next generation.

Step (11): Go to step (2), if some criterion is not reached.

5. Numerical Experiments

Some computer experiments were carried out to evaluate performances of this method.

5-1 Static Nonlinear Mappings

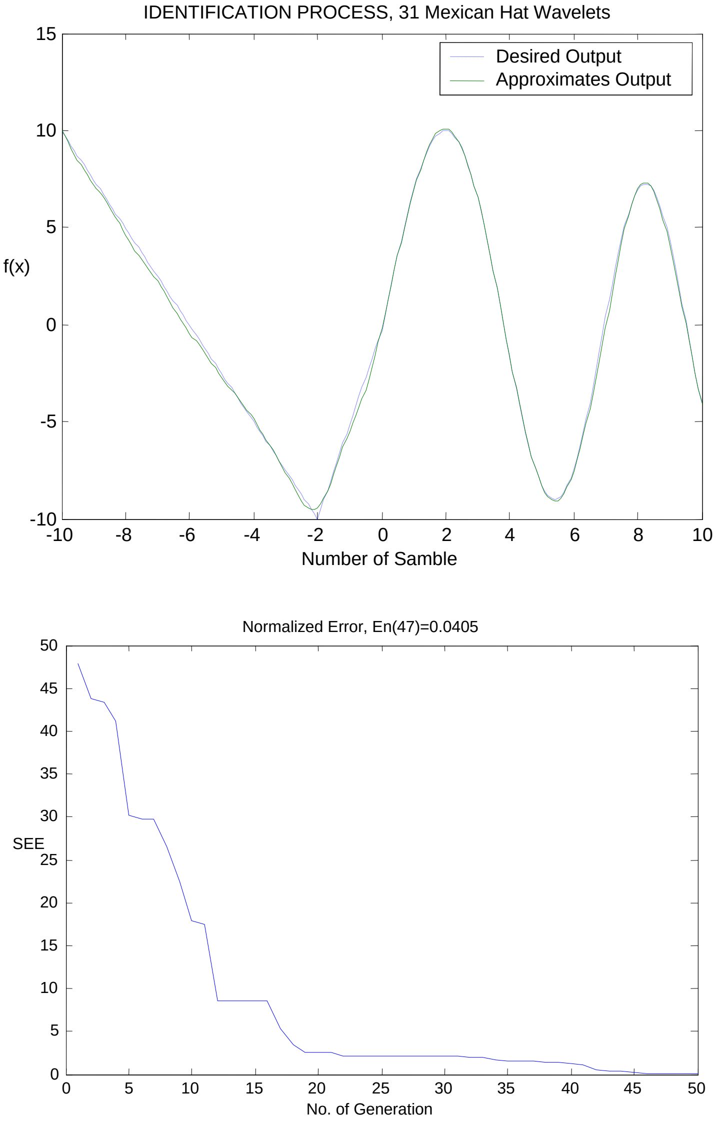
The first example is the approximation of a single variable function, without noise, given by:

$$f(x) = \begin{cases} -2.186x - 12.864 & \text{if } -10 \le x < -2 \\ 4.246x & \text{if } -2 \le x < 0 \\ 10e^{-0.05x-0.5}\sin\left((0.03x+0.7)x\right) & 0 \le x \le 10 \end{cases} \quad \text{................} \; (23)$$

This example was first proposed in [2], which is one of the seminal papers on wavelet networks (wavenets).

The following simulations will describe the results of the wavenet network performance employing Mexican Hat super-mother wavelets. GAs is used for searching an optimal arrangement of the weights. The following parameters are set, population size=100, crossover probability (Pc)=0.87, and mutation probability (Pm)= 0.20.

Figure (4) capture the learning performance of the wavenet network using (31) Mexican Hat wavelets at level (5), SSE of 0.0405 is reached for only (47) generations.



Figures (4) Simulation results for static nonlinear system using GA

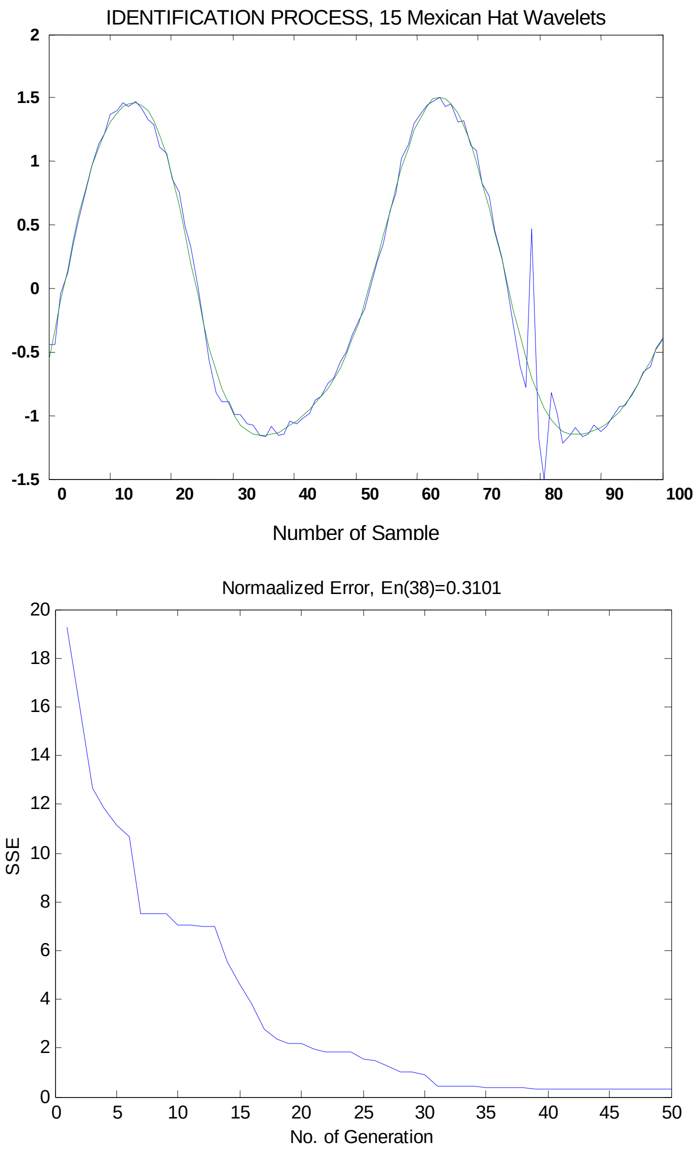
5-2 Identification of a Noisy Impulse Corrupted Model

The system $y(x)$ resumed with 10 db SNR (noise $\sigma^2 = 0.1$) and an outlier at x = 80.

$$\mathbf{y(x) = \frac{y(n-1)y(n-2)}{1 + y^2(n-1) + y^2(n-2)} + u(n)} \quad \textbf{.. (24)}$$

where $u(n) = 0.8\sin(2\pi n/50)$, *random noise with normal distribution.*

Figure (5) presents the noisy impulse corrupted system data, and the resynthesized system signal using the estimated optimum parameters corresponding to (15) Mexican Hat wavelets. As a result, the network can tolerate some noisy effects and impulse corruption. SSE of 0.3101 is reached for only (38) generations. The following parameters are settings, population size= 100, crossover probability (Pc)= 0.7, and mutation probability (Pm)= 0.2.



Figures (5) Simulation results for the nonlinear dynamic system using GA

6. Conclusion

The idea of combining wavelet theory with neural networks resulted in a new type of neural network called wavelet networks or wavenets. The wavenets use wavelet functions as hidden neuron activation functions.

The wavenet networks can be used for function approximation to static and dynamic nonlinear input-output modeling of processes, The wavenet improves the performance of the trained network for fast convergence, robustness to noise interference, and high complex ability to learn and track of unknown/undefined complex systems.

Genetic algorithm has been used successfully for training wavenet because GAs reaches quickly the region of the optimal solutions. Tests shows that GA obtain best weight vector and produces a lower sum square error (SSE) in a short period of time (No. of iterations).

7. References

1. Q., Zhang, and A., Benveniste, ***"Wavelet Network"***, IEEE Trans. Neural Networks, Vol. 3, No. 6, Nov.1992, pp. 889-898.

2. Q., Zhang, ***"Using Wavelet Network in Nonparametric Estimation"***, IEEE Trans. on Neural Networks, Vol. 8, No. 6, 1997.

3. I., Daubechies, ***"Ten Lectures on Wavelets"***, SIAM, 1992.

4. H. K., Roberto, and B. M., Victor, ***"Linear-Wavelet Models for System Identification"***, 2002, Available on line at http://www.csd.abdn.ac.uk.

5. G., Lekutai, ***"Adaptive Self-Tuning Neuro Wavelet Network Controllers"***, html 1997, Available on line at http://www.Iims. mech.

6. I., Daubechies, ***"The Wavelet Transform, Time-Frequency Localization and Signal Analysis"***, IEEE Transactions on Information Theory, Vol. 36-5, 1990, pp. 961-1005.

7. S., Mallat, ***"A Theory for Multiresolution Signal Decomposition: The Wavelet Transform"***, IEEE Trans. Pattern Anal. Machine Intell., Vol. 7, 1989, pp. 674-693.

8. Y., Oussar, and G., Dreyfus, ***"Initialization by Selection for Wavelet Network Training"***, Neurocomputing, Vol. 34, 2000, pp. 131-143, Available on line at http://www.neurones.espci.fr/Francais.Docs.

9. K., Kobayashi, and T., Torioka, ***"Designing Wavelet Networks Using Genetic Algorithms"***, 1998, Available on line at http:// www.nn.csse.yamaguchi-u.ac.jp/k.

10. A., Mathur, ***"Wavelet Self Organizing Maps and Wavelet Neural Networks: A study"***, 2000, Available on line at http://www.systemtechnik.tuilmenau.de.

11. Q., Zhang, ***"Neural Network Structures"***, 1999, Available on line at http:// www.Ieee. Cz/ knihovn a/ zhang.

12. G., Vachtsevanos, and P., Wang, ***"A Wavelet Neural Network Framework for Diagnostics of Complex Engineered Systems"***, 1999, Available on line at http://www.shef.ac.uk.