

Designing Wavelet Networks Using Genetic Algorithms

Kunikazu Kobayashi and Toyoshi Torioka
Department of Computer Science and Systems Engineering,
Faculty of Engineering, Yamaguchi University
Phone: +81-836-35-9479
Fax: +81-836-29-0053
E-mail: k@csse.yamaguchi-u.ac.jp

ABSTRACT: A new method to design wavelet networks using genetic algorithms is proposed. Designing wavelet networks is regarded as a combinatorial optimization problem to arrange the windows of wavelets and then genetic algorithms are used to derive their optimal arrangement. The effectiveness of the proposed method is verified through computer experiments.

1 INTRODUCTION

In designing neural networks, it is important to determine an optimal network size, the number of hidden layers and units, to realize desired mappings.

Backpropagation method proposed by Rumelhart et al is widely used as a learning rule for multi-layered perceptrons (MLPs) [Rumelhart, Hinton, and Williams 1985]. It has, however, some problems such as trapping into local minima and slow convergence. Since there is no established method to determine the network size of MLPs one have to do it empirically. If the network size is smaller than the optimal size it is difficult to realize desired mappings. On the other hand, if the network size is larger generalization error increases. Therefore, it is crucial to determine the optimal network size for realizing desired mappings.

Recently, the methods for designing the networks using wavelet analysis have been proposed. Wavelet networks have the following merits [Pati and Krishnaprasad 1993; Zhang and Benveniste 1992; Kobayashi and Torioka 1994; Kobayashi, Torioka and Yoshida 1994].

- (1) It is guaranteed that wavelet networks can sufficiently approximate any functions in the Hilbert space $L^2(R)$.
- (2) The internal expressions are readily understandable because of their multi-resolutional structures.
- (3) It is not difficult to add or delete the hidden units not affecting the performances because the basis function is localized in the time-frequency plane and has no bias.

Pati et al proposed a method for network synthesis using discrete affine wavelet transform[Pati and Krishnaprasad 1993]. However, their method does not guarantee an optimal arrangement of windows because it uses only information of expansion coefficients in network synthesis.

In the present paper, we regard the arrangement of windows as a combinatorial optimization problem [Ueda, Kobayashi and Torioka 1997]. Then we propose a new method to optimally arrange windows using genetic algorithms. Through computer experiments, it is verified that our method provide an optimal arrangement.

2 PRELIMINARIES

In this section, the basics of wavelets is briefly introduced[Chui 1992]. Throughout the present paper, it is assumed that we treat a class of functions in the Hilbert space $L^2(R)$. The inner product and norm are defined as follows.

$$\langle f, g \rangle \triangleq \int_R f(x) \overline{g(x)} dx, \quad (1)$$

$$\|f\| \triangleq \langle f, f \rangle^{1/2}, \quad (2)$$

where $\bar{g}$ denotes the complex conjugate of g .

2.1 WAVELET TRANSFORM

The integral wavelet transform of function $f \in L^2(R)$, $T(a, b)$, is defined as

$$T(a, b) \triangleq \langle f, \psi^{(a,b)} \rangle, \quad (3)$$

where ψ is a basis function called *mother wavelet* or *analyzing wavelet* in wavelets but we named it *fitting wavelet (FW)* [Kobayashi and Torioka 1994] and then $\psi^{(a,b)}$ is obtained by translation of b and dilation by a .

$$\psi^{(a,b)}(x) = \frac{1}{\sqrt{a}} \psi\left(\frac{x-b}{a}\right). \quad (4)$$

The FW must satisfy the following equation.

$$c_\psi = \int_R \frac{|\Psi(\omega)|^2}{|\omega|} d\omega < \infty, \quad (5)$$

where Ψ is the Fourier transform of ψ .

2.2 INVERSION FORMULA

In realizing wavelet networks, parameters a and b should be discretized. In the present paper, we consider the following discretization.

$$\begin{cases} a = 2^{-n} \\ b = 2^{-n}mb_0 \end{cases} \quad (6)$$

where b_0 is a constant called *sampling rate* and $n, m \in \mathbf{Z}$. The value of b_0 should be determined such that ψ generates a frame of $L^2(R)$ (See [Chui 1992]).

In this case, the inversion formula is defined as:

$$f(x) \triangleq \sum_{n,m \in \mathbf{Z}} \langle f, \psi_{n,m} \rangle \psi^{n,m}(x), \quad (7)$$

where $\psi_{n,m}(x) = 2^{n/2} \psi(2^n x - mb_0)$ and $\{\psi^{n,m}\}$ refers to the dual basis of $\{\psi_{n,m}\}$.

The FW should satisfy the following stability condition because f must be reconstructed from partial information of $T(a, b)$.

$$A \|f\|^2 \leq \sum_{n,m \in \mathbf{Z}} |\langle f, \psi_{n,m} \rangle|^2 \leq B \|f\|^2, \quad (8)$$

where $0 \leq A \leq B < \infty$ ($A, B \in R$).

2.3 TIME-FREQUENCY WINDOW

In wavelets, one of the most important concepts is *window* of a FW. The FW is localized in the time-frequency plane. That is, the FW can see such a time-frequency region and not other region. This property results in the identification of localization.

For FW ψ , center $\bar{x}$ and radius Δ_ψ are defined as:

$$\begin{aligned} \bar{x} &\triangleq \frac{1}{\|\psi\|^2} \int_R x |\psi(x)|^2 dx, \\ \Delta_\psi &\triangleq \frac{1}{\|\psi\|} \left\{ \int_R (x - \bar{x})^2 |\psi(x)|^2 dx \right\}^{\frac{1}{2}}. \end{aligned} \quad (9)$$

Similarly, for the Fourier transform of ψ , Ψ , center $\bar{\omega}$ and radius Δ_Ψ are defined as:

$$\begin{aligned} \bar{\omega} &\triangleq \frac{1}{\|\Psi\|^2} \int_{R^+} \omega |\Psi(\omega)|^2 d\omega, \\ \Delta_\Psi &\triangleq \frac{1}{\|\Psi\|} \left\{ \int_{R^+} (\omega - \bar{\omega})^2 |\Psi(\omega)|^2 d\omega \right\}^{\frac{1}{2}}, \end{aligned} \quad (10)$$

where Ψ^+ is Ψ defined for $\omega \in R^+ = [0, \infty)$. Since we only consider the positive frequency, the FW must satisfy more strict condition than eq.(5), i.e.

$$\frac{c_\psi}{2} = \int_{R^+} \frac{|\Psi(\omega)|^2}{\omega} d\omega < \infty. \quad (11)$$

Therefore, ψ has a time-frequency window,

$$W_\psi = [\bar{x} - \Delta_\psi, \bar{x} + \Delta_\psi] \times [\bar{\omega} - \Delta_\Psi, \bar{\omega} + \Delta_\Psi]. \quad (12)$$

In general, $\psi_{n,m}$ has a window,

$$W_{n,m} = [2^{-n}(\bar{x} + mb_0 - \Delta_\psi), 2^{-n}(\bar{x} + mb_0 + \Delta_\psi)] \times [2^n(\bar{\omega} - \Delta_\Psi), 2^n(\bar{\omega} + \Delta_\Psi)]. \quad (13)$$

Note that the size of the window is constant for any combination of n and m , i.e.

$$\begin{aligned} 2^{-n+1}\Delta_\psi \times 2^{n+1}\Delta_\Psi &= 4\Delta_\psi\Delta_\Psi \\ &= \text{const.} \end{aligned} \quad (14)$$

3 WAVELET NETWORK

It is assumed that the targets are restricted in both time and frequency domains. Thus, the inversion formula can be approximately realized using the networks with finite hidden units.

Consider a three-layered network ($1 \times N \times 1$) shown in Fig.1. The input and output units are linear elements and the output function of the hidden units which satisfies both admissibility and stability conditions, i.e. eq.(11) and eq.(8). It is assumed that the network sufficiently approximates the target. Intuitively, this means that the time-frequency region is effectively covered by the N windows. The estimate of the network, $\hat{f}$, is represented by:

$$\begin{aligned} \hat{f}(x) &= \sum_{n,m} c_{n,m} \psi_{n,m}(x) \\ &= \sum_{n,m} c_{n,m} 2^{n/2} \psi(2^n x - mb_0), \end{aligned} \quad (15)$$

where $c_{n,m}$ is a weight parameter between the hidden layer and the output unit.

4 DESIGN ALGORITHM

In this section, an algorithm to design wavelet networks using genetic algorithms is proposed. In eq.(15), The rough flow of our algorithm is as follows.

- 1) Given a data set S from a time series.

$$S = \{(x_k, f_k) | x_k \in [x_{\min}, x_{\max}], k = 1 \sim M\}. \quad (16)$$

- 2) Estimate the frequency bands $\omega \in [\omega_{\min}, \omega_{\max}]$ of the data set using Fourier analysis.

- 3) Pick up all the windows whose centers are involved in the time-frequency concentration (See Fig.2):

$$Q = [x_{\min}, x_{\max}] \times [\omega_{\min}, \omega_{\max}]. \quad (17)$$

- 4) An optimal arrangement of the windows, i.e. optimal combination of n and m , is searched using genetic algorithm. The parameter $c_{n,m}$'s are updated using gradient descent method. In our coding scheme, each string is represented by binary digit, which shows whether the corresponding unit exist, 1, or not, 0. Thus, all the strings consist of N binary digits. Then we employ three genetic operators, reproduction, crossover and mutation. The fitness function is defined by the following equation.

$$\text{fitness}_i = \frac{1}{E_i} + K \frac{1}{N_i}, \quad (18)$$

where E_i and N_i denote the approximation error (root mean squared error defined below) and the number of hidden units for individual i , respectively and K is a trade-off parameter between two terms.

$$E_i = \sqrt{\frac{1}{M} \sum_{k=1}^M \{\hat{f}(x_k) - f_k\}^2}. \quad (19)$$

5 EXPERIMENTS

Some computer experiments were carried out to evaluate performances of our method. The following basis function was used for a FW (See Fig.3).

$$\psi(x) = x \exp(-x^2/2). \quad (20)$$

The centers and radii in both time and frequency domains are calculated using eqs.(9) and (10)

$$\begin{cases} \bar{x} = 0 \\ \Delta_\psi = \sqrt{\frac{3}{2}} \end{cases} \quad \begin{cases} \bar{\omega} = \frac{2}{\sqrt{\pi}} \\ \Delta_\psi = \sqrt{\frac{3}{2} - \frac{4}{\pi}} \end{cases}$$

Then we prepared the following function as a target.

$$f(x) = x \sin(x) \cos(5x) \sin(10x) \cos(30x) \sin(50x). \quad (21)$$

The target function is sampled in $x \in [0, 1]$ at sampling rate $\omega_s = 128\pi$. Consequently, we have 64 samples.

The frequency bands of the samples are estimated using FFT (fast Fourier transform):

$$\omega \in [2\pi, 12\pi] \cap [20\pi, 32\pi]. \quad (22)$$

There are 248 FWs whose centers are involved in the above bands.

The followings are parameter settings:

$$\begin{aligned} b_0 &= 1.0 \\ \text{population size} &= 20 \\ K &= 10 \\ \text{crossover probability} &= 0.8 (\text{one-point crossover}) \\ \text{mutation probability} &= 0 \sim 0.3 (\text{adaptive mutation}) \end{aligned}$$

After 50 generations, we had 139 FWs. Namely, the number of FWs could be reduced using genetic algorithms.

Figure 4 shows the result by our method. The solid line shows the approximation and the broken line is the target. Table 1 shows the evaluation values of approximation error (RMSE1) and generalization error (RMSE2). In this table, WN1 refers to our method, WN2 is a result without optimization [Pati and Krishnaprasad 1993] and MLP is a result of MLP with 139 hidden units.

Table 1: Results

	N	RMSE1	RMSE2
WN1	139	3.48×10^{-3}	7.22×10^{-2}
WN2	248	4.60×10^{-4}	1.05×10^{-2}
MLP	139	2.49×10^{-2}	2.38×10^{-2}

6 CONCLUSION

The present paper has proposed the method for designing wavelet networks. Genetic algorithms are used for searching an optimal arrangement of the hidden units. Through computer experiments, it is shown that our method could reduce the number of hidden units.

REFERENCES

- Chui 1992. An Introduction to Wavelets. Wavelet Analysis and Its Applications (Volume 1), Academic Press.
- Kobayashi and Torioka 1994. A Wavelet Neural Network for Function Approximation and Network Optimization, Vol.4, pp.505–510, Proceedings of ANNIE'94, AMSE Press.
- Kobayashi, Torioka and Yoshida 1994. A wavelet neural network with network optimizing function. Trans. on Institute of Electronics, Information and Communication Engineerings (In Japanese), Vol.J77-D-II, No.10, pp.2121–2129.

Pati and Krishnaprasad 1993. Analysis and synthesis of feedforward neural networks using discrete affine wavelet transformations. IEEE Trans. on Neural Networks, Vol.4, No.1, pp.73–85.

Rumelhart, Hinton and Williams 1985. Learning internal representations by error propagation. In Parallel Distributed Processing: Explorations in the Microstructures of Cognition (Volume 1), pp.318–362. MIT Press.

Ueda, Kobayashi and Torioka 1997. A wavelet neural network with evolutionary generated structure. Trans. on Institute of Electronics, Information and Communication Engineerings (In Japanese), Vol.J80-D-II, No.2, pp.652–659.

Zhang and Benveniste 1992. Wavelet networks. IEEE Transactions on Neural Networks, Vol.3, No.6, pp.889–898.

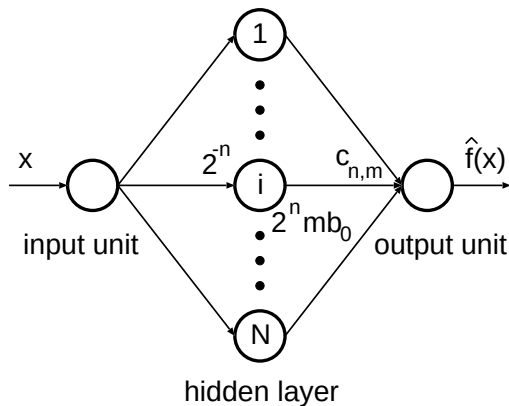


Figure 1: A wavelet network

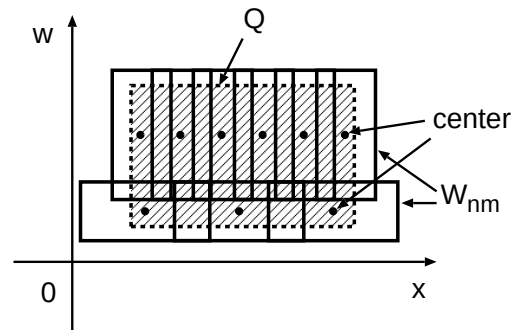


Figure 2: Time-frequency concentration Q and windows $W_{n,m}$

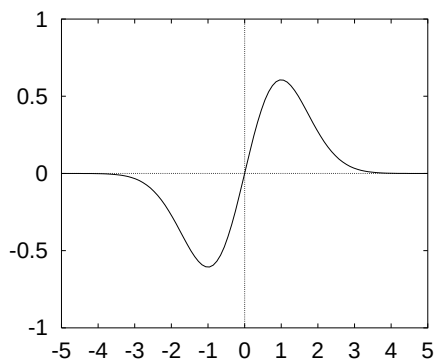


Figure 3: A fitting wavelet used in our experiments

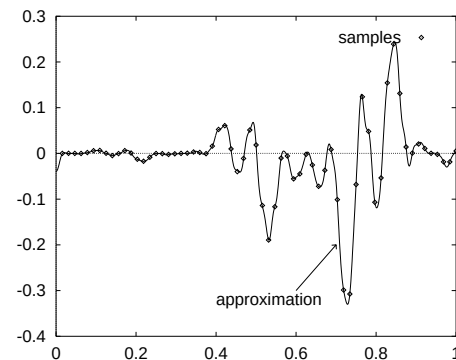


Figure 4: Approximation by our method