FUZZY WAVE-NETS: AN ADAPTIVE, MULTIRESOLUTION, NEUROFUZZY LEARNING SCHEME

Marc Thuillard
Siemens Building Technologies
Alte Landstrasse
CH-8708 Maennedorf
Switzerland
Email: Marc.thuillard@CCH.CERBERUS.CH

ABSTRACT: The complementarity of fuzzy logic and multiresolution analysis has been now recognized. Successful applications of fuzzy-wavelet techniques in commercial products (flame detectors) are on the market. We present here a new technique that combines wavelet and fuzzy logic under the form of a fuzzy wave-net. The main novelty of the algorithm is that the evolution equation consists of a neural version of the frame algorithm that permits to generalize wave-nets to frames, and more specifically to develop a fuzzy-wave net. The fuzzy-wave net approach furnishes an automatical procedure to determine adaptively the best membership functions and rules during on-line learning. With only a few points, not much information is known and the system uses a small number of rules with low resolution membership functions. As the number of points increased, the number of rules is raised and membership functions with a better resolution are introduced if necessary.

KEYWORDS: wavelet, fuzzy, neurofuzzy, on-line learning, wave-net

INTRODUCTION

The need for the integration of multiresolution techniques in soft computing has lead to two novel signal processing methods: wavelet networks (Zhang (1992), Szu (1992)) and fuzzy wavelets (Thuillard (1997, 1998a, 1998b)).

Wavelet networks combine neural networks and wavelet theory. In its simplest form, wavelet networks uses the similarity existing between the structure of a perceptron and a wavelet decomposition. In the one-dimensional case, the output f(x) of a 3-layers perceptron is

$$f(x) = \sum_{i=1}^{k} d_i \cdot \psi(a_i \cdot x + b_i) \qquad (1)$$

with y the activation function and a_i, b_i, c_i the network parameters (weights) that are optimized during learning. The similarity existing between a wavelet decomposition structure and a perceptron is obvious if f(x) is regarded as a weighted sum of wavelets:

$$f(x) = \sum_{m,n} d_{m,n} \cdot \psi(2^m \cdot (x - n)) + \bar{f} \qquad (2)$$

with $\bar{f}$ the main of f(x), $d_{m,n}$ the wavelet coefficients and y the wavelet.

In its simplest version, a wavelet network corresponds to a 3-layers perceptron using wavelets as activation functions.

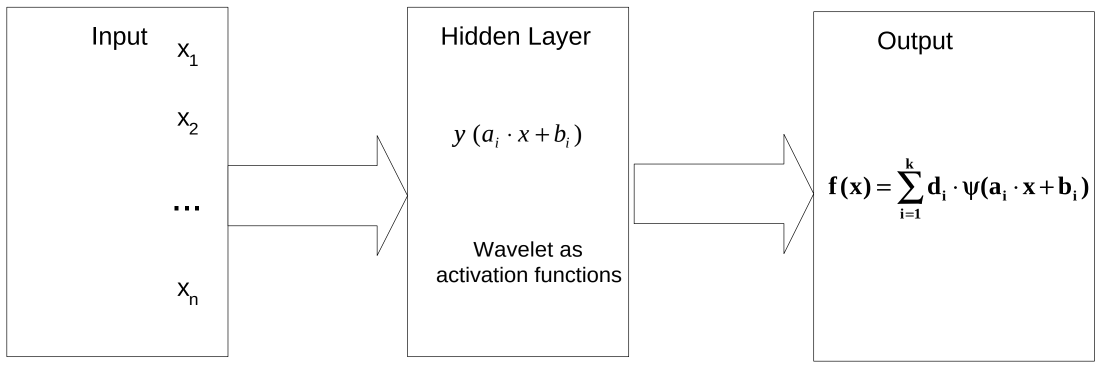

Figure 1: Structure of a wavelet network.

A number of other methods uses wavelet coefficients as input to a conventional neural network. A further method has been proposed by Bakshi (1992). The method, called wave-net uses a central property of orthogonal wavelets, namely the orthogonality of the different wavelets and scaling functions. We will show below that wave-nets can be extended to neurofuzzy.

ON-LINE LEARNING WITH FUZZY-WAVELET: A NEUROFUZZY APPROACH

The basic idea behind fuzzy-wavelet is to use wavelet theory to determine the most appropriate resolution to describe locally a control surface with fuzzy rules. The method furnishes also the best rules with their confidence levels. Fuzzy-wavelet techniques represent an alternative to neurofuzzy techniques for off-line learning or classification problems.

Fuzzy-wavelet can be also applied in a modified form to on-line learning problems. The method is quite different from the fuzzy-wavelet methods described in the introduction. Under this new form, it belongs to the neurofuzzy class of neural networks and can be classified as a fuzzy wave-net. The main feature of the fuzzy-wave net is that the most appropriate membership functions and rules are modified adaptively during learning using a multiresolution technique. With only a few points, not much information on the control surface is known and the control surface is better described with a small number of rules. As the number of points increased, the number of rules is raised if necessary. The method furnishes an automatical procedure to determine adaptively the best membership functions and rules. As starting point, we will use the basic architecture of a wave-net.

Figure 3 shows the architecture of the learning algorithm. It consists of a series of neural networks, using both wavelets $\psi_{j,n}$ (x) and scaling functions $\phi_{j,n}$ (x) as activation functions. Each neural network uses activation functions of a given resolution. The jth neural network optimizes the coefficients $c_{j,n}$ and $d_{j,n}$, with $f_j(x)$ the output of the jth neural network.

$$f_j(x)=\sum_n d_{j,n} \cdot \psi_{j,n}(x)+\sum_n c_{j,n} \cdot \phi_{j,n}(x) \qquad (3)$$

The structure of the network is up to here very similar to the wave-net. In a wave-net the coefficients $c_{j,n}$ and $d_{j,n}$ are obtained with a gradient descent. The method works well only with orthogonal wavelets. We will present now a new method that permits to extend to biorthogonal wavelets. The main motivation behind our approach is that once the wave-net works with biorthogonal wavelets, then it is straightforward to transform it into a multiresolution Neurofuzzy method, by using for instance biorthogonal splines. This is achieved by using the following evolution equation:

$$d_{j,n}(k) = d_{j,n}(k-1) + LR\,(f_j(x)-y(x))\;\tilde{\psi}_{j,n}(x) \qquad (4)$$

$$c_{j,n}(k) = c_{j,n}(k-1) + LR\,(f_j(x)-y(x))\;\tilde{\phi}_{j,n}(x) \qquad (5)$$

with $y_k(x)$, the k^{th} input point , $c_{j,n}(k)$ $d_{j,n}(k)$ the weights at step k and LR the learning rate, $\tilde{\psi}_{j,n}(x)$ $\tilde{\phi}_{j,n}(x)$, the dual functions to $\psi_{j,n}(x)$ and $\phi_{j,n}(x)$.

The algorithm defined by eq.(3-5) is very similar to the frame algorithm (Pei (1997)), that permits the iterative reconstruction of a signal from its frame projection. In other words, eq.(3-5) is a neural network implementation of the frame algorithm (also called Richardson method).

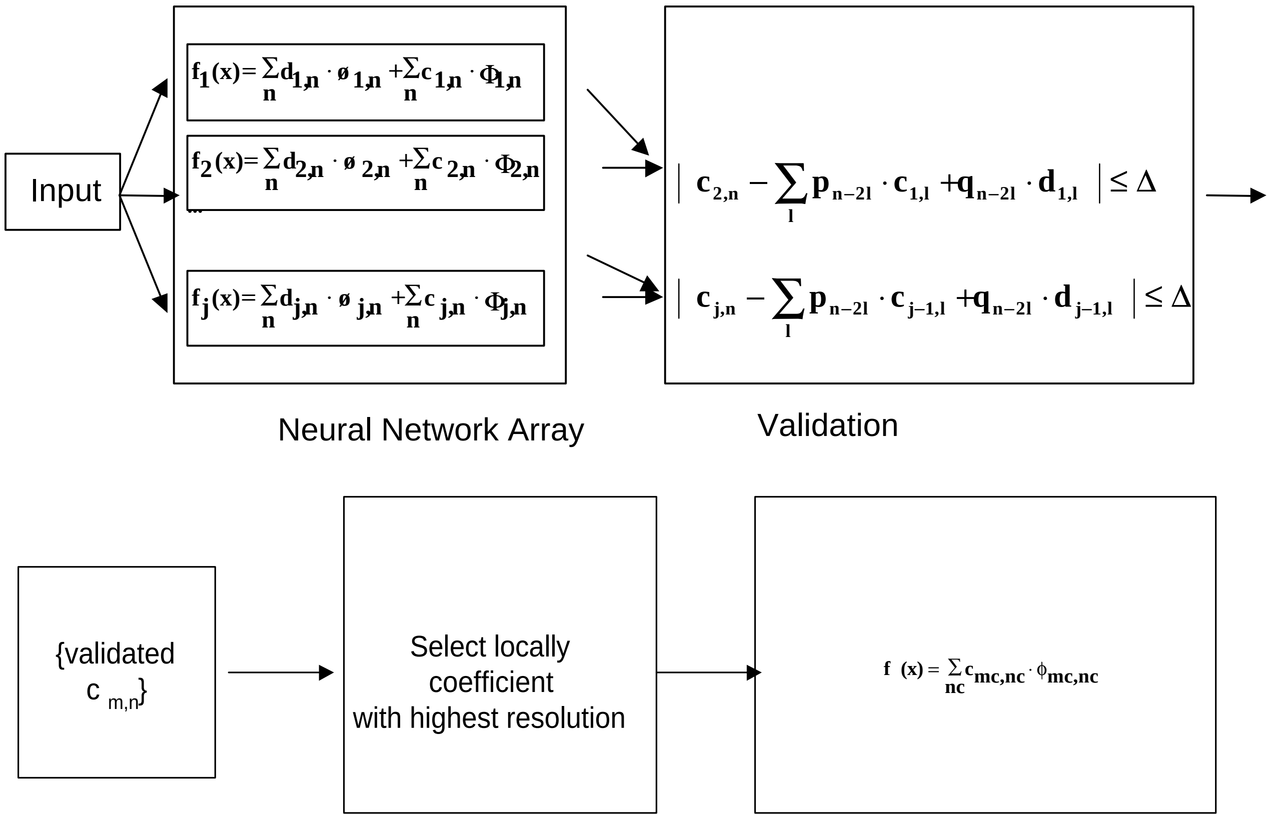

Figure 3: Structure of a fuzzy-wave net. The input signal is approximated at several resolution as a weighted sum of wavelets $\psi_{n,m}$ and scaling functions $\phi_{n,m}$ at a given resolution. The validation module compares the approximation coefficients $c_{j,n}$ to the approximations and wavelet coefficients at one level of resolution lower.

At each iteration step, the weights from the different networks are validated using a property of the wavelet decomposition, namely that the approximation coefficients c_j at level j can be computed from the approximation and wavelet coefficients at level j-1:

$$c_{j,n}=\sum_l p_{n-2l}\cdot c_{j-1,l}+q_{n-2l}\cdot d_{j-1,l} \qquad (6)$$

with p_{n-2l} and q_{n-2l} the filter coefficients corresponding to the reconstruction algorithm (see Mallat (1998) for details on the reconstruction algorithm and the filter coefficients).

In order for a coefficient to be validated, the difference between the weight of the membership function (model j) and the weight computed from the approximation and wavelet coefficients at one level of resolution lower (model j-1) must be smaller than a given threshold. As validation criterium for the coefficient $d_{j,n}$, we require

$$\left| \; c_{j,n} - \sum_{l} p_{n-2l} \cdot c_{j-1,l} + q_{n-2l} \cdot d_{j-1,l} \; \right| \le \Delta \qquad (7)$$

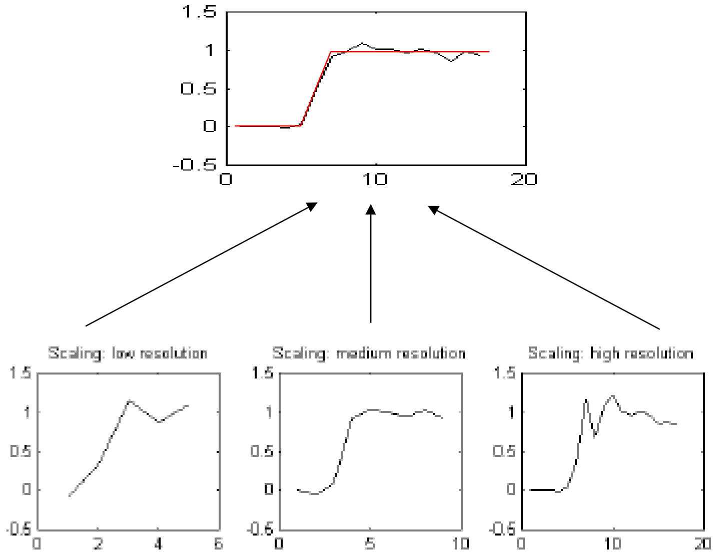

Figure 4: input function and output of the fuzzy-wavelet network after 60 steps (above). Below:output of the 3 neural networks at step 60.

Figure 4 shows an example, using this strategy. Biorthogonal spline wavelets and scaling functions (biorthogonal 3.2) proposed by Cohen (1982) are used activation functions. The model consists of an array of 3 neural networks, each corresponding to a different resolution level.

Figure 4 shows the input and output functions after 60 steps (k=60). The error of each neural network is larger than the error of the output of the wave-net (figure 4b). This is so, because the best coefficient is chosen locally out of the 3 possible choices. The decision on which coefficient to use is obtained from the validation eq.(9) (Δ=0.1). The best coefficients are chosen adaptively among the set of validated coefficients**.** The selection of the best coefficients is made according to a simple rule. The validated coefficients corresponding to the highest possible resolution are taken (default coefficient=$d_{1,n}$). The transformation of the results into a fuzzy controller is straightforward and can be carried out using one of the methods in Thuillard (1997, 1998a, 1998b).

REFERENCES

Bakshi, B.B., and Stephanopoulos G.,1992, “Wavelets as Basis Functions for Localized Learning in a Multi-Resolution Hierarchy“, IJCNN Int. Joint Conf. On Neural Networks, vol. 1, IEEE, Baltimore, II-141.

Cohen, A., Daubechies, I., and Feauveau, J.-C.,1982,“Biorthogonal Bases of Compactly Supported Wavelets.“ Commun. On Pure and Appl. Math. 45, 485.

Pei, S.-C. and Yeh, M.-H.,1997,“An Introduction to Discrete Finite Frames“, IEEE Signal Proc. Mag. 14, N°6, 84.

Szu, H. H. Telfer B., and Kadambe,S.,1992,“Neural Network Adaptive Wavelets for Signal Representation and Classification“, Opt. Eng. 31, 1907.

Mallat, S., 1998,“A Wavelet Tour of Signal Processing“, Academic Press (San Diego).

Thuillard, M.,1997,“Adaptive Fuzzy-Wavelet Modelling“, European Symposium on Intelligent Techniques (Bari), 244.

Thuillard M.,1998a,“Fuzzy-Wavelets: Theory and Applications“, Proc. EUFIT’98, Sixth European Congress on Intelligent Techniques and Soft Computing, Sept.8-10,1998 (Aachen), Ed. H.-J. Zimmermann, Mainz Verlag, Vol. 2, 1149.

Thuillard, M.,1998b,“Fuzzy Logic expressed in the Wavelet Framework“, Wavelets Applications Workshop, Tessin (abstract).

Zhang Q. and Benveniste,A.,1992,“Wavelet Network“, IEEE Trans. Neural Networks 3, 889.

APPENDIX A: BIORTHOGONAL WAVELET DECOMPOSITION

A bi-orthogonal wavelet decomposition necessitates two families of wavelets $\{y_{m,n}\}$ and $\{\tilde{y}_{m,n}\}$ satisfying the biorthogonality condition $<y_{m,n},\tilde{y}_{k,l}>=d(m-k)\cdot d(n-l)$. Any function f in $L^2(\Re)$ can be decomposed on the two wavelet families $\{y_{m,n}\}$ and $\{\tilde{y}_{m,n}\}$, with m, and n standing for dilation and shift:

$$f(x)=\sum_m\sum_n <\psi_{m,n}(x),f(x)>\cdot\tilde{\psi}_{m,n}(x)=\sum_m\sum_n <\tilde{\psi}_{m,n}(x),f(x)>\cdot\psi_{m,n}(x)= \sum_m\sum_n d_{m,n}\cdot\psi_{m,n}(x) \quad (A1)$$

The notation $<y_{m,n},f>$ describes the integral $2^{m/2}\cdot\int_{-\infty}^{\infty}y^*(2^m\cdot x-n)\cdot f(x)\cdot dx$.

If the decomposition is stopped at the j[th] level of decomposition, the function f(x) is

$$f(x)=f_1(x)+f_2(x)+...+f_j(x)+r(x) \quad (A2)$$

with

$$f_j(x)=\sum_n d_{j,n}\cdot\psi_{j,n}(x) \quad (A3)$$

$$r(x)=\sum_n c_{j,n}\cdot\phi_{j,n}(x) \quad (A4)$$

Both $c_{j,n}$ and $d_{j,n}$ can be obtained from c_{j-1} by moving average schemes:

Decomposition algorithm:

$$c_{j-1,n}=\sum_l h_{l-2n}\cdot c_{j,l} \quad (A5)$$

$$d_{j-1,n}=\sum_l g_{l-2n}\cdot c_{j,l} \quad (A6)$$

Inversely, the coefficient $c_{j,n}$ can be obtained from the coefficients $c_{j-1,n}$ and $d_{j-1,n}$ with the reconstruction algorithm.

Reconstruction algorithm:

$$c_{j,n}=\sum_l p_{n-2l}\cdot c_{j-1,l}+q_{n-2l}\cdot d_{j-1,l} \quad (A7)$$

If $y_{m,n} = \tilde{y}_{m,n}$, the wavelet decomposition corresponds to the decomposition relation for orthogonal wavelets:

$$f(x) = \sum_m \sum_n < \psi_{m,n}(x), f(x) > \cdot \psi_{m,n}(x) \tag{A8}$$

Figure 1 shows the dual wavelets $y(x)$ and $\tilde{y}(x)$ together with the corresponding scaling functions $f(x)$ and $\tilde{f}(x)$ for a biorthogonal (3.2) wavelet.

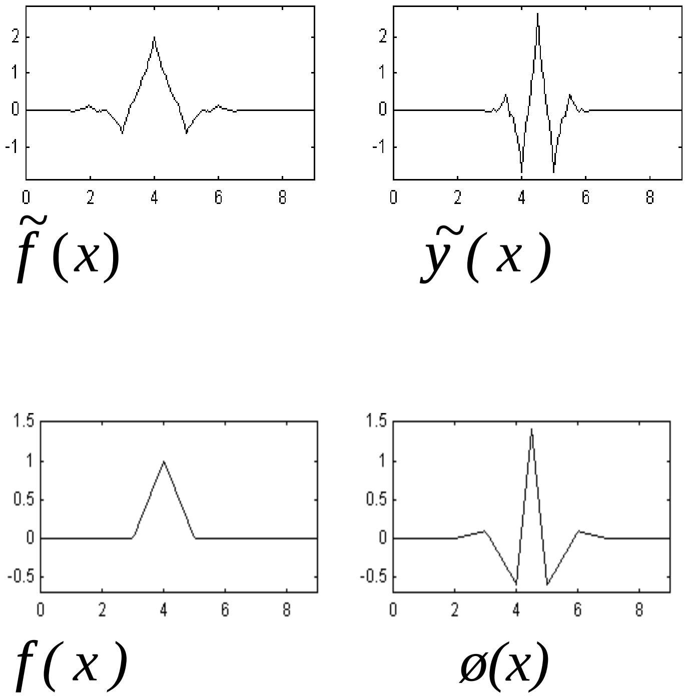

Figure A1: The biorthogonal 3.2 spline family