Initialization by Selection for Multi Library Wavelet Neural Network Training

Wajdi bellil[1], Mohamed Othmani[2] and Chokri Ben Amar[3]

[1] Faculty of sciences, University of Gafsa, City Zarroug, Gafsa, Tunisia
wajdi.bellil@ieee.org
[2] Department of informatics, High Institute of Technological Studies of Gafsa, Tunisia
mohamed.othmani@ieee.org
[3] Department of Electrical Engineering, University of Sfax, Tunisia
Chokri.benamar@ieee.org

Abstract. This paper presents an original architecture of Wavelet Neural Network based on multi Wavelets activation function and uses a selection method to determine a set of best wavelets whose centers and dilation parameters are used as initial values for subsequent training library WNN for one dimension and two dimensions function approximation. Every input vector will be considered as unknown functional mapping and then it will be approximated by the network.

1 Introduction

Wavelet Neural Networks (WNN) were introduced by Zhang and Benveniste [1-3] in 1992 as a combination of artificial neural networks and wavelet decomposition. WNN have recently attracted great interest, because of their advantages over radial basis function networks (RBFN) as they are universal approximators but achieve faster convergence and are capable of dealing with the so-called "curse of dimensionality." In addition, WNN are generalized RBFN. However, the generalization performance of WNN trained by least-squares approach deteriorates when outliers are present.

The task of training WNN involves estimating parameters in the network by minimizing some cost function, a measure reflecting the approximation quality performed by the network over the parameter space in the network. The least squares (LS) approach is the most popularly used in estimating the synaptic weights which provides optimal results.

Feed forward neural networks such as multilayer perceptrons (MLP) and radial basis function networks (RBFN) have been widely used as an alternative approach to functions approximation since they provide a generic black-box functional representation and have been shown to be capable of approximating any continuous function defined on a compact set in R^n with arbitrary accuracy [4]. Following the concept of locally supported basis functions such as RBFN, a class of wavelet neural networks (WNN) which originate from wavelet decomposition in signal processing has become more popular lately [5, 6, 7, 8, 9]. In addition to the salient feature of approximating any non-linear function, WNN outperforms MLP and RBFN due to its capability in dealing with the so-called "curse of dimensionality" and non-stationary signals and in faster convergence speed [10]. It has also been shown that RBFN is a special case of WNN.

This paper comprises four sections. Section 2 discusses the architecture of Multi Library Wavelet Neural Networks (MLWNN) and its performance function approximation. Section 3 contributes to Beta MLWNN and to discuss the implementation and results. Finally, Section 4 gives conclusions and summary for present research work and other possibilities of future research directions.

2 Theoretical Background

2.1 Classical Wavelet Neural Network Architecture

Wavelets occur in family of functions and each is defined by dilation a_i which controls the scaling parameter and translation t_i which controls the position of a single function, named the mother wavelet ψ(x). Mapping functions to a time-frequency phase space, WNN can reflect the time-frequency properties of function. Given an n-element training set, the overall response of a WNN is:

$$\hat{y}(w) = w_0 + \sum_{i=1}^{N_p} w_i \Psi_i \text{ , } where \ \Psi_i = \Psi\left(\frac{x - t_i}{a_i}\right) \tag{1}$$

where N_p is the number of wavelet nodes in the hidden layer and w_i is the synaptic weight of WNN.

This can also be considered as the decomposition of a function in a weighted sum of wavelets, where each weight w_j is proportional to the wavelet coefficient scaled and shifted by a_i and t_i. This establishes the idea for wavelet networks [11, 12].

This network can be considered composed of three layers: a layer with Ni inputs, a hidden layer with N_p wavelets and an output linear neuron receiving the weighted outputs of wavelets. Both input and output layers are fully connected to the hidden layer.

2.2 Multi Library Wavelet Neural Network (MLWNN)

A MLWNN can be regarded as a function approximator which estimates an unknown functional mapping:

$$y = f(x) + \varepsilon \tag{2}$$

where f is the regression function and the error term ε is a zero-mean random variable of disturbance. Constructing a MLWNN involves two stages: First, we should construct a wavelet library $W=\{W_1, W_2,\ldots,W_n\}$ of discretely dilated and translated versions of some mothers wavelets function $\Psi^1, \Psi^2,\ldots, \Psi^n$:

$$W_j = \left\{ \begin{array}{l} \Psi_i^j : \Psi_i^j(x) = \alpha_i \Psi^j\left(a_i(x - t_i)\right), \\ \alpha_i = \left(\sum_{k=1}^{N} \left[\Psi^j(a_i(x_k - t_i))\right]^2 \right)^{-\frac{1}{2}} \\ i = 1,\ldots, L \text{ and } j = 1,\ldots, n \end{array} \right\} \quad (3)$$

Where x_k is the sampled input, and L is the number of wavelets in each sub library W_j. Then select the best M wavelets based on the training data from multi wavelet library W, in order to build the regression. The architecture of multi library wavelet network is given in figure 1.

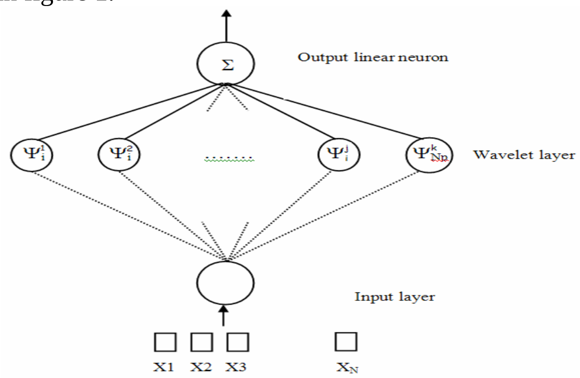


Fig. 1. MLWNN architecture.

$$\hat{y}(x) = \sum_{i\in I} w_i \Psi_i^1(x) + \sum_{i\in I} w_i \Psi_i^2(x) + \ldots + \sum_{i\in I} w_i \Psi_i^n(x) \quad (4)$$

2.3 An Initialization Procedure using a Selection Method

It is very inadvisable to initialize the dilations and translations randomly, as is usually the case for the weights of a standard neural network with sigmoid activation function. In the case of wavelet neural network and due to the fact that wavelets are rapidly vanishing functions, 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.

We propose to make use of multi library wavelet using a selection method to initialize the translation and dilation parameters of wavelet networks trained using gradient-based techniques. The procedure comprises five steps:

1- Adapt mother wavelets support as input signal support,

2- Generate a multi library of wavelets, using some families of wavelets described by relation (3),

3- Compute the mean square error for every wavelet output, 4- Choose, from the library, the Np wavelets that have the weakest error.

5- Use the translations and dilations of the Np relevant wavelets as initial values and use gradient descent algorithms like least mean squares (LMS) to minimize the mean-squared error:

$$J(W) = \frac{1}{N}\sum_{i=1}^{N}\left(y_i - \hat{y}(W)\right)^2 \quad (5)$$

where *J(W)* is the real output from a trained MLWNN at the fixed weight vector *W*.

3 BETA MLWNN for Function Approximation

In this section, we illustrate the new initialization procedure using a selection method on a multi library wavelet neural network based on Beta wavelet family and and compare its effectiveness to that of the classical procedure.

3.1 Beta Wavelet Family

The Beta function [14] is defined as:
if p>0, q>0, (p, q) ∈ IN

$$\beta(x) = \begin{cases} \left(\frac{x-x_0}{x_c-x_0}\right)^p \left(\frac{x_1-x}{x_1-x_c}\right)^q \; if \; x\in\left[x_0, x_1\right] \\ 0 \quad else \end{cases} \quad (6)$$

Where, $x_c = \dfrac{px_1 + qx_0}{p+q}$

We prove in [15] that all the derivatives of Beta function $\in L^2(\Re)$, are of class C^∞ and satisfy the admissibility wavelet condition.

3.2 Example 1: 1-D Function Approximation

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

$$f(x) = \begin{cases} -2.186x - 12.864 \;\; for \; x\in[-10,-2[\\ 4.246x \;\; for \; x\in[-2,-0[\\ 10\exp(-0.05x-0.5)\sin(x(0.03x+0.7)) \;\; for \; x\in[0,10[\end{cases} \quad (7)$$

The graph of this function is shown on Figure 2.

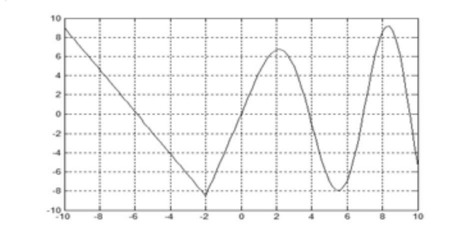


Fig. 2. The function output in the domain of interest.

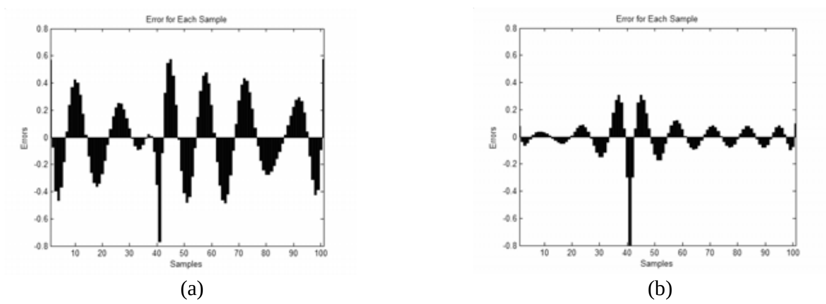
First, simulations on the 1-D function approximation are conducted to validate and compare the proposed MLWNN with the classical WNN. The input x is constructed by the uniform distribution on [-10 10]. The training sequence is composed of 101 points. The performance of the model is estimated using a test set of 101 equally spaced examples different from the training set.

We define the NMSE (Normalized Mean Square Error) as evaluation criteria.

$$NMSE = \frac{1}{N}\sum_{k=1}^{N}\left(\hat{f}(x_k) - y_k\right)^2 \quad (8)$$

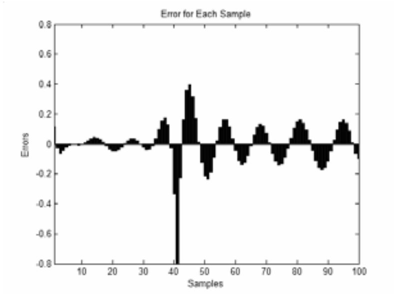
In the following, we present the results obtained with a network of 12 Beta wavelets, chosen as mother wavelets (second and third derivative of Beta function), for training network. Figure 3 shows the initial error histogram (a) obtained when the 101 input patterns are initialized with the classical architecture and the final error histogram (b) obtained when the 101 input patterns are training after 1000 iterations. Figure 4 shows the initial error histogram (c) obtained when the 101 trainings are initialized with the initialization by selection procedure using MLWNN and the final error histogram (d) obtained when the 101 input patterns are training after 1000 iterations. Comparing figures 3 and 4 shows clearly that the initialization by selection using MLWNN leads to:

- The best result in term of NMSE,
- Less scattered results both on the training set and on the test set.
- Using multi wavelet mothers as activation function gives best approximation.

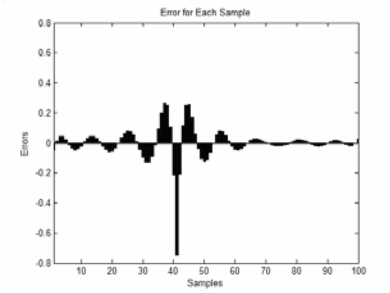


(a) (b)

Fig. 3. (a) Initial error for each sample after initialization using classical WNN architecture.(b) Final error for each sample after initialization using classical WNN architecture after 1000 iterations.



(c)



(d)

Fig. 4. (c) Initial error for each sample after initialization using MLWNN architecture. (d) Final error for each sample after initialization using MLWNN architecture after 1000 iterations.

Figure 5 shows the evolution of the NMSE according to the iteration; we can see the superiority of the proposed initialization selection algorithm based on multi wavelet library over the classical WNN based on one mother wavelet.

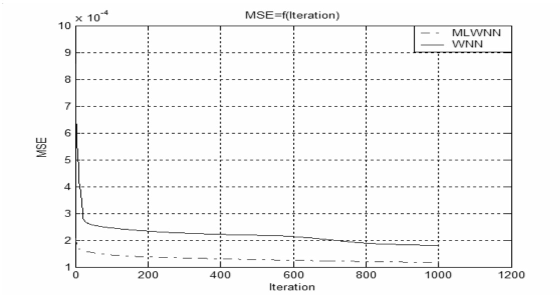


Fig. 5. Evolution of the NMSE according to the iteration.

3.3 Example 2: 2-D Function Approximation

The process to be modeled is simulated by a function of two variables without noise. The expression of this function is given by:

$$f(x_1, x_2) = \left(\frac{x_1+2}{2}\right)^5 \left(\frac{x_2+2}{2}\right)^5 \left(\frac{2-x_1}{2}\right)^5 \left(\frac{2-x_2}{2}\right)^5 \qquad (9)$$

Figure 6 is a plot of the surface defined by relation (9).

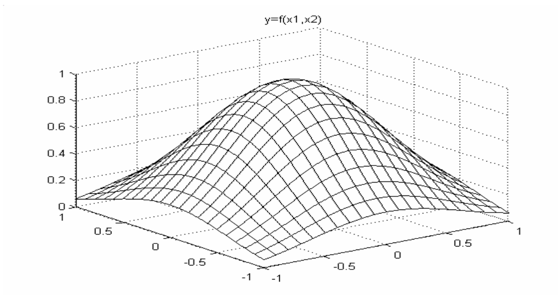
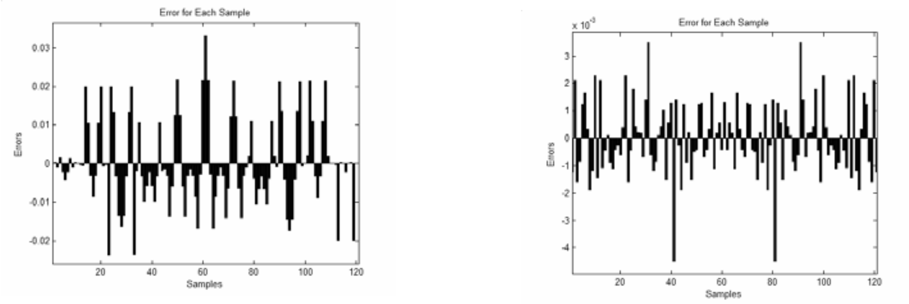


Fig. 6. 2-D data to be approximated.

In the following, we present the results obtained with a network of 9 Beta wavelets, chosen as mother wavelets (second and third derivative of Beta function), for training network. The training set contains 11x11 uniform spaced points. The test set V is constructed by 21x21 stochastic points on [-1,1]x[-1,1]. Figure 7 shows the final error histogram (a) obtained when the 121 trainings are initialized with the classical architecture initialization and the final error histogram (b) obtained when the 121 trainings are initialized with a selection procedure using MLWNN.



(a) (b)

Fig. 7. (a) Final error for each sample after initialization using classical WNN architecture. (b) Final error for each sample after initialization using MLWNN architecture.

These results show that the effect of the classical WNN initialization is much smaller than when the wavelet centers and dilations are initialized by selection using a multi library WNN, used together with Beta wavelets, it makes wavelet neural network training very efficient because of the adjustable parameters of Beta function.

4 Conclusions

Wavelet networks are a class of neural networks consisting of wavelets. In this paper, we have proposed a new Initialization by Selection algorithm for Multi library Wavelet Neural Network Training for the purpose of modeling processes having a small number of inputs.

We have shown that, when used a multi library wavelet networks and a selection procedure leads to results that are much more interesting than the classical architecture initialization. The selection of "relevant" wavelets within a regular wavelet lattice can also be performed by the technique of shrinkage. However, wavelet shrinkage is usually studied with orthonormal (or biorthonormal) wavelet bases, restricted to problems of small dimension.

As future research directions, we propose to use MLWNN in the case of adaptive self tuning PID controllers. The MLWNN is needed to learn the characteristics of the plant dynamic systems and make use of it to determine the future inputs that will minimize error performance index so as to compensate the PID controller parameters.

References

1. S. Mallat, 1998. *A wavelet tour of signal processing*. academic press.
2. Q. Zang, 1997. *Wavelet Network in Nonparametric Estimation*. IEEE Trans. Neural Networks, 8(2):227-236.
3. Q. Zang et al., 1992. *Wavelet networks*. IEEE Trans. Neural Networks, vol. 3, pp. 889-898.
4. H. Bourlard, Y. Kamp, 1988. *Autoassociation by multilayer perceptrons and singular values decomposition*, Biol. Cybernet. 59 (1988) 291-294.
5. A. Averbuch, D. Lazar, 1996. *Image compression using wavelet transform and multiresolution decomposition*, IEEE Trans. Image Process. 5 (1) 4-15.
6. G. Basil, J. Jiang, 1997. *Experiments on neural network implementation of LBG vector quantization*, Research Report, Depart of Computer Studies, Loughborough University.
7. G. Candotti, S. Carrato et al., 1994. *Pyramidal multiresolution source coding for progressive sequences*, IEEE Trans. Consumer Electronics 40 (4) 789-795.
8. S. Carrato, 1992. *Neural networks for image compression*, Neural Networks: Adv. and Appl. 2 ed., Gelenbe Pub, North-Holland, Amsterdam, pp. 177-198.
9. O.T.C. Chen et al., 1994. *Image compression using self-organisation networks*, IEEE Trans. Circuits Systems For Video Technol. 4 (5) (1994) 480-489.
10. M.F. Barnsley, L.P. 1993. Hurd, *Fractal Image Compression*, AK Peters Ltd.
11. Y. Oussar, 1998. *Réseaux d'ondelettes et réseaux de neurones pour la modélisation statique et dynamique de processus*, Thèse de doctorat, Université Pierre et Marie Curie.
12. C. Foucher and G. Vaucher, 2001. *Compression d'images et réseaux de neurones*, revue Valgo n°01-02, Ardèche.
13. Q. Zhang, 1997. *Using Wavelet Network in Nonparametric Estimation*, IEEE Trans. Neural Network, Vol. 8, pp.227-236.
14. C. Aouiti, M.A Alimi, and A. Maalej, 2002. *Genetic Designed Beta Basis Function Neural Networks for Multivariable Functions Approximation, Systems Analysis, Modeling, and Simulation*, Special Issue on Advances in Control and Computer Engineering, vol. 42, no. 7, pp. 975-1005.
15. C. Ben Amar, W. Bellil and A. Alimi. 2005-2006 *Beta Function and its Derivatives: A New Wavelet Family.* Transactions on Systems, Signals & Devices Volume 1, Number 3.
16. D. L. Donoho and I. M. Johnstone, 1995. *Adapting to unknown smoothness via wavelet shrinkage*, J. Amer. Statist. Assoc., vol. 90, no. 432, pp. 1200–1224, Dec.