# Wavelets Based Neural Network for Function Approximation

Yong Fang[1] and Tommy W.S. Chow[2]

[1] School of Communication and Information Engineering, Shanghai University, Shanghai, 200072, China
yfang@staff.shu.edu.cn
[2] Department of Electronic Engineering, City University of Hong Kong, Hong Kong, China
eetchow@cityu.edu.hk

**Abstract.** In this paper, a new type of WNN is proposed to enhance the function approximation capability. In the proposed WNN, the nonlinear activation function is a linear combination of wavelets, that can be updated during the networks training process. As a result the approximate error is significantly decreased. The BP algorithm and the QR decomposition based training method for the proposed WNN is derived. The obtained results indicate that this new type of WNN exhibits excellent learning ability compared to the conventional ones.

## 1 Introduction

The approximation of a general continuous function by neural network (NN) has been widely studied because of its outstanding capability of fitting nonlinear models for input/output data. A three-layer NN is usually represented by the following finite sums of the form:

$$g(\boldsymbol{x}) = \sum_{i=1}^{N} w_i \sigma(\boldsymbol{a}_i^T \boldsymbol{x} + b_i)\,, \tag{1}$$

where $w_i, b_i \in \boldsymbol{R}$, $\boldsymbol{a}_i \in \boldsymbol{R}^n$, $\sigma(\cdot)$ is a given function from $\boldsymbol{R}$ to $\boldsymbol{R}$, $\boldsymbol{x} \in \boldsymbol{R}^n$ is the input vector. It has been proved that the output, $g(\boldsymbol{x})$, is dense in the space of continuous function defined on $[0,1]^n$ if $\sigma(\cdot)$ is a continuous, discriminating function. Generally, $\sigma(\cdot)$ is adopted as a sigmoid function that is discriminatory.

Because wavelet decomposition has been emerged as a new powerful tool for representing nonlinearity, a class of network combining wavelets and neural networks have recently been investigated [1-10]. It has shown that this class of wavelet networks can provide better function approximation ability than ordinary basis function networks.

It is noticed that the development of the above WNNs are all theoretically based upon the wavelet frame theory given by Daubechies [4] in one-dimensional (1-D) case and generalized by Kugarajah and Zhang [7] in a multi-dimensional (M-D) case.

J. Wang et al. (Eds.): ISNN 2006, LNCS 3971, pp. 80 – 85, 2006.

A rather complex network architecture is inevitably resulted when the construction of an MWNN is solely in accordance with the wavelet frame theory. For practical applications, a relative large network size is often required.

In this paper, we study the possibility of using a 1-D wavelet function for the development of an MWNN. Based upon the theories of approximation of nonlinear functions using neural networks [8], our proposed WNN theorems are derived in accordance with the theories that the sigmoid function can be replaced by any continuous or discontinuous function [8], if certain conditions are satisfied (the wavelet function satisfy these conditions too) [8]. Following the development of this new MWNN, we based upon the discrete wavelet transforms to further develop another new type of MWNN, called MWNN-DWT. The activation function of the MWNN-DWT is a linear combination of wavelet bases rather than the sigmoid function or the wavelet function.

## 2 Wavelets and Wavelet Neural Networks (WNN)

Wavelets are functions whose dilations and translations form a frame of $L^2(\boldsymbol{R})$. That is, for some $a > 0, b > 0$, the family

$$\psi_{lk}(x) = a^{-l/2}\psi(a^{-l}x - bk)\text{, for } l,k \in Z \tag{2}$$

satisfies the frame property [4]. The sufficient conditions of wavelet frames were fiven in [4]. For instance, the "Morlet" wavelet with $a = 2, b = 1$ can build a frame of $L^2(\boldsymbol{R})$. Hence, the collection of all linear combination of elements of the frame $g(x) = \sum_{(l,k)\in \mathrm{I}} w_{lk}\psi_{lk}(x)$ (where $I \subset Z^2$) is dense in $L^2(\boldsymbol{R})$. This implies that for any $f \in L^2(\boldsymbol{R})$ and $\varepsilon > 0$, there exists a positive integer N and constants such that

$$\left| f(x) - \sum_{i=1}^{N} w_i \psi(a_i x + b_i) \right| < \varepsilon\,. \tag{3}$$

Pati and Krishnaprasad [3] connected the wavelet with NN by applying Daubechies' results [4]. They proposed the following network form

$$g(x) = \sum_{(l,k)\in \mathrm{I}} w_{lk}\psi_{lk}(x)\,, \tag{4}$$

where the index set I is the integer translation and the integer dilation, which are determined by using the time-frequency localization properties under a given accuracy [3-4], [6]. Kugarajah and Zhang [7] firstly built the M-D wavelet frames by a single mother wavelet as following form

$$\psi_{l,\boldsymbol{k}}(\boldsymbol{x}) = a^{-nl/2}\psi(a^{-l}\boldsymbol{x} - b\boldsymbol{k}) \quad \text{for } l \in Z, \boldsymbol{k} \in Z^n \tag{5}$$

where $\psi(x) \in L^2(\boldsymbol{R}^n)$, $\boldsymbol{x} \in \boldsymbol{R}^n$, $a, b \in \boldsymbol{R}$ and $a > 1$. Also, they gave the sufficient conditions of wavelet frames by generalizing Daubechies' theorem [4]. In the sequence, the dilation index l is a scalar and the scalar dilation parameter $a^l$ is shared by all the dimensional of a wavelet. They proposed a methodology to construct a M-D wavelet function leading to frames [7], but not all 1-D wavelet can be extended as a M-D wavelet. The special conditions can be found in [7]. Zhang and Benveniste [1], and Kugarajah and Zhang [7] respectively gave the M-D wavelet frames in the following multi-scaling forms:

$$\psi_{l,k}(\boldsymbol{x}) = (\det D_j)^{1/2}\psi(D_j\boldsymbol{x} - b\boldsymbol{k}) \quad \text{for } \boldsymbol{j},\boldsymbol{k} \in Z^n , \tag{6}$$

$$\psi_{l,k}(\boldsymbol{x}) = \det D_j^{1/2}\psi(D_j\boldsymbol{x} - T\boldsymbol{k}) \quad \text{for } \boldsymbol{j},\boldsymbol{k} \in Z^n , \tag{7}$$

where $D_j = diag(a^{j_1},\cdots,a^{j_n})$ , $j = (j_1,\cdots j_n)^T \in Z^n$ and $T = diag(b_1,\cdots,b_n)$ , $a > 1$, $b_i > 0, i = 1,\cdots,n$ , $b > 0$ . They proved that if a 1-D wavelet function $\psi(x)$ can constitute frames, the tensor product of the 1-D wavelet function can also constitute frames. Zhang and Benveniste [1] presented a network structure in the form of

$$g(x) = \sum_{i=1}^{N} w_i\psi[D_iR_i(x - \boldsymbol{t}_i)] + \bar{g} , \tag{8}$$

where $R_i$ are rotation matrices, and $\bar{g}$ is introduced to deal with nonzero mean functions on finite domains.

## 3 MWNN Based on Discrete Wavelet Transform (MWNN-DWT)

In the MWNN described in (12), the activation function is a 1-D mother wavelet function. As it is well known that the dilations and translations of wavelet can provide a good representation of nonlinearity [4],[6], a 1-D dilations and translations for the wavelet in the MWNN is then developed to enhance the approximation capability of the network. In this section we derive a new type of MWNN, called MWNN-DWT,

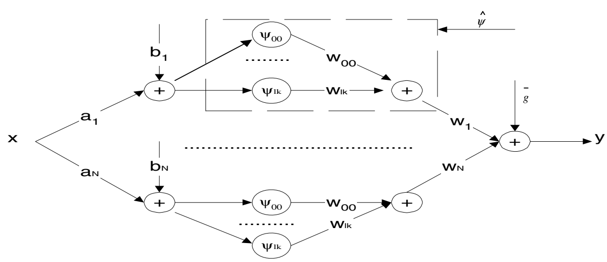


**Fig. 1.** The architecture of the MWNN

which is based upon the discrete wavelet transform. The output, $g(x)$, of MWNN-DWT depicted in Fig. 1 is represented by the following form:

$$g(\boldsymbol{x}) = \sum_{i=1}^{N} w_i \hat{\psi}(\boldsymbol{a}_i^T \boldsymbol{x} + b_i) + \bar{g} , \tag{9}$$

where $\hat{\psi}(x) = \sum_{(l,k)\in \mathrm{I}} w_{lk}\psi(a^l x - kb)$, $\{a^{l/2}\psi(a^l x - kb) | l,k \in Z\}$ is a frame for the $L^2(\boldsymbol{R})$ and $\mathrm{I}$ is the index set of pairs (l,k) of integer translation and integer dilation. The parameter $\bar{g}$ is introduced so that the approximation of functions with nonzero average is possible [1]. It is clear that the activation function in MWNN-DWT (9) is expressed in a linear combination of the dilating and translating wavelets. Through adaptively adjusting the 1-D wavelet frames, the MWNN-DWT is capable of providing an excellent approximation capability.

The (13) can be rewritten as the following two equivalent forms:

$$g(\boldsymbol{x}) = \sum_{i=1}^{N} w_i \sum_{(l,k)\in \mathrm{I}} w_{lk} \psi[a^l(\boldsymbol{a}_i^T \boldsymbol{x} + b_i) - kb] + \bar{g} \tag{10}$$

$$g(\boldsymbol{x}) = \sum_{(l,k)\in \mathrm{I}} w_{lk} \sum_{i=1}^{N} w_i \psi[a^l(\boldsymbol{a}_i^T \boldsymbol{x} + b_i) - kb] + \bar{g} \ . \tag{11}$$

To demonstrate the approximate capability of the MWNN-DWT, it is applied to approximate the function $f(x) = 0.5e^{-x}\sin(6x)$ over the domain [-1,1]. The parameter $\bar{g}$ was initialized by the mean of the function observations, and other parameters were simply randomized between -0.5 and 0.5. It should also be noted that a rather special procedure in initializing the parameters is used in [1], while the parameters of the proposed MWNN-DWT is simply randomized in a general fashion. In order to compare the approximation capability of the networks described in (8) and (11), the initial parameters of these networks are all similarly randomized. Note that no rotation parameter is required for the wavelet network (8) in a 1-D case [1]. The training set consist of 100 points uniformly sampled in $f(x)$. The function approximation performance of an MWNN-DWT in (11) with 5 neurons and 9 wavelet coefficients is compared to an WNN in (8) with 8 wavelons. In both networks, 25 parameters are used and trained by the standard BP algorithm. Fig.2 is the total squared errors over 2000 iterations, where the solid line shows the error of the MWNN-DWT and the dashed line represents that of the WNN. Clearly, the proposed MWNN-DWT provides much better results compared to that of the WNN.

It is noted that the parameters of the MWNN-DWT consists of two parts: the weights of the network and the coefficients of the wavelets. In order to determine the optimal index set I and to speed up the convergence of the MWNN-DWT, we divide the training process into two stages. Firstly, let $w_{00} = 1$, $w_{lk} = 0$, $l,k \neq 0$ and apply the BP algorithm for the network training. In this case, $\hat{\psi}(x) = \psi(x)$, the network is the same as the MWNN in (13) but less parameters is required. In this training stage, the activation function is fixed and the standard BP algorithm is used. After a number of iterations and the network converges to a specified level of error, all the parameters are fixed. Therefore, the (11) can be expressed as

$$g(\boldsymbol{x}) = \sum_{(l,k)\in I} w_{lk} G_{lk}(\boldsymbol{x}) + \bar{g} \ , \tag{12}$$

where $G_{lk}(\boldsymbol{x}) = \sum_{i=1}^{N} w_i \psi[a^l(\boldsymbol{a}_i^T \boldsymbol{x} + b_i) - kb]$. The second training stage is to determine the coefficients of the wavelets $w_{lk}$. As $w_{lk}$ appears to be linear in the output equation (12) of the network, most optimization techniques can be used to minimize the measure function if the index set is given. In this paper, the QR decomposition is used to adjust the activation function according to the form in (16). This process is equivalent to solving a least squares minimization problem for a given training set $T_P = \{(\boldsymbol{x}_i, f(\boldsymbol{x}_i)\}_{i=1}^{P}$. By choosing an order of the index set the equation (12) can be written as

$$F = GW \,, \tag{13}$$

where $F = (f(\boldsymbol{x}_1), \cdots, f(\boldsymbol{x}_P))^T$ is a column vector, $G = (\cdots, \hat{\boldsymbol{G}}_{lk}, \cdots, \boldsymbol{1})$ is the $P \times \#(I)$ matrix, #(I) denotes the number of the elements of the index set I, $\hat{\boldsymbol{G}}_{lk} = (G_{lk}(\boldsymbol{x}_1), \cdots, G_{lk}(\boldsymbol{x}_P))^T$, and $W$ is the coefficient vector which needs to be determined. To obtain the matrix G, search the $\hat{\boldsymbol{G}}_{lk}$ in accordance with the order: l, $k$ = $0, \pm 1, \pm 2, \cdots$. If the $rank(G, \hat{\boldsymbol{G}}_{lk}) > rankG$ and $\max_j G_{lk}(\boldsymbol{x}_j) > \varepsilon$, then $G \Leftarrow (G \;, \hat{\boldsymbol{G}}_{lk}, \boldsymbol{1})$, where $G = (G \;, \boldsymbol{1})$ and the given $\varepsilon$ is a very small number. Following the above procedures, QR decomposition is used to determine the coefficients and the averaged W. The weights can then evaluated. Obviously, the overall convergence rate is substantially speeded up. In the above example, it required 5000 iterations to converge to a total squared error of 0.0112. In order to speed up the rate of convergence and to enhance the approximation capability of our network, the proposed training scheme was used for the function $f(x) = 0.5e^{-x}\sin(6x)$. The proposed network with 5 hidden units was firstly trained by BP algorithm. After 100 iterations, we obtain the matrix G with 12 columns, which implies that there are 12 coefficients of wavelets required to be determined by QP decomposition. Our results show that a total squared error of $7.9954 \times 10^{-5}$ was obtained by only 100 training iterations together with the QR decomposition. The computational time required by this training scheme is substantially reduced compared to other MWNN training schemes.

For the case of a 2-D function given in above section: $f(x, y) = \sin(x)/x \sin(y)/y$, an MWNN-DWT with 50 hidden units was trained by the proposed training scheme. In this example, only 400 training iterations (with BP algorithm) together with the QR decomposition was required to provide a total squared error of 0.1688. Fig. 3 and Fig. 4 show the original function and approximation results respectively.

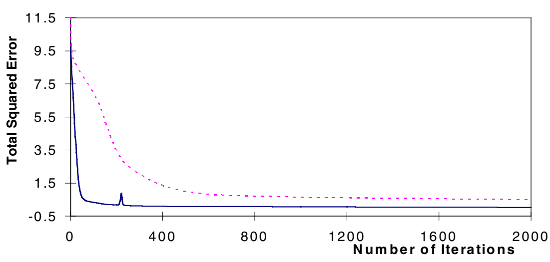


**Fig. 2.** The total squared errors of the WNN (8) and the proposed network for first function

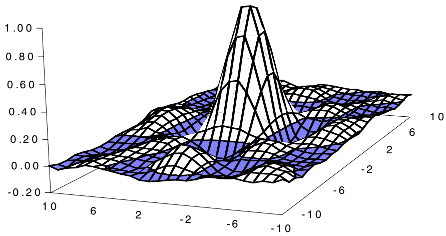


**Fig. 3.** Original 2-D function

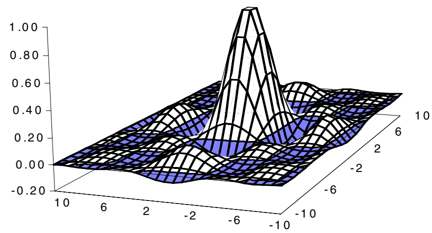


**Fig. 4.** Resulting approximation

## 4 Conclusion

In this paper, we use Discrete Wavelet Transform to extend the MWNN to a new type of network called, MWNN-DWT. This enables us to minimize the measure function by adjusting the wavelet bases activation function. Our results indicate that the proposed MWNN-DWT can deliver an enhanced function approximation capability. The proposed training algorithm, which is based on the standard BP algorithm and QR decomposition, has an outstanding convergence rate compared to other MWNN training algorithms.

## Acknowledgement

This work is supported by the National Natural Science Foundation of China (60472103, Shanghai Excellent Academic Leader Project (05XP14027), and Shanghai Leading Academic Discipline Project (T0102).

## References

1. Zhang, Q., Benveniste, A.: Wavelet Networks. IEEE Trans. on Neural Networks 3(1992) 889-898
2. Yamakawa, T., Uchino, E., Samatsu, T.: Wavelet Neural Networks Employing Overcomplete Number of Compactly Supported Non-orthogonal Wavelete and Their Applications. IEEE Int. Conf. on Neural Network 1(1994) 1391-1396
3. Pati, Y. C., Krishnaprasad, P.S.: Analysis and Synthesis of Feedforward Neural Networks Using Discrete Affine Wavelet Transformations. IEEE Trans. on Neural Networks 4(1993) 73-85
4. Daubechies, I.: The Wavelet Transform, Time-frequency Localization and Signal Analysis. IEEE Trans. on Information Theory 36(1990) 961-1005
5. Zhang, J., Walter, G.G., Miao, Y., Wayne, W.N.: Zhang, Q., Benveniste, A.: Wavelet Neural Networks for Function Learning. IEEE Trans. on Signal Processing 43(195) 1485-1496
6. Daubechies, I, Ten Lectures on Wavelets. Philadelphia, PA: SIAM Press(1992)
7. Kugarajah, T. Zhang, Q.: Multidimensional Wavelet Frames. IEEE Trans. on Neural Networks 6(19995) 1552-1556
8. Chen, T., Chen, H.: Approximarions of Continuous Functionals by Neural Networks with Application to Dynamic System. IEEE Trans. on Neural Networks 4(1993) 910-918
9. Chow, T.W.S., Fang, Y.: Two-dimensional Learning Strategy for Multiplayer Feedforward Neural Network. Neurcomputing 34(2000) 195-206