Comparative Analysis of the Generalization of Dyadic Wavelet based Neural Networks (DWNN) towards Optimal Network Structure for Non-linear Function Estimation

Pushpalatha M P and Nalini N

Abstract— **An algorithm for constructing and training the Dyadic Wavelet Neural Network is proposed combining the theory of multiresolution analysis (MRA) of wavelet transforms and the conventional neural networks. The focus is to mainly improve function approximation accuracy in terms of dilation and translation parameters of wavelets, meanwhile not increasing the number of wavelet bases leading to optimal network structure. The proposed activation functions are drawn from a family of orthonormal basis functions. The good localization characteristics of the basis functions, both in the input and frequency domain allow hierarchical, multi-resolution learning of input-output function from experimental data. Compared with Wavelet Neural Networks of previous works, the model accuracy and generalization capability of the DWN are improved by adjusting the resolution parameter as it plays a significant role in dyadic wavelet analysis and approximation of a given function. In the learning process exhaustive experimentations are conducted to illustrate the impact of learning rate with respect to the generalization of the output function. All these advantages have been reflected in our experimentations. Two benchmark functions are simulated to illustrate the effectiveness of the method.**

Index Terms—**Dyadic Wavelet Neural Networks (DWN), Function learning, Orthonormal scaling functions, Wavelet Transforms (WT).**

I. Introduction

Wavelets are a new family of basis functions that combine powerful properties such as orthogonality, compact support, localization in time and frequency, and fast algorithms [1, 2 and 3]. Cybenkos research has made the neural network into a powerful tool for signal representation particularly in function estimation problems [5].

The wavelet theory has found many applications in numerical analysis and signal processing [4], the Wavelet Transform (WT) of which provides another novel approach towards the function approximation problem [6,7,18]. Incorporating the time-frequency localization properties of wavelets and the learning abilities of general neural network (NN), WNN has shown its advantages over the regular methods utilizing wavelets as the basis function to construct a network [8, 9,10,12, 13, 14, 15, 17, 19, 20, 21].

The WNN's are characterized based on their activation functions used. When the continuous wavelets are selected as activation functions, and the parameters can take on any continuous real value, the translation, dilations and the network weights are updated in the process of training to optimize the network output. In the case of Dyadic wavelet series decomposition the network weights are optimized through learning process while the translation and dilation parameters are discretized in prior according to the dyadic sampling of the time-frequency plane. However, in order to improve computational efficiency, the values of the shift and scale parameters are often limited to some discrete lattices and are termed as the Discrete Wavelet Transform (DWT). The proposed methodology takes the advantage of the DWT in the learning process [11].

Though functional approximators other than wavelets may have the universal approximation property, in general, they do not possess the multiresolution property needed for approximation problems. The wavelets with coarse resolution capture the global (low frequency) behavior easily, while the wavelets with fine resolution can capture the local behavior (higher frequency) of the function accurately [1, 2]. This unique characteristic leads to the wavelet based neural network to be of the advantages of achieving fast convergence, easy learning and high accuracy.

The learning algorithm in the analysis phase validates the statistical measures used for the optimal network structure and parameter compatibility. The tuning of simulation parameters (learning rate and resolution) helps to arrive at an optimal network model through a training process.

This paper is organized as follows. The notations used and the basic concepts of wavelet subspace are introduced in section II and section III respectively. The learning algorithm for constructing the DWN is introduced in section IV. Two benchmark functions are simulated to illustrate the performance of DWN in comparison with other methods reported are given in section V. Finally, a brief conclusion is drawn in section VI.

M.P.Pushpalatha is with Department of Computer Science and Engineering Sri Jayachamarajendra College of Engineering Mysore -570 006, India (phone: 091 821-2548285; fax: 091 821-2548280 e-mail: mppvin@gmail.com.)

Nalini N is with Department of Computer Science and Engineering, Nitte Meenakshi Institute of Technology, Bangalore, India. (email: nalinaniranjan@hotmail.com).

II. Notations

The notations used are as in Table I.

TABLE I NOTATIONS USED IN DYADIC WAVELET NEURAL NETWORK

Term	Meaning
$L^2(R)$	Space in which squares of all functions are $f(.)$ are integrable on real data set R
$c_{i,m}$	WNN weights
$f(.)$	non-linear system
i	scaling parameter for scaling function
m	translation parameter for scaling function
n_j	scaling parameters for wavelet function
R	set of real numbers
$\{S\}_{i\in Z}$	closed subspace of $L^2(R)$
V_0	Subspace for rough approximations
W_j	Subspace for detail approxiamtions
$\phi(.)$	Scaling function
$\phi_{i,m}(.)$	Orthonormal base for Subspace S_i
$\psi(.)$	Wavelet function
$\psi_{j+i-1,n_j}(.)$	Orthonormal base for Subspace W_j
P	Number of training Patterns (Samples)
l	Index for pattern
N	Number of input units
k^l	Index for input unit
x_k^l	Input in vector notation i.e. l^{th} pattern presented at the k^{th} input
$\Phi^l_{a_j,b_j}$	Activation of Scaling unit j with pattern l
D_i^l	Desired Output for unit i with pattern l
O_i^l	Actual Output for unit i with pattern l
i, j, k	Index for output, hidden and input layers
b_j	Translation parameter of j^{th} scaling unit
a_j	Dilation parameter of j^{th} scaling unit
L	Resolution level
θ	Vector of $W_{i,j}, b_j, a_j$
η	Learning Rate (LR)
W_{ij}	Weight between i^{th} output and j^{th} wavelet unit
e^l	Error for the l^{th} pattern

III. Wavelet Subspace In Space L^2 (R)

In this section, we briefly review the theoretical results from wavelet theory that are relevant to this work.

The space $\mathbb{L}^2(R)$ of the system $f(.)$, must be divided in to a closed subspace V_0 and a sequence of subspaces $W_j, j = 1, 2, 3, \ldots, +\infty$. Based on this space division, the definition of the subspace V_0 is inherited from the theory of multi-resolution approximation [1, 2]. The subspace V_0 which characterizes the rough approximation modeling at Resolution-0, can be any subspace S_i.

The subspace $W_j, j = 1, 2, 3, \ldots, +\infty$ characterizes the modeling of the system details at Resolution-1, Resolution-2, Resolution $+\infty$, are then defined as the difference between the subspace S_j and the subspace S_{j-1} as described by equation given below:

$$W_j = S_j - S_{j-1}, j = 1, 2, 3, \ldots, +\infty \quad (1)$$

The properties of the subspaces S_j are specified in equations given below

$$\forall (j,k) \in Z^2, \hat{f}(t) \in W_j \Leftrightarrow \hat{f}(t - 2^j k) \in W_j \quad (2)$$

$$\forall (j) \in Z^2, \hat{f}(t) \in W_j \Leftrightarrow \hat{f}(\frac{t}{2}) \in W_{j-1} \quad (3)$$

$$\forall p, j \in [1, +\infty], p \neq j, W_p \cap W_j = 0 \quad (4)$$

$$V_0 + lim \sum_{j=1}^{+\infty} W_j = L^2(R) \quad (5)$$

Equation (2) describes the invariant property of the subspaces W_j. This property characterizes the system detail modeling on the subspace W_j. Equation (3) represents a system modeling on the subspace W_{j-1} that is carried out by dilating the wavelet bases on the subspace W_j by 2. The property described in equation (4) guarantees that the system detail modeling on the subspace W_j is independent from the system detail modeling on the subspace W_p. Equation (5) ensures that when the subspace W_j goes to positive infinity, i.e., the sum of the system rough modeling on the subspace V_0 and the system detail modeling on subspaces $W_j, j = 1, 2, 3, \ldots, +\infty$ converges to the exact function [1,2,16].

From the above analysis, the system modeling can be carried out on the set of subspaces $\left\{V_0 + \sum_{j=1}^{+\infty} W_j\right\}_{j\in Z}$.

IV. Training with Dyadic Wavelet Network

A fixed Dyadic Wavelet Network system model is in structure similar to a wavelet multiresolution model. The network structure is illustrated in Figure 1. It forms a wavelet multiresolution decomposition similar to the conventional Multi-Resolution Analysis (MRA) which involves not only a wavelet but also the associated scaling function.

The proposed model is developed by using orthonormal bases functions yielding the potential advantage of minimum usage of wavelet activation units. The number of activation units at each resolution can be determined as by shifting the wavelet (scaling) function in parallel with a step of the full support length. The DWN construction is detailed as follows.

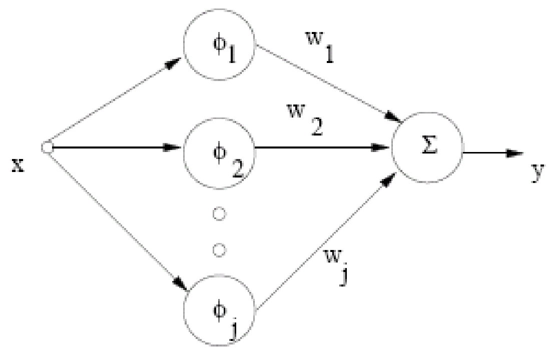

Fig. 1. Dyadic Wavelet Network

The orthonormal scaling functions $\phi_{i,m}(x)$, which are generated from the scaling function $\phi(x)$ by translation and dilation constitute orthonormal bases for the subspace S_i, as described in Equation (6)

$$\phi_{i,m}(x) = 2^{\frac{i}{2}} \phi\left(\frac{x - 2^i.m}{2^i}\right) \quad (6)$$

The system rough approximation $f(.)|_{V_0}$ is modeled through the orthonormal bases $\phi_{i,m}(x)$ on the subspace V_0 as described in equation

$$f(.)|_{V_0} = f(.)|_{S_i} = \Sigma_m c_{i,m}\phi_{i,m}(.) \quad (7)$$

where $c_{i,m}$ are coefficients of the system rough approximation and '*m*' are translation parameters of the orthonormal bases on the subspace V_0. The orthonormal wavelet functions Ψ_{j+i-1,n_j}, which are generated from the wavelet function $\Psi(x)$ by translation and dilation, represent the orthonormal bases for the subspace W_j [1, 16, 18]. The orthonormal property of the wavelet function and the scaling function ensures that the inner dot product among any two different WNN's is always zero. The well trained WNN parameters at low resolution can be directly used in higher resolutions without any change. The learning algorithm of the DWN is given in the following section. The construction of DWN for multiresolution representation corresponds to wavelet decomposition and as a consequence the proposed representation of a system presents an advantage of resolution levels that can be separately trained. Consequently we can also show that the resolution levels are independent.

A. Learning Algorithm

The algorithm consists of two phases, the initialization phase and training phase.

Network Initialization

- Form vector θ of network parameters $W_{i,j}, b_j, a_j$

Assign weights $W_{ij} = 0$ to links between hidden layer units and output layer.

- Choose a resolution level L
- Calculate the dyadic position b_j of wavelon within the given resolution.
- Set $a_j = 2^L$
- Finally initialize a wavelon with b_j, a_j.

Training

- For each input output pattern (x^l, D^l) i.e., $(x, f(x))$
- compute the output of the DWN by $O_i^l = \sum_{j=1}^{j_{max}} w_{i,j}\phi_{a_j,b_j}^l$

where ϕ_{a_j,b_j}^l as given in equation (A)

- Compute the output with parameter vector θ given by $O_{\theta_i}^l = \sum_{i=1}^{j_{max}} w_{i,j}\sqrt{a_{i,j}}\phi(a_{i,j}x - b_{i,j})$
- Compute error using $e^l = |D^l - O^l|$
- if $e^l > e_{min}$
- Weights W_{ij} in vector θ are modified in the opposite direction of the gradient $C(\theta, x_k, f(x_k))$
- Compute the Gradient using $\frac{\partial c}{\partial W_j} = e_k\sqrt{a_j}\phi(Z_{k,j})$

 Where $e_k = D_k^l - O_k^l$
- Update weights by $\frac{\partial c}{\partial W_j}\eta$
- Repeat till error for all the patterns fall below the minimum prescribed value.
- Else save weights, network parameters and exit.

V. Experimentation Results And Discussion

In this section, to investigate the feasibility and efficiency of the proposed algorithm for Dyadic Wavelet Network, two benchmark datasets for non-linear one-dimensional function approximations are considered. To assess the approximation results, the ***figure of merit*** δ **[*8*]** is applied.

$$\delta = \sqrt{\frac{\Sigma_{i=1}^P[O_i^l - D_i^l]^2}{\Sigma_{i=1}^P(D_i^l - \overline{y})^2}} \quad (8)$$

Where $\overline{y} = \frac{1}{P}\Sigma_{i=1}^P D_i^l \quad (9)$

A. Experiment (1) Benchmark dataset of function 1

The benchmark data function chosen is a piecewise function defined as follows:

$$f(x) = \begin{cases} -2.186x - 12.864 & \text{if -10} \leq \text{x} \leq \text{-2} \\ 4.246x & \text{if -2} \leq x < 0 \\ 10exp^{(-0.05x-0.5)} & \\ sin[(0.03x + 0.7)x] & \text{if } 0 \leq x \leq 10 \end{cases}$$

over the domain D = [-10,10].

We sampled 200 points, distributed uniformly over [-10, 10] as training dataset to demonstrate the Dyadic wavelet network for approximation of static functions in terms of performance evaluation and efficiency of function learning.

To implement the wavelet decomposition network, scaling function from quadratic B-splines as given by eqn (A) is used

$$\phi(x) = \left\{ \begin{array}{ll} 1/2(x+1)^2 & \text{for } -1 \leq x < 0 \\ \frac{3}{4} - (x - 1/2)^2 & \text{for } 0 \leq x < 1 \\ \frac{1}{2}(x-2)^2 & \text{for } 1 \leq x < -2 \\ 0 & \text{else} \end{array} \right\} Eqn(A)$$

During network learning, several experimentations have been conducted by varying the network simulation parameters, i.e. learning rate and iterations at different resolution levels. The Performance index δ of the network output function approximates to 0.0342 with less hidden units. Figures 2 to 5 are the results of the desired output and the proposed model output through the learning process of 1000 training epochs for approximation of the piecewise defined function for different Resolutions (L). The generalization of output function at different levels has greater impact on the accuracy level of the output function. The optimal network output is obtained with good generalization for 23 wavelons and is as shown in Figure 5 for Resolution L = 0. The network parameter values obtained for the optimal wavelon is given in Table II. During experimentations learning rate η of 0.1 has been chosen and for analysis the figure of merit δ for ach approximation is computed on a uniformly sampled test set of 200 points [8, 21].

Table III depicts the network performance comparative results of DWNN and other reported WNN models [8, 16]. Figures 6 to 9 and from Figures 10 to 13 shows the network output for the optimal wavelon '23' by varying number of iterations and learning rate respectively. Exhaustive experimentations demonstrated that the proposed DWN method resulted in an enhanced performance in terms of reducing the number of iterations and hidden nodes when compared with conventional neural networks and WNN [8]. It is observed that the change in error is negligible on increasing the number of iterations greater than 1000, has not resulted in any considerable improvement. Even varying learning rate parameter has least impact on the generalization of the network as indicated in the respective plots. This clearly reveals that the proposed learning algorithm for DWN is considerably robust and effective for function learning applications.

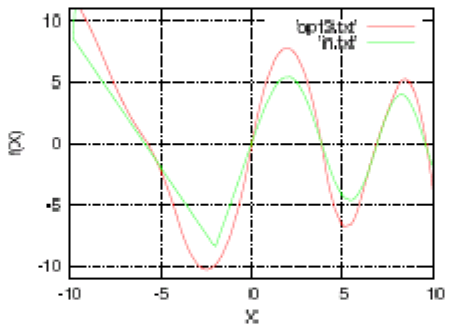

Fig. 2. Network output with 13 units and L = -1

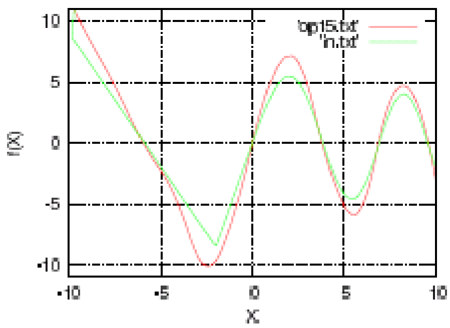

Fig. 3 . Network output with 15 units and L = -0.75

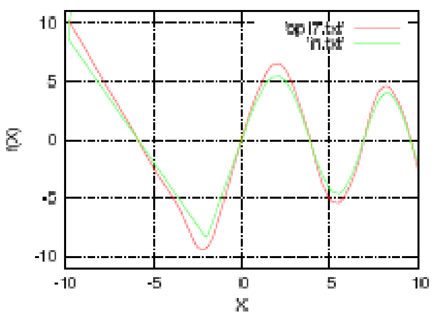

Fig. 4. Network output with 17 units and L = -0.5

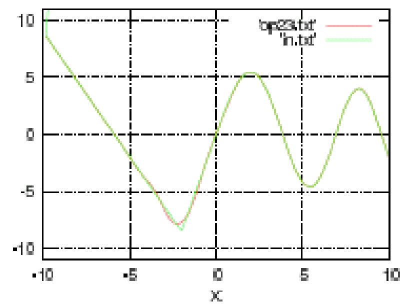

Fig. 5. Network output with 23 units and L = 0

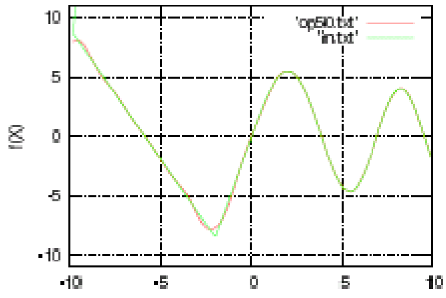

Fig. 6. Network output with 23 units and Learning Iterations = 50

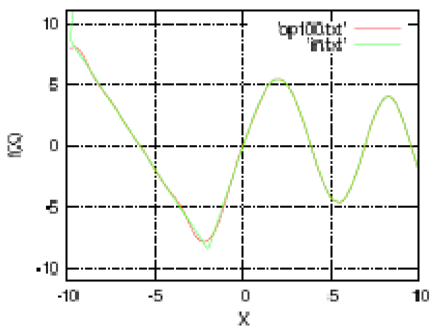

Fig. 7. Network output with 23 units and Learning Iterations = 100

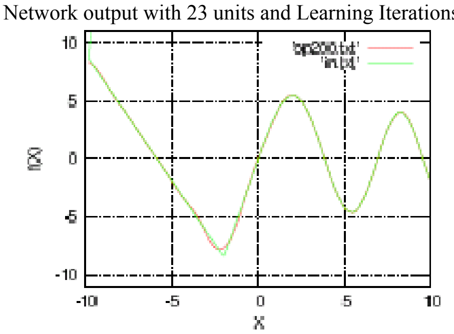

Fig. 8. Network output with 23 units and Learning Iterations = 200

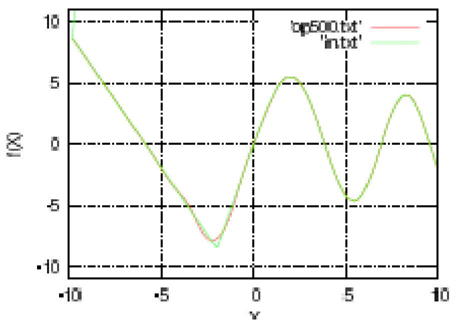

Fig. 9. Network output with 23 units and Learning Iterations = 500

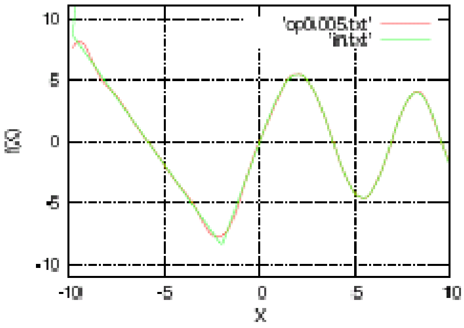

Fig. 10. Network output with 23 units and Learning rate = 0.005

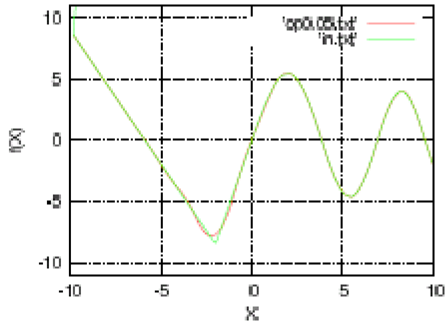

Fig. 11. Network output with 23 units and Learning rate = 0.05

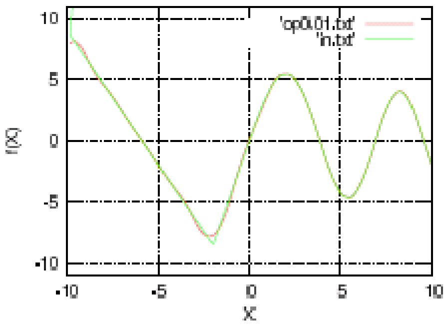

Fig. 12. Network output with 23 units and Learning rate = 0.01

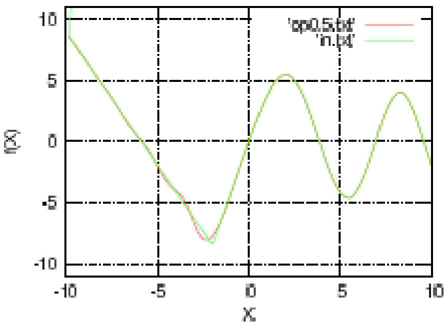

Fig. 13. Network output with 23 units and Learning rate = 0.5

TABLE II WEIGHT MATRIX FOR 23 WAVELONS WITH L =0

Weights	Translation	Dilation
10.054390	-11.000000	1.000000
7.916141	-10.000000	1.000000
5.702116	-9.000000	1.000000
3.558355	-8.000000	1.000000
1.302456	-7.000000	1.000000
-0.760919	-6.000000	1.000000
-3.195432	-5.000000	1.000000
-4.852712	-4.000000	1.000000
-8.250127	-3.000000	1.000000
-7.148192	-2.000000	1.000000
-1.856585	-1.000000	1.000000
1.978178	0.000000	1.000000
5.546448	1.000000	1.000000
5.444451	2.000000	1.000000
1.797312	3.000000	1.000000
-3.162258	4.000000	1.000000
-5.255840	5.000000	1.000000
-2.267699	6.000000	1.000000
2.988380	7.000000	1.000000
4.607416	8.000000	1.000000
0.343351	9.000000	1.000000
-4.507716	10.000000	1.000000

TABLE III APPROXIMATION RESULTS OF FUNCTION 1

Network	Units	Parameters	Iterations	Performance index (δ)
Proposed DWN Wavelet	23	24	1000	0.0342
Decomposition[8]	31	32	10000	0.06350
WNN[21]	20	42	1000	0.06775
Proposed G-RBF	20	42	1000	0.1142
NN-BP [19]	11	33	1000	1.318

B. *Experiment (2) Benchmark dataset of function (2)*

To demonstrate the validity of the proposed Dyadic Wavelet Network for approximating a single variable function [13], training points are obtained from the function defined by

$$f(x) = sin\frac{(2\pi x)}{e^{x}}, x \in (0,10) \tag{10}$$

TABLE IV STATISTICAL RESULTS WITH FIXED LEARNING

Number of Wavelons	Resolution(L)	Mean	Variance	Standard Deviation
5	-2	-1.599513	0.000000	0.000000
8	-1	0.011345	0.009004	0.094888
13	0	0.007602	0.015944	0.126271
25	1	0.005899	0.010227	0.101129
36	1.5	0.004870	0.007574	0.087028
49	2.0	0.004108	0.005458	0.073880
58	2.25	0.003767	0.004589	0.067740
69	2.5	0.003457	0.003853	0.062072

The approximation results for function (2) with increase in number of wavelons and Resolution (L) with fixed learning rate are shown in Table IV. In general the observations from the plots shown in Figures 14 to 19 indicate that with choice of decomposition level for network model, the low frequency values are approximated better than the high frequency values. Figure 14 and 15 shows the learning characteristics with less number of wavelons and different resolution values indicate poor generalization. With resolution L= 0 and 13 wavelons in the hidden layer the high frequency values are generalized better which is observable in Figure 16. Table IV provides the statistical results with choice of wavelons from 5 to 69. Figures 17 to 19 show that the network output for 25, 36 and 69 wavelons results in better generalization for low frequency values with decrease in variance. The generalization response obtained attributes to the characteristics of orthonormal scaling function chosen, which has better smoothness and convergence properties.

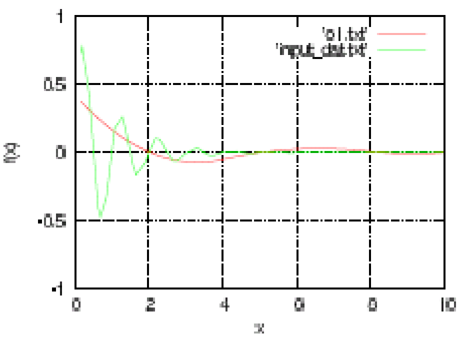

Fig. 14. Network output with 5 units and L= -2

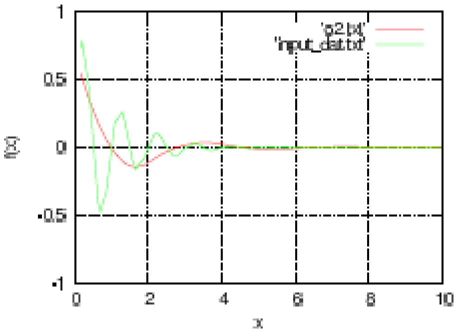

Fig. 15 . Network output with 8 units and L= -1

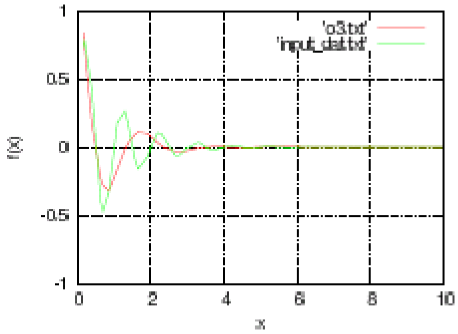

Fig. 16. Network output with 13 units and L=0

TABLE V STATISTICAL RESULTS WITH FIXED 400 SAMPLES, 49 WAVELONS WITH DIFFERENT LEARNING RATES

Learning rate	0.05	0.01	0.005
Mean	0.409220	0.409317	0.409419
Variance	0.246531	0.246365	0.246140
Standard Deviation	0.496519	0.496352	0.496125

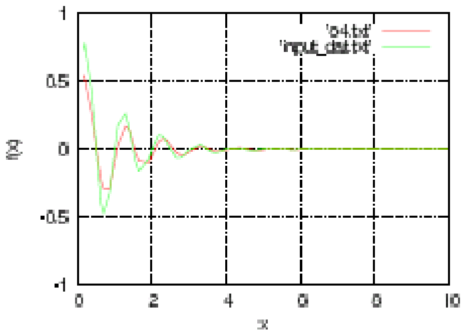

Fig. 17. Network output with 25 units and L=1

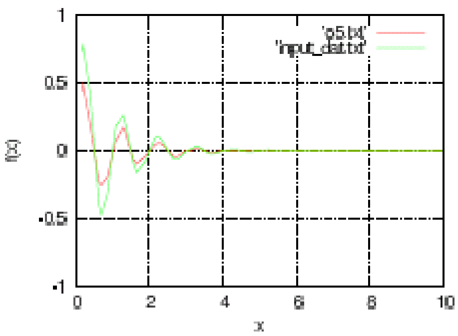

Fig. 18. Network output with 36 units and L=1.5

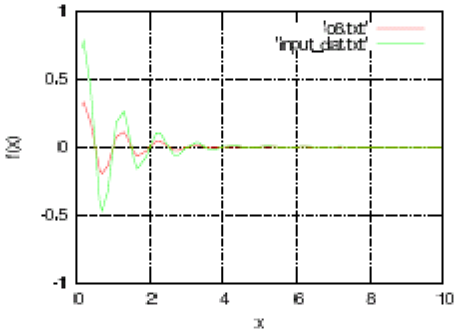

Fig. 19. Network output with 69 units and L=2.5

For function (2) experimentations are conducted considering 400 samples. Table V shows the various statistical measures i.e., mean, variance and standard deviation for 49 wavelons with constant Resolution L=2 and varying learning rates. Observations indicate that there is a subtle difference in the statistical measures obtained by varying the learning rate. The network responses are shown in Figures 20 to 22. Figure 23 indicates poor generalization performance of network with increase in wavelons, and reduction in learning rate. Similarly the network output for fixed Resolution 0 and 13 wavelons with different learning rates are shown in Figures 24 to 27. The observation of plots clearly indicates the effect of varying learning rate on the generalization characteristics and also the choice of optimal learning rate for training the network. The experimentations indicate that smaller values of learning rate results in poor generalization.

Table VI shows the Weight matrix i.e., translation and dilation parameters for the network model with 13 wavelons and resolution L=0. Since, it is a dyadic network the scaling function parameters remain constant with changes only in the weights.

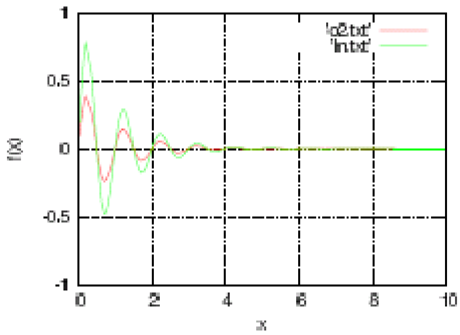

Fig. 20. Network output with 49 units, L=2 and LR =0.05

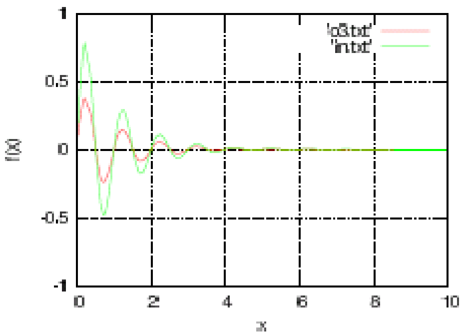

Fig. 21. Network output with 49 units, L=2 and LR =0.01

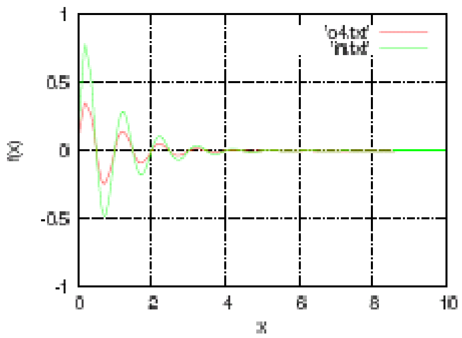

Fig. 22. Network output with 49 units, L=2 and LR =0.005

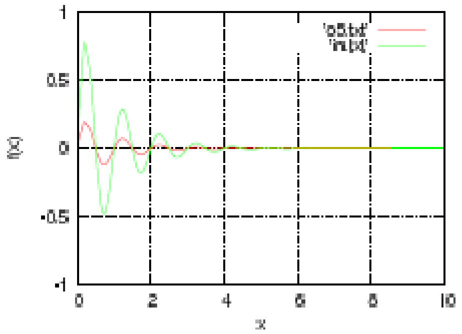

Fig. 23. Network output with 49 units, L=2 and LR =0.0005

TABLE VI WEIGHT MATRIX FOR 13 WAVELONS WITH L=0

Weights	Translation	Dilation
3.282858	-1.000000	1.000000
-0.627016	0.000000	1.000000
0.208237	1.000000	1.000000
-0.062971	2.000000	1.000000
0.010895	3.000000	1.000000
-0.006189	4.000000	1.000000
0.006977	5.000000	1.000000
-0.001643	6.000000	1.000000
0.000369	7.000000	1.000000
-0.000168	8.000000	1.000000
0.000147	9.000000	1.000000
-0.000328	10.000000	1.000000
0.000000	11.000000	1.000000

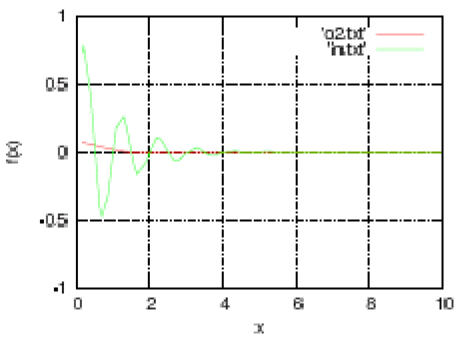

Fig. 24. Network output with 13 units, L=0 and LR =0.0005

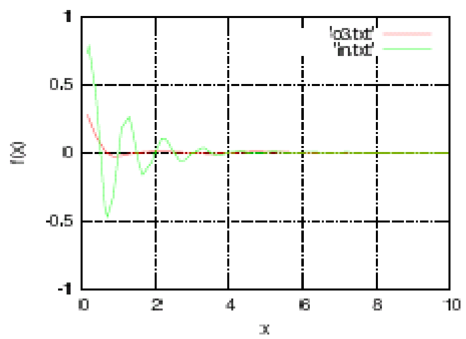

Fig. 25. Network output with 13 units, L=0 and LR =0.005

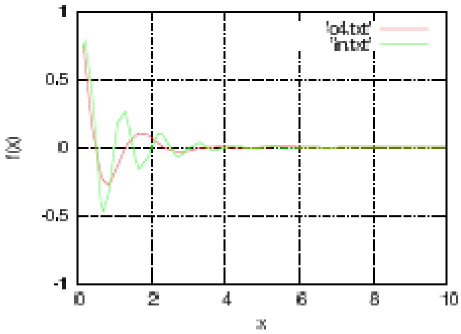

Fig. 26. Network output with 13 units, L=0 and LR =0.04

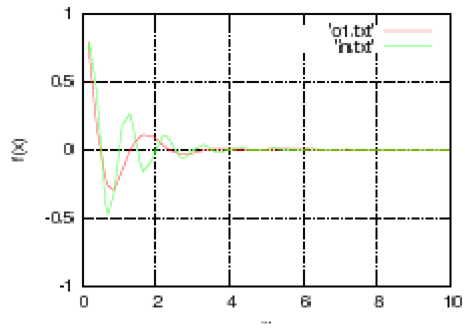

Fig. 27. Network output with 13 units, L=0 and LR =0.05

VI. CONCLUSIONS

Wavelet Neural Networks have been successfully applied to the problem of function representations. Nevertheless, the analysis and design of WNN still remains a difficult task, since it requires a clear analysis of simulation parameters that are crucial to obtain a good approximation. The success of the proposed algorithm resides on the choice of generic activation function that keeps a balance between the outputs of the target function and the coordinates of the input vector space. In this paper the analysis of DWN is done considering the various statistical measures tuning with the simulation parameters (learning rate and resolution) so as to ensure the convergence capabilities. The study results also indicate that the orthonormal scaling functions considered resulted in smoothness and regularity on the convergence of network model.

In our work we have experimented considering orthonormal basis function such as Battle-lemarie function, as the orthonormal wavelet-based networks can provide a unique and efficient representation for the given function, this also means that the approximation accuracy of this kind of DWN highly relies on the selected wavelet bases. We can also choose wavelet frames and Daubechies wavelets to construct the proposed DWN. Further study on extending of the proposed algorithm on DWN under additive noise is an interesting area under investigation in terms of generalization abilities.

In summary, the presented DWN not only reserves the multiresolution capability of WNN, but also has the advantages of high approximation accuracy and good generalization performance.

REFERENCES

[1] S G Mallat, "A Theory for Multi Resolution Signal Decomposition: the Wavelet Representation", IEEE Transactions on Pattern Analysis and Machine Intelligence, vol.11, pp. 674-693, July 1989.

[2] I.Daubechies. "The Wavelet transforms time-frequency localization and signal analysis." IEEE Trans. Information Theory, Vol 36: 961 -1005, 1990.
[3] I Daubehies, "Ten Lectures on Wavelets", Society for Industrial and Applied Mathematics, Philadelphia, 1992.
[4] Chui C.K. Wavelets, "A tutorial in Theory and Applications." New York: Academic 1992.
[5] Cybenko G "Approximation by Superposition of a sigmoidal function", IEEE Transactions Neural networks, 1989: 2(1): 303-314.
[6] T. Poggio and F. Girosi, "Networks for approximation and learning", .A. N. Expert, Proceedings of IEEE, vol.78, pp.1481-1497, Sept. 1990.
[7] Delyon B and Juditsky A. Benveniste A "Accuaracy analysis for wavelet approximations ", IEEE Trans. Neural Netwroks, 1995; 6(2): 332-48.
[8] Q. Zhang and A. Benveniste, "Wavelet Networks", IEEE Transactions Neural Networks, Vol.3. November 1992.
[9] J. Zhang, G.G. Walter and W.N. Wayne Lee, "Wavelet Neural Networks for Function Learning", IEEE Transactions on Signal Processing, Vol.43, No. 6, pp 1485-1490, June 1995.
[10] Y.C. Pati and P.S. Krishna Prasad, "Analysis and synthesis of feed forward network using discrete affine wavelet transformations," IEEE Transactions Neural Networks , vol.4, Jan. 1993.
[11] S. Sitharama Iyengar, E.C. Cho, Vir V.Phoha, "Foundations of Wavelet Networks and Applications", A CRC Press Company, Newyork Washington, D.C.
[12] Yacine Oussar, "Initialization by selection for wavelet Network Training" Neurocomputing, pp 131-143, 2000.
[13] Alberto,I Rogas,Gonzalez and others "Output value-based initialization for radial basis function neural networks", Neural Processing Letters, Vol 25, pp 209-225, 2007.
[14] Q. H. Zhang "Using Wavelet Network in Nonparametric Estimation", IEEE Trans, Neural Networks, Vol 8, no.2, pp 227-236 , March 1997.
[15] Daniel W.C.Ho, Ping-An Zhang, Jinhua Xu, "Fuzzy Wavelet Networks for Function Learning" IEEE Transactions on Fuzzy Systems, pp 200-211, Vol 9, No1, Feb 2001.
[16] Song, Y. Ponci F, "An incremental modeling approach with a hierarchical structure and progressive accuracy using dyadic wavelet-based neural networks" Proceedings of the 13th International Conference on Intelligent Systems Application to Power Systems, pp 194-200, Nov. 2005.
[17] Kai-Chiu Kan,woK-wo Wong "Self Constructing algorithm for synthesis of wavelet networks" Electronics Letters pp-1953-1955,Vol 34 No.20, Oct 1998.
[18] N Chihab, A Zergainoh, P Duhamel, J P Astruc " Orthonormal Non-Uniform B-Spline scaling and wavelet bases on non-equally spaced knot sequence for multiresolution signal approximations .
[19] Yongyong He, Fulei Chu, Binglin Zhong "A Hierarchical Evolutionary Algorithm for Constructing and Training Wavelet Networks" Neural Computing and Applications, pp-357-366, 2002.
[20] Pushpalatha M.P. and Nalini N "Generalization Analysis Of Wavelet Frame Based Neural Network For Function Representations Using Compactly Supported Gaussian Wavelets", International Joint Journal Conference in Computer Electronics and Electrical, CEE 2009 International Journal of Recent Trends in Engineering [ISSN 1797-9617] by the Academy Publishers, Finland.
[21] Pushpalatha M.P. and Nalini N " A Compact network with improved generalization using Wavelet basis function network for Static Function Representation", Proceedings of the 2009 International Joint Conference on Neural Networks, June IJCNN'09, IEEE Press.

Pushpalatha M P received her BE degree in Computer Science and Engineering and MTech from University of Mysore , India in 1989 and 1997. Presently she is working as faculty in the department of Computer Science and Engineering, SJ College of Engineering, Mysore, India and a research scholar in Visvesvaraya Technological University, Belgaum, India. Her research interests are in soft computing.

Nalini N received the BE degree in Computer Science and Engineering from Kuvempu University, Karanataka, India in 1996, MS in Software Engineering from BITS, Pilani, India in 1999, PhD from Visvevaraya Technological University, Belgaum, India in 2007. Presently she is working as Professor and Head of the department of Computer Science and Engineering, Nitte Meenakshi Institute of Technology, Bangalore, India. Her research interests are in soft computing and security aspects in sensor networks. She has authored /co-authored more than 30 papers on these topics. She is an Associate Editor of Journal of Information Technology www. maxwellsci.com.