

# Wavelet Multi-Layer Perceptron Neural Network for Time-Series Prediction

Kok Keong Teo, Lipo Wang\* and Zhiping Lin

School of Electrical and Electronic Engineering  
Nanyang Technological University  
Block S2, Nanyang Avenue  
Singapore 639798

<http://www.ntu.edu.sg/home/elpwang>  
{p7309881e, elpwang, ezplin}@ntu.edu.sg

\*To whom all correspondence should be addressed

**Abstract.** In this paper, we investigate the effectiveness of wavelet Multi-Layer Perceptrons (MLP) neural network for temporal sequence prediction. It is essentially a neural network with input signal decomposed to various resolutions using wavelet transform. Wavelet transform can expose the time-frequency information that is normally hidden. We show that wavelet MLP network provides prediction performance comparable to the conventional MLP. After the less important inputs are eliminated, the wavelet MLP shows more consistent performance for different weight initialization in comparison to the conventional MLP.

## 1. Introduction

In many instances, the desire to predict the future is the driving force behind the search for laws to explain certain phenomena. Example range from forecasting weather and Newton's laws of motion.

The oldest and most studied method, a linear autoregression (AR) is to fit the data using the following [1]:

$$y(k) = \sum_{i=1}^T a(i)y(k-i) + e(k) \quad (1)$$
$$= \hat{y}(k) + e(k) .$$

where

$y(k)$  actual value of the time series

$a(i)$  weightage

$e(k)$  prediction error

$\hat{y}(k)$  predicted value of  $y(k)$

This AR model forms  $y(k)$  as a weighted sum of past values of the sequence. This model can provide good performance only when the system under investigation is linear or nearly linear. However the performance may be very poor for cases in which system dynamics is highly nonlinear.

Neural network has demonstrated great potential for time-series prediction where system dynamics is nonlinear. Lapedes and Farber [2] first proposed using MLP for nonlinear signal prediction. It led to an explosive increase in research activities in examining the approximation capabilities of MLP [3]-[6].

Neural networks are developed to emulate the human brain that is powerful, flexible and efficient. However, conventional networks only process the signal on its finest resolution. It is however not the case for human brain. For example, the retinal image is likely to be processed in separate frequency channels [8].

The introduction of wavelet decomposition [7]-[11] provides a new tool for approximation. Inspired by both the MLP and wavelet decomposition, Zhang and Benveniste [12] invented a new type of network, call a *wavelet network*. This has caused rapid development of a new bred of neural network model integrated with wavelets. Most researchers used wavelets as radial basis functions that allow hierarchical, multi-resolution learning of input-output maps from experimental data [13]-[16]. Liang and Page [17] proposed a new learning concept and paradigm for neural network, called multiresolution learning based on multiresolution analysis in wavelet theory.

In this paper, we use wavelets to break the signal down into its multiresolution components before feeding them into a MLP. We show that the wavelet MLP neural network is capable of utilizing the time-frequency information to improve its consistency in performance.

## 2. WAVELET

Wavelet theory provides a unified framework for a number of techniques that had been developed independently for various signals processing application, e.g., multiresolution signal processing used in computer vision; subband coding, developed for speech and image compression; and wavelet series expansions, developed in applied mathematics. In this section, we will concentrate on the multiresolution approximation that is utilized in the presented model.

### 2.1 Multiresolution [10][11]

Wavelet  $\psi$  can be constructed such that the dilated and translated family

$$\left\{ \psi_{j,i}(t) = \sqrt{2^j} \psi(2^j(t-i)) \right\}_{(j,i) \in \mathbb{Z}^2} \quad (2)$$

where  $\psi$  (mother wavelet) is an orthonormal basis of  $L^2(\mathbf{R})$ , where  $L^2(\mathbf{R})$  denote the vector space of square-integrable, one-dimensional function  $f(x)$ , and let  $\mathbf{V}_j$  denote a close subspace in  $L^2(\mathbf{R})$ . Orthogonal wavelets dilated by  $2^j$  carry signal variations at the resolution  $2^j$ . Thus wavelet can be used to compute the approximation of signal at various resolutions with orthogonal projections on different spaces  $\{\mathbf{V}_j\}_{j \in \mathbb{Z}}$ . Each subspace contains the approximation of all function  $f(x)$  at resolution  $2^j$ . The approximation of signal at resolution  $2^{j+1}$  contains all information necessary to compute the signal at the lower resolution. Thus they are a set of nested vector subspace,

$$\subset V_j \subset V_{j+1} \subset V_{j+2} \subset \quad (3)$$

Therefore when computing the approximation of function  $f$  at resolution  $2^j$ , some information about  $f$  is lost. As the resolution increases to infinity, the approximate

signal converges to the original signal. When the resolution approaches zero, the signal vanishes. If  $P_{V_j}$  denotes the orthogonal projection operator from  $L^2(\mathbf{R})$  onto  $V_j$

$$\lim_{j \rightarrow -\infty} \|P_{V_j} f\| = 0 \quad (4)$$

On the other hand when the resolution  $2^j$  approaches  $+\infty$ , the signal approximation converges to the original signal:

$$\lim_{j \rightarrow +\infty} \|f - P_{V_j} f\| = 0 \quad (5)$$

(5) guarantee that the original signal can be reconstructed using decomposed signals at lower resolution.

## 2.2 Signal Decomposition

A tree algorithm can be used for computing wavelet transform by using the wavelet coefficients as filter coefficients. Assume that vector  $s^m$  represents the sampled signal  $f$  at finest resolution  $2^m$ . Lowpass filter  $L$  is employed to produce a coarser approximation at resolution  $2^{m-1}$ . Thus

$$s^{j-1} = L s^j \quad j=1,2,\dots,m \quad (6)$$

The detail signal  $d^j$  at resolution  $2^j$  is obtained by applying a highpass filter  $H$  to  $s^j$ . That is

$$d^{j-1} = H s^j \quad j=1,2,\dots,m \quad (7)$$

Thus the process can be repeated to produce signals at any desired resolution (Fig.1).

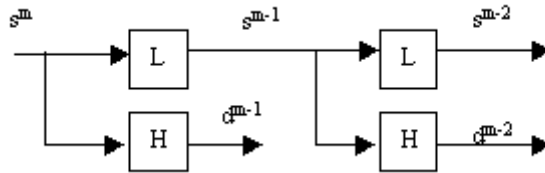


Fig. 1. Decomposition Process

The signal can be reconstructed using two synthesis filter  $L^*$  and  $H^*$  (the transposed matrices of  $L$  and  $H$ , respectively). Thus reconstruction is given by (Fig. 2).

$$s^j = L^* s^{j-1} + H^* d^{j-1} \quad (8)$$

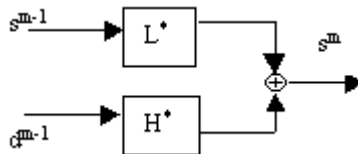


Fig. 2. Reconstruction

Therefore, any original signal can be represented as

$$f = s^m = s^0 + d^0 + d^1 + \dots + d^{m-1} + d^m \quad (9)$$

### 3. WAVELET MLP NEURAL NETWORK

Fig.3 shows the Wavelet MLP Neural Network used in this paper. The input signal is passed through a tapped delay line to create short-term memory that retains aspect of the input sequence relevant to making predictions. This is similar to time lagged MLP except that the delayed data is not sent directly into the network. Instead it is decomposed by wavelet transform to form the input of the MLP. Fig. 4 shows an example of two level decomposition of the tapped delay data  $x$ . Data  $x$  is decomposed to coarser (CA1) and detailed (CD1) approximation. The coarser (CA1) is further decomposed into it coarser (CA2) and detailed (CD2) approximations.

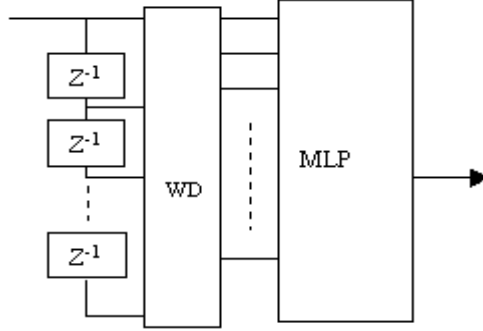


Fig. 3. Model of Network (WD=Wavelet Decomposition)

Furthermore, we are looking into possibility of discarding certain wavelet-decomposed data that is of little use in the mapping of input to output. The mapping is expected to be highly nonlinear and dependent on the characteristic of individual signal.

Let

$$s_i = \sum_j^n |w_{ij}| \quad (10)$$

represent the importance of input  $x_i$

where

$w_{ij}$  weightage of the of input  $i$  to neuron  $j$   
 $n$  number of hidden neurons

$$s_i' = \frac{s_i}{\max(s_i)} \quad (11)$$

serves as indicator of the relative importance of inupt  $x_i$ .

where

$s_i'$  normalized input strength  
 $\max(s_i)$  maximum of  $s_1, s_2, \dots, s_I$ ,  $I$  is the number of inputs

Input point having small  $s_i$  will be considered to be trivial and maybe discarded without affecting the prediction performance.

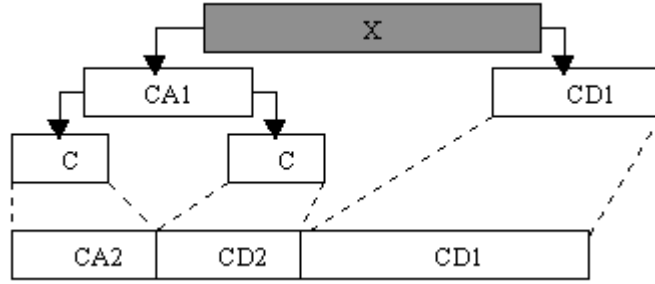


Fig. 4. The two level decomposition to form input to the neural network

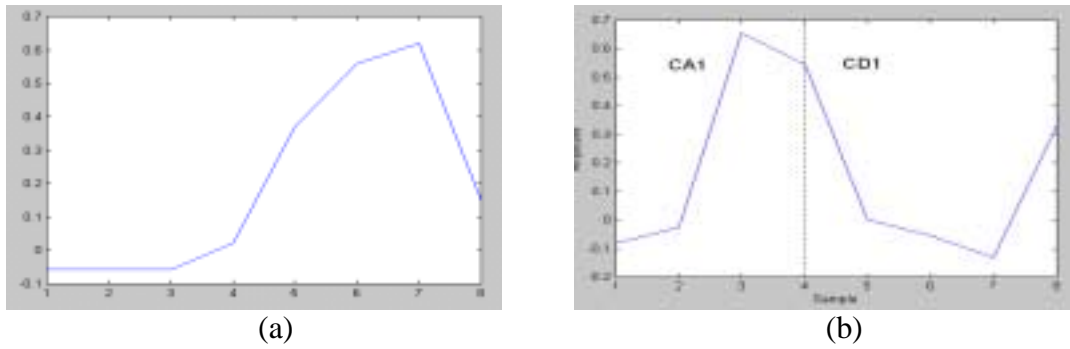


Fig. 5. (a) the first 8 data points of Mackey-Glass (b) decomposed by Daubechies1 wavelet

#### 4. Simulation

The Mackey-Glass time-series prediction is frequently used as a benchmark in time-series studies.

The Mackey-Glass time-delay differential equation is defined by

$$\frac{dx(t)}{dt} = \frac{0.2x(t-\pi)}{(1+x(t-\pi))^{10}} - 0.1x(t) \quad (12)$$

The MLP used in our simulations consists of an input layer, a hidden layer of two neurons and one output neuron, and is trained by the backpropagation algorithm using the Levenberg-Marquardt for fast optimization [18]. All neurons use the conventional sigmoid activation function; however, the output neuron employed a linear activation function as frequently used in forecasting applications.

In order to compare our result, the normalized mean squared error (NMSE) is used to assess forecasting performance. The NMSE is computed as

$$NMSE = \frac{1}{\sigma^2} \frac{1}{N} \sum_{t=1}^N [x(t) - \hat{x}(t)]^2 \quad (13)$$

where

$x(t)$  actual value of the time series

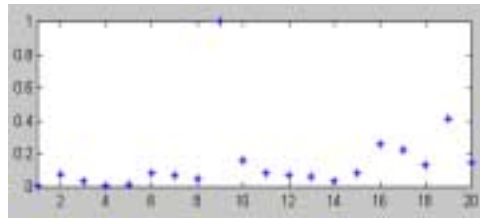
$\hat{x}(t)$  predicted value of  $x(t)$ ;

$\sigma^2$  variance of the time series over the predicting duration.

$N$  is the number of elements

The data is divided into three sections, the training, validation and testing. The training data is of length 220, follow by validation and testing data, each of length 30. Validation NMSE is evaluated every 20 epochs. When there is an increase in the validation NMSE, training stops. Test data is used to test the generalization performance of the network and has not been seen by the network during training or validation.

Early stopping by monitoring validation error often shows multiple minima as a function of training time and results are also sensitive to the weight initialization [6]. In order to have a fair comparison; simulation is carried out for each network with different random weight initialization over 100 trials. The 50 lowest NMSE is kept for calculations of mean and standard deviation, which are then used for comparisons.



**Fig. 6.** Distribution of relative importance of 20 inputs for the wavelet MLP network with decomposition level one in one of the simulations, which is similar to the results in other simulations

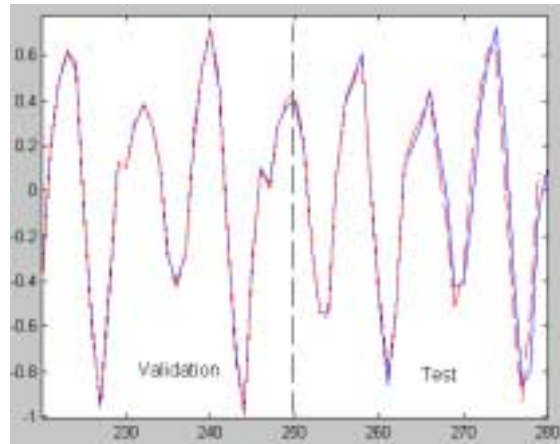
The simulations indicate that the input points 1,4 and 5 is consistently less important than other inputs (Fig.6). Simulations are re-run after these less important inputs are eliminated. This results a network of size 17:2:1 (17 inputs, 2 hidden neurons and 1 output neuron). We denote this wavelet MLP Neural Network (FNN) by (17:2:1) = (20:2:1) - [1,4,5]. Simulations are done on other network sizes to reduce the number of inputs when possible.

**Table 1.** Result of the three networks on different architecture of network

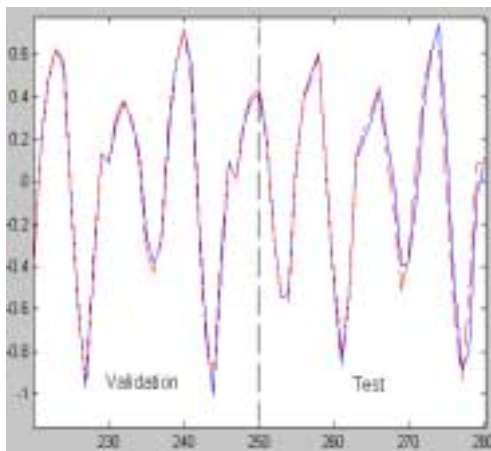
| Architecture | Type             | Normalized Mean Square Error (NMSE) |                        |         |
|--------------|------------------|-------------------------------------|------------------------|---------|
|              |                  | Mean                                | Standard Deviation     | Minimum |
| 20:2:1       | Conventional MLP | 0.063                               | 0.016                  | 0.0047  |
|              | Wavelet MLP      | 0.054                               | 0.0026                 | 0.036   |
| 17:2:1       | Conventional MLP | 0.043                               | 0.0093                 | 0.0069  |
|              | Wavelet MLP      | 0.045                               | $1.38 \times 10^{-6}$  | 0.045   |
|              | (20:2:1)-[1,4,5] | 0.027                               | $8.051 \times 10^{-6}$ | 0.027   |

|        |                  |       |         |        |
|--------|------------------|-------|---------|--------|
| 12:2:1 | Conventional MLP | 0.021 | 0.0107  | 0.0017 |
|        | Wavelet MLP      | 0.032 | 0.0030  | 0.013  |
| 11:2:1 | Conventional MLP | 0.015 | 0.0078  | 0.0017 |
|        | Wavelet MLP      | 0.032 | 0.0029  | 0.017  |
|        | (12:2:1)-[7]     | 0.021 | 0.00064 | 0.017  |

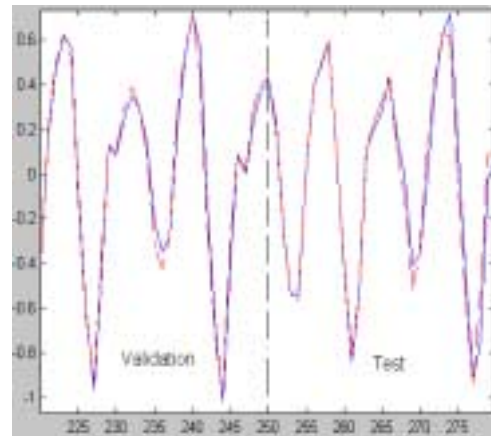
Table 1 shows that wavelet MLP network provides prediction performance comparable to the conventional MLP. After less important inputs are eliminated, the wavelet MLP shows more consistent performance for different weight initialization to the conventional MLP.



(a)



(b)



(c)

**Fig. 7.** Result for network size 17:2:1 on Mackey Glass time series prediction (a) MLP with test and validation NMSE=0.0028 and 0.0481, respectively. (b) Wavelet MLP with decomposition level one, verification and test NMSE=0.0068 and 0.0450, respectively. (c) Wavelet MLP (17:2:1) = (20:2:1) - [1,4,5], test and validation NMSE=0.0126 and 0.0273, respectively. Dotted line is the actual data, whereas continuous line is the predicted data.

## 5. CONCLUSION

In this paper, we used a wavelet MLP, consisting of a wavelet decomposition layer and a conventional MLP, for time series prediction. We analyzed the relative

importance among the input wavelets. After less important wavelets are eliminated, the modified wavelet MLP network provides a more consistent and stable network that is evident in its low mean and standard deviation for NMSE. This is in contrast to the conventional MLP network that has large performance swing and is sensitive to weight initialization.

However the wavelet MLP without input elimination did not show significance improvement over the conventional MLP. It is suspected that different signals have different time-frequency compositions. Thus the decomposition level, type of wavelet or decomposition type may vary significantly with signal. Therefore more work is required to equip the network with the ability to adapt to different signals without human intervention.

## REFERENCE

- [1] C. Chatfield, *The Analysis of Time Series, An Introduction*, Chapman & Hall 1989 Fourth Edition.
- [2] A. Lapedes and R. Farber, "Nonlinear signal processing using neural network: Prediction and system modeling," *Los Alamos Nat. Lab, Tech. Rep, LA-UR-872662*, 1987
- [3] Simon Haykin; *Neural Networks: A Comprehensive Foundation*, Macmillan College Publishing Company, Inc. 1994
- [4] Eric A. Wan, "Time Series Prediction by using a connectionist network with internal delay lines," in *Time Series Prediction: Forecasting the Future and Understanding the Past*, SFI Studies in the sciences of complexity, Proc Vol XV, pp:195-217, Addison-Wesley, 1993
- [5] Michael C. Mozer, "Neural Net Architecture for Temporal Sequence Processing," in *Time Series Prediction: Forecasting the Future and Understanding the Past*, SFI Studies in the sciences of complexity, Proc Vol XV, pp: 243-264, Addison-Wesley, 1993
- [6] Andreas S. Weigend, Bernardo A. Huberman and David E. Rumelhart, "Predicting Sunspot and Exchange Rates with Connectionist Networks," in *Nonlinear Modeling and Forecasting*, SFI studies in sciences of complexity, Proc. Vol. XII, pp: 395-432, Addison-Wesley, 1992
- [7] Mallat, S.G, "A Theory for Multiresolution Signal Decomposition: The Wavelet Representation," *IEEE Transactions on Pattern Recognition* Vol: 11 No. 7, pp. 674-693, , July. 1989
- [8] Mallat, S.G, "Multifrequency channel decompositions of images and wavelet models," *IEEE Transactions on Acoustics, Speech and Signal Processing* [see also *IEEE Transactions on Signal Processing*] Vol: 37 No. 12, pp. 2091-2110, , Dec. 1989
- [9] Vetterli, M.; Herley, C, "Wavelets and filter banks: theory and design," *IEEE Transactions on Signal Processing*, Vol: 40 No.9 , pp, 2207-2232 Sept. 1992
- [10] Mallat, S.G, *A Wavelet Tour of Signal Processing*, Academic Press, 1998
- [11] Gilbert Strang and Truong Nguyen, *Wavelet and Filter Banks*, Wellesly-Cambridge Press, 1996
- [12] Zhang, Q.; Benveniste, A. "Wavelet networks," *IEEE Transactions on , Neural Networks*, Vol: 3 6, pp. 889 - 898, Nov. 1992
- [13] Bakshi R Bakshi and George Stephanopoulos, "Wavelets as Basis for Localized Learning in a Multi-Resolution Hierarchy," *IJCNN., International Joint Conference Neural Networks on Vol 2.*, pp: 140 -145, 1992
- [14] Ciuca, I.; Ware, J.A., "Wavelet networks as an alternative to neural networks," 6th International Conference on Emerging Technologies and Factory Automation Proceedings 1997. ETFA '97, pp: 353 -358
- [15] Felipe Miguel Aparicio Acosta and Jean-Marc Vesin, "Nonlinear Prediction of Time Series using Radial Wavelet Networks," *Time-Frequency and Time-Scale Analysis*, Proceedings of the IEEE-SP International Symposium , 1992 , pp: 189 -192
- [16] Chun-Ta Chen, Azimi-Sadjadi, M.R and Chunhua Yuan, "Signal Representation Using Adaptive Wavelet-Net," International Conference on Neural Networks 1997, Vol: 4, pp: 2225 -2230
- [17] Yao Liang and Edward W. Page, "Multiresolution Learning Paradigm and Signal Prediction," *IEEE Transactions on Signal Processing*, Vol: 45, No. 11, 1997
- [18] Martin T. Hagan and Mohammad B. Menhaj, "Training Feedforward Networks with the Marquardt Algorithm," *IEEE Transactions on Neural Networks*, Vol:5, No. 6, 1994