ICIC Express Letters
Volume **2**, Number **4**, December **2008**
 ISSN 1881-803X
pp. **401–407**

BI-DIRECTIONAL FUNCTION LEARNING METHOD FOR TIME SERIES PREDICTION

Hiroki Tamura[1], Koichi Tanno[1], Hisasi Tanaka[1] and Zongmei Zhang[2]

[1]Faculty of Engineerng
University of Miyazaki
1-1Gakuen Kibanadai Nishi, Miyazaki, 889-2192 Japan
htamura@cc.miyazaki-u.ac.jp

[2]Alcatel-lucent Qingdao R&D Center
Qingdao city, Laoshan Area, Shandong Province, P. R. China
zongmeizhang@alcatel-lucent.com

Received July 2008; accepted September 2008

Abstract. *The local linear wavelet neural network is an improvement of wavelet network and commonly used learning algorithm is gradient descent method. In this paper, we attempt to predict sunspots, Mackey-Glass time series and Box-Jenkins data using a local linear wavelet neural network. Furthermore, we propose a technique using bi-directional function learning method. The simulation results show the effectiveness of the proposed method.*
Keywords: Local linear wavelet neural network, Local search, Bi-directional function, Sunspots data, Mackey-glass time series, Box-Jenkins data

1. **Introduction.** The wavelet theory which offers efficient algorithms for numerical analysis, signal processing and other applications, are usually limited to applications of small dimension wavelets because of the dimension problems. Artificial neural networks which are powerful tool for handling problems of large dimension suffer from the lack of efficient constructive methods. The wavelet neural networks (abbr. WNN), first mentioned by Zhang [1, 2], overcame the disadvantages of both wavelet theory applications and neural networks. The local linear wavelet neural network (abbr. LLWNN) [3, 4] is an improvement of the WNN, in which the connection weights between the hidden layer neurons and output neurons are replaced by a local linear model.

One of the most frequently used learning algorithm for WNN is the gradient descent method. Weight Perturbation [5] is a neural network training technique based on gradient descent. The gradient of the Mean Square Error (MSE) with respect to a weight is approximated by applying a small perturbation. Local Search Method (abbr. LS) is a heuristic algorithm like the gradient descent method. However, many models and learning algorithm for time series prediction have the over-learning problem.

In this paper, we attempt to predict sunspots, Mackey-Glass time series and Box-Jenkins data using LLWNN and LS. Furthermore, we propose a technique using bi-directional function learning method. The over-learning problem can be eased by using bi-directional function learning method. The simulation results show that the propsed method using bi-directional function learning method improves from original LLWNN.

2. **Local Linear Wavelet Neural Network.** According to wavelet transformation theory, wavelets are a family of functions generated from one function $\psi(\mathrm{x})$ (called the mother wavelet) by the operation of dilation and translation as follows:

$$\Psi = \left\{ \Psi_{\mathrm{i}} = |\mathbf{a}_i|^{-1/2} \psi \left(\frac{\mathbf{x} - \mathbf{b}_i}{\mathbf{a}_i} \right) : \mathbf{a}_i, \mathbf{b}_i, \mathbf{x} \in R^n, i \in Z \right\} \tag{1}$$

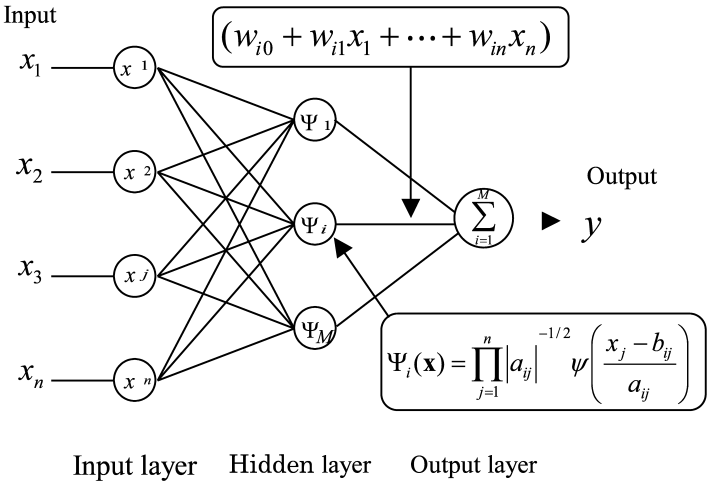

FIGURE 1. Local linear wavelet neural network

$$\mathbf{x} = (x_1, x_2, \cdots, x_n), \mathbf{a}_i = (a_{i1}, a_{i2}, \cdots, a_{in}), \mathbf{b}_i = (b_{i1}, b_{i2}, \cdots, b_{in})$$

where $\mathbf{x}$ is the input vector, $\mathbf{a}_i$ is a scale parameter and $\mathbf{b}_i$ is a translation parameter. The mother wavelet is orthogonal to all functions which are obtained by shifting the mother right or left by an integer amount. Furthermore, the mother wavelet is orthogonal to all functions which are obtained by dilating (stretching) the mother by a factor of 2^j (2 to the jth power) and shifting by multiples of 2^j units.

The WNN can be considered as one-hidden-layer neural network with wavelets as activation functions of its hidden layer neurons. The output of the WNN is given by

$$f(x) = \sum_{i=1}^{M} w_i \Psi_i(x) = \sum_{i=1}^{M} w_i \left|a_i\right|^{-1/2} \psi\left(\frac{x - b_i}{a_i}\right) \tag{2}$$

where Ψ_i is the wavelet activation function of ith neuron of the hidden layer. w_i is the weight connecting the ith neuron of the hidden layer to the output layer neuron. For the n-dimensional input space, the multivariate wavelet basis function can be calculated by the tensor product of n single wavelet basis functions as follows

$$\psi(x) = \prod_{i=1}^{n} \psi(x_i) \tag{3}$$

A feature of the WNN is the localized activation of the hidden layer neuron. Therefore, the connection weights associated with the neurons can be viewed as local parameters. This feature provides better learning efficiency and structure transparency. However, the drawback of the WNN is that for higher dimensional problems many hidden layer neurons are required.

In order to take advantage of the local capacity of the WNN while not having too many hidden neuron, Chen [4] proposed an alternative type of WNN called LLWNN. The architecture of the proposed LLWNN is shown in Figure 1. The output in the output layer is given by

$$\begin{aligned} y &= \sum_{i=1}^{M} (w_{i0} + w_{i1}x_1 + \cdots + w_{in}x_n)\Psi i(x) \\ &= \sum_{i=1}^{M} (w_{i0} + w_{i1}x_1 + \cdots + w_{in}x_n) \left|a_i\right|^{-1/2} \psi\left(\tfrac{x-b_i}{a_i}\right) \end{aligned} \tag{4}$$

A linear model

$$v_i = w_{i0} + w_{i1}x_1 + \cdots + w_{in}x_n \tag{5}$$

is introduced to take the place of the straightforward weight w_i. This LLWNN is proved to have as good performances as WNN and provides a solution for high-dimension problems.

3. Learning Algorithm.

3.1. **Local search method.** The basic principle of Weight Perturbation [6] is the introduction of noise into the weight parameters. If the addition of noise lowers the energy function, the perturbation to the weight parameters is accepted and this means that lower energy for that input in the future.

For the LLWNN, we can define a vector V whose elements include all parameters (weights connect the hidden layer neurons and the output layer neuron, scale parameters and translation parameters of the wavelet function) as:

$$V = [w_{10}, w_{11}, ...w_{ij}..., a_{11}, a_{12}, ...a_{ij}..., b_{11}, b_{12}, ...b_{ij}...]^T \tag{6}$$

The energy function to be used, can be written as:

$$E = E(V) = \sqrt{\frac{1}{N}\sum_{t=1}^{N}(y_t^{teacher} - y_t^{output})^2} \tag{7}$$

where y^{output} is the output of LLWNN. y^{output} is the teacher signal. N is training data size. The local search with weight perturbation as a neighbor search method adjusts the vector V iteratively to minimize the function $E(V)$. First, the search starts at an initial point V_0 and moves along $3n$ directions. $3n$ denotes the element number of vector V. For example, the h-th direction becomes:

$$e_h = (\underset{1}{0}, ..., 0, \underset{h}{1}, 0, ..., \underset{3n}{0})^T \tag{8}$$

Given initial iteration V_0, the perturbation size Δ_0, the sequence of iterations V_0, V_1, $\cdots$, V_{3n} in R^{3n} can be obtained by repeating the following process. At iteration k $(0 < k \leq 3n)$ whose initial value is 1, the search starts at $V_k \in R^{3n}$, and the objective at iteration k is to find a direction e_h (positive or negative) such that $E(V_k + \Delta_k e_h) < E(V_k)$ or $E(V_k - \Delta_k e_h) < E(V_k)$, $h \in (1, ..., 3n)$, where $\Delta_k > 0$ is the perturbation size. If such a direction is found, then this iteration is declared successful, and the next iteration V_{k+1} is either $V_{k+1} := V_k + \Delta_k e_h$ or $V_{k+1} := V_k - \Delta_k e_h$. However, if such a direction cannot be found, then this iteration is declared unsuccessful, and the next iteration $V_{k+1} := V_k$ which is the same as the point before. The next perturbation size parameter Δ_{k+1} is reduced to $d\Delta_k$, where $0 < d < 1$ is constant over all iterations. Therefore, if the gradient $\nabla E(V_k)$ is non-zero, and the step size Δ_k is small enough, then the iteration will finally find the minimum solution either locally or globally.

3.2. **Bi-directional function learning method.** In this section, we propose a technique which use bi-directional function. Neural networks usually learn by one directional data only. However, it has over-learning problem which inclines toward training data. Neural networks are learning the function near training data simultaneously with training data, and it knows experientially that over-learning problem can be reduced. In a prediction problem, opposite-directional data are chosen as a function near training data. It is thought by learning direction data and opposite-directional data simultaneously that the deviation to direction data can be reduced. However, if opposite-directional data are too large, it will also be expected that the deviation to opposite-directional data can be reduced.

We propose a technique of judging the generalization of LLWNN by adding a bi-directional function to the energy function E. The new function which is used for judging

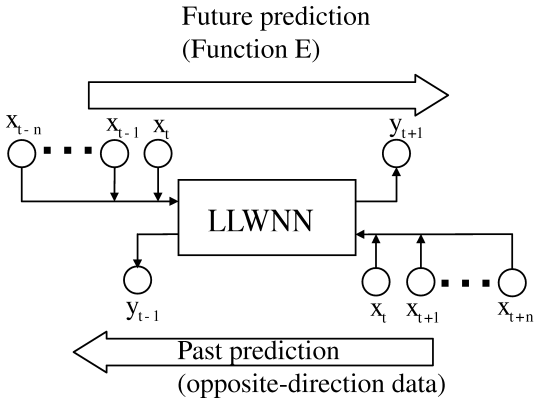

FIGURE 2. The technique using bi-directional function

the generalization as information criterion. Thus, the proposed new function G is as follows:

$$G = E + \lambda \cdot \sqrt{\frac{1}{L}\sum_{t=1}^{L}(y_{N-t}^{teacher} - y_{N-t}^{output})^2} \tag{9}$$

λ is a constant. L is the opposite-directional data size. It predicts using a LLWNN when the function G is minimum. The image figure of the technique using bi-directional function is shown in Figure 2. In this research, our purpose is to examine the validity of proposed technique in computer simulations. Moreover, the action of λ is investigated.

4. **Simulations.** The scale and translation parameters and weight parameters are randomly initialized at the beginning and are optimized by LS discussed in the above section. For all our experiments, the selected mother wavelet is as follows:

$$\psi(x) = -x\exp(-\frac{x^2}{2}) \tag{10}$$

4.1. **Application to sunspot data.** The sunspot series is an annual record of visible spots on the face of the sun from the year 1700. The data set is nonlinear, non-Gaussian, and has traditionally been used to measure the effectiveness of some nonlinear statistical models. The use of artificial neural networks has been recognized recently as a promising way of predicting sunspot series [7]. Our experiments use the sunspot data which are records of monthly sunspot activities from the Jan.1749 to Dec.1999. Figure 3 shows the monthly sunspot data (3000 dots) of this period. We use the average obtained by dividing input data (Figure 3: data from 1 to 2500) by 10 for training (N=250) and opposite-directional data size is 50 (L=50). We then applied this trained model to forecast the next 50 data (the average is obtained by dividing data from 2501 to 3000 by 10).

For comparison, we performed LS method (λ=0), the proposed method using bi-directional function ($\lambda > 0$), and the proposed method ($\lambda > 0$) with technique of input data adding random noise. We select a LLWNN with 10 inputs, one hidden layer with 7 hidden units and one output unit to predict sunspot numbers. We predict the $x(t+1)$ data using the input variables $x(t), x(t-1), x(t-2)...x(t-8)$ and $x(t-9)$, respectively. For each of the method, 20 training process were performed with a random set of initial parameters in order to remove the effects of the initial values of parameters on the final solutions. According to the experiments, 20 test errors were found for each method. We compared the performances of the proposed method by calculating the average test error (RMSE) of the 20 experiments. The results show that the proposed method was greatly enhanced in LLWNN. Figure 4 shows the result of the changing value of λ (0.1-0.5,1.0) and

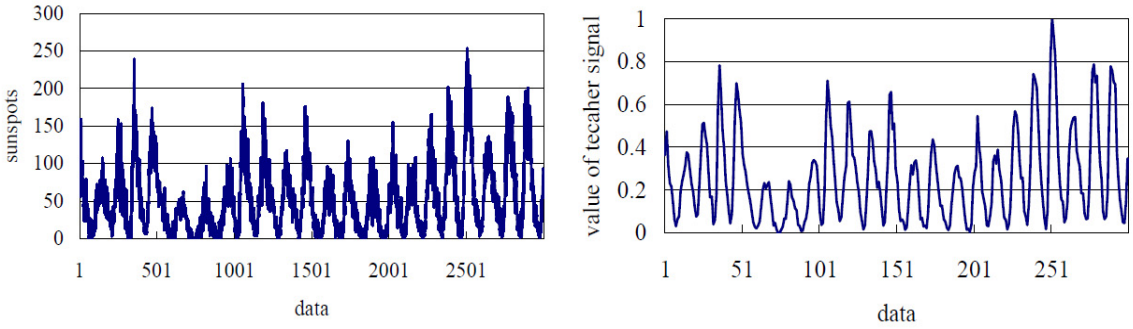

FIGURE 3. The monthly sunspot data (3000 dots) from year 1749 to year 1999 and teacher signal for LLWNN

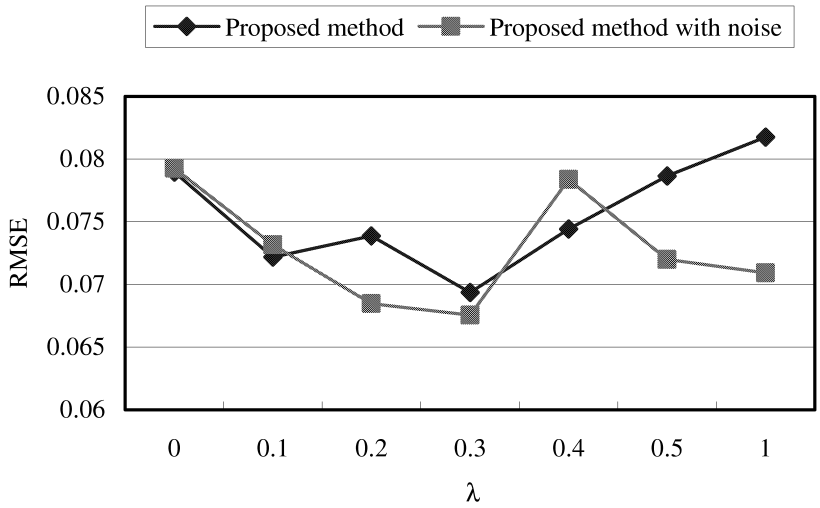

FIGURE 4. Simulation results (sunspots data)

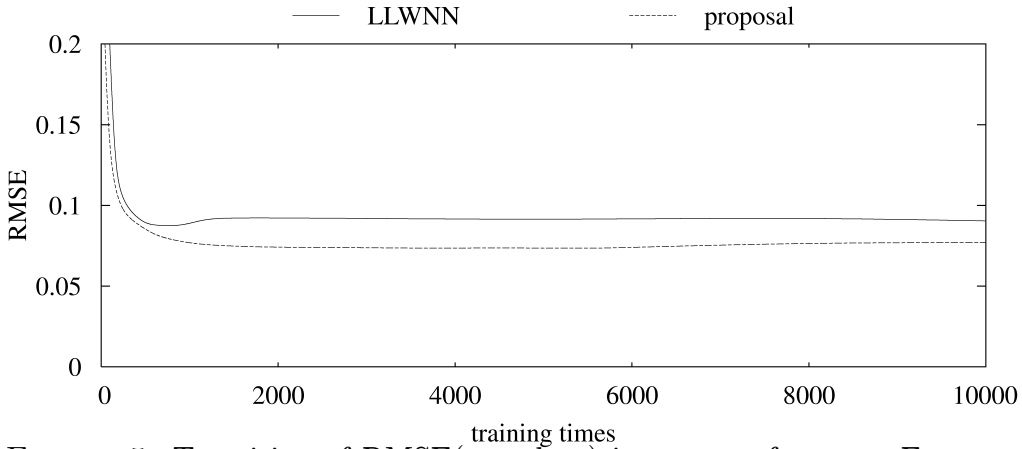

FIGURE 5. Transition of RMSE(test data) in sunspot forecast. For comparison, we performed LLWNN and proposed method (λ=0.3).

it is clear that the average error of λ value 0.3 is the best result in this experiment. Proposed method (λ=0.3) had statistical significance (t-test:$p = 0.05$). And Figure 5 shows the transition of RMSE(test data) of LLWNN(λ=0) and proposed method (λ=0.3). From Figure 5, proposed method does not have over-learning from LLWNN(λ=0).

4.2. **Application to Mackey-Glass time series.** The chaotic Mackey-Glass differential delay equation is recognized as a benchmark problem that has been used and reported

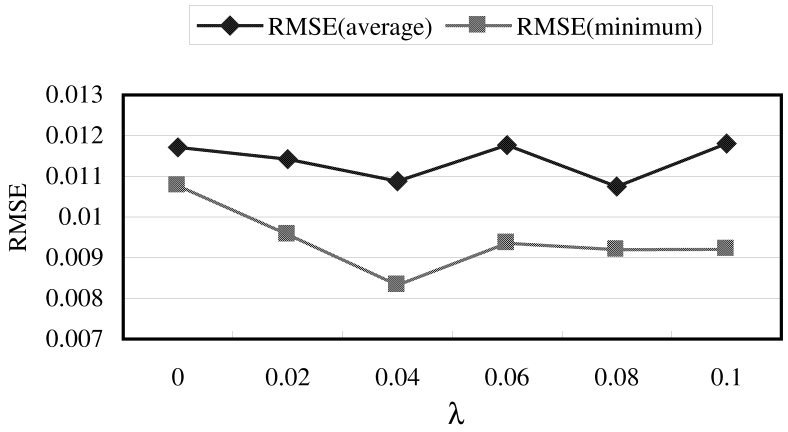

FIGURE 6. Simulation results (Mackey-Glass time-series)

by a number of researchers for comparing the learning and generalization ability of different models.

For comparison, we also performed the proposed method and LS. We select a LLWNN with 4 inputs, one hidden layer with 10 hidden units and one output unit to predict Mackey-Glass data. We predict the $x(t+6)$ using the input variables $x(t), x(t-6), x(t-12)$ and $x(t-18)$, respectively. In this simulation, 500 sample points are used. The first 250 data pairs of the series were used as training data (N=250) and the opposite-directional data size is 250 (L=250), while the remaining 250 data were used to validate the model identified. We compared the performances of the proposed methods by calculating the average test error and the best test error (using LLWNN of minimum function G) of the 20 experiments which are listed in Figure 6. Figure 6 shows the result of the changing value of λ (0.00-0.10) and it is clear that the best RMSE with λ value 0.04 is the best result in this experiment. But, the proposed method did not have statistical significance when λ=0.04 (t-test:$p = 0.05$ and $p = 0.10$).

4.3. **Application to Box-Jenkins data.** The gas furnace data (series J) of Box and Jenkins (1970) is well known and frequently used as a benchmark example for testing identification and prediction algorithms. The data set consists of 296 pairs of input-output measurements.

For this simulation, 6 inputs variables are used for constructing a proposed model. We select a LLWNN with 6 inputs, one hidden layer with 10 hidden units and one output unit. We predict the $x(t+1)$ using the input variables $x(t), x(t-1), x(t-2), x(t-3), x(t-4)$ and $x(t-5)$, respectively. In this simulation, 296 sample points are used. The first 200 data pairs of the series were used as training data (N=200) and the opposite-directional data size is 96 (L=96), while the remaining 96 data were used to validate the model identified. We compared the performances of the proposed methods by calculating the average test error and best test error of the 20 experiments which are listed in Figure 7. Figure 7 shows the result of the changing value of λ (0.00-0.10) and it is clear that the best RMSE with λ value 0.08 is the best result in this experiment. Also, the proposed method (λ=0.08) had statistical significance (t-test:$p = 0.10$).

5. **Conclusions.** In this paper, we proposed a LLWNN using bi-directional function learning method for time series prediction. Simulation results show the LLWNN for time series prediction improves a little by using bi-directional function. The model that generalization is good can be obtained by using bi-directional function from some experiment results. Especially, proposed method was effective for time series data with noise. The proposed method is simple, but it worked effectively for the time series prediction.

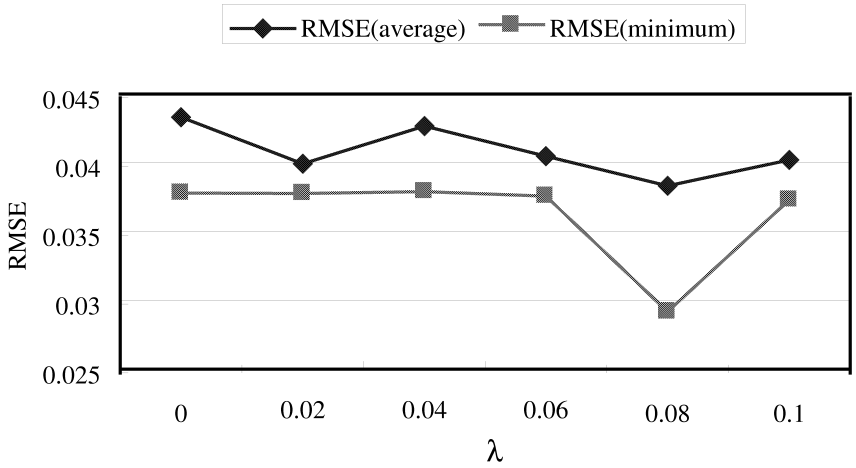

FIGURE 7. Simulation results (Box-Jenkins data)

REFERENCES

[1] Q. Zhang and A. Benveniste, Wavelet networks, *IEEE Transactions on Neural Networks*, vol.3, no.6, pp.889-898, 1992.

[2] Q. Zhang, Using wavelet network in nonparametric estimation, *IEEE Transactions on Neural Networks*, vol.8, no.2, pp.227-236, 1997.

[3] Y. Chen, B. Yang and J. Dong, Time-series prediction using a local linear wavelet neural network, *Neurocomputing*, vol.69, no.2006, pp.449-456, 2006.

[4] Y. Chen, J. Dong, B. Yang and Y. Zhang, A local linear wavelet neural network, *Proc. of the 5th World Congress on Intelligent Control and Automation*, pp.1954-1957, 2004.

[5] M. Jabri and B. Flower, Weight perturbation: An optimal architecture and learning technique for analog VLSI feedforward and recurrent multi-layer networks, *IEEE Transactions on Neural Networks*, vol.3, no.1, pp.154-157, 1992.

[6] J. Werfel, X. Xie and H. Sebastian Seung, Learning curves for stochastic gradient descent in linear feedforward networks, *Neural Information Processing Systems Foundation*, 2003.

[7] J. Kyngas and J. Hakkarainen, Predicting sunspot numbers with evolutionary optimized neural networks, *Proc. of the Second Nordic Workshop on Genetic Algorithms and Their Applications*, J. Alander (ed.), pp.173-180, 1996.