RECURRENT NEURAL NETWORKS AND DISCRETE WAVELET TRANSFORM FOR TIME SERIES MODELING AND PREDICTION

Fu-Chiang Tsui[1,2] *Mingui Sun*[1,3] *Ching-Chung Li*[2] *Robert J. Sclabassi*[1,2,3]

[1]Laboratory for Computational Neuroscience
[2]Departments of Electrical Engineering and [3]Neurological Surgery
University of Pittsburgh, Pittsburgh, PA 15261, U.S.A
lcn@neuronet.pitt.edu

ABSTRACT

A new approach is presented for time-series modeling and prediction using *recurrent neural networks*(RNNs) and a *discrete wavelet transform*(DWT). A specific DWT, based on the cubic spline wavelet, produces a set of wavelet coefficients from coarse to fine scale levels. The RNN has its current output fed back to its input nodes, forming a nonlinear autoregressive model for predicting future wavelet coefficients. A predicted trend signal is obtained by constructing the interpolation function from the predicted wavelet coefficients at the coarsest scale level, V_0. This method has been applied to intracranial pressure data collected from head trauma patients in the intensive care unit. The method has been shown to be more efficient than one which uses raw data to train the RNN.

1. INTRODUCTION

To observe and predict physiological conditions for patients in the intensive care unit(ICU), we have developed an automatic ICU monitoring system[1]. This system acquires, displays, and analyzes raw data in real time. Intracranial pressure(ICP), one of our acquired signals, sampled at 200Hz, is an important indicator for judging the physiological conditions of head trauma patients. Our past experience shows that the training of a recurrent neural network(RNN) for modeling several minutes of data would take hours to complete. On the other hand, if we utilize data obtained over a very short time interval, the system is not well-represented, and the resulting model does not adequately characterize the patient's physiological state. Thus we decided to investigate the prediction problem using a neural network in wavelet coefficient space. The approach is based on the well known data reduction properties of the wavelet transform allowing us to both rapidly train the network and predict trends in the underlying data.

This work was supported by the Ben Franklin Foundation of Pennsylvania, Computational Diagnostics, Inc. and NIH(NS30318)

In this work, we assume that ICP may be represented by a nonlinear autoregressive(AR) model and is cyclostationary over a certain time interval. Our goal is to construct a system model which is capable of predicting the ICP and eventually the patient's physiological condition.

This paper is organized as follows. In Section 2, we briefly explain the concept of the DWT, computation of wavelet coefficients, and present an example. Section 3 shows the structure of a RNN. In Section 4, we apply our algorithm to predict the ICP trend based on predicting the wavelet coefficients. Finally, the results are discussed in Section 5.

2. A DISCRETE WAVELET TRANSFORM

A discrete wavelet transform(DWT)[2] is applied in our method to reduce the training time of the neural network. The motivation for using this DWT is the computation sequence, which allows us to compute wavelet coefficients from coarse to fine scale levels efficiently.

The goal of the DWT is to approximate as closely as possible any function in *homogeneous sobolev space*[2], $H_0^2(I)$, by summing interpolation functions at several scale levels which are computed from a *Multiresolution Analysis*(MRA) where I is the interval, $[0, L]$, for $L \in Z_+$ and $L > 4$.

The MRA is constructed by shifting and dilating interior and boundary scaling functions($\phi(x)$, $\phi_b(x)$)[2] defined by,

$$\phi(x) = \frac{1}{6}\sum_{j=0}^{4}\left(\begin{array}{c}4\\ j\end{array}\right)(-1)^j(x-j)_+^3 \qquad (1)$$

and

$$\phi_b(x) = \frac{3}{2}x_+^2 - \frac{11}{12}x_+^3 + \frac{3}{2}(x-1)_+^3 - \frac{3}{4}(x-2)_+^3 \qquad (2)$$

where the subscript "+" of x_+ denotes $x > 0$.

By subsampling the raw data and interpolating with the scaled cubic spline wavelet basis, we obtain the wavelet coefficients at each scale level. Therefore, fewer number of wavelet coefficients are obtained at coarse scale levels.

 0-7803-2431-5/95 $4.00 © 1995 IEEE

2.1. Example

We apply the DWT to the ICP waveforms from coarse to fine scale levels for four levels, *i.e.*, V_0, V_1, V_2, V_3. Since the ICP is digitized, we take every 128 samples as an integer index set with respect to the interval I. Fig-1 shows the DWT of approximately 30-second of ICP data. The solid line in Fig-1 denotes the trend of the ICP.

From this figure, it may be observed that the projected function at finer scale levels produces a better approximation of the ICP waveforms. On the other hand, finer scale levels employ more wavelet coefficients to construct their interpolation functions.

3. RECURRENT NEURAL NETWORK(RNN)

There are two network topologies in artificial neural networks, *i.e.* feedforward and recurrent neural networks(RNN). The RNN which has feedback connections between layers allows the dynamics of the signal to be captured.

The RNN employed in this research is shown in Fig-2. In this network, we feed back only the output to the input nodes with a time delay. Both nonlinear and linear functions are employed in two hidden and one output layer. The RNN in Fig-2 can be described by a nonlinear AR model(Eq.-3)[3][4] which matches our assumption

$$y(n) = F(y(n-1), y(n-2), y(n-3), ..., y(n-k)) \quad \text{for} \quad n,\ k \in Z_+ \tag{3}$$

where $y(n)$ is the current output of the system and F is a nonlinear function including hidden and output layers.

The choice of the order for the nonlinear AR model is also important. It is more difficult to formulate the order of a nonlinear model than a linear one. In this paper, we select the order based on the experimental results. As to the structure of the multilayered network, we have utilized the *geometric pyramid rule*[5]. The network structure we used has 40 inputs, 1 output, 4 nodes in first hidden layer, and 2 nodes in second hidden layer, *i.e.*, $N_{40,4,2,1}$.

3.1. Network Training and Prediction

In a supervised learning scheme, we utilized the *back propagation method*[4] to optimize the 327 weights in this RNN. The stopping criteria of the network training is based on an objective function defined by the mean square error between observed and estimated output where the observed output is the data in the training set and the estimated output is the output computed from the network. If the error is smaller than a predetermined threshold, the network stops learning. During the learning process, the network feeds back the observed output data to input layer.

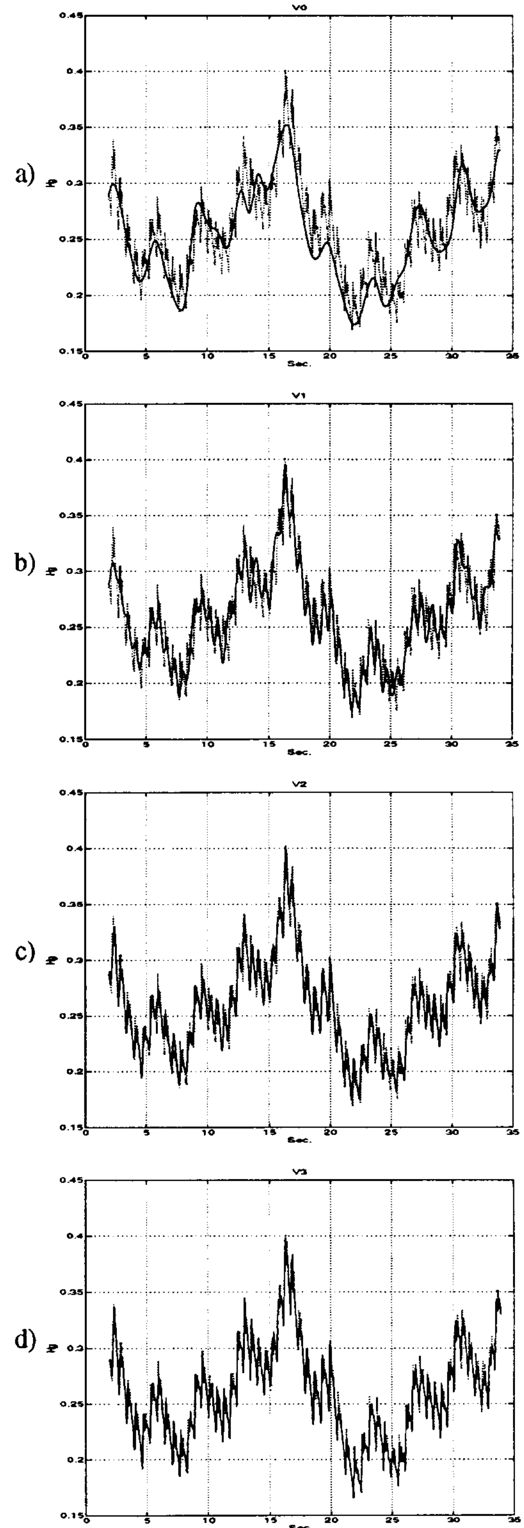

Figure 1: Projected functions of a sampled ICP signal using DWT in (a) V_0, (b) $V_1 = V_0 \oplus W_0$, (c)$V_2 = V_1 \oplus W_2$, and (d) $V_3 = V_2 \oplus W_3$ levels, where the solid-line and dash-line denote the projected functions and the raw ICP signal respectively.

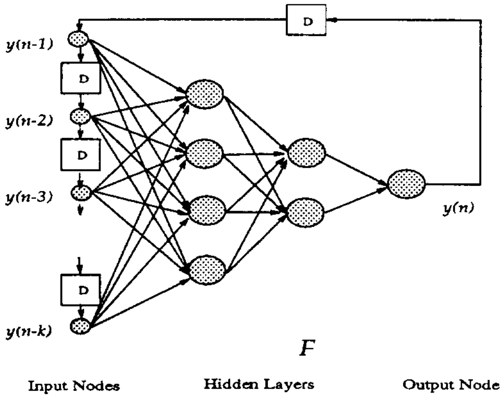

Figure 2: A Recurrent Neural Network for Prediction: the first and second hidden layers contain four and two nodes respectively. Only one node is present at the output.

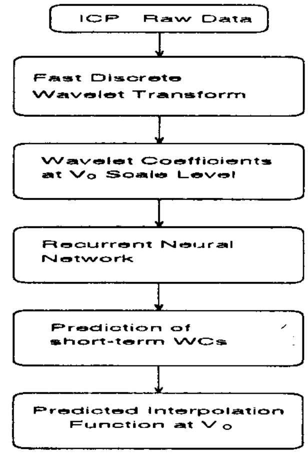

Figure 3: Block Diagram of Modeling and Prediction for Projected ICP at V_0 scale level.

In this approach, we utilize the RNN in wavelet coefficient space, *i.e,* the network predicts future wavelet coefficients based on the information in previous wavelet coefficients.

After the network stops learning, the desired weights in the network are obtained. During the prediction process, the network feeds back the estimated output data to the input layer.

For the prediction process, the network starts to predict the first output data(a wavelet coefficient) based on the initial input data at input nodes which are obtained from the last 40 elements in the given training sequence. Next, the network feeds back the estimated output data to input nodes at the input layer with a time delay. We obtain predicted data by repeating the above prediction process.

Then we construct the interpolation function based on the predicted wavelet coefficients. Note that the interpolation function in V_0 is constructed by summing scaling functions multiplied by appropriate coefficients. It is observed that if the prediction error for wavelet coefficients is small, the predicted error in the interpolation function, *i.e.,* predicted trend function, is also small. This approach is shown in Fig-3.

4. EXPERIMENTAL RESULTS

We have applied this approach to predict the trend of ICP signals. In the ICP case, the predicted signal trend is the interpolation function in V_0. We predict the ICP trend by predicting the wavelet coefficients. The wavelet coefficients are obtained from interpolating samples of every other 128 points by scaling functions in V_0.

Fig-4a displays the training result of the wavelet coefficients in V_0, *i.e.,* the coarsest level. The training sequence consists of 501 wavelet coefficients(starting from index 50 to index 550) representing the raw ICP data over approximately 320 seconds as shown in Fig-4a. The first predicted wavelet coefficient(index 551) is computed from the initial wavelet coefficients at input layer(index 511 - 550 in the training sequence). Then the network feeds back the estimated wavelet coefficient to the input layer recursively.

We predict the next 100 wavelet coefficients (corresponding to approximately 64 seconds, index 551 - 600) as shown in Fig-4b where the mean prediction error is approximately 6.5%. The total training time of this experiment is approximately 120 seconds on a HP715/75 workstation.

By comparison, if we predict this same 64-seconds of raw ICP waveform which contains 12800 samples, and utilize the training sequence containing 320-seconds of ICP waveform, corresponding to 64000 samples, the network takes eight hours to train. Therefore, this approach significantly reduces the training time.

In the prediction process, only the predicted wavelet coefficients are fed back. The error between the predicted and observed data is not utilized.

Once we know the coefficients $\{c_{-1,k}\}_{k=m}^{n}$ defined in [2] for $m, n \geq 0$ from a RNN prediction, we construct the interpolation function in $[m+3, n]$ for $m < n$, and $m, n \in Z_+$, *i.e.,* the support interval of the predicted interpolation function; note that the supports of boundary and interior scaling functions are $[0, 3]$ and $[0, 4]$ respectively.

Fig-5 demonstrates the projected functions using predicted and observed coefficients in V_0. The mean percent error in Fig-5b is approximately 4.51%.

5. DISCUSSION

In this paper, we utilize a RNN in wavelet coefficients space. We take advantage of both a MRA in DWT and the nonlinear dynamic model contributed by the RNN.

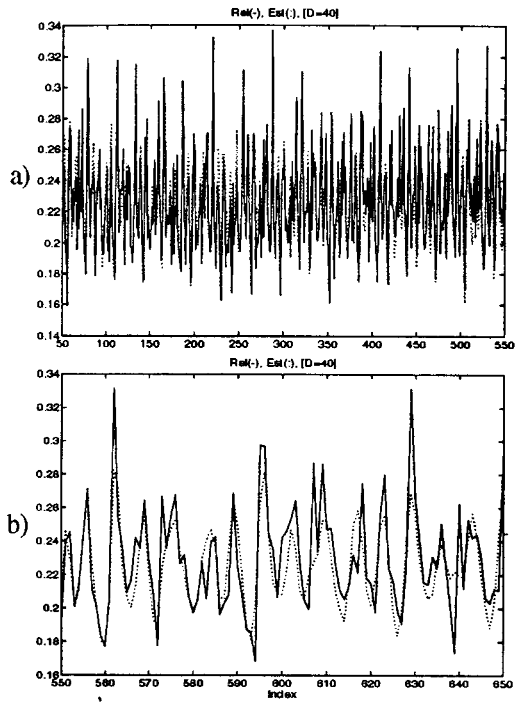

Figure 4: (a)Training result of wavelet coefficients in V_0, (b) Predicted(dashed) and observed(solid) values for the next 100 wavelet coefficients.

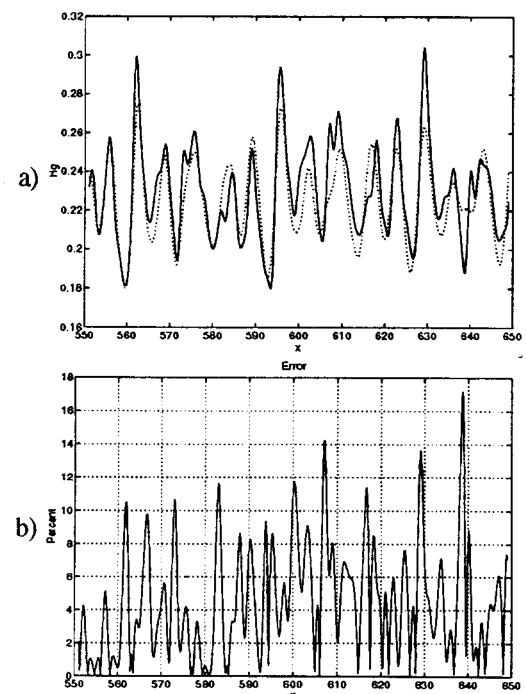

Figure 5: (a)ICP Trend functions from predicted and observed coefficients at V_0 scale level. (dash and solid lines: predicted and observed functions respectively.) (b) Percent Error between two functions.

Based on this approach, we can predict longer time intervals of future data. However the prediction performance also depends on the dynamics of the training set. If the training set does not contain the entire system dynamics, the prediction for a nonlinear dynamic system may be inadequate.

Since we use fewer coefficients to represent the raw data over longer periods of time, our computational efficiency is increased. Based on the experiments, this approach provides a fast prediction of time-series trend; *i.e.*, the computation time in trend prediction time is reduced from hours to minutes by using the wavelet coefficients instead of raw data in the RNN. our approach also appears experimentally to be stable in predicting the physiological status of a patient. This approach provides efficient prediction. Since this DWT computes wavelet coefficients from coarse to fine levels, the computation of coefficients at finer scale levels, which are not utilized, are avoided.

In the future, this approach may have the capability for parallel training of multi-RNNs if wavelet coefficients are available at several scale levels and each scale level has its own network to train and predict the wavelet coefficients.

6. ACKNOWLEDGEMENT

We acknowledge valuable discussions with Dr. Jianzhong Wang on the *Discrete Wavelet Transform*. Thanks also go to Dr. Donald W. Marion for his valuable discussions on the patients studied.

7. REFERENCES

[1] M. Sun, F.-C. Tsui, D. W. Marion and R. J. Sclabassi, "The Wavelets and Their Applications to the ICU Monitoring," in *Intelligent Engineering Systems Through Artificial Neural Networks*, Vol. 4, C.H. Dagli, B.R. Fernandez, J. Ghosh and R.T.S. Kumara ed., AMSE Press, 1994, pp. 541-546.

[2] W. Cai and J.-Z. Wang. "Adaptive Wavelet Collocation Methods for Initial Value Boundary Problems of Nonlinear PDE's," To appear in *SIAM Numer. Anal.*, 1994.

[3] A. Lapedes, R. Farber. "Nonlinear Signal Processing Using Neural Networks: Prediction and System Modeling," *Technical Report LA-UR-87-2662*, Los Alamos National Laboratory, Los Alamos, New Mexico 87545, 1987.

[4] S. Haykin. "Neural Networks. A Comprehensive Foundation," *Macmillan College Publishing Co.*, New York, 1994.

[5] T. Masters, "Practical Neural Network Recipes in C++," *Academic Press, Inc.*, Boston, 1993.