Time Series Analysis with Wavelets and Artificial Neural Networks

Zhiqiang Bi ZB26@CS.CMU.EDU
Carnegie Mellon University, Dept. of Physics and CALD 5000 Forbes Ave., Pittsburgh, PA 15213-3890

Abstract

Time series analysis has attracted a lot of attention for a long time because of its applications in a wide range of fields like science, economy, management, medicine, etc. We propose a new algorithm to perform time series analysis and prediction in order to achieve better accuracy than current techniques. Our algorithm combines wavelet transform and artificial neural networks. Noticing that treatment of boundary condition for wavelet transform might play an important role in predicting the future and previously known boundary conditions, such as periodic boundary condition or reflective boundary condition are arbitrary, we give a new method, the correctional technique. Preliminary experiments has been carried out on a synthetic dataset and some real dataset. Some promising results have been obtained.

1. Introduction

In many fields of natural and social sciences, people are interested in predicting the future from the observed phenomena. More specifically, if we observe a sequence of n events, $y_1, \ldots, y_n$ at past times $t_1, \ldots, t_n$, how can we predict the events $y_{n+1}, \ldots, y_{n+m}$ that will happen at future times $t_{n+1}, \ldots, t_{n+m}$? Examples include studying the origin and the future evolution of the universe, forecasting the weather, predicting the stock price or economical growth, managing the real (and virtual) traffic, predicting the sales of various products in a supermarket, or even continuing a great composer's unfinished work. These problems can all be classified as time series analysis (TSA).

In this project, we combine the state-of-the-art signal processing technique, the wavelet transform and artificial neural networks to perform the time series prediction. In wavelet transform, we apply a correctional method to treat the boundary condition. Our algorithm has the following characteristics:

- Wavelet transform is used to decompose the original signal into several components in multiple scales. Because wavelets are able to represent a finite time sequence in multiple scales and wavelet transform is a non-parametric method, i.e. very few *ad hoc* assumptions are made about the original signal, wavelets have proven to be a useful tool in time series analysis.
- Artificial neural networks are used to predict the event at the next moment using a sequence of past events as the input. Because neural networks with one hidden layer and arbitrary number of hidden nodes can approximate any nonlinear function, they are very useful for generalization.
- A more careful treatment of boundary condition for wavelet transform is applied. We believe the boundary condition is important in time series analysis and prediction; therefore, more careful treatment is needed.

In this paper, we first go over some prior work for time series prediction, we than propose our algorithm, which followed by some experiments on artificial data and real data, and we conclude by some discussions.

2. Prior Work

2.1 Linear Time Series Models

Time series are usually modeled as stochastic processes. Every observation in a stochastic process is a random variable, so a stochastic process can be defined as a collection of random variables ordered in time. A classical techniques for time series analysis is autoregressive moving average model (ARMA). According to ARMA(M, N) model:

$$x_t = \sum_{m=1}^{M} a_m x_{t-m} + \sum_{n=0}^{N} b_n e_{t-n} \qquad (1)$$

where a_m and b_n are constants to be determined by the data and e_t is some random variable. The first term is called autoregression first propose by Yule[2] while the second term is called moving average.

In order to capture long term trend of a time series, a technique called "integration" [3] is also used. First, the original time series is transformed into its dth-order differences:

$$\begin{aligned} y_t^{(0)} &= x_t \\ y_t^{(d+1)} &= y_{t+1}^{(d)} - y_t^{(d)} \end{aligned}$$

Analysis and prediction can be done on dth order difference of the original signal, then they can be "integrated" d times to recover the prediction to the original signal. A model combined dth-order integration to the ARMA(m, n) is called ARIMA(m, d, n) model.

These linear models are very easy to understand and straightforward to implement. However, because they are over-simplified models and have very limited expressing power, they can fail for even moderately nonlinear systems. Some examples of nonlinear sys tems for which the linear model fail are given in [4]. Therefore, for more accurate predictions, more complicated nonlinear models should be investigated.

2.2 Nonlinear Models

Despite the obvious disadvantages of linear models, they remained to be the major techniques for time series analysis until 1980, when Tong and Lim proposed the threshold autoregressive model (TAR)[5]. Under TAR, the value of the next step is chosen from one of two linear functions, which easily creates a global nonlinear system. Another extension to the linear models can be carried out by including quadratic or even higher order powers, as proposed by Volterra[6].

Recently, modern machine learning techniques have been applied to the time series analysis. Artificial neural networks[7, 8] are among the most significant. Neural network is a useful machine learning algorithm which enables us to systematically learn the response function of a system purely from input and output without making any assumptions. Because neural network with a hidden layer has been shown[10] to be able to approximate any function with arbitrary accuracy and backpropogation provides an easy implementation[9], it is naturally more preferred than the linear models and previous nonlinear models. Kolarik and Rudorfer[11] use a sequence of data points, $x_{t-s}, \ldots, x_{t-1}$ as an input vector for the neural network and x_t is the object value. Here s is a window size subject to appropriate choice. After the window slides over the whole sequence of the observed time series of length N, $N-s-1$ training examples are obtained which can be used to train the neural network which in turn can be used to make prediction by sliding further into future. Lin, et al in their work[12] applied Baum-Haussler Rule[13] to decide the number of hidden nodes and carefully studied the short term and long term forecasting, which differ in that the residuals are not available for long term forecasting, so the feedback loop is not necessary. Both works showed that neural network gives better performance than traditional linear models.

Among variations of neural networks is finite impulse response (FIR) network (also known as time delayed neural network (TDNN))[14, 15]. In FIR network, in every layer, the inputs are first fed into a filter. The filter produces a weighted sum of past n values of the input. This elegantly incorporates the idea of autoregression into the neural network. The output of the filter then is transferred to a sigmoid function and the output be-comes the input of next layer. Temporal backpropagation can be used to train the FIR network. FIR network has been shown as a successful technique for time series prediction[4]

2.3 Wavelet Methods

Instead of working on the original signal directly, we work on the transformed signals. The basic idea is to first decompose the original signal into components according to some generalized Fourier transform and then apply some predicting method to these individual components. Intuitively, different components capture different characters of the signal, so the prediction they yield should also differ. High frequency components can be used to predict the near future while low frequency components can usually tell the long term trend. If they were treated differently, better results would be expected. Fourier transformation is usually a natural choice to decompose the signal into components. However, recent studies[16] show that they are outperformed by another kind of transform, wavelet transform.

Wavelets are a set of orthogonal basis functions in $L^2(\mathcal{R})$. These basis generated from two specially constructed and mutually orthonormal functions: a scaling function ϕ and the mother wavelet, ψ. Other wavelet functions can be obtained by translations of scaling function ϕ, and dilations and translations of the mother wavelet ψ from the following relationships:

$$\phi_{j_0k} = 2^{j_0/2}\phi(2^{j_0/2}t-k) \qquad (2)$$

$$\psi_{jk} = 2^{j/2}\psi(2^jt-k), \;\; j>j_0;\; k\in Z \qquad (3)$$

The fundamental reason why wavelet transform has become more popular than Fourier transform is its ability to specify both location (via translation) and frequency (via dilation). The traditional Fourier transform can only express the global frequency spectrum without providing any localized information, such as when or where the signals are occurring. For wavelets, however, their intrinsic multiresolution feature can automatically adjust the window size to resolve local information. As a result of their nice properties outlined above, wavelets have been successfully used in time series analysis. General reviews can be found in, for example, Morettin[19, 20], Priestley[21] or Percival and Walden[22]

Because wavelets are useful in decomposing a signal and neural networks are powerful for generalization, it is natural to combine these two to perform time series analysis. As a matter of fact, some work has been done by Aussem and Murtagh[23]. They decomposed the times series into components with wavelets and let each component grow with a neural network, then combine the predictions together to obtain an overall prediction. However, we notice that some improvement can be made on their approach. This gives rise to the topic of the research project we are proposing.

3. Correctional time series prediction using stationary wavelet transform and neural networks

When it comes to wavelet transform, we suspect that boundary treatment may play a significant role in time series forecasting problems for the following reasons. First, if using wavelets decomposition would ever be able to improve the forecasting quality, it's indicated that the better we isolate the components at different scales, i.e., the more uniform scale each component represents, the better the resultant prediction is. However, using *ad hoc* boundary conditions such as periodic or reflective extensions does not guarantee that the tail part of each component really represents the same resolutional information as represented by the rest of this component. Since forecasting extends the time series from the tail, boundary handling may thus have a significant influence on where the series goes. Based on this reasoning, we believe a more careful handling of the boundary will be helpful.

We propose a Time Series Analysis with Correctinoal Wavelet Transform and Artificial Neural Network, a combination of TSA, Wavelet Transform, Neural Network, and Correctional treatment of boundary conditions. For the boundary condition, we decide to use the idea of predictor-corrector from numerical analysis. Suppose we have a time series, $y_1, \ldots, y_N$, and we want to predict $y_{N+1}, \ldots, y_{N+M}$. Our algorithm works as follows:

1. We make initial prediction using neural network on the original time series and use the predicted values as the boundary condition for the wavelet decomposition.
2. Using linear method and artificial neural network to predict each component from y_{N+1} to y_{N+M}.
3. Decompose $y_1, \ldots, y_N$ using wavelet transform again, but this time boundary is treated with predicted values in step (2)
4. Iterate step (2) and (3) until some converging criterion is met.

To test goodness of fit, we can first split an observed data sequence into two parts, one for training and one for testing. We then apply a prediction algorithm to grow to the same length as the testing sequence. A distance, L^2 for examples, between the testing sequence and the predicted sequence is computed and can be taken as a score for testing the goodness of fit. A lower score represents a better fit.

Intuitively, the end of the observed signal plays an important role in predicting the future, thus it has to be treat carefully. Our iterative scheme removes the arbitrariness of the standard periodic or reflective boundary condition in wavelet transform. It is reasonable to believe better results can be obtained.

4. Experiments and Results

4.1 Artificial Data

In order to get some hints of how our method works, we first apply it to a synthetic dataset as shown in Fig.(1)

This dataset is made up of one linear function and two sinusoidal waves. We use a base linear function because we suspect that the usual reflective boundary condition will fail to work for this long term trend. We first train the neural network on

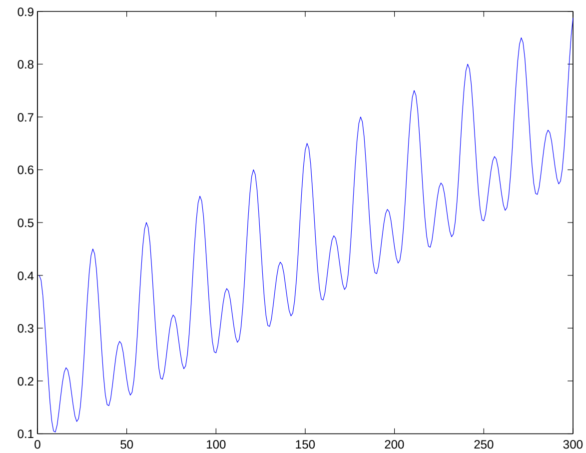

Figure 1. The synthetic data used to test our algorithm

the original signal and obtain the first order prediction, shown in Fig.(2).

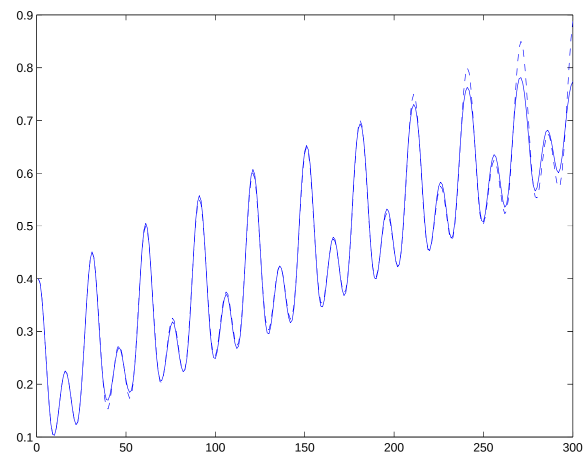

Figure 2. Training the neural network on original time series and make the first order prediction. The dashed line is the original signal and the solid line is the prediction.

We then perform a wavelet transform on the time series with boundary condition taken from the first order prediction. We plot the first four components and the residue in Fig.(3). We see that the first four components have retained perfect periodicity while the overall trend is captured by the residue component. If we first work on each of these components, we expect that each of these components can be predicted very well, therefore a good overall prediction can be achieved. The result is shown in Fig.(4). In this relatively trivial task, we obtained almost perfect prediction.

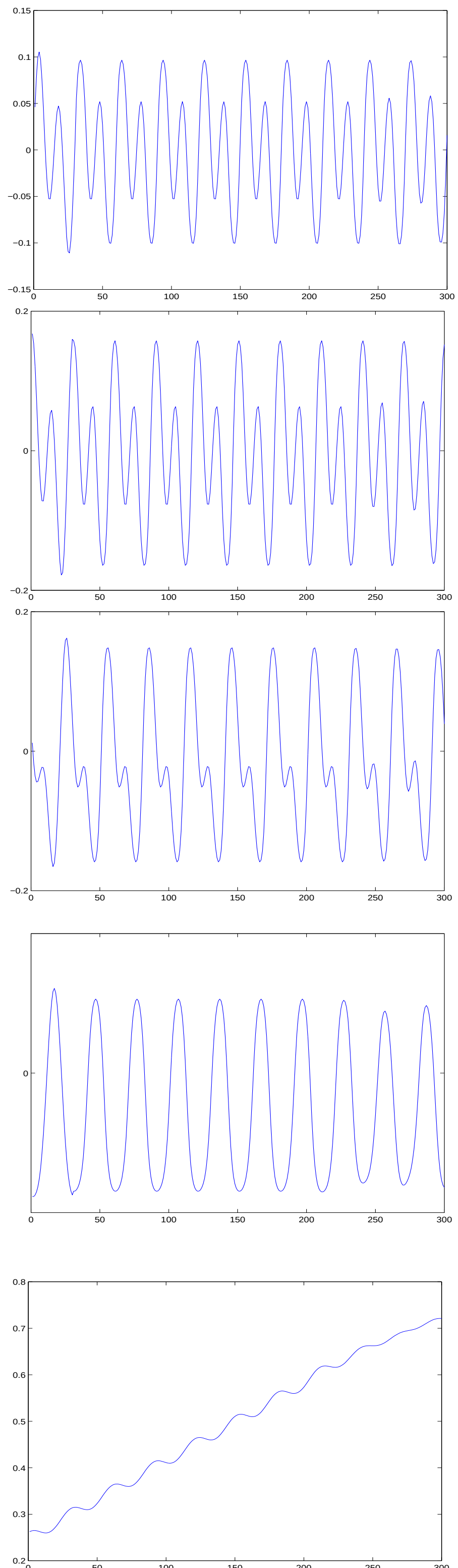

Figure 3. First four components obtained by wavelet transform and the residue

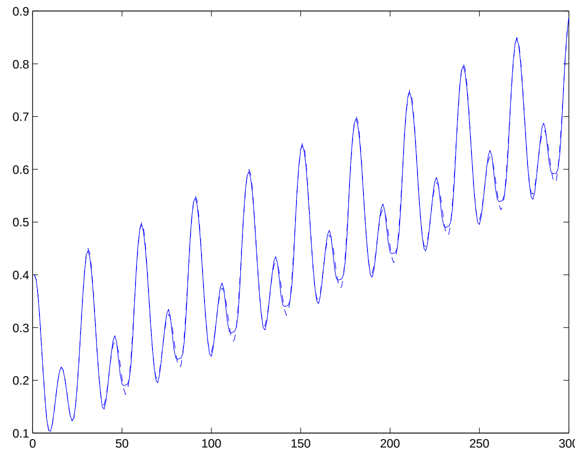

Figure 4. The overall prediction obtained by neural network and wavelet transform with correctional boundary condition.

If we use a usual boundary condition, such as the reflective boundary condition, for the wavelet transform, we obtain the prediction shown in Fig.(5), which is clearly a failure. We can con-

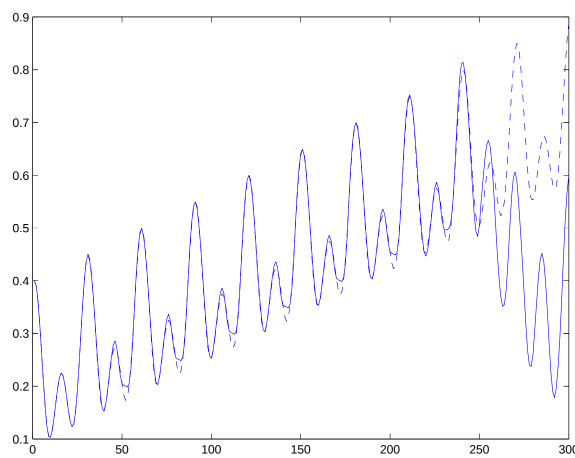

Figure 5. The overall prediction obtained by neural network and wavelet transform with reflective boundary condition.

sider the boundary condition as some kind of hint to the predictor about what is going to happen in the immediate future. As this example shows, the hint is sometimes much more important than what the predictor has already seen. When doing time series analysis and prediction, although it is a good idea to first decomposed the signal with some transforms, most transforms involve boundary condition treatment, which has to be treated carefully.

4.2 Real Data

We also apply our method to a real dataset, the Mickey-Glass Equation dataset from the Santa Fe competition[4]. The whole dataset has 1500 data points. In our experiment, we use the first 1400 points for training and predict the next 40 points. To give a general idea of this time series, we show the first 200 points in Fig.(6)

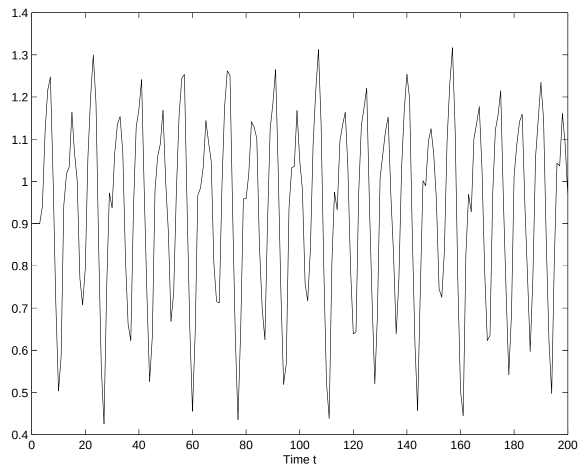

Figure 6. First 200 points from the Mackey-Glass Equation dataset

We then decompose the training signal into five components using wavelet transform and use FIR network to predict the next 40 points to each of these components. The results are shown in Fig.(7)

We can then sum up the prediction to each of these components to obtain the overall prediction as shown in Fig.(8]. The mean square error (MSE) is 0.0976.

However, if we use the reflective boundary condition, we obtain the prediction shown in Fig.(9). The MSE is 0.1189.

Therefore, the correctional boundary condition gives 21.8% increase of prediction accuracy.

5. Discussion

We designed a new algorithm to perform time series analysis and prediction. We use wavelet

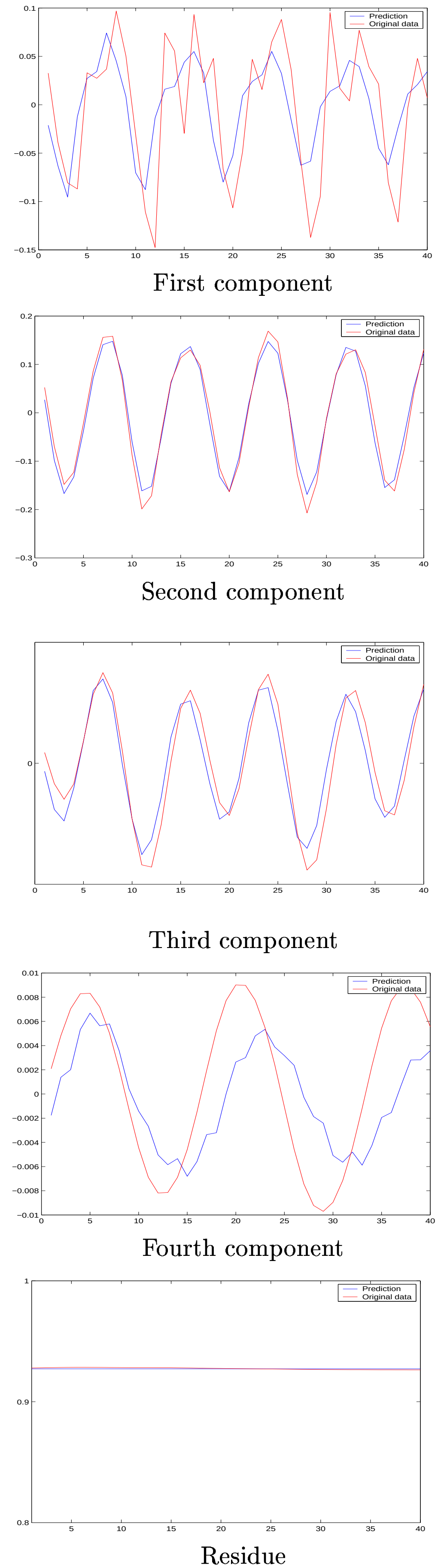

Figure 7. Prediction on each of the five component obtained by wavelet transform on the Mackey-Glass Equation dataset with correctional boundary condition

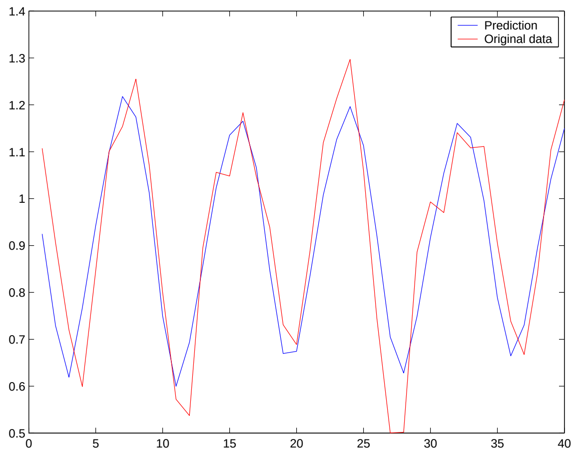

Figure 8. Overall prediction obtained by correctional boundary condition

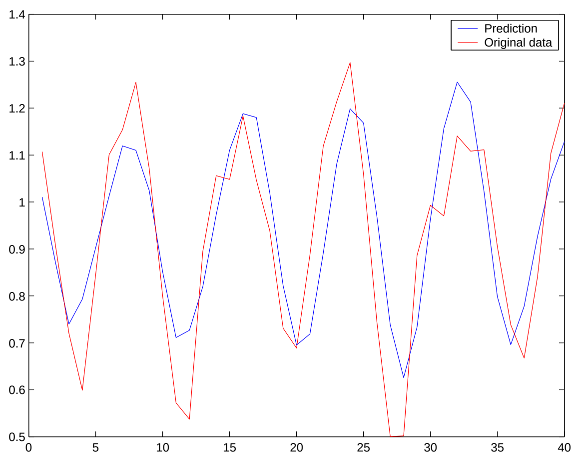

Figure 9. Prediction obtained by reflective boundary condition

transform to do signal decomposition and neural network to make prediction. However, wavelet transform involves the specifications of boundary conditions. All literature we have read treated the boundary condition in a convenient but arbitrary manner. The most usual choices are reflective boundary conditions or periodic boundary condition. Intuitively, the values at the boundary, or the most recent observations should play a very important role in time series prediction. We therefore proposed a method for which we initially make a first order predict in time domain. The first order prediction is then fed back into the time series as values beyond the boundary. This somewhat avoids the arbitrariness of the usual boundary condition treatment. In our experiment, we see we obtain better prediction results consistently with our boundary condition.

However, there are still problems. In our experiments, sometimes we observe that the first order prediction, or the prediction from the original time series, achieves better accuracy than the prediction made by wavelet transform with whatever boundary condition, which means wavelet transform doesn't necessarily improves the prediction accuracy. According to the results up to now, this usually happens when the first order prediction is already very good. When first order prediction doesn't give good result, wavelet transform can improve the prediction accuracy. This seems to suggest that the goodness of prediction by our method is dataset dependent, just like every method else. It seems to be a more sensible algorithm to set up a criterion for acceptance. If the first order prediction on testing dataset gives good accuracy, then it is accepted; otherwise, we perform wavelet transform to pursue better results. More experiments are needed to back this claim. However, the main idea of this paper is that a more careful boundary condition treatment is needed if we want to do wavelet transform, and it is consistently verified by all of our experiments.

6. Conclusion

In this paper, we designed a new algorithm for time series prediction, which combines wavelet transform and neural network. In wavelet transform, we apply a correctional boundary condition treatment, which has been shown to give better prediction than usual boundary condition, such as reflective boundary condition. More experiments are needed in the future in order to understand how to adapt our algorithm to a wider range of datasets.

7. Acknowledgment

The author would like to thank Peng Yu and Anna Goldenberg for helpful discussions.

References

[1] Lai, T. L. (1974) Control charts based on weighted sums. Ann. Statist. **2**, 134-147.

[2] Yule, G. (1927) On a Methed of Investigating Periodicity in Disturbed Series with Special Reference to Wolfer's Sunspot Numbers. Phil. Trans. Roy. Soc. London **A 226**:267-298.

[3] Box, G.E.P. and D.A. Pierce (1970), Distribution of residual autocorrelations in autoregressive integrated moving average time series models, Journal of the American Statistical Association, **65**, 1509-1526.

[4] Weigend, A.S. and Gershenfeld, N. A., Time Series Prediction, Forestcasting the Future and Understanding the Past, A Proceedings Volume in the Santa Fe Instaitute Studies in the Sciences of Complexity, 16-17.

[5] Tong, H. and Lim, K.S. (1980) Threshold Autoregression, Limit Cysles and Cyclical Data. J. Roy. Stat. Soc. B **42**:245-292.

[6] Volterra, V. (1959) THeory of Functionalsand of Integral and Integro-Differential Equations. New York:Dover

[7] Hu, M.J.C. (1964) Application of the Adaline System to Weather Forecasting. E.E. Degree Thesis. Technical Report 6775-1, Stanford Electronic Lab, Stanford, CA.

[8] Lang, K.J., Waibel, A.H. and Hinton, G.E. (1990) A time-delay neural network archiecture for isolated word recognition. Neural Networks, **3** 33-43.

[9] Mitchell, T. (1997) Machine Learning. McGraw-Hill.

[10] Cbenko, G. (1989) Approximation by superpositions of a sigmoidal function. Mathematics of Control, Signals, and Systems, **2**, 303-314

[11] (1994) Time series forecasting using neural networks. Time seris & Neural Networks, **86**

[12] Lin, F., Yu, X.H., Gregor, S. and Iroms, R. (1995) Time Series Forecasting with Neural Networks. Complexity International, **2** (http://www.csu.edu.au/ci/vol02/cmxhk/cmxhk.html)

[13] Baum E. B. and Haussler D.(1988), "What size net gives valid generalization?", Neural Computation, **1** 151-160.

[14] E. Wan, "Temporal backpropagation for FIR neural networks", International Joint Conference on Neural Networks, Vol. 1, San Diego, 575-580

[15] E. Wan, Temporal backpropagation: an efficient algorithm for finite impulse response neural networks", Proceedings of the 1990 Connectionist Models Summer School, Morgan Kaufmann, 131-140, 1990

[16] Daubechies, I. (1992) Ten Lectures on Wavelets. Philadelphia: SIAM

[17] Nason, G.P. and Silverman, B.W. (1995) The discrete wavelet transform is S. J. Comput. Graph. Statist., **3**, 163-191

[18] Nason, G.P. (1996) Wavelet shrinkage using cross-validation. J. R. Statist. Soc. B, **58**, 463-479

[19] Morettin, P.A. (1996) From Fourier to wavelet analysis of time series. In Proceedings in Computational Statistics, Pratt, A. (Ed.), 111-122, New York: Physica-Verlag.

[20] Morettin, P.A. (1997) Wavelets in Statistics. Resenhas, **3**, 211-272.

[21] Priestley, M.B.(1996) Wavelets and time-dependent spectral analysis. J. Time Series Anal., **17**, 85-104

[22] Percival, D. B. and Walden, A.T.(1999), Wavelet Methods for Time Series Analysis. Cambridge:Cambridge University Press.

[23] Aussem, A. and Murtagh, F. (1997) Combining Neural Network Forecasts on Wavelet-transformed Time Series. Connection Science Vol. 9, **1** 113-121.