Time Series Analysis with Wavelets and Artificial Neural Networks

Zhiqiang Bi

March 10, 2001

Abstract

Time series analysis attracts a lot of attention for a long time because of its applications in a wide range of fields like science, economy, management, medicine, etc. Various methods, such as linear models, nonlinear models, wavelets have been used to analyze time series and make predictions. We propose a new algorithm to perform time series analysis and prediction in order to achieve a better accuracy than current techniques. Our algorithm combines wavelet transform and artificial neural networks. We first apply wavelet transform to the original signal and decompose the signal into several components. We then perform prediction on each component with a linear model or with neural networks. Then predictions from all components are summed up to give the overall prediction. 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. Rather than using the *ad hoc* periodic or reflective boundary conditions, we append the predicted values from previous steps to the observed series for wavelet transform and then iterate. Experiments will be carried out on a variety of time series datasets, including financial data, scientific data and medical data, etc. A goodness-of-fit criterion will be proposed in order to evaluate performance of several models. The work will be carried out in two months. If successful, this algorithm can be applied to almost any field of time series analysis and prediction.

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 virtue) traffic, predicting the sales of various products in a super market, or even continuing a great composer's unfinished work. These problems can all be classified as time series analysis (TSA).

We propose a research project on a new method to do time series analysis, the Correctional Wavelet Transform with Neural Networks. Our approach combines the state-of-art signal processing technique, the wavelet transformation 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 nonparametric method, i.e. very few *ad hoc* assumptions are made about the original signal, wavelets have proven to be a useful tool in time seris 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 lay 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 treatment is important in time series analysis and prediction. With our innovative way of treating the boundary condition, we expect more accurate predictions.

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. The classical techniques for time series analysis are the linear time series models, including the moving average models (MA), the autoregressive models (AR) and their mixture, autoregressive moving average models (ARMA).

For Nth-order moving average model[1], or MA(N), the relationship between input and output is assumed to be

$$x_t = \sum_{n=0}^{N} b_n e_{t-n} \tag{1}$$

where e_t is a sequence of independent random variables drawn from a distribution with zero mean and constant variance. x_t is the output at time t. The coefficients $b_0, \ldots, b_n$ define a Nth-order convolution filter. Given an observed sequence, x_t, maximum likelihood estimation can be used to estimate these coefficients.

The autoregressive method was invented by Yule[2] when he was trying to predict the annual number of sunspots. According to a Mth-order autoregressive model, or AR(M), output at time t can be written as a linear combination of M previous outputs:

$$x_t = \sum_{m=1}^{M} a_m x_{t-m} + e_t \tag{2}$$

Here e_t can be either a control input or noise. Parameters a_m can be estimated by either maximum likelihood or least square approximation.

If both the sequence of inputs (or noise) and the observations from previous steps are taken into account, a mixture model of MA and AR can be obtained; this model is called ARMA(M, N) model:

$$x_t = \sum_{m=1}^{M} a_m x_{t-m} + \sum_{n=0}^{N} b_n e_{t-n} \tag{3}$$

Similar methods can be used to estimate coefficients a_m and b_n.

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 systems for which the linear model fail are given in [3]. 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)[4]. 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[5].

Recently, modern machine learning techniques have been applied to the time series analysis. Artificial neural networks[6], [7] 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[9] to be able to approximate any function with arbitrary accuracy and backpropogation provides an easy implementation[8], it is naturally more preferred than the linear models and previous nonlinear models. Kolarik and Rudorfer[10] 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 sides 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[11] applied Baum-Haussler Rule[12] 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 work showed that neural network gives better performance than traditional linear models.

2.3 Wavelet Methods

In previous sections, we discussed linear and nonlinear techniques for time series analysis. These techniques can be applied to the original time series, or in other words, to the signal in time domain; they can also be applied to the signal in frequency domain. 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 are expected. Fourier transformation is usually a natural choice to decompose the signal into components. However, recent studies[13] 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:

$$\phi_{j_0 k} = 2^{j_0/2}\phi(2^{j_0/2}t - k) \tag{4}$$

$$\psi_{jk} = 2^{j/2}\psi(2^j t - k), \;\; j > j_0; \; k \in Z \tag{5}$$

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. A refined Fourier transform, the windowed Fourier transformation, has been invented trying to give some localized information. In this method, a window slides through the a whole signal sequence. The Fourier transformation is performed to the portion of the signal inside the window only. Therefore, localized information is obtained after the window passes the whole time series. The flaw of this method is that the window size is fixed for all frequencies in the spectrum, thus the amount of localization is fixed for all frequencies, while preferably, large window size is required for low frequency components and small window size is required for high frequency components. 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[16],[17], Priestley[18] or Percival and Walden[19]

Because wavelets are useful to decompose 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[20]. 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 implementing the classical discrete wavelet transform, a low-pass filter $\mathcal{H}$ and a high-pass filter $\mathcal{G}$ are needed. They are applied to the digital signal to yield a smooth version of the signal and a detail version respectively. Then, a decimation is carried out so that only half of the detail version is kept as the wavelet transform coefficients at this level and half of the smooth version is recursively processed using $\mathcal{H}$ and $\mathcal{G}$ again for even coarser resolution levels. Due to the decimation, the number of wavelet coefficients is halved every time we move to a coarser level. The consequence is that the coarser level we are at, the less information is available to train the forecasting model, which may lead to the overall predicting inaccuracy. To overcome this problem, we are mainly going to exploit the stationary(redundant) wavelet transform though we will still implement the classical wavelet transform for comparison. Below, we will follow the definition of the stationary wavelet tranform given in[20].

The basic idea of stationary wavelet transform is to fill in the gap caused by decimation using redundant information obtained from the original signal. Assuming the original signal is $x(t)$ where $t = 1, 2, ..., n$, we define the smooth versions of $x(t)$ at different scales as

$$\begin{aligned} c_0(t) &= x(t) \\ c_i(t) &= \Sigma_{l=-\infty}^{\infty} h(l) c_{i-1}(t + 2^{i-1} l) \end{aligned} \tag{6}$$

where i goes from 1 to p and h is a low-pass filter with compact support, such as the B_3 spline, defined as $(1/16, 1/4, 3/8, 1/4, 1/16)$. Notice here the difference from the classical DWT is that the decimation is left out, which renders the components at different scales of the same length. We define the detail versions of $x(t)$ as

$$d_i(t) = c_{i-1}(t) - c_i(t)$$

The set $\{d_1, d_2, ..., d_p, c_p\}$ represents the wavlet transfrom of the data up to the resolution level p. The term c_p is the residual. The inverse transform is given by

$$x(t) = c_p(t) + \Sigma_{i=1}^{p} d_i(t)$$

Asssuming that we use the B_3 spline filter, computing the last element of c_1 according to Eq. (6) requires the knowledge of $x(n+1)$, $x(n+2)$, $x(n+3)$ and $x(n+4)$. Similarly, computing $c_2(n)$ requires $x(n+1)$, $x(n+2)$, ..., $x(n+12)$, and computing $c_i(n)$ would require $x(t)$ where $t = n+1, n+2, ..., n+(2^{i+2}-4)$. Since all these x values are in the "future", some guess has to be provided to these values. The typical boundary treatment includes periodic treatment where $x(n+k) := x(k)$, and reflective treatment where $x(n+k) := x(n-k+1)$.

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 garantee 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 first decompose the sequence into several components using wavelet transform with either periodic boundary condition or reflective boundary condition.
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 Work Plan

4.1 Data collection

Initially, we are going to use varies datasets from the Sante Fe Time Series Competition[3], such as the physiological data from a patient with sleep apnea, high-frequency exchange rate data, etc. If the algorithm works as expected and time permits, we are going to analyze several datasets from different research groups at CMU, such as disk access data, traffic data, supermarket sales data, or even music data. Data collection and preprocessing will take about a week.

4.2 Software and Programming

For the classical wavelet transform part, we use matlab. we have already downloaded, installed and tested the WaveLab toolbox for matlab from Stanford University. We will implement the redundant wavelet transform and neural networks in C. These take about two weeks.

4.3 Testing and Writing

Next, we are going to perform extensive testing on various datasets, including

- Prediction using linear models in time domain.
- Prediction using neural networks in time domain.
- Prediction using classical wavelet transform and neural networks with traditional boundary conditions.
- Prediction using redundant wavelet transform and neural networks with traditional boundary conditions.
- Prediction using redundant wavelet transform with correctional boundary treatment.

Comparison will be made among these models. This will take three to four weeks. we will spend the rest of the time writing the final report.

5 Summary

To summarize, we will go over several time series techniques, such as the linear models, nonlinear models, wavelet transform and artificial neural network. A new scheme of treating the boundary condition, the correctional method is proposed for the wavelet transform in order to exploit the importance of the end of the sequence in prediction. The goal is to achieve a better fit than previously known methods in terms of less prediction error.

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

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

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

[6] 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.

[7] 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.

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

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

[10] (1994) Time series forecasting using neural networks. Time seris & Neural Networks, **86**

[11] 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)

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

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

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

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

[16] 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.

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

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

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

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