AN APPLICATION OF WAVELET NETWORKS TO NONLINEAR NONSTATIONARY TIME SERIES ANALYSIS

Minu. K. K.1, C. Jessy John2
Department of Mathematics, National Institute of Technology Calicut, NIT Campus P O-673601, Kerala, India.
minukk@gmail.com, jessy@nitc.ac.in

Abstract: Nonlinear-nonstationary time series analysis has become highly imperative with the advances in Time series analysis. Existing methods are selected specific to the application and hence not suitable for general purpose. Wavelet Neural Networks (WNN) or wavelet networks are universal approximators and achieve faster convergence. Various methods of study currently used are affected by the presence of Outliers in the input data and also by the approximation error in case of discontinuities in the data. Wavelet Networks provide a solution to such existing problems. [Nature and Science 2010;8(8):27-30]. (ISSN: 1545-0740).

Key Terms: Wavelet Neural Network, Wavelet Transform, Outliers Elimination, Function approximation

I. INTRODUCTION

A time series is stationary when its properties are statistically invariant over time. There are various forms of stationarity like weak sttionarity which requires constant mean for the series. Linear models are widely used to study time series. Linear models under stationarity assumption are a powerful tool in the analysis. For instance, Moving Average or Auto-Regressive models are well known models. When the underlying series is nonlinear and nonstationary, new methods are to be developed.

A fundamental problem in Time Series Analysis is developing models from an observed series which is frequently referred to as function approximation. Artificial Neural Networks (ANN) is seen to be a good choice because they provide flexible nonlinear models whose parameters are adapted according to the available data. ANN are valuable for under-standing the fundamental problem of function learning or developing models from observed data as they are capable of learning rather complicated functions. Mathematically, the unknown mapping between the observed input-output data pairs is expressed as a combination of activation functions of the nodes. If the activation functions are local basis functions, the network is more suitable for learning functions with local variations and discontinuities. If the basis functions are orthogonal also, then the representation in their terms is unique.

Wavelets are functions that have excellent approximation properties and can be used for non stationary case also [1]. Replacing the basis function in a feed forward neural network by an orthonormal basis consisting of a family of wavelets a WNN is formed. Wavelet decomposition makes use of the theory of functional analysis where the ordinary basis functions are substituted by members of a wavelet family. This approach is so helpful that now there is the concept of adaptive wavelet shape as per the training data instead of adapting the parameters of a fixed shape basis function. [2], [3]. Networks having multi resolution hierarchies are also capable of providing answers to the problems of non–uniformly distributed input –output spaces.

In many applications Time Series obtained are non-stationary and non linear. There are methods like ANN and Nonlinear Autoregressive Moving Average with Exogenous Inputs (NARMAX) model to handle non linear modeling. When the given time series is nonlinear and nonstationary WNN provides is a good choice.

There are interruptive events such as strike, out breaks of war or even unnoticed errors of typing and recording which influence time series observation. Consequently there will be spurious observations that are inconsistent with the rest of the series; such observations are usually referred to as outliers [4]. When the timing and causes of interruption are known their effects can be accounted for by using intervention models in time series analysis. But usually they are unknown. Outliers make the resultant inference in data analysis unreliable or even invalid. Hence much effort has been devoted to tackle their presence in experimental input and output data. In an experimental observation outliers may be due to the offset of sensors, analog to digital conversion errors or even by the malfunctioning of transmission device. Such errors are difficult to pick up before processing the data.

There is a non–ignorable error in function approximation in the presence of discontinuities in the input time series similar to the Gibbs phenomenon in Fourier series [5]. As a function approximator, Wavelet network is successful in capturing the shapes of the function accurately. Increasing the size of the network or increasing the network parameters do not quite solve the problem in ANN. How this can be handled in a wavelet network is discussed below.

In this paper the methods for eliminating the effects of outliers and eliminating the ripples in function approximation due to discontinuity are discussed. In section II, a review of wavelets and WNN are considered, in section III a discussion of outlier problems is included, in section IV various methods are proposed, experimental verification is included in section V and conclusions are included in section VI.

II. WAVELETS AND WAVELET NEURAL NETWORK

A family of wavelets is derived from the translations and dilations of a single function. If $\varphi(x)$ is the starting function, the members of the family in the discrete domain are given by $2^{-\frac{m}{2}}\varphi(2^{-m}x-k)$ for $m,k\in Z$, the set of all integers. The function $\varphi(x)$ is called a wavelet. Next, consider a continuous, square integrable function $F(x)\in L^2(R)$. Taking 2^m as the sampling interval, denote $F_m(x)=A_mF(x)$ as the approximation of $F(x)$ at the resolution m. So with increasing m, the approximation $F_m(x)$ becomes coarser. If V_m denotes the vector space containing all possible approximations of F(x) at the resolution 2^m , then A_m is a projection operator on the space V_m. [6], [7].

Mallat has shown that translations and dilations of a scaling function $\phi(x)\in V_0$ form an orthonormal basis for V_m. Since, $A_mF(x)\in V_m$ we have $F_m(x)=A_mF(x) = \sum a_{mk}\phi_{mk}(x),$ the coefficients a_{mk} being the projections of F(x) onto the orthonormal basis function ϕ_{mk}. Suppose Wm denotes the orthogonal complement of V_m in V_{m-1}. Then the (m–1)th approximation of F(x) is $A_{m-1}F(x)=A_mF(x)\oplus D_mF(x)$ where D_m is a projection operator on W_m. Thus the difference of information contained in the two approximations at resolutions m and m – 1 is given by $D_mF(x)$ which is called the detail of F(x) at the resolution m. Mallat has shown that these exists a unique function $\varphi(x)$, called a wavelet, $\varphi_{mk}(x)$ whose translations and dilations form an unconditional orthonormal basis of Wm. So,

$D_mF(x)=\sum_k d_{mk}\varphi_{mk}(x)$; where d_{mk} are wavelet coefficients.

Thus $F(x)=\sum_{k=1}^{N_p} a_{pk}\phi_{pk}(x)+\sum_{k=1}^{N_p} d_{pk}\varphi_{pk}(x).$

The multi resolution analysis is a sequence … V_{-2}, V_{-1}, V_0, V_1 …of spaces of functions defined on R such that the following conditions are satisfied:

1. V_n is a closed subspace of L_2(IR) for every $n\in Z$.
2. $V_{n+1}\subset V_n$ for every $n\in Z$.
3. $\bigcup_{n=\infty}^{\infty} V_n$ is dense in L_2 (R)
4. $\bigcap_{n=-\infty}^{\infty} V_n=\{0\}.$
5. $f\in V_n$ if and only if $f(2^n.)\in V_0$ for all $n\in Z.$
6. There exists a φ in V_0 such that $\{\varphi_{0,k}:k\in Z\}$ is an orthonormal basis in V_0 , φ is the scaling function or father wavelet [8] . Given an n–element training set the over all response of a WNN is $\hat{y}(w)=W_0+\sum_{i=1}^{N_p} w_i\varphi_i\left(\frac{x-ti}{a_i}\right)$ where N_p the number of wavelet nodes in the hidden layer and w_i is the synaptic weight of wavelet network. There are different ways of constructing a wavelet network [9]. After the initial network is constructed it is further trained by the gradient descent algorithms like Least Minimum Squares (LMS) to minimize the mean-squared error:

$J(w)=\frac{1}{n}\sum_{i=1}^{N}[y_i-\hat{y}(w)]^2$ where $\hat{y}(w)$ is the real output from a trained wavelet network at the fixed weight vector W [9].

A feed forward neural network with wavelets as activation functions is a wavelet network. They have universal approximation properties, i.e., if $\Im$ is a set of functions on R^d where $\Im=\bigcup\Im_n$ where $\Im_n$ are subsets of functions. In the case of wavelet networks $\Im_n$ is the set of all wavelet networks with scale $n=2^M$. Then $\Im$ is said to possess the property of universal approximation if it is dense in the space of continuous functions C(U) supported on a compact subset U of R^d. This means for any f in C(U) there is a sequence $f_n\in\Im_n$ such that $f_n\to f$ uniformly [10]. In addition wavelet networks for certain classes of problems achieve the same quality of approximation as neural networks with a considerably reduced size [11].

III. Outlier Problem

The presence of outliers causes serious problems in data analysis. The outliers may be due to the contamination in the input space or due to the contamination in the output space. In the former case they do not contribute directly to the residuals but in the latter case they do. It is very difficult to eliminate them from the data. The reason for this can be explained as follows.

For simplicity, consider a network with a simple input node $f(x)$. Assume that θ is the parameter set of the network whose parameters are adjusted at each time step by minimizing a given function E and

$$\theta_{k+1} = \theta_k - \eta \sum_{p=1}^{M} \frac{\partial E(r_p)}{\partial \theta_k}$$

where $r_p = t_p - f(x_p)$, the residual for the pth training pattern with desired value t_p, Here η is a step size parameter and $E(r_p)$ is often referred to as the objective function of the network. The gradient is $\sum_{p=1}^{M} \frac{\partial E(r_p)}{\partial \theta_k} = \sum_{p=1}^{M} \frac{\partial E(r_p)}{\partial r_p} \frac{\partial r_p}{\partial \theta_k}$ where $\frac{\partial E(r_p)}{\partial r_p}$ is known as the influence function. To accept the performance of the networks, the difference between the out put of the network and the desired output should approach zero for all training patterns, i.e. $r_p \equiv 0$. For terminating training, the criterion is $\sum \frac{\partial E(r_p)}{\partial \theta_k} \equiv 0$. i.e., value of the parameter network tend to be nearly the same. In the Least Square Criterion, $\frac{\partial E(r_p)}{\partial r_p} = r_p$.

When the underlying error distribution is Gaussian, Least Square approach provides the optimal results. But this may not be true in real situations [12].

IV. Approach to Solution

Suppose we modify the objective function E discussed in section III. Following the discussion in [12], a class of objective functions can be formed such that (1) they pass through the origin (2) they have a unique maximizing point 'a' for $r > 0$. They have a unique minimizing point '–a' for $r < 0$. We take the interval with these two extreme points as end points called the confidence interval of the residual. Now define the class of objective functions with the form $E_R(r_p) = \sum_{p=1}^{P} [\phi(r_p) - \phi(0)]$ where $\phi(r_p)$ is a continuous function, $\phi(0)$ is a constant and P is the total number of inputs.

Assuming that average of the residuals of all training patterns should be capable of representing the residual distribution, the average of all residuals represent the error incurred by the approximation. This reduction of confidence interval is similar to designing a low–pass filter for reduction of noise effect in signal processing applications. It may be noted that the detail coefficients in the wavelet expansion act as a low pass filter. So an idea about the form of the objective function may be obtained from it. After calculating the average of all residuals as $r_{ave} = \sum_p [t_p - f(x_p)] / p$, where p is the total number of training patterns, the confidence interval of the residual is $[-c.r_{ave}, c.r_{ave}]$ where c is a constant. Let Ψ(r) = derivative of ϕ(r). Selecting Ψ(r) as the first derivative of Gaussian function, $e^{-r^2/2\sigma}$ the objective function is obtained as $E(r) = \sigma(1 - e^{-r^2/2\sigma})$. This is a robust objective function as explained above and in [14]. Thus the problem of outliers can be resolved with derivative of Gaussian wavelet as activation function in the wavelet network.

Now, the approximation $S_{\varepsilon,\rho} = \int_{\varepsilon}^{\rho} h_a * g_a * S . \frac{da}{a}$

where h is the analyzing wavelet and g is the synthesizing wavelet,

$$S_{\varepsilon,\rho} = r_\varepsilon * S, \text{where } r(t) = \frac{1}{t} \int_0^t h * g(\tau) \frac{d\tau}{\tau}.$$

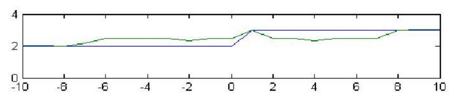
Hence $2G(t) = r * Sign(t) = 2\int_{-\infty}^{t} r(u) du - 1$. [14].

Whenever r(t)<0 ,for $t>t_N$, the largest zero of r(t), the above approximations $S_{\varepsilon,\infty}$ shows the rippling in the approximation similar to Gibbs phenomenon. For $-1 \le 2G(t) \le 1$ no such error is observed. So by conveniently choosing the wavelets g and h, as the first derivative of Gaussian wavelet, the error in approximation can be made to disappear. As wavelet network basically uses a wavelet expansion, using this Gaussian wavelet in wavelet network we can eliminate the error.

V. Experimental Verification

The above assertions can be verified using a function which is discontinuous at zero in a wavelet network. In figure (1), the plot of the input function having discontinuity at zero and the predicted function using wavelet network are shown.

The original function is shown as solid line.



Figure(1) Input function having discontinuity and its output function

For verification in the outlier elimination case, a signal is formed by adding outliers to the sine function.
The resultant signal is predicted using wavelet network. The plot of the original in the solid line and contaminated signal are shown in figure (2). It shows the robustness of wavelet network.

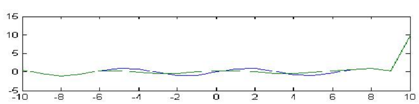


Figure (2) input with Outlier and its output function

VI. CONCLUSION

ANN are preferred for function approximations. But they have major draw backs when handling outliers especially when localized activation functions are used. Also, ripples arise in the approximation due to the presence of discontinuity. Wavelet networks can be used to tackle these problems efficiently. A method used in the ANN is modified to handle outliers in wavelet networks and properties of wavelets are exploited to better function approximation.

REFERENCES

[1] Wei, H.L. and Billings, S.A., "Identification of time-varying systems using multiresolution wavelet models", International Journal of Systems Science,33,(15),2002

[2]. Konomopoulos, A and Endou, A., "Wavelet Decomposition and Radial Basis Function Network for System monitoring", IEEE Transactions on Nuclear Science, Vol.45, No.5, October,1998

[3] Zhang, O. and Benveniste, A., "Wavelet Networks", IEEE Transactions on Neural Networks, Vol.3, No.6, November,1992

[4]. Wei, W. S. William, "Time Series Analysis: Univariate and Multivariate Methods", Addison Wesley Publishing Company, 1994

[5]. Fatemi, M., Roopaei, M. and Shabaninia, F., "New Enhanced Method for Radial Basis Function Neural Networks in Function approximation", Proceedings of the fifth international conference on Hybrid Intelligent Systems, 2005

[6]. Daubechies. I., Ten "Lectures on Wavelets", Philadelphia, PA: SIAM Press,1992

[7]. Mallat, S. G., "A Theory for Multiresoltuion Signal Decomposition: The Wavelet Representation", IEEE Transaction Pat. Anal.Mach.Int., 11,7,1989

[8]. Debnath, L. and Mikusinski, P., "Hilbert Spaces with Applications", Acxademic Press,2006

[9]. Zhang, O. and Benveniste, A., "Wavelet Networks", IEEE Transactions on Neural Networks, Vol.3, No.6, November,1992

[10]. Zhang, J., Walter, G. G. et. al., "Wavelet Neural Network for Learning", IEEE Transaction on Signal Processing, Vol.43, No.6, June,1995

[11]. Iyengar. S. S., Cho, E.C. and Phoha, Vir.v,, "Fondations of Wavelet Networks and Applications", Chapman and Hall/CRC ,2002

[12]. Lee, C. C., Cheng, P. C., et. al., "Robust Radial Function Neural Networks", IEEE Transactions on Systems, Man and Cybernetics, Vol. 29, No.6, December,1999.

[13]. Chien-Cheng Lee, Pau –Choo Cheng et al "Robust Radial Basis Function Neural Networks", IEEE Transactions on Systems ,Man and Cybernetics, Vol.29, No.6, December,1999

[14]. Holschneider.H., wavelets, "An Analysis Tool", Clarendon Press,1998

[15] Bakshi, B. K. and Stephanopoulous, G., "Wavelets as Basis Functions for Localised Learning in a Multi-resolution Hierarchy, Neural Networks", IJCNN,Vol.2, 1992 June 7-11

[16] Lineesh.M.C. and C.Jessy John,"Analysis of Non-stationary Time Series using Wavelet Decomposition",Nature and Science,2010,volume 1.

8-5-2010