Neurocomputing 149 (2015) 1074–1084

Contents lists available at ScienceDirect

Neurocomputing

journal homepage: www.elsevier.com/locate/neucom

Wavelet Volterra Coupled Models for forecasting of nonlinear and non-stationary time series

R. Maheswaran*, Rakesh Khosa

Department of Civil Engineering, IIT Delhi, New Delhi 110016, India

ARTICLE INFO

Article history:
Received 31 July 2013
Received in revised form
14 May 2014
Accepted 13 July 2014
Communicated by Bijaya Ketan Panigrahi
Available online 19 September 2014

Keywords:
Wavelets based forecasting
Nonlinear time series
Non-stationary time series

ABSTRACT

This paper provides a simple forecasting framework for nonlinear and non-stationary time series using Wavelet based nonlinear models. The proposed method exploits the ability of wavelets to detect non-stationarities that may be present in a given time series in combination with higher order nonlinear Volterra Models. The utility of the proposed model is verified using two examples: the first based on a synthetically generated times series with nonlinear and non stationary features; the second case study examined in the paper pertains to forecasting of number of pilgrims visiting the well known religious shrine at Katra in the state of Jammu and Kashmir in India. Further, the proposed model was applied to 3 time series from M3 competition. The results show that the proposed models perform better when compared with the performance of some well known benchmark models. The long term predictive capability of the wavelet based nonlinear models has also been studied separately.

1. Introduction

Forecasting the state of any system requires modelling of the underlying physical mechanism responsible for their generation. In practice, many of the real-world systems exhibit nonstationarity (in the sense that statistical characteristics change over time due to either internal or external dynamics) as well as nonlinear behaviour [6]. Explicit examples of such signals are the Sunspot numbers and the Nile River flow series. Evidence of presence of nonstationarity in some existing time series has raised a number of questions as to the adequacy of the conventional statistical methods for long-term regional forecasting [6]. The fundamental assumption in most empirical or statistical approaches is stationarity over time [2]. Pre-processing of non-stationary time series, such as through differencing, is often a pre-requisite for the next stage of model building and, as a disadvantage, is accompanied by an amplification of high frequency noise in the data [23]. Young, [23] proposed the Dynamic Regression methods as an alternative and, while these latter methods may indeed be useful for the analysis of non-stationary time series, however, their inherent static regression feature cannot fully explain many complex geophysical datasets [6]. The identification of a nonlinear dynamical model that relates directly to the underlying dynamics of the system being modelled remains a challenge and is an active research topic. Several models have been proposed for modelling nonlinear time series and range from traditional NARMAX [5] models to recent developments such as Neural Networks and Support Vector Machines (SVM). Many past studies including Kim and Waldes [1,8,9]; Nourani et al. [15] concluded that even though AI methods seem to perform well for a given data set, they lack the ability to discern and identify non-stationary components that are inherent in the natural time series. Further, it is not always possible to determine significance of the input variables prior to the exercise and it is important to identify and eliminate redundant input variables and those that contribute information in amounts deemed as insignificant. Finally, the knowledge contained in the trained networks is difficult to interpret because it is distributed across the connexion weights in a complex manner without lending itself to objective interpretation.

Recent developments in wavelet theory have resulted in the development of numerous applications [3,4,7,14], and Renaud et al. [20] for modelling non-stationary time series.

The intrinsic advantages offered by wavelets provided a fillip to initiatives on nonlinear modelling and saw an emergence of combination based approaches where wavelets were coupled with AI based techniques like ANN (see, for example, Kim (2003), Cannas and Sias (2005), Renaud et al. [20], Partal and Kisi [17], Kisi [9], and Adamoswki (2010)) and Support Vector Machines (SVM), Osowski and Garanty [16]). However, in the context of these AI models, Kisi [9] comments that these models do not overcome the disadvantages that are normally attributed to ANN based model and proposed a simple wavelet based linear regression model that the latter author applied for modelling the time series of monthly stream flows.

* Corresponding author.
E-mail address: maheswaran27@yahoo.co.in (R. Maheswaran).

http://dx.doi.org/10.1016/j.neucom.2014.07.027
0925-2312/

Nomenclature

AI artificial intelligence.
ANN artificial neural network
MISO multiple input single output.
SVM support vector machines.
NARMAX (nonlinear autoregressive moving average with eXogenous inputs)

As an alternative to this linear approach, Maheswaran, Khosa [11] proposed a Wavelet Volterra Coupled (WVC) model for modelling stream flow time series and has been shown to have the capability to accommodate non-linear attributes that the time series may be naturally vested with. As an extension of that work, this paper presents a proposal that combines the aforementioned WVC model with a suitably designed Kalman Filter for modelling a non-stationary time series along with an evaluation of its performance when applied on different kinds of synthetically generated time series having some prescribed attributes as well as on actual observed time series.

2. Wavelet transform

The well-known Fourier transform involves the projection of a series onto an orthonormal set of trigonometric components. In particular, Fourier series have infinite energy (they do not fade away) and finite power (do not change over time). In contrast, wavelets have finite energy with a very compact support and, by implication, they grow and decay in a limited time period. As a key concept, basic Fourier transform highlights the spectrum of any given signal but, importantly, this frequency decomposition is global rather than localized. Wavelet transform, on the other hand, offers the capability of localized multi-resolution decomposition (see, for example, Percival and Walden [18]) and, therefore, provides information not only on the frequency components that are present in a signal but also their occurrence in time.

The wavelet series approximation of a series $y(t)$ is defined by

$$y(t)=\sum_k c_{J,k}\phi_{J,k}(t)+\sum_{j=0}^{J}\sum_{k=0} w_{j,k}\psi_{j,k}(t) \tag{1}$$

where $\phi_{(.)}$ denotes the father wavelet and captures the smooth and low-frequency part of the signal, $\psi_{(.)}$ is the mother wavelet and captures the details and other high-frequency components, J is the number of multi-resolution levels (or scales) and k ranges from one to the number of coefficients present in the corresponding component.

The coefficients $c_{J,k}$ and $w_{j,k\ j=0,1,\ldots J}$ are known the scaling and wavelet coefficients which are given by

$c_{Jk}=\langle y,\phi_{Jk}\rangle$ and $w_{jk}=\langle y,\psi_{jk}\rangle$

where $\langle .,.\rangle$denotes the standard L^2 inner product.

The functions $\phi_{J,k}(t)$ and $\psi_{j,k}(t)$ are the approximating wavelet functions and are generated from ϕ and ψ functions through scaling and translation operators as follows:

$$\phi_{J,k}(t)=2^{-J/2}\phi\left(\frac{t-2^Jk}{2^J}\right) \tag{2a}$$

$$\psi_{j,k}(t)=2^{-j/2}\phi\left(\frac{t-2^jk}{2^j}\right),\quad j=1,2,\ldots J \tag{2b}$$

The transformation of time series $y(t)$ into the coefficients $c_{J,k}$ and $w_{j,k\ j=0,1,\ldots J}$ is termed as the discrete wavelet transform of the series y and the output of a discrete wavelet transform can take various forms. The above wavelet transform is called discrete wavelet transformation. Even though it is advantageous in data compression but not useful in time series analysis due to shift variance. As an alternative, non-decimated, (or undecimated) wavelet transforms such as the Maximal Overlap Discrete Wavelet Transform (MODWT) by Percival and Walden [19] and the '*a*' '*trou*' wavelet transform proposed by Renaud et al. [20] significantly overcome the difficulties that arise on account of lack of shift invariance associated with decimation based approaches to wavelet decomposition. In this work we have used MODWT provided by Percival and Walden [19] as WMTSA Toolbox.

3. Nonlinear integrating framework – MISO Volterra representation

The Volterra series is a model used for representing nonlinear time-invariant systems with memory. The memory of the system denotes the how far in time, the past events are influencing the system behaviour in the future. The Volterra model is based on a simple extension of the Taylor series expansion for nonlinear autonomous causal systems with memory and any given function $Y(t)$ may be written as

Or, alternatively, in terms of the Volterra operators as

$$\begin{aligned}Y(t)=&\int_0^t h_1(\tau_1)X(\tau-\tau_1)d\tau_1\\&+\int_0^t\int_0^t h_2(\tau_1,\tau_2)X(\tau-\tau_1)Y(\tau-\tau_2)d\tau_1\,d\tau_2\\&+\int_0^t\int_0^t\int_0^t h_3(\tau_1,\tau_2,\tau_3)X(\tau-\tau_1)X(\tau-\tau_2)X(\tau-\tau_3)d\tau_1\,d\tau_2\,d\tau_3+\ldots\end{aligned} \tag{3}$$

$$Y(t)=H_1[x(t)]+H_2[x(t)]+H_3[x(t)]+\ldots H_n[x(t)]+\ldots \tag{4}$$

The functions $h_n(\tau_1,\tau_2,\ldots,\tau_n)$ are called the *Volterra kernels* of the system, and the transformation $H_n[x(t)]$, which represents the convolution integral, is known as the nth order *Volterra operator*.

The Volterra kernels can be assumed to be symmetric without any loss of generality and is defined to be one which is a symmetric function of its arguments so that any of the $n!$ possible interchanges of the arguments $\tau_1,\tau_2,\ldots,\tau_n$ leave the kernel value $h_n(\tau_1,\tau_2,\ldots,\tau_n)$ unchanged. For example, when $n=2$, $h_2(\tau_1,\tau_2)$ is symmetric if $h_2(\tau_1,\tau_2)=h_2(\tau_2,\tau_1)$.

Generally, it has been seen that the expansion up to second order is sufficient to explain the non-linearity in most natural systems [10] and for these cases, the second-order Volterra expansion reduces to

$$Y(t)=\int_0^t h_1(\tau_1)X(\tau-\tau_1)d\tau_1+\int_0^t\int_0^t h_2(\tau_1,\tau_2)X(\tau-\tau_1)X(\tau-\tau_2)d\tau_1\,d\tau_2 \tag{5}$$

The above equation can be expressed in discretized form as shown below

$$\begin{aligned}&y(t)=\sum_{j=1}^{M_k}H_1(j)x(t-j)+\sum_{i=1}^{M_k}\sum_{j=1}^{M_k}H_2(i,j)x(t-i)x(t-j)\\&y(t)=y_1(t)+y_2(t)\\&t=1,2,3\ldots n.\end{aligned} \tag{6}$$

where M_k represents the memory of the system and H_1 and H_2 represents the discrete form of the first and second order Volterra kernels (h_1 and h_2) that have to be estimated. The Volterra series