Contents lists available at ScienceDirect

Digital Signal Processing

www.elsevier.com/locate/dsp

Fast wavelet transform assisted predictors of streaming time series

Marco Stocchi [a], Michele Marchesi [b]

[a] Dept. of Electronic Engineering, University of Cagliari, v. Marengo 9, Italy
[b] Dept. of Mathematics and Computer Science, University of Cagliari, v. Porcell 4, Italy

ARTICLE INFO

Article history:
Available online xxxx

Keywords:
Streaming datasets
Time series forecast
Fast Wavelet Transform
Shift variance theorem

ABSTRACT

We explore the shift variance of the decimated, convolutional Discrete Wavelet Transform, also known as Fast Wavelet Transform. We prove a novel theorem improving the FWT algorithm and implement a new prediction method suitable to the multiresolution analysis of streaming univariate datasets using compactly supported Daubechies Wavelets. An effective real value forecast is obtained synthesizing the one step ahead crystal and performing its inverse DWT, using an integrated group of estimating machines. We call Wa.R.P. (Wavelet transform Reduced Predictor) the new prediction method. A case study, testing a cryptocurrency exchange price series, shows that the proposed system can outperform the benchmark methods in terms of forecasting accuracy achieved. This result is confirmed by further tests performed on other time series. Developed in C++, Standard 2014 conformant, the code implementing the FWT and the novel Shift Variance Theorem is available to research purposes and to build efficient industrial applications.

1. Introduction

The problem of forecasting streaming datasets has been largely explored in the past, especially in the financial field; the main reason being the availability of large price time series, which are better suited to test any newly contrived predictor system. We also believe that such research field will become a core issue with the advancement of technologies such as the Internet of Things, giving the emerging need of forecasting the data generated by sensors and surveillance devices.

Recently, many researchers [42,32,23] are heading towards novel multiscale analysis approaches, motivated by recent findings about the powerful methods of wavelets, the latter being applied either alone, or in conjunction with other prediction models. Wang et al. [54] studied the possibility to improve the forecast of financial time series integrating a wavelet analysis denoising preprocessing module, and back propagation neural networks to perform regression. Their results show that such approach is promising, and motivated us to develop an implementation of the wavelet denoising neural network (WDNN) in order to benchmark the herein proposed inference engine's performance.

More recently, hybrid wavelet-based machine learning systems have been proposed. In Huang et al. [29] the wavelet analysis is combined with Support Vector Regression to forecast prices. The authors report improvements in the precision of the results, comparing them with the outputs of other standard SVR methods. Fang et al. [21] propose a prediction system based on genetic algorithms and wavelet neural networks, achieving a better accuracy than other benchmarking methods. A wavelet analysis is also used by Andrieş et al. [2] to investigate the behavior of exchange rates of several national currencies of eastern european countries.

Quite recently, many discrete wavelet transform (DWT algorithms) where proposed; the majority of them are classified as undecimated, shift invariant transforms, hence they are immediately applicable to the analysis of streaming datasets. Examples are the Stationary DWT [45], the "à trous" [52,39], the Maximum Overlap DWT [24]. These methods are alternative to the decimated, convolutional Discrete Wavelet Transform (also know as the Fast Wavelet Transform – FWT), which is implemented by iteratively filtering and downsampling a source series using two quadrature mirror filters. However, the FWT is known to be a *non* shift-invariant algorithm. Such feature causes major difficulties when performing the DWT of shifting time series. In particular the lack of shift invariancy, in signals processed using the FWT, impairs the possibility to compare directly two DWT crystals, calculated before and after a shift-insertion operation. Moreover, in order to extrapolate the nth step ahead of a time series, there must be full availability of the past data, hence the possibility to contrive a predictor based on the FWT analysis is difficult as well. It is then logical that the lack of any law describing the FWT coefficients transposition in the shift domain implies the need of endowing a forecasting system with

E-mail address: marco.stocchi@diee.unica.it (M. Stocchi).

https://doi.org/10.1016/j.dsp.2017.09.014
1051-2004/

a large number of regressor machines (one for each DWT coefficient). Such redundant configuration would cause lengthy training operations and it would inevitably degrade the accuracy achievable. It is our opinion that the challenge of shifting time series prediction, assisted by the wavelet analysis, would be better approached if the number of machines (deputed to perform real value forecasts) could be reduced to a minimum. Also, retaining the FWT algorithm, instead of adopting a shift invariant wavelet transform, would allow to implement a much more efficient and parsimonious system.

To this purpose, we explore the shift variance properties of the FWT of streaming input datasets, proving a novel Shift Variance Theorem – SVT. We integrate the SVT in the implementation of our predictor, greatly reducing the number of the required estimator machines. We name such predictor system a Wa.R.P. – Wavelet transform Reduced Predictor. Results show that the suggested method can outperform the benchmark models when applied to the Bitcoin-US Dollar hourly exchange rates. In order to evaluate the degree of generalization of the predictor, we also test the Wa.R.P. using three currency exchange pair datasets and three statistical time series. The predictor confirms its accuracy when compared with artificial neural networks and support vector regression.

The novel SVT is not only useful to devise highly efficient predictor systems, but it also allows to speed up the computation of the DWT of streaming univariate datasets. We studied the computational complexity of the reduced FWT, showing that an asymptotic reduction of 50% of the operations needed is achieved. We also tested a CPU implementation of the reduced FWT, observing an effective, significant improvement in computation time, proportional to the size of the sample used.

The software, available for download from a public repository (see Appendix A), is currently used for research purposes; the projected pathway to achieve an industrial impact has been considered at the Agile Group of the Department of Mathematics and Computer Science of the University of Cagliari. As such, an efficient SaaS application, suitable to provide DWT calculation services, is presently under development. Our ambition is to endow it with the highest feasible number of external univariate data sources.

This paper is organized as follows: Sec. 2 provides a background of the discrete wavelet transform, introducing the notation used; it addresses the shift variance problem and proofs the associated theorem. It also proposes a framework suitable to employ regression machines, using the DWT crystals of a sampled series. Sec. 3 describes the case-study application (the Bitcoin – US Dollar hourly exchange rates prediction), benchmarking the results obtained by the Wa.R.P. engine with other forecasting systems. Sec. 4 reports the prediction results obtained using further sets of source time series. Sec. 5 addresses several aspects of the calibration of the system, useful to achieve improvements in the forecasting performance. A final discussion and the conclusions are reported in Sec. 6.

2. Method

Let us denote $f(t):\mathbb{R}^+ \to \mathbb{R}$ a source signal, to be sampled at constant time intervals. At each new sampled data insertion, a multiresolution analysis (MRA) of a fixed size window of the sampled 1D series is performed, using the Fast Wavelet Transform algorithm; the resulting output, called *crystal*, is stored as a row entry of a matrix of appropriate size. The prediction system performs a forecast of the one-step ahead crystal, using dedicated regressor machines. In the present work, the Wa.R.P. engine is endowed with a group of multilayer perceptrons, each trained to perform the forecast of a single coefficient of the DWT. Eventually, the prediction of the input series is obtained by inverting the FWT of such estimated crystal.

In order to give a brief review of the discrete wavelet transform theory and, contextually, to establish notation, let us recall that the following families of functions $h_{m,n}(t)$, generated by shifting and scaling an appropriately chosen mother wavelet h:

$$h_{m,n}(t) = \left\{ 2^{-m/2}\, h(2^{-m}t - n) : m, n \in \mathbb{Z} \right\}, \tag{1}$$

constitute an orthonormal basis of the vector space of measurable square integrable 1D functions $L^2(\mathbb{R})$ [14]. They have been applied, in the past, to the analysis of signals pertaining to many scientific disciplines, and recently they were also applied to the study of financial time series. Their elements have good localization properties in both the spatial and Fourier domains. If the source time series is continuous, its DWT is defined as:

$$T_{m,n}(f) = \langle h_{m,n}, f \rangle = 2^{-m/2} \int_{\mathbb{R}} f(t)\, h(2^{-m}t - n)\, \mathrm{d}t. \tag{2}$$

Mallat [40] proved that the functions $f(t) \in L^2(\mathbb{R})$ can be considered as a limit of successive approximations (smoothed versions of $f(t)$), and that it is possible to find the wavelet coefficients $T_{m,n}$ as the difference of two approximations of $f(t)$ at consecutive different scales. To obtain this, it is necessary to define two families of scaling and wavelet functions, $\phi_{m_0,n}(t)$ and $\psi_{m,n}(t)$, respectively, which enable to express $f(t)$ by means of the following series:

$$f(t) = \sum_n c_{m_0,n}\phi_{m_0,n}(t) + \sum_m \sum_n d_{m,n}\psi_{m,n}(t). \tag{3}$$

Eq. (3) is the key to signal reconstruction (also called inverse DWT – IDWT), whereas the forward DWT allows to retrieve the $c_{m_0,n}$ and $d_{m,n}$ sets of coefficients. Moreover, a multiresolution decomposition of a sampled series can be performed efficiently in a recursive way, by means of filtering and downsampling operations, hence the algorithm name of *fast* wavelet transform (FWT). Usually, these filters are denoted by h and g. Because of their properties, in signal analysis they are referred to as quadrature mirror filters. The interested reader can rely on the original description of the Multiresolution Analysis, introduced in Mallat [40]. A detailed review of the wavelet based decomposition and reconstruction algorithms can be found in Daubechies [14].

Let us denote a source digital signal X of size n_X, composed of the last consecutive sampled values of $f(t)$:

$$X = \{ x_{n_X-1}, \ldots, x_1, x_0 \}. \tag{4}$$

In eq. (4) the x_0 element is a newly inserted source element. The size n_X is kept constant by popping (removing) the first and older element from vector X (first in last out). Let us also denote n_h the size of both h and g filters, m the recursion depth of the procedure, 2^{-m} being the resolution $\mathcal{R}_m$ to which the signal is analyzed at depth m. If $m=0$ ($\mathcal{R}_0 = 2^0 = 1$), the input series is of course the source series itself. The forward FWT starts by convolving the source series with both filters h and g, and retaining (separately) one sample out of two. This allows to obtain, respectively, two sets of coefficients, herein denoted by c_1 and d_1, of the next level $m=1$ (lower resolution $\mathcal{R}_1 = 2^{-1}$), respectively representing a smoothed version of X and the difference of information between the two adjacent series $c_0 \triangleq X$ and c_1. The resulting vector of c_1 coefficients is used as input to the operation of the next level $m=2$, and the process can be repeated up to the maximum depth M. The last set of coefficients c_M is retained. The maximum depth M depends on both n_X and n_h:

$$M = \log_2 n_X - \lceil \log_2 n_h \rceil. \tag{5}$$

Please cite this article in press as: M. Stocchi, M. Marchesi, Fast wavelet transform assisted predictors of streaming time series, Digit. Signal Process. (2017), https://doi.org/10.1016/j.dsp.2017.09.014