Wavelet Based Exchange Rate Forecasting with Improved Instance Based Learning

Pushpalatha M P
1
, and Nalini N
2
,

1 Department of CSc and Eng,Sri Jayachamarajendra College of Engineering, Mysore ,Karnataka, INDIA
2 Department of ISc and Eng, Nitte Meenakshi Institute of Technology, Bangalore, Karnataka, INDIA

Abstract—In this paper we present a novel wavelet based exchange rate forecast model integrating wavelet filters for denoising and Improved Instance Based Learning(IIBL) approach. The proposed model implements a novel technique that extends the nearest neighbor algorithm to include the concept of pattern matching so as to identify similar instances thus implementing a nonparametric regression approach. The work demonstrates the feasibility of integrating with suitable non-redundant orthogonal wavelet filters at the preprocessing stage to achieve accurate forecasting. The multi-scaling property of the wavelet transform enhances the prediction with high accuracy for volatile time series. The impact of using Discrete Wavelet Transform (DWT) has been systematically illustrated in the preprocessing stage on the accuracy of forecasting. The analysis of simulations demonstrate that the proposed wavelet based IIBL model results in accurate predictions and encouraging results for exchange rate series when compared with the conventional neural network, wavelet and wavelet denoising methods.

Keywords: Instance based learning, Wavelet transforms, Exchange rate forecast.

1. Introduction

Time series forecasting is a challenging task due to its high volatility and noisy environment. Classically, the time series data are assumed to be stationary: their characterising quantities behave homogeneously over time. Multi-step ahead time series forecasting has become an important activity in various fields of science and technology due to its usefulness in future events management. In the context of time series, the long-term or multi-step prediction problem is an interesting problem since it obtains predictions several steps ahead into the future starting from information at current instant [7], [4], [5], [7], [8]. Nearest Neighbour (NN) method is widely known for its computational simplicity. The Single Nearest Neighbour (SNN) method involves computing the proximity of the current value with all the data in the training period. It uses Euclidean distance to compute the proximity of current value with all the data in the estimation period. The K-Nearest Neighbour (KNN) method extends the SNN method, chooses k neighbours that have the k highest proximity values and involves computing the proximity of the current value with all the data in the training period. The KNN method uses a simple averaging technique to predict multi-steps ahead which usually does not result in an accurate forecast. A novel wavelet based prediction model integrating wavelet filters for denoising and Improved Instance Based Learning approach developed has been presented. The proposed model implements a novel technique that extends the nearest neighbour algorithm to include the concept of pattern matching so as to identify similar instances thus implementing a non-parametric regression approach. A hybrid distance measure combining correlation and Euclidean distance to select similar instances has been proposed. Hence modification in the standard nearest neighbour method results in improved forecasting accuracy without affecting the simplicity of the nearest neighbour method. Daubechies wavelets is applied at the preprocessing stage mainly to suppress the noise in the signals, to result in better prediction values in terms of performance indices used [29].

2. Multi-step Prediction

In many time series applications, one-step prediction schemes are used to predict the next sample of data x(k+1), based on previous samples. The disadvantage of one-step prediction is that, it may not provide enough information especially in situations where a broader knowledge of the time series behaviour is desirable to anticipate the behaviour of the time series process. Hence the long-term or multi-step prediction model is used as it obtains predictions several steps ahead into the future i.e., x(k+1), x(k+2), x(k+3), starting from information at current instant k as in [4],[9],[11],[2],[19],[22].

Many techniques exist for the approximation of the underlying process of a time series, linear methods such as Auto Regressive external input (ARX), Auto Regressive Moving Average (ARMA) model etc., and nonlinear ones such as Artificial Neural Networks (ANN). In general, these methods try to build a model of the process. The model is then used on the last values of the series to predict the future values. The common difficulty of all the methods is the determination of sufficient and necessary information for an accurate prediction as stated in [11],[18].

Accuracy of multi-step ahead forecasting using nearest neighbour depends upon the quality of the search for the best
matched pattern. Hence, it is very important to search for the most similar pattern from the stored patterns. Euclidean distance based search used in the standard nearest neighbour method minimizes the distance between the reference pattern and the patterns stored in the database without considering the similarity of the shape of the patterns [11],[12]. Further, model based multi-step ahead forecasting fails in many cases because the model cannot learn the dynamics of the system completely. And most of the existing methods are useful for single-step ahead time series forecasting and their accuracy degrades in the case of multi-step ahead forecasting.

Hence the proposed work incorporates a hybrid distance measure combining correlation and Euclidean distance to select similar instances from the stored patterns. It has been motivated by the effective preprocessing capability of wavelet filters and the predictive power of improved instance based learning system, to represent a hybrid prediction system [10],[23].

3. Wavelet Theory

Let $L^2(\mathbb{R})$ denote the space of all square integrable functions in $\mathbb{R}$. Let $\psi(t) \in L^2(\mathbb{R})$ be a fixed function. The function $\psi(t)$ is said to be a wavelet if and only if its Fourier Transform (FT) $\hat{\psi}(\omega)$ satisfies

$$ C_\psi = \int_0^\infty \frac{|\hat{\psi}(w)|^2}{|w|} dw < \infty $$

(1)

The equation (1) is called the admissibility condition [29],[?], which implies that the wavelet must have a zero average

$$ \int_{-\infty}^{\infty} \psi(t) dt = \hat{\psi}(0) = 0 $$

(2)

and therefore must be oscillatory.

Let us define the function $\psi_{a,b}$ by

$$ \psi_{a,b}(t) = \frac{1}{\sqrt{a}} \psi\left(\frac{t-b}{a}\right) $$

(3)

where $b \in \mathbb{R}$ is a translation parameter, whereas $a \in \mathbb{R}^+ (a \neq 0)$ is a dilation or scale parameter. The factor $a^{-\frac{1}{2}}$ is a normalization constant such that $\psi_{a,b}$ has the same energy for all scales $a$. It is observable that the scale parameter $a$ in equation (3) rules the dilations of the spatial variable ($t-b$).

In the same way, factor $a^{-\frac{1}{2}}$ rules the dilation in the values taken by $\psi$.

With equation (3), it is able to decompose a square integrable function $f(t)$ in terms of dilated-translated wavelets.

We define the continuous wavelet transform (CWT) of $f(t) \in L^2(\mathbb{R})$ by

$$ T_\psi[f](a,b) = \langle f, \psi_{a,b} \rangle = \int_{-\infty}^{\infty} f(t) \overline{\psi_{a,b}(t)} dt $$

$$ = \frac{1}{\sqrt{a}} \int_{-\infty}^{\infty} f(t) \overline{\psi\left(\frac{t-b}{a}\right)} dt, $$

(4)

where $\langle , \rangle$ is the scalar product in $L^2(\mathbb{R})$ defined as

$$ \langle f, g \rangle := \int f(t) \overline{g(t)} dt, $$

and the symbol bar denotes complex conjugation. The CWT (4) measures the variation of $f$ in a neighborhood of point $b$, whose size is proportional to $a$.

Reconstructing $f$ from its wavelet transform, the reconstruction formula given by [?] is

$$ f(t) = \frac{1}{C_\psi} \int_{0}^{+\infty} \int_{-\infty}^{+\infty} T_\psi[f](a,b) \psi_{a,b}(t) \frac{dadb}{a^2} $$

(5)

the above equation implies the need of equation (1).

However, some data are represented by finite number of values hence it is important to consider a discrete version of CWT of equation(4). In our proposed work the orthogonal (Discrete) wavelet bases are employed. This method associates the wavelet with orthonormal bases of $L^2(\mathbb{R})$.

The expansion of an arbitrary signal $x(t)$ on an orthonormal wavelet basis takes the form

$$ x(t) = \sum_m \sum_n x_{m,n} \psi_{m,n}(t) $$

(6)

$$ x_{m,n}^m = \int_{-\infty}^{+\infty} x(t) \psi_{m,n}(t) dt $$

(7)

where the orthonormal wavelet basis functions are related according to

$$ \psi_{m,n}(t) = 2^m \psi(2^m t - n) $$

(8)

with both $m$ and $n$ as the dilation and translation indices, respectively. The family of equation (8) can be obtained from equation (3), setting the parameters $a = 2^{-m}$ and $b = \frac{n}{2^m}$.

The contribution of the signal at a particular wavelet level $m$ is given by

$$ x_m(t) = \sum_n x_{m,n} \psi_{m,n}(t) $$

(9)

Equation (9) gives us information of the time behaviour of the signal within different scale bands, and gives their contribution to the total signal energy.

4. Wavelet Denoising

The main advantage of modelling with wavelets lies in the fact that is possible to represent the transitory characteristics of the time series in more efficient way. This advantage derives from the fact that wavelets are limited duration functions, moreover the shape of the wavelets used for modelling can be chosen according to the characteristics and behaviour of the time series to be modelled. A filter based on wavelet transform can be implemented to obtain a more accurate signal of the process under interest as in [23][25].

Multiscale analysis varying from one to several levels of decompositions have been performed and the improvement
in accuracy is seen at higher level of decompositions as described in \cite{27,28,29,10,13,17}. One way to model a time series is to consider it as a deterministic function with noise incorporated. When the noise element in a time series is carefully minimized by a process called denoising, a better model can be obtained for that series.

Many denoising methods have been compared in the literatures of \cite{30,26,3,12,14}. A good denoising approach consists in setting the smallest coefficients to zero and shrinking the remaining ones above a certain threshold as in \cite{24}.

Hence, the idea behind wavelet denoising is to threshold the wavelet coefficients at every multiresolution level so that the amounts of noise present in the detail coefficients are removed. Of the wide choice of wavelet filters, orthogonal wavelets have been considered such as Haar and Daubechies at the preprocessing stage mainly to suppress the noise in the signals, which results in better prediction values in terms of performance indices used \cite{27,21}.

The denoising objective is to suppress the noise part of the signals and to recover function \( f \). It is reported that soft thresholding is more effective than hard thresholding approach, hence soft thresholding has been employed as in \cite{13,24}.

## 5. PROPOSED ALGORITHM

Instance Based Learning is a framework and methodology that can be applied to generate time series prediction using specific instances \cite{4}. The data preprocessing method adopting wavelet filters facilitates to the process of data representation and is able to deal with the non-stationary involved in most of the real time series \cite{1}. The de-noising objective is to suppress the noise part of the signals and to recover function \( f \). The denoising procedure is adapted considering appropriate wavelet filters \cite{2}, \cite{7}. The proposed approach extends the nearest neighbor algorithm to include the concept of pattern matching to identify similar instances. Pattern matching in the context of time-series forecasting refers to the process of matching current state of the time series with its past states. The specific instances chosen by the proposed approach are combined using non-linear regression to generate multi-step ahead predictions. We extend IIBL with a significant test to distinguish noisy instances by cascading with wavelet filters. Given a time series data set the instance based classification aims at locating similar pieces of information independently of their location in time. While locating the most similar neighbors the method tries to eliminate outliers present in the time series data. In our approach the time series data set is partitioned into training period set and test period set. The training period is then subpartitioned into \( N \) instances termed as ‘windows’ each of specific number of observations \( \text{L} \) and referred as sliding window size. The number of nearest neighbors \( \text{H} \) refer to the best instances out of \( \text{N} \) instances required for the process of forecasting. The optimal values of \( \text{H} \) and \( \text{L} \) are determined by conducting several trials so as to obtain optimal results in terms of RMSE and MAPE. The instance that lies just before the time value to be predicted is chosen as the critical instance (reference pattern) against which the similarity of other instances (candidate patterns) is to be estimated. Euclidean correlation is used as the similarity metric to choose similar instances. Once the similar instances (patterns in this case) are chosen they are combined using nonlinear regression method to predict future forecast pattern. The forecasted instance is then added to the training set, now treated as the new reference pattern, against which the similarity of other candidate patterns is estimated. This process is continued till the required number of forecast value is generated.

The proposed learning algorithm is given below.

1) Opt suitable train_ period (TR) and test period (TS) from total data set size.
2) Initialise \( L \) and \( H \) to suitable values.
3) Select Ref_pattern = \( X_\text{cur} \).
4) Set Candidate_patterns = \( X_{i} \) where \( i = 1, 2, 3 \ldots n-1 \)
5) for \( j = 1 : L \)
   if \( \text{corr}[t] > 0 \)
   \( \text{Euclid corr}[t] = \sqrt{\sum(x_{j,\text{cur}} - x_{j,i=1,2,3,\ldots})^2} \)
   \( j = 1, 2, 3, \ldots \) for \( t = 1, 2, 3 \ldots \) (size/L) where size is the total number of observations in the training set.
6) Choose \( H \) lowest values from Corr[i]
   for \( j = 1, \ldots (\text{size}/L) \)
   \( \text{Low}[i] = \min(\text{Corr}[j]); \) where \( i = 1:H \)
7) for \( i = 1:L \)
   for \( j = 1:H \)
   \( X_{\text{cur}+1} = \alpha + \beta X_{i,j} \)
8) Add \( X_{\text{cur}+1} \) to train_ period and set \( X_{\text{cur}} \) to \( X_{\text{cur}+1} \)
9) Set Candidate_patterns = \( X_i \) where \( i = 1, 2, 3 \ldots n \)
10) Repeat steps 5 to 9 until all values in test_ period (TS) are generated.

## 6. Experimentation and Results

The importance of real exchange rate time series is characterized with high volatility while Mackey-Glass(MG) time series changes direction slowly. And MG series volatility curve has same regular patterns whereas the exchange rate volatility curve does not have any regular pattern. Hence MG time series prediction tasks cannot fully reflect the forecasting ability and not predominantly used as benchmark problem in the fields of economics and finance. There are a large number of publications reported in \cite{14,15} dealing with the new approach.
With issue of exchange rate forecasting. However, different publications use different data sets and experimental set-ups. To make the comparison with reported works, we have carried out experiments on Japanese yen, Canadian and Australian dollar exchange rate series as defined in Chong Tan [15] to evaluate the robustness of the proposed method.

Table 2: Performance Results for Japanese Yen, Canadian and Australian Dollar rate series

<table>
<thead>
<tr>
<th>Filter</th>
<th>Steps L</th>
<th>Train</th>
<th>Fore</th>
<th>MSE</th>
<th>RMSE</th>
<th>MAPE</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>1</td>
<td>100</td>
<td>100</td>
<td>0.3891</td>
<td>0.0238</td>
<td>0.0034</td>
</tr>
<tr>
<td></td>
<td>6</td>
<td>126</td>
<td>126</td>
<td>0.0194</td>
<td>0.1392</td>
<td>0.00085</td>
</tr>
<tr>
<td></td>
<td>6</td>
<td>126</td>
<td>126</td>
<td>0.2760</td>
<td>0.5258</td>
<td>0.0025</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>1</td>
<td>100</td>
<td>100</td>
<td>0.0000031</td>
<td>0.00056</td>
<td>0.0030</td>
</tr>
<tr>
<td></td>
<td>6</td>
<td>100</td>
<td>100</td>
<td>0.0000044</td>
<td>0.00066</td>
<td>0.0035</td>
</tr>
<tr>
<td></td>
<td>1</td>
<td>100</td>
<td>100</td>
<td>0.00000396</td>
<td>0.00222</td>
<td>0.0012</td>
</tr>
<tr>
<td></td>
<td>6</td>
<td>1000</td>
<td>1000</td>
<td>0.00000353</td>
<td>0.0020</td>
<td>0.0010</td>
</tr>
<tr>
<td></td>
<td>6</td>
<td>1000</td>
<td>1000</td>
<td>0.00000353</td>
<td>0.0025</td>
<td>0.0022</td>
</tr>
<tr>
<td></td>
<td>1</td>
<td>1000</td>
<td>1000</td>
<td>0.00000353</td>
<td>0.0025</td>
<td>0.0022</td>
</tr>
</tbody>
</table>

Figures 1 and 2 show the forecast responses for Japanese yen rate series without filter and with db3 filter for L=1 step ahead prediction considering 100 for training and 100 points for forecasting respectively. It is evident from the table 2 and figures 1 and 2 that prediction with db3 filter results in good performance with MAPE of 0.0015, whereas without filter a MAPE of 0.0044.

Figures 3 and 4 show the forecast responses with filter db3 and L=6 and for db1 and L= 6 considering 126 points each for training and forecasting respectively. For this rate series, db3 wavelet filter shows better prediction with MAPE value of 0.00085 than db1 filter with MAPE of 0.0025.

Forecasting for Japanese Yen rate series was done considering various compositions of data points and adopting different wavelet filters in the preprocessing stage. We have achieved considerably good forecasting performance for Japanese Yen rate series with and without filters and better performance with adopting filters.

Figures 5 and 6 show the forecast responses for Canadian dollar rate series without filter for L=1 and L=6 steps, considering 100 points each for training and forecasting respectively. It is observed from the table 2 and figures 5 and 6 that there is no significant improvement in prediction performance and also with MAPE values tabulated. The difference in MAPE for L=1 and 6 steps are negligible.

Figure 7 shows the forecast response with filter db3 and...
L=1 considering 100 data points for training and forecasting respectively. Figure 8 shows the forecast response with filter db3 and L=6 considering 1000 points each for training and forecasting. Observations based on the responses obtained for this rate series indicate good forecasting performance for various compositions of data points. The results obtained on experiments are tabulated in table 2.

Figures 9 and 10 show the forecast responses for Australian dollar rate series without and with db3 filter for L=6 steps ahead prediction considering 100 points for training and forecasting respectively. It is evident from the table 2 and figures 9 and 10 that prediction with db3 filter results in good performance with MAPE being 0.0012, whereas without filter MAPE being 0.0055.

Figures 11 and 12 show the forecast responses with db3 filter and L=6 and L=1, considering 1000 points each for training and forecasting respectively. For this dollar rate
series, from results in table 2 it is observed that with db3 wavelet filter for L=6 steps shows better prediction with MAPE of 0.0015 than for L=1 with MAPE of 0.0022.

Forecasting for Australian dollar rate series was done considering various compositions of data points and adopting db3 wavelet filters in the preprocessing stage and achieved considerably good forecasting performance.

Table 3 shows the comparative prediction performance results for Japanese Yen, Canadian and Australian dollar rate series obtained for the proposed method with and without filters with the approaches from [15].

From observations of results on comparative performance, the proposed hybrid model provides far more accurate forecasts than the approaches reported in [15] and this method with and without filters adaptation has also resulted in lesser Root mean square error(RMSE).
Table 3: Comparative Prediction performance

<table>
<thead>
<tr>
<th>Method</th>
<th>RMSE</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Japan</td>
</tr>
<tr>
<td>NN+Wavelet Denoising</td>
<td>3.3492</td>
</tr>
<tr>
<td>NN+Wavelet Packet Denoising</td>
<td>4.1822</td>
</tr>
<tr>
<td>New Method(DWT)</td>
<td>0.5182</td>
</tr>
<tr>
<td>without Statistical Feature</td>
<td>1.1507</td>
</tr>
<tr>
<td>New Method(SWT)</td>
<td>0.9987</td>
</tr>
<tr>
<td>Proposed IIBL</td>
<td>0.07146</td>
</tr>
<tr>
<td>Proposed WIIBL with db3</td>
<td>0.1392</td>
</tr>
</tbody>
</table>

* values are taken from [15]

7. Conclusions

Integrated multi-step prediction systems for predicting the exchange rate time series has been proposed. The work demonstrates the feasibility of integrating with daubechies wavelet filters at the preprocessing stage to achieve accurate forecasting. The proposed work incorporates a hybrid distance measure combining correlation and Euclidean distance to select similar instances from the stored patterns. WIIBL has been motivated by the effective preprocessing capability of wavelet filters and the predictive power of improved instance based learning system, to represent a hybrid prediction system. The multiscaling property of the wavelet transform enhances the prediction with high accuracy for volatile time series. To observe the impact of size of dataset used and also to analyse the learning characteristics of the proposed model, experiments were conducted considering various partition size of data set for training and testing input sample values. The most important conclusion of this study is that the proposed WIIBL is a useful tool for forecasting exchange rate series behaviour, with use of appropriate filters during preprocessing for accurate predictions.

References


