A Wavelet Based Prediction Method for Time Series

Cristina Stolojescu
Alexandru Isar
Politehnica University
Timisoara, Romania

Ion Railean
Technical University
Cluj-Napoca, Romania

Sorin Moga
Philippe Lenca
Institut TELECOM,
TELECOM Bretagne, France

Stochastic Modeling Techniques and Data Analysis International Conference,
Chania, Crete.
8 - 11 June 2010

Contents

Subject

- **Wavelet-based forecasting method for time series.**
- Multi-resolution decomposition of the signal implemented using the Stationary Wavelet Transform.
- four prediction methods:
  - Artificial Neural Networks
  - ARIMA
  - Linear regression
  - Random walk.

Context of study

- Time-series forecasting
- Two types of real data series:
  - WiMAX network traffic data.
    - 8 weeks of complete data
    - 67 Base Stations
  - financial data- the total number of EUR-USD currency exchanges.
    - 15 weeks of complete data
    - 1 data set

Objectives

- to evaluate the best prediction method.
- to evaluate the prediction accuracy by using different types of mother wavelets.

Introduction

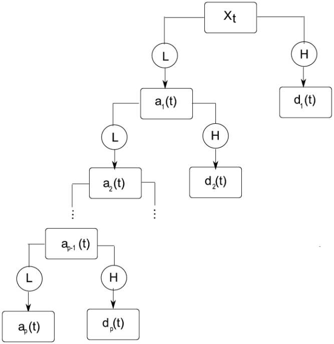
- **Forecasting** = the process of estimation in unknown situations based on the study of the past data behavior over time, in order to take decisions.
- **Time-series forecasting** : the historical values are collected and analyzed in order to develop a model describing the behavior of the series.
- If the time series are non-stationary an efficient way is to use the wavelet decomposition technique in the preprocessing step.

Wavelet Analysis

- Wavelets provide a useful decomposition of the time series, in terms of both time and frequency.
- Wavelets are localized in time (or space) which makes them suitable for analyzing non-stationary signals.
- Multi-resolution analysis (MRA) = signal processing technique that takes into account the signal's representation at multiple time resolutions.
- We have implemented the MRA using à trous algorithm, which corresponds to the computation of the Stationary Wavelet Transform (SWT).

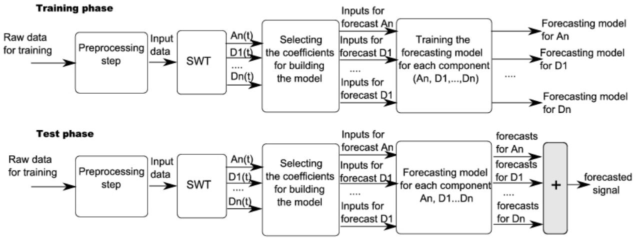
The Stationary Wavelet Transform

- two parameters:
  - the mother wavelet which generates the decomposition.
  - the number of decomposition levels.
- The SWT decomposition tree:



Forecasting framework

Our forecasting framework is presented in the next figure:



Forecasting methodologies

- **Artificial Neural Networks.**
  - the basic idea is to train the ANN with past data and then to use it to predict future values.
  - we used Time-delayed neural networks(TDNN).
- **ARIMA processes.**
  - are the natural generalizations of ARMA processes.
  - are based on Box-Jenkins methodology.
- **Linear Regression.**
  - represent a simple statistical tool for modeling the output as a linear combination of the input.
- **Random walk.**
  - are based on the hypothesis that from one period to the next, the original time series takes a random “step” away from its last recorded position.

Evaluation criteria

X_t - original data values, $\overline{X_t}$ - mean of X_t, F_t - predicted values
To evaluate the prediction performance between ANNs and genetically optimized ANNs, we used the following evaluation criteria:

Evaluation criteria

X_t - original data values, $\overline{X_t}$ - mean of X_t, F_t - predicted values
To evaluate the prediction performance between ANNs and genetically optimized ANNs, we used the following evaluation criteria:

- Symmetrical Mean Absolute Percentage Error (SMAPE)

$$SMAPE = \frac{2}{n}\sum_{t=1}^{n}\frac{|X_t - F_t|}{(X_t + F_t)}$$

Evaluation criteria

X_t - original data values, $\overline{X_t}$ - mean of X_t, F_t - predicted values
To evaluate the prediction performance between ANNs and genetically optimized ANNs, we used the following evaluation criteria:

- Symmetrical Mean Absolute Percentage Error (SMAPE)

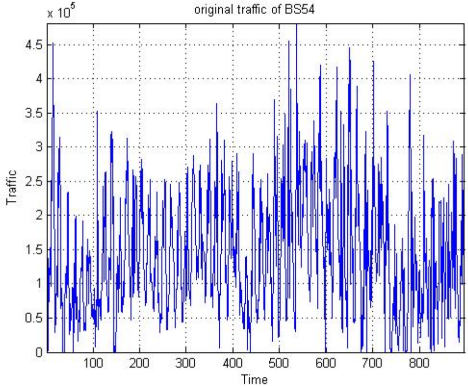
$$SMAPE = \frac{2}{n} \sum_{t=1}^{n} \frac{|X_t - F_t|}{(X_t + F_t)}$$

- R-Square (RSQ)

$$R^2 = \frac{\sum_{t=1}^{n} (F_t - \overline{F_t})^2}{\sum_{t=1}^{n} (X_t - \overline{X_t})^2}$$

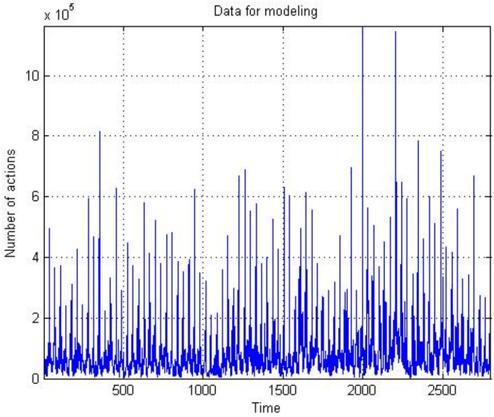
Original data

- Wimax traffic data
  - 8 weeks of complete data
  - 67 Base Stations
    - up-link and down-link (bytes and packets)
    - sampled each 15 minutes



Original data

- financial data
  - 15 weeks of complete data
    - sampled each 15 minutes



Comparison between wavelets, WiMAX traffic.

- wavelet families: Daubechies, Coiflet, Symmlet, Biorthogonal and Reverse Biorthogonal.

Wavelet	RSQ	SMAPE	MAPE	MSE	RMSE	MAE	SMAPE L	MAPE L	MAE L
coif 1	1.445	1.09	**0.2113**	11.72	2.80	1.0304	0.890	0.0020	0.9599
coif 2	1.493	1.22	0.2285	12.95	2.83	0.8748	0.837	0.0019	**0.7191**
db 1	**1.168**	**1.08**	0.2367	**8.06**	**2.43**	**0.7685**	**0.812**	**0.0016**	0.7327
db 2	1.364	1.15	0.2451	10.52	2.69	0.8408	0.855	0.0019	0.7768
db 3	1.358	1.12	**0.2117**	**9.76**	**2.64**	0.8193	0.857	0.0018	0.7678
db 4	1.490	1.11	0.2159	10.61	2.58	0.7985	0.834	0.0018	0.7563
db 5	1.435	1.11	0.2190	12.56	2.75	0.8339	0.823	0.0019	0.7730
bior 3.1	0.695	1.13	0.3152	9.86	2.52	84020	0.860	0.0018	**0.7071**
rbio 1.1	**1.200**	**1.08**	0.2215	10.00	2.61	**0.7948**	**0.820**	**0.0017**	0.8947
rbio 2.2	1.482	1.19	0.3202	10.29	2.71	0.8747	0.891	0.0018	0.7690
rbio 3.3	1.952	1.21	0.2623	10.33	2.88	0.9509	0.907	0.0022	1.0690
sym 2	1.365	1.26	0.2146	13.20	2.89	0.8854	0.895	0.0019	0.7412

Table: Comparison between wavelets, WiMAX traffic.

Comparison between wavelets, financial data

Wavelet	RSQ	SMAPE	MAPE	MSE	RMSE	MAE	SMAPE L	MAPE L	MAE L
coif 1	0.688	0.556	0.1152	1.627	1.220	0.3586	0.516	0.0821	0.4240
coif 2	0.455	0.522	**0.0793**	**1.089**	**1.038**	**0.2967**	**0.453**	0.0732	0.3799
db 1	0.625	0.520	**0.0839**	1.356	1.126	0.3175	0.454	**0.0713**	**0.3690**
db 2	0.715	0.578	0.1088	1.610	1.219	0.3550	0.497	0.0812	0.4199
db 3	0.586	0.585	0.1156	1.499	1.188	0.3618	0.531	0.0864	0.4461
db 4	**0.871**	0.600	0.1114	1.604	1.239	0.3745	0.527	0.0863	0.4459
db 5	**0.808**	0.587	0.1121	1.557	1.225	0.3700	0.546	0.0912	0.4710
bior 3.1	0.628	0.534	0.1137	1.552	1.173	0.3390	0.433	**0.0712**	**0.3681**
rbio 1.1	0.615	**0.519**	0.0958	1.286	1.096	0.3208	0.457	0.0716	0.3704
rbio 2.2	0.541	0.555	0.1087	1.440	1.149	0.3406	0.491	0.0790	0.4081
rbio 3.3	0.772	0.595	0.0993	1.418	1.167	0.3480	0.455	0.0716	0.3705
sym 2	0.476	**0.499**	0.0890	**1.117**	**1.037**	**0.2962**	**0.453**	0.0736	0.3813

Table: Comparison between wavelets, EUR-USD currency exchanges.

Results

- We propose the following methods for two weeks prediction:
  - three types of methods based on ANN
    - ANN No Sliding
    - ANN Known Sliding
    - ANN UnKnown Sliding
  - ARIMA
  - Linear Regression
  - Random walk

Forecasting techniques comparison

- In the case of WiMAX traffic, the comparison was made using Daubechies1 mother wavelets.

Forecasting Model	SMAPE L	MAPE L	MAE L
ANN No Sliding	**0.472**	0.0011	0.4428
ANN Known Sliding	0.509	**0.0009**	0.4241
ANN UnKnown Sliding	0.722	0.0017	0.6681
ARIMA	0.772	0.0027	0.9990
Linear Regression	0.523	0.0031	**0.3868**
Random Walk using Wavelets	4.440	0.0030	1.3633

Table: Forecasting techniques comparison for WiMAX traffic.

Forecasting techniques comparison

- For the financial data base, we used the Coiflet2 mother wavelets.

Forecasting Model	SMAPE L	MAPE L	MAE L
ANN No Sliding	0.169	0.020	0.1079
ANN Known Sliding	**0.153**	**0.0178**	**0.957**
ANN UnKnown Sliding	0.267	0.0344	0.1812
ARIMA	1.135	0.3245	1.7054
Linear Regression	0.191	0.0243	0.1109
Random Walk using Wavelets	0.940	0.2670	1.4173

Table: Forecasting techniques comparison for financial data.

Conclusions

- Regarding the Wavelet transform, our results show that Haar (Daubechies1) and Reverse biorthogonal 1 improve the performance of the prediction technique.
- as much the support of the mother wavelets increases, the performance of the wavelet transform deteriorates
- ANN outperforms the other forecasting methods.
- Our model can be used for different time series.

Thank you for your attention!