Multi-Resolution Time Series Forecasting Using Wavelet Decomposition

N. J. de Vos

***Abstract*—A forecasting approach based on wavelet decomposition is presented. Separate forecasting models for the decomposed signals are built, after which a final model combines the various forecasts. An ensemble approach is followed in which multi-layer perceptrons, radial basis function networks and support vector regression models are used.**

I. Introduction

Univariate time series forecasting requires an adequate extraction of information from data in order to be able to make accurate and reliable forecasts. Identification of noise, seasonality and trends, for example, can therefore significantly improve the effectiveness of forecasting models. Methods that combine analysis of the time and frequency domain are proven to be effective in retrieving additional information. It is for this reason that a combination of wavelet decomposition and computational intelligence forecasting techniques has been tested in the NN3 Neural Forecasting competition.

II. Methods

A. Wavelet Decomposition

Wavelet analysis can be used to analyze a time series in both time and scale (i.e., frequency). For mathematical details of the wavelet decomposition procedure, the reader is referred to [1]. In short, the procedure of wavelet decomposition is as follows. For each level of decomposition, convolutions with a discrete low-pass filter for approximation and a high-pass filter for detail are made, followed by downsampling of both results.

The wavelet that was used is the symlet of order 12 (see [2] for more information). To avoid border distortion, symmetric replication of the boundary values was applied. A three-level decomposition procedure was followed, resulting in three detail time series and a residual approximation time series.

B. Forecasting Approach

After the decomposition of the original series in three detail signals and an approximation signal, separate forecasting models are formulated for each of these signals (an idea after [3] and [4]). Each one of these forecast models only use lagged inputs from the current signal as model input. A fifth model was formulated that makes forecasts using previous values of the time series only. The five forecasts serve as input to a final model that makes the final forecast. An overview of this modeling approach is presented in Fig. 1.

A single cross-validation approach was taken by separating the last 18 values of the original time series. Since these were not used for training, they could be used for testing the accuracy of the various forecasting models. This approach has been used for determining various parameters of the different forecasting techniques discussed in the following subsections. The objective function that was used for the cross-validation procedure was the Scaled Mean Absolute Percentage Error (SMAPE).

The inputs to all forecast models were chosen to be the five lag times for which the linear correlation was highest. The input to all aggregation models consisted of the single last forecasts of each of the forecast models.

The final step in the forecasting approach is to make a weighted average of the two best-performing models. These weights are determined by the performance on the cross-validation data.

C. Multi-Layer Perceptron networks

The MLP models that were used have one hidden layer. The MLP models for all time series have 3 hidden neurons, a number that was determined using a trial-and-error approach. The transfer function in the neurons was chosen to be the logistic sigmoid function. The Levenberg-Marquardt training algorithm was used in combination with the mean absolute error objective function to train the MLP. A regularization component was added to the objective function to avoid overfitting of the network. Because of sensitivity to the initialization of the MLP weights, the training algorithm sometimes gets stuck in local minima. For this reason, the training during cross-validation was repeated 10 times, after which mean values of the SMAPE were taken.

D. Radial Basis Function networks

RBF networks are variants of artificial neural networks that emerged during the 1980s. For more details on these networks, the reader is referred to [5].

When defining and training the RBF network models, the two main parameters that needed to be chosen were the spread of the RBF and the maximum number of hidden neurons. By simulating a grid of possible values for these parameters, the optimal values were chosen for each of the 111 time series.

N. J. de Vos is with the Water Resources Section of Delft University of Technology, P.O. Box 5048, 2600GA, Delft, The Netherlands (phone: 0031152786857; fax: 0031152785559; e-mail: n.j.devos@ tudelft.nl).

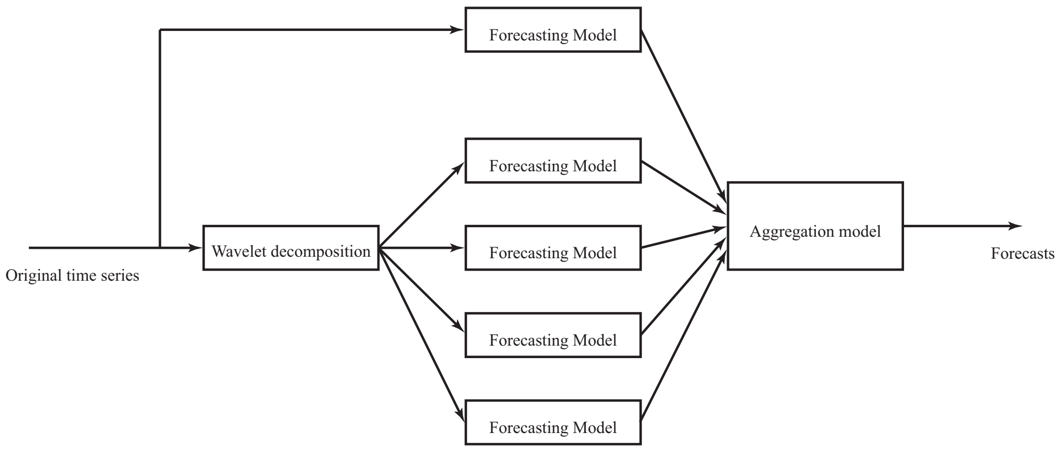

Fig. 1. Overview of the forecast approach.

E. Support Vector Regression models

In [6], a neural network variant called support vector machines was introduced, which were originally applied to pattern recognition problems. SVR models [7] are variants that are capable of dealing with nonlinear regression problems.

So-called epsilon-SVR models were used with a radial basis function kernel type. Two parameters needed to be defined for these models: the cost parameter C and the γ parameter in the kernel function. Similarly to the RBF networks, a grid of possible combinations of the two model parameters was simulated, resulting in approximate optimal values for each of the 111 time series.

III. Summary

Within the framework of the NN3 Neural Forecasting competition, a forecasting approach based on wavelet decomposition and subsequent model ensemble forecasting was developed. The general idea behind this approach is to decompose the original time series into signals of different resolutions (i.e., from high-frequency noise to low frequency trends) using a wavelet decomposition, and to separately forecast each of these signals using several techniques from the field of computational intelligence. Subsequently, an aggregating model is used to combine these forecasts into a forecast of the original time series. Three computational intelligence model types were used for the forecasting and aggregation tasks: the Multi-Layer Perceptron (MLP), the Radial Basis Function (RBF) network, and the Support Vector Regression (SVR) model. Finally, a simple weighting approach was taken to combine the forecasts of the MLP, RBF and SVR into a final forecast.

References

[1] S. G. Mallat, "A theory for multiresolution signal decomposition: the wavelet representation," *IEEE Trans. Pattern Anal.*, vol. 11, no. 7, pp. 647–693, July 1989.

[2] I. Daubechies, *Ten Lectures on Wavelets*, Society for Industrial and Applied Mathematics, 1992.

[3] B. L. Zhang, R. Coggins, M. A.Jabri, D. Dersch, B. Flower, "Multiresolution forecasting for futures trading using wavelet decompositions," *IEEE Trans. Neural Networks*, vol. 12, no. 4, pp. 765–775, July 2001.

[4] A. Aussem, J. Campbell, F. Murtagh, "Wavelet-based feature extraction and decomposition strategies for financial forecasting," *J. Computational Intelligence in Finance*, vol. 6, no. 2, pp. 5–12, March/April 1998.

[5] J. E. Moody, C. Darken, "Fast learning in networks of locally-tuned processing units," *Neural Computation*, 1, pp. 281–294, 1989.

[6] V. N. Vapnik, *The Nature of Statistical Learning Theory*, Springer-Verlag, New York, 1995.

[7] V. N. Vapnik, S. E. Golowich, A. J. Smola, "Support vector method for function approximation, regression estimation, and signal processing," *Advance in Neural Information Processing Systems*, vol. 9, pp. 281–287, 1996.