Available online at www.sciencedirect.com
ScienceDirect
International Journal of Forecasting 27 (2011) 700–707
international journal of forecasting
www.elsevier.com/locate/ijforecast

Forecasting the NN5 time series with hybrid models

Jörg D. Wichard*

Leibniz-Institut für Molekulare Pharmakologie (FMP), Molecular Modeling Group, Robert-Roessle-Str. 10, 13125 Berlin, Germany

Available online 21 May 2010

Abstract

We propose a simple way of predicting time series with recurring seasonal periods. Missing values of the time series are estimated and interpolated in a preprocessing step. We combine several forecasting methods by taking the weighted mean of forecasts that were generated with time-domain models which were validated on left-out parts of the time series. The hybrid model is a combination of a neural network ensemble, an ensemble of nearest trajectory models and a model for the 7-day cycle. We apply this approach to the NN5 time series competition data set.

Keywords: Forecasting competitions; Combining forecasts; Nonlinear time series; Seasonality; Neural networks

1. Introduction

Time series forecasting is a growing field of interest, with applications in nearly all fields of science. There is a lot of interest in building models that can give an indication of trends and changes in the economy or in climatology. Seasonal time series are of particular interest, because the recurring pattern could be used to build predictive models.

The NN5 time series competition (Crone, 2008) includes daily time series of cash money withdrawals from cash machines at different locations. All time series show a strong 7-day cycle, as well as some recurring seasonal periods such as the summer holidays or the changing date of Easter. Almost all time series have some missing values, and some time series show long term trends or remarkable irregularities such as bursts or gaps. Before we train our forecasting model, we perform a preprocessing step wherein we close the gaps using a simple but effective interpolation scheme.

The objective of the competition was to predict the next 56 days for each time series. The day of the year was provided as additional information, indicating that the forecasting period includes an Easter event.

We suggest a hybrid strategy in order to cope with the different seasonal features of the time series. We train three different types of models and combine the individual forecasts. The motivation for this approach is based on the observation that hybrid models are both more accurate and more stable than typical individual forecasts. In an empirical study, Armstrong (2001) pointed out that combining forecasts derived from different methods could improve the forecast accuracy.

* Tel.: +49 30 94793 279; fax: +49 30 94793 169.
E-mail addresses: wichard@fmp-berlin.de, joergwichard@web.de.

0169-2070/$ - see front matter
doi:10.1016/j.ijforecast.2010.02.011

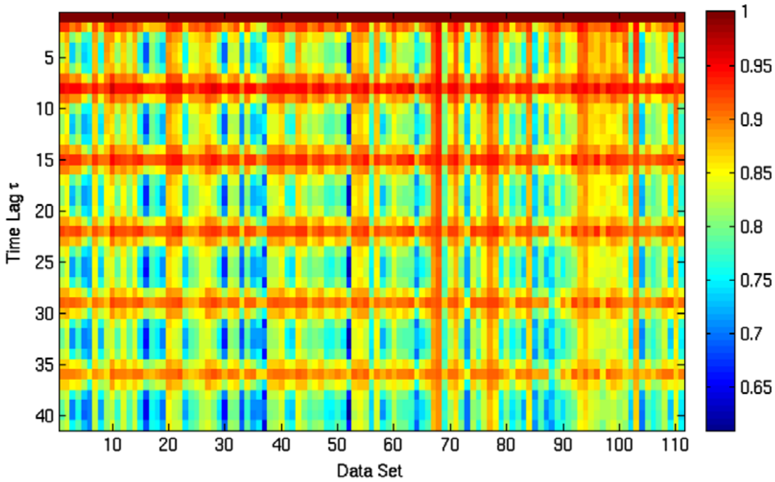

Fig. 1. Color coded values of the autocorrelation function for the 111 data sets with a time lag τ. The 7-day cycle is indicated by the periodicity of the autocorrelation function.

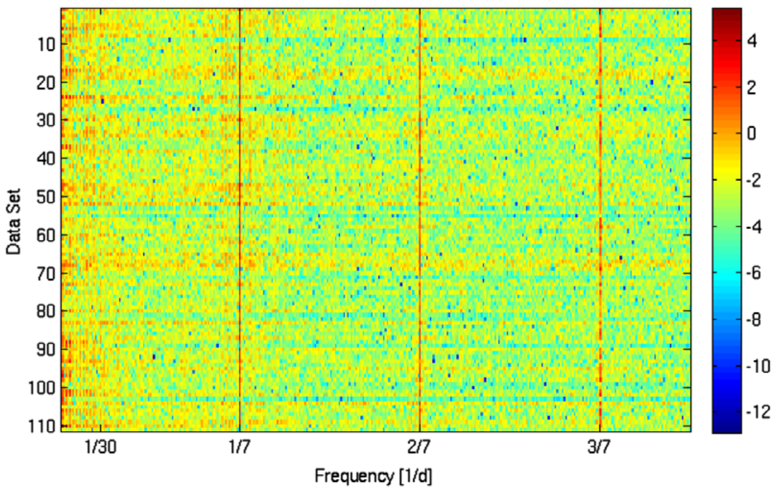

Fig. 2. Color coded values of the logarithmic power spectrum for all 111 data sets. The 7-day peak is presented for all data sets, together with the high order harmonics.

Our contribution to the NN5 time series competition came sixth out of 19 competitors among the computational intelligence methods and ninth among 27 competitors overall (Crone, 2008).

2. Data and preprocessing

The data are provided on the NN5 competition website (Crone, 2008). The task is to forecast a set of 111 daily time series of cash money withdrawals from cash-machines at different locations in England. The time series consist of 2 years of daily cash money demand, and the forecasting horizon is 56 days. All of the time series show a weekly seasonality, as is indicated by the periodicity of the autocorrelation function in Fig. 1. Fig. 2 shows the logarithmic power spectra for the time series. They all share the dominating 7-day peak and the corresponding high order harmonics. We also expected to find a day-of-the-month periodicity, but only a few of the time series seem to have the characteristic peak in this part of the power spectrum (see Fig. 2).