Thesis - Temporal template

Oleg Ostashchuk

March 18, 2016

1 Introduction

1.1 Times Series and its Analysis

Time series is a common mathematical expression that can be observed in many fields. Very often it can be read in texts about statistics, signal processing or finance. The term denotes a data storing format, which consists of a sequence of values recorded over some time intervals. Values should represent the same meaning and there can be at most one value for each time unit. Therefore, not all sequences of values belongs to time series format.

- Time series examples: Sequence of daily sales for one specific product.
- Non–times series examples: Sequence of sales for group of products.

Time series analysis is a set of methods for analyzing time series data, the main goal of which is to understand the behavior or structure of data points, in order to extract potentially useful information. Time series forecasting is one of the most important analysis, performed over the time series data. The general idea is based on the fact that information about the past events can be exploited to build a forecasts about the future events. In the case of time series data, this means that forecasting models use already observed values to predict future values before they are measured. A typical example is prediction of the product's future price, based on the last price changes.

There is a lot of processes, where people collect big amounts of data, which are often stored in time series format. Usually, it could be different technical measurements from sensors, performance or load measurements. Another large group represents different economic indicators like electricity prices, number of sales, stock prices, interest ratings. Forecasting of all these mentioned processes is an important task.

Today, in the age of computers, the abilities of time series forecasting had increased. It is believed and observed that computers can make predictions much more precisely than humans. There have also been invented new forecasting methods, based on machine learning algorithms.

1.2 Aim and Structure of Thesis

The main aim of this thesis is to compare effectiveness of specific time series forecasting methods on the various practical test data from different fields. The specified aim is going to be reached by fulfilling following steps:

1. Make a research on time series forecasting methods and choose the most perspective methods.
2. Select practical sets of test data from different fields.
3. Implement forecasting models for each combination pair of test data x method.
4. Perform tests and compare prediction quality of methods for each set of test data.
5. Make a conclusion about methods' effectiveness for concrete set of test data and opportunities of improvements.

2 Forecasting - formal definition

2.1 Formal task definition

Let's assume that observations of some specific process are available at discrete units of time $t = 1, 2, \ldots,$ T. Then the sequence of values $Z_t = Z_1, Z_2, \ldots, Z_T$ is denoted as a time series.

At the moment of time unit T, it is necessary to make a forecast of L future values of the given process Z. In other words, it is needed to forecast time series values for each time unit $(T+1), (T+2), \ldots\ (T+L)$. Time unit T is a moment, when the forecast is performed and it is called **origin**. The parameter L is denoted as a **lead time** and it represents the amount of future values to be forecasted.

In order to calculate the time series values for future time units, it is necessary to denote functional dependency that describes a relationship between past and future values of the given time series.

$$\begin{bmatrix} Z_{T+1} \\ Z_{T+2} \\ \vdots \\ Z_{T+L} \end{bmatrix} = f(Z_T, Z_{T-1}, Z_{T-2}, ...) + \begin{bmatrix} e_{T+1} \\ e_{T+2} \\ \vdots \\ e_{T+L} \end{bmatrix}$$

The functional dependency is usually called a forecast function or also a forecast model. The objective is to find forecast function such that the mean square of differences between

the actual and forecasted values is as small as possible for each lead time l.

$$E = \frac{1}{L} \sum_{t=T+1}^{T+L} e_t^2 \rightarrow min$$

In addition to calculations of future values, sometimes it is also good to specify the limits of accuracy. The accuracy of the forecasts may be expressed by calculating probability limits on either side of each forecast.

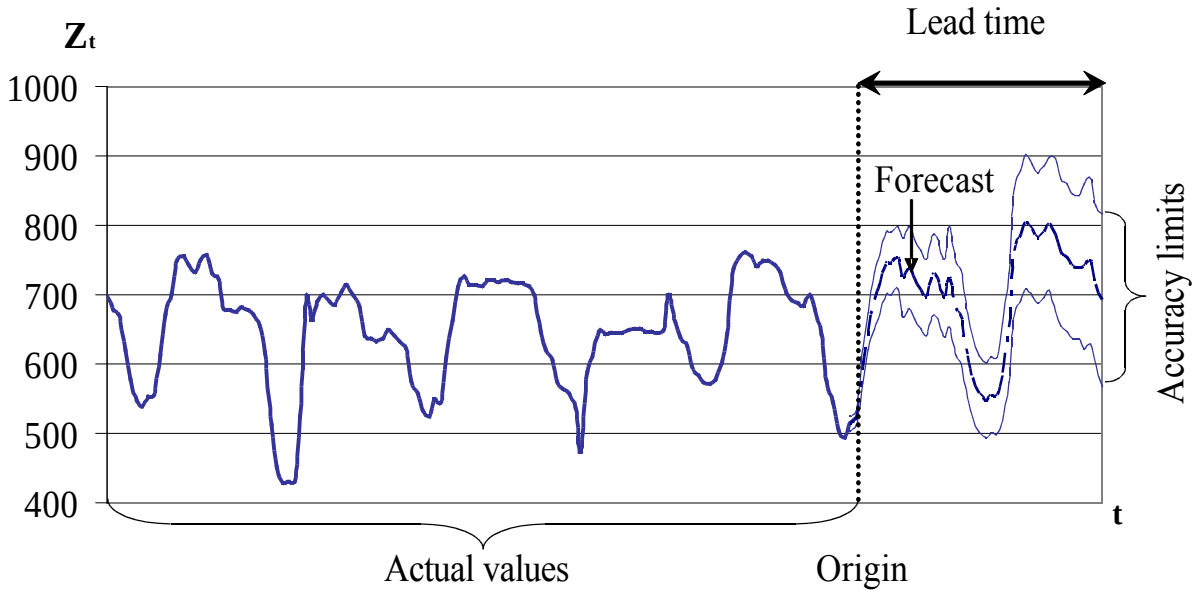

Figure 1: Time series forecasting example