Machine Learning Strategies for Time Series Forecasting

Gianluca Bontempi, Souhaib Ben Taieb, and Yann-Aël Le Borgne

Machine Learning Group
Computer Science Department, Faculty of Sciences
ULB, Université Libre de Bruxelles
Bd Triomphe, 1050, Brussels, Belgium
{gbonte,sbentaieb,yleborgn}@ulb.ac.be
http://mlg.ulb.ac.be

Abstract. The increasing availability of large amounts of historical data and the need of performing accurate forecasting of future behavior in several scientific and applied domains demands the definition of robust and efficient techniques able to infer from observations the stochastic dependency between past and future. The forecasting domain has been influenced, from the 1960s on, by linear statistical methods such as ARIMA models. More recently, machine learning models have drawn attention and have established themselves as serious contenders to classical statistical models in the forecasting community. This chapter presents an overview of machine learning techniques in time series forecasting by focusing on three aspects: the formalization of one-step forecasting problems as supervised learning tasks, the discussion of local learning techniques as an effective tool for dealing with temporal data and the role of the forecasting strategy when we move from one-step to multiple-step forecasting.

Keywords: Time series forecasting, machine learning, local learning, lazy learning, MIMO.

1 Introduction

A *time series* is a sequence S of historical measurements y_t of an observable variable y at equal time intervals. Time series are studied for several purposes such as the forecasting of the future based on knowledge of the past, the understanding of the phenomenon underlying the measures, or simply a succinct description of the salient features of the series. In this chapter we shall confine ourselves to the problem of forecasting. Forecasting future values of an observed time series plays an important role in nearly all fields of science and engineering, such as economics, finance, business intelligence, meteorology and telecommunication [43]. An important aspect of the forecasting task is represented by the size of the horizon. If the one-step forecasting of a time series is already a challenging task, performing multi-step forecasting is more difficult [53] because of

M.-A. Aufaure and E. Zimányi (Eds.): eBISS 2012, LNBIP 138, pp. 62–77, 2013.

additional complications, like accumulation of errors, reduced accuracy, and increased uncertainty [58,49].

The forecasting domain has been influenced, for a long time, by linear statistical methods such as ARIMA models. However, in the late 1970s and early 1980s, it became increasingly clear that linear models are not adapted to many real applications [25]. In the same period, several useful nonlinear time series models were proposed such as the bilinear model [44], the threshold autoregressive model [56,54,55] and the autoregressive conditional heteroscedastic (ARCH) model [22] (see [25] and [26] for a review). However, the analytical study of nonlinear time series analysis and forecasting is still in its infancy compared to linear time series [25].

In the last two decades, machine learning models have drawn attention and have established themselves as serious contenders to classical statistical models in the forecasting community [1,43,61]. These models, also called black-box or data-driven models [40], are examples of nonparametric nonlinear models which use only historical data to learn the stochastic dependency between the past and the future. For instance, Werbos found that Artificial Neural Networks (ANNs) outperform the classical statistical methods such as linear regression and Box-Jenkins approaches [59,60]. A similar study has been conducted by Lapedes and Farber [33] who conclude that ANNs can be successfully used for modeling and forecasting nonlinear time series. Later, other models appeared such as decision trees, support vector machines and nearest neighbor regression [29,3]. Moreover, the empirical accuracy of several machine learning models has been explored in a number of forecasting competitions under different data conditions (e.g. the NN3, NN5, and the annual ESTSP competitions [19,20,34,35]) creating interesting scientific debates in the area of data mining and forecasting [28,45,21].

This chapter aims to present an overview of the role of machine learning techniques in time series forecasting by focusing on three aspects: the formalization of one-step forecasting problems as supervised learning tasks, the discussion of local learning techniques as an effective tool for dealing with temporal data and the role of the forecasting strategy when we move from one-step to multi-step forecasting.

The outline of the chapter is as follows. Section 2 introduces some basic notions of time series modeling and the formalization of the forecasting task as an input-output problem. Section 3 discusses the role of machine learning techniques in inferring accurate predictors from observed data and introduces the local learning paradigm. Section 4 presents several strategies for multi-step forecasting which have been proposed so far in literature. Section 5 reviews how local learning techniques have been integrated with multiple-step strategies to perform accurate multi-step forecasts.

2 Forecasting and Modeling

Two main interpretations of the forecasting problem on the basis of historical dataset exist. Statistical forecasting theory assumes that an observed sequence