Forecasting Big Time Series: Theory and Practice

Christos Faloutsos
CMU and Amazon Research
christos@cs.cmu.edu

Valentin Flunkert
AWS AI Labs
flunkert@amazon.com

Jan Gasthaus
AWS AI Labs
gasthaus@amazon.com

Tim Januschowski
AWS AI Labs
tjnsch@amazon.com

Yuyang Wang
AWS AI Labs
yuyawang@amazon.com

ABSTRACT

Time series forecasting is a key ingredient in the automation and optimization of business processes: in retail, deciding which products to order and where to store them depends on the forecasts of future demand in different regions; in cloud computing, the estimated future usage of services and infrastructure components guides capacity planning; and workforce scheduling in warehouses and factories requires forecasts of the future workload. Recent years have witnessed a paradigm shift in forecasting techniques and applications, from computer-assisted model- and assumption-based to data-driven and fully-automated. This shift can be attributed to the availability of large, rich, and diverse time series data sources and result in a set of challenges that need to be addressed such as the following. How can we build statistical models to efficiently and effectively learn to forecast from large and diverse data sources? How can we leverage the statistical power of “similar” time series to improve forecasts in the case of limited observations? What are the implications for building forecasting systems that can handle large data volumes?

The objective of this tutorial is to provide a concise and intuitive overview of the most important methods and tools available for solving large-scale forecasting problems. We review the state of the art in three related fields: (1) classical modeling of time series, (2) modern methods including tensor analysis and deep learning for forecasting. Furthermore, we discuss the practical aspects of building a large scale forecasting system, including data integration, feature generation, backtest framework, error tracking and analysis, etc. While our focus is on providing an intuitive overview of the methods and practical issues which we will illustrate via case studies and interactive materials with Jupyter notebooks.

KEYWORDS

Forecasting; Tensor Analysis; Neural Network; Time Series

ACM Reference Format:
Christos Faloutsos, Valentin Flunkert, Jan Gasthaus, Tim Januschowski, and Yuyang Wang. 2019. Forecasting Big Time Series: Theory and Practice. In *The 25th ACM SIGKDD Conference on Knowledge Discovery and Data Mining (KDD ’19), August 4–8, 2019, Anchorage, AK, USA.* ACM, New York, NY, USA, 2 pages. https://doi.org/10.1145/3292500.3332289

KDD ’19, August 4–8, 2019, Anchorage, AK, USA

ACM ISBN 978-1-4503-6201-6/19/08.
https://doi.org/10.1145/3292500.3332289

1 INTRODUCTION

Time series data occur naturally in countless domains including medical analysis, real estate, financial analysis, sensor network monitoring and social activity mining. Of all the time series related data mining tasks, forecasting is one of the most sought-after applications of data-driven methods (and arguably the most difficult one) due to its importance in industrial, social and scientific applications. For example, forecasting plays a key role in automating and optimizing operational processes in most businesses and enables data driven decision making. Forecasts of product supply and demand can be used for optimal inventory management, staff scheduling and topology planning, and are more generally a crucial technology for most aspects of supply chain optimization. Outside of the retail use-case, the increasing volume of online, time-stamped activities represents a vital new opportunity for data scientists and analysts to measure the collective behavior of social, economic, and other important evolutions. Facing rapidly growing data sets, the most fundamental requirements are the efficient and effective forecasting of “big” time series sequences.

Time series forecasting is a well known topic that has attracted interest in various research communities (e.g., statistics, machine learning, econometrics, operational research, databases, data mining, networking) for several decades. In the statistics and econometrics communities, the prevalent forecasting methods in use today have been developed in the setting of forecasting individual or small groups of time series with complex models designed and tuned by domain experts. Simultaneously, data mining and database researchers have been focusing on finding patterns in thousands or millions of related time series. Examples include forecasting the energy consumption of individual households, forecasting the load for servers in a data center, or forecasting the demand for all products that a large retailer offers. In these scenarios, a substantial amount of data on past behavior of similar, related time series can be leveraged for making a forecast for an individual time series. Using data from related time series not only allows fitting more complex (and hence potentially more accurate) models without overfitting, it can also alleviate the time and labor intensive human selection and preparation of co-variates and model selection steps required by classical techniques.

This tutorial aims to bring together classical forecasting techniques, time series data mining techniques, and neural forecasting methods through a concise and intuitive overview of the most important tools and techniques that we can use to help us understand

and forecast time series. We will provide a comprehensive overview of proven and current directions for time series forecasting, and deal specifically with the following key topics: (1) classical linear and non-linear modeling of time series, (2) scalable tensor methods, (3) deep learning for forecasting, and (4) practical aspects of developing large scale forecasting systems. We shall supply IPython notebooks to illustrate commonly used forecasting techniques that are covered in this tutorial.

2 OUTLINE

(1) **Introduction to Forecasting (15 mins)**
   - Basic (explanatory) analysis and decomposition of time series, i.e., trend, level, seasonality, etc.
   - Point forecast vs. probabilistic forecast
   - Forecast accuracy metrics

(2) **Classical Methods: Linear and Non-linear Models (45 mins)**
   - Linear regression
     - Parameter estimation, least squares (LS), recursive LS
     - Time series transformations, i.e., power transform, Box-Cox, etc.
   - Linear dynamical systems and exponential smoothing
     - Exponential smoothing (ES), Holt-Winters, and general Innovation state space models (ISSM)
     - ISSM with features, missing data, and different likelihood functions
   - Gaussian processes and other Bayesian time series models
   - Non-linear dynamical systems
     - lag-plots, fractal dimension, power-law, and non-linear equations
     - non-linear dynamical system for online activities
     - information diffusion in social networks

(3) **Modern Methods: Tensor Analysis and Deep Learning (90 mins)**
   - Scalable Tensor Analysis
     - Basics of matrix and tensor factorization
     - Decomposition of higher-order tensors
     - Big sparse tensors and forecasting of complex time-stamped events
   - Deep learning for forecasting
     - Multi-layer perceptron (feedforward neural networks)
     - Recurrent neural networks (RNN)s: classic, Sequence-to-Sequence and other architectures
     - Others: Convolution, Wavenet, Generative Adversarial Networks (GANs), and all that

(4) **Time Series Forecasting in Practice (30 mins)**
   - Building large scale forecasting systems for real-world problems
     - data cleaning and integration
     - feature extraction and generation
     - forecasting models and backtest components
     - visualization dashboard and forecasting accuracy tracking
   - Developing Deep Autoregressive Network (DeepAR) and other models in AWS Sagemaker and Amazon Forecast

1. Introduction to Forecasting. In the opening chapter of the tutorial, we introduce the basic forecasting concepts and terminology. The classical time series analysis tools such as time series decomposition, lag plots, autocorrelations, etc. are also introduced. We discuss how to evaluate the accuracy of a forecast with metrics such as mean absolute percentage error (MAPE) and quantile losses. In particular, we discuss the problem of assessing the quality of a *probabilistic* forecast and introduce the notion of a *proper scoring rule* and show why it is important for producing a calibrated forecast.

2. Classical Methods: Linear and Non-linear models. We next provide a comprehensive review of the classical methods for forecasting, which mainly focus on individual time series (local). First, we cover the classic linear methods for forecasting, including the linear regression, exponoential smoothing (ETS), autoregressive and moving average models (ARIMA), as well as useful tools in data management system such as MUSCLES and AWSOM. Next, we introduce non-linear dynamic models. We explain the importance of non-linear equations and the concept of gray-box non-linear mining. We also review recent work on understanding the non-linear time evolution of online user activities. Analyses of epidemics, blogs, social media, propagation and the cascades they create have attracted much interest.

3. Modern methods: Tensor Analysis and Deep Learning. In this part, we move to the territory of modern forecasting techniques, in particular with tensor methods and deep learning models. In contrast to the classical methods, the modern approaches learn across multiple related time series (global). In the first part, we present large-scale studies of complex time-stamped events and big sparse tensors. Then we move to Neural Networks (NNs) for forecasting. In the 90s, Feedforward NNs were popular among forecasters with applications in electrical load, financial time series, and others. Recent ground-breaking successes of deep neural network in other areas of machine learning have brought revived interests in applying deep learning techniques, especially recurrent neural networks and their variants, to time series forecasting. In this part, we first introduce the multi-layer perceptron (feedforward NN) as an extension of the linear regression models introduced in part 1, and then we give an overview of different types of Recurrent Neural Networks (RNNs), which capture the sequential nature of time series data. Different RNN forecasters are introduced and we explain the intuitions behind different structures and demonstrate their performances on a variety types of time series. Finally, we discuss new directions for deep generative models for forecasting, in particular, with models that combines the strengths of both RNNs and classical probabilistic graphical models.

4. Time Series Forecasting in Practice. In the final chapter of the tutorial, we shall discuss about different aspects of building a forecasting system in the real world. In particular, we shall share the lessons learned developing the scalable forecasting system for retail within Amazon and deep learning based forecasting algorithms (DeepAR) in AWS SageMaker and Amazon Forecast.