Chapter 6
Large Multivariate Time Series Forecasting:
Survey on Methods and Scalability

Youssef Hmamouche
Aix-Marseille Université, France

Hana Alouaoui
Aix-Marseille Université, France

Piotr Marian Przymus
Aix-Marseille Université, France

Alain Casali
Aix-Marseille Université, France

Lotfi Lakhal
Aix-Marseille Université, France

ABSTRACT

Research on the analysis of time series has gained momentum in recent years, as knowledge derived from time series analysis can improve the decision-making process for industrial and scientific fields. Furthermore, time series analysis is often an essential part of business intelligence systems. With the growing interest in this topic, a novel set of challenges emerges. Utilizing forecasting models that can handle a large number of predictors is a popular approach that can improve results compared to univariate models. However, issues arise for high dimensional data. Not all variables will have direct impact on the target variable and adding unrelated variables may make the forecasts less accurate. Thus, the authors explore methods that can effectively deal with time series with many predictors. The authors discuss state-of-the-art methods for optimizing the selection, dimension reduction, and shrinkage of predictors. While similar research exists, it exclusively targets small and medium datasets, and thus, the research aims to fill the knowledge gap in the context of big data applications.

DOI: 10.4018/978-1-5225-4963-5.ch006

Copyright © 2019, IGI Global. Copying or distributing in print or electronic forms without written permission of IGI Global is prohibited.

INTRODUCTION

Time series analysis and time series data mining aim to describe patterns and evolutions occurring in data over the time. Among the many useful applications of time series data mining and analysis, time series forecasting is especially salient as it contributes crucial information to corporate and/or institutional decision-making. Thus, to no surprise, is often an important part of business intelligence (BI) systems, which allow a company to gather, store, access, and analyze corporate data to aid in decision-making.

In today's information-driven world, countless numbers of numerical time series are generated by industry and researchers on any given day. For many applications -- biology, medicine, finance, industry, among others -- high dimensional time series are required. Modern time series analysis systems are expected to process and store millions of such high dimensional data points per minute, twenty-four hours a day, seven days a week, generating terabytes of logs. Needless to say, dealing with such voluminous datasets raises various new and interesting challenges.

The first models developed for time series forecasting were univariate models based on auto-regression principle. In such models, historic observations are used to make future forecasts. The most popular of these univariate models include the Auto-Regressive (AR), Auto-Regressive Moving Average (ARMA), and the Auto-Regressive Integrated Moving Average (ARIMA) models (Box, 2013). Let us detail the ARIMA model, which takes three integer parameters (p, d, q).

Where p is the lag parameter of the auto-regressive part, d is the non-stationarity order of the time series, and q is lag parameter of the moving average part.

The non-stationarity of time series is allowed with this model. Consider a time series that is non-stationary of order d.

The ARIMA (p, d, q) model consists in applying the ARMA (p, d) model after transforming the time series to stationary by differencing it d times, where d is the order of integration or non-stationarity. The ARMA (p, q) model expresses a stationary times series y(t) according to the q last error terms and the p past observations. It can be expressed as follows:

$$y\left(t\right) = \alpha_0 + \sum_{i=1}^{p} \alpha_i y\left(t-i\right) + \sum_{i=1}^{q} \beta_i \in \left(t-i\right) + \in \left(t\right)$$

where $\in\left(t\right)$ are the error terms, α_i and β_i are the parameters of the model.

Despite their advantages, univariate forecasting approaches have some drawbacks: for one, they do not take into account potentially exploitable data of other time series in the same dataset.

26 more pages are available in the full version of this document, which may be purchased using the "Add to Cart" button on the publisher's webpage: www.igi-global.com/chapter/large-multivariate-time-series-forecasting/209572

Related Content

A Novel Approach for the Customer Segmentation Using Clustering Through Self-Organizing Map
Debaditya Barmanand Nirmalya Chowdhury (2019). *International Journal of Business Analytics (pp. 23-45).*
www.irma-international.org/article/a-novel-approach-for-the-customer-segmentation-using-clustering-through-self-organizing-map/226971

Performance Measures and RTB Optimization
Wenxue Huang, Yuanyi Panand Jianhong Wu (2014). *Encyclopedia of Business Analytics and Optimization (pp. 1847-1855).*
www.irma-international.org/chapter/performance-measures-and-rtb-optimization/107373

Visualization of High Dimensional Data
Gokmen Zararsiz, Cenk Icozand Erdener Ozcetin (2014). *Encyclopedia of Business Analytics and Optimization (pp. 2653-2664).*
www.irma-international.org/chapter/visualization-of-high-dimensional-data/107444

Big Data Business Intelligence in Bank Risk Analysis
Nayem Rahmanand Shane Iverson (2015). *International Journal of Business Intelligence Research (pp. 55-77).*
www.irma-international.org/article/big-data-business-intelligence-in-bank-risk-analysis/149262

A Conceptual Model of Metadata’s Role in BI Success
Neil Foshay, Andrew Taylorand Avinandan Mukherjee (2014). *Information Quality and Governance for Business Intelligence (pp. 1-19).*
www.irma-international.org/chapter/a-conceptual-model-of-metadatas-role-in-bi-success/96142