Forecasting at scale
by Sean J. Taylor and Benjamin Letham

Summary by Konstantinos Konstantinidis

April 2021

1 Introduction

Forecasting at scale is very usual in data organization and solves may problems that the engineer/scientist may with find in front of him.This article tries to explain some of the common problems of the subject in various situations.It is divided in 5 parts that we will see below.

2 Features of Business Time Series

Business forecasting problems come in huge variety,on the other hand we can use some features to make our lives easier.Here is an overview:

Auto ARIMA, has the ability to prone to large trend errors and fail to capture any seasonality

Exponential smoothing and seasonal naive, can capture weekly seasonality, but miss longer-term seasonality

3 The Prophet Forecasting Model

The Prophet Forecasting Model is designed to handle the common features of business time series.Here are some common attributes of the model.

Designed to have intuitive parameters that can be adjusted without knowing the details of the underlying model

Decomposable time series model with three main model components: trend, seasonality and holidays Also an error term: any idiosyncratic changes which are not accomodated by the model

Similar to a generalized additive model (GAM): class of regression models with potentially non-linear smoothers applied to the regressors

Frame the forecasting problem as a curve-fitting exercise != time series models that explicitly account for the temporal dependence structure in the data

3.1 Trend Model

Saturating growth model

Piecewise linear model

Changepoints can be specified or may be automatically selected given a set of candidates. A parameter controls the flexibility of the model altering its rate

The uncertainty in the trend is measured assuming the future will see the same average frequency and magnitude of rate changes that were seen in the history

3.2 Seasonability

Fourier series to provide a flexible model of periodic effects

3.3 Holidays and Events

Analyst provide a custom list of past and future events, identified by the event or holiday's unique name

Include additional parameters for the days surrounding the holiday, essentially treating each of the days in the window around the holiday as a holiday itself

3.4 Model fitting

Benefit of a decomposable model: allow us to look at each component of the forecast separately

Default values are appropriate to most forecast problems

3.5 Analyst-in-the-Loop Modeling

Analysts making forecasts often have extensive domain knowledge about the quantity they are forecasting, but limited statistical knowledge.

Parameters to increase/decrease trend flexibility and strength of the seasonality component

4 Automating Evaluation of Forecasts

Use of baseline forecasts Modeling forecast accuracy: mean absolute percentage error (MAPE) -¿ for interpretability

4.1 Simulated Historical Forecasts

Produce K forecasts at various cutoff points in the history Simulate the errors we would have made hade we used this forecasting method at those points in the past Simple, easy to explain and relatively uncontroversial for generating insight into forecast errors Issues to be aware:

The more simulated forecasts we make, the more correlated their estimates of error are Forecasting methods can perform better or worse with more data

4.2 Identifying Large Forecast Errors

When the forecast has large errors relative to the baselines, the model may be misspecified. Analysts can adjust the trend model or the seasonality, as needed Large errors for all methods on a particular date are suggestive of outliers. Analysts can identify outliers and remove them. When the SHF error for a method increases sharply from one cutoff to the next, it could indicate that the data generating process has changed. Adding changepoints or modeling different phases separately may address the issue.

5 Conclusion

Some valuable results that we get for Forecasting at scale is that it's a simple, modular regression model that often works well with default parameters, and that allows analysts to select the components that are relevant to their forecasting problem and easily make adjustmens as needed. It also has a system for measuring and tracking forecast accuracy, and flagging forecasted that should be checked manually to help analysts make incremental improvements