Forecasting at Scale

Sean J. Taylor,Benjamin Letham

Facebook

2017

Outline

Outline

Background knowledge

Motivation

Demand for high quality forecasts often far outstrips the pace at which they can be produced.

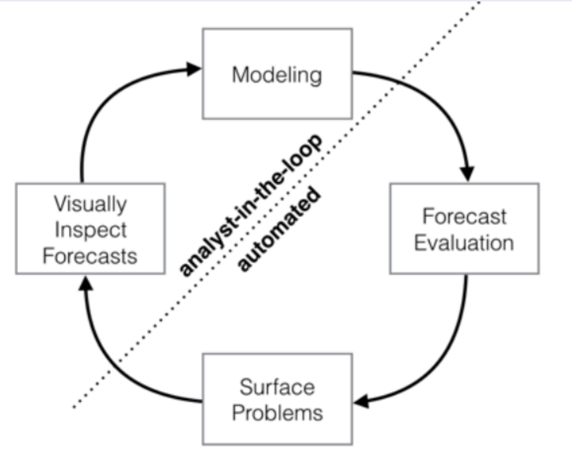
- Provide some useful guidance for producing forecasts at scale,for several notions of scale.

Three Scales try to address

- Large number of people making forecasts without training in time series methods.
- Large variety of forecasting problems with potentially idiosyncratic features
- Large number of forecasts are created, necessitating efficient, automated means of evaluating and comparing them, as well as detecting when they are likely to be performing poorly.

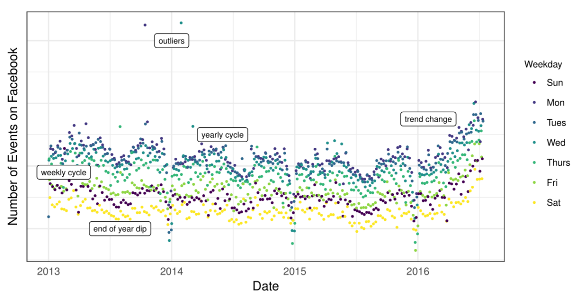
Analyst-in-the-loop approach

Analyst-in-the-loop approach



Features of Business Time Series

Features of Business Time Series

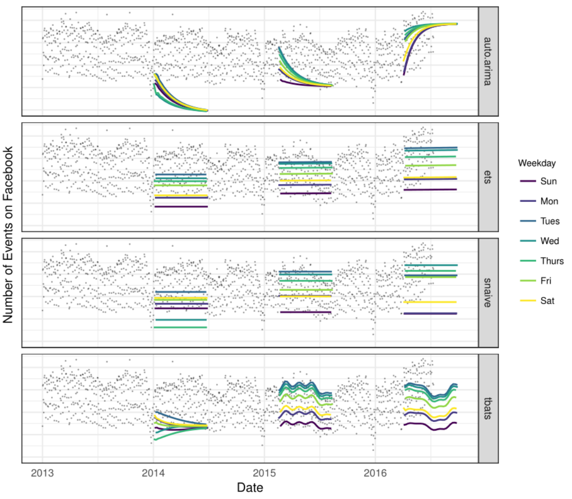


Several Features

1. Multiple strong seasonalities,weekly and yearly cycles
2. Trend changes,trend in the last six months
3. Outliers
4. Holiday effects.

Comparisons with different time series models

Comparisons with different time series models



Outline

1. Introduction
2. Features of Business Time Series
3. Prophet Forecasting Model
   - Decomposable time series model
   - The Trend Model: Nonlinear & Linear
   - Seasonality
   - Holidays and Events
   - Model fitting & Performance
   - Analyst-in-the-Loop Modeling
4. Automating Evaluation of Forecasts
5. Conclusion

Decomposable time series model

[Harvey & Peters, 1990]

$$y(t) = g(t) + s(t) + h(t) + \epsilon_t$$

- $g(t)$: trend function,models non-periodic changes
- $s(t)$: periodic changes,weekly and yearly seasonality
- $h(t)$: represents the effects of holidays which occur on potentially irregular schedules over one or more days
- ϵ_t: idiosyncratic changes which are not accommodated by the model(usually assumed to be normally distributed)

Advantages

- Easily accommodate seasonality with multiple periods.
- Measurements do not need to be regularly spaced.
- Fitting fast & interpretable parameters can be changed by analysts to impose assumptions on the forecast.

Nonlinear, Saturating Growth

Logistic growth model

$$g(t) = \frac{C}{1+exp(-k(t-m))}$$

- C: carrying capacity.
- k: the growth rate.
- m: offset parameter.

Drawbacks of above model

- The carrying capacity C is not constant
  - Replace C with $C(t)$
- The growth rate k is not a constant
  - Incorporate a varying rate in order to fit historical data.

Nonlinear, Saturating Growth

Strategies

Incorporate trend changes in the growth model by explicitly defining changepoints where the growth rate is allowed to change.

Changepoints

Suppose there are S changepoints at times $s_j, j = 1, ..., S$. We define a vector of rate adjustments $\delta \in R^S$, where δ_j is the change in rate that occurs at time s_j.

Indicator function

$$a_j(t) = \begin{cases} 1, & \text{if } t \geq s_j, \\ 0, & \text{otherwise.} \end{cases}$$

Nonlinear, Saturating Growth

Rate at time t

The rate at time t is the base rate k, plus all of the adjustments up to that point: $k + \sum_{j:t>s_j} \delta_j = k + a(t)^T \delta$.

Offset parameter at changepoint j

$$\gamma_j = \left(s_j - m - \sum_{l<j} \gamma_l\right)\left(1 - \frac{k + \sum_{l<j} \delta_l}{k + \sum_{l \le j} \delta_l}\right).$$

Piecewise logistic growth model

$$g(t) = \frac{C(t)}{1 + \exp(-(k + \mathbf{a}(t)^\intercal \boldsymbol{\delta})(t - (m + \mathbf{a}(t)^\intercal \boldsymbol{\gamma})))}.$$

- $C(t)$:Analysts often have insight into market sizes and can set these accordingly.

Linear Trend with Changepoints

Linear Trend with Changepoints

$$g(t) = (k + \mathbf{a}(t)^\intercal \boldsymbol{\delta})t + (m + \mathbf{a}(t)^\intercal \boldsymbol{\gamma}),$$

- k: growth rate
- δ: rate adjustment
- m: offset parameter
- $\gamma_j = -s_j \delta_j$: to make the function continuous

Automatic Changepoint Selection

Two Strategies for Automatic Changepoint Selection

1. The changepoints s_j could be specified by the analyst using known dates of product launches and other growth-altering events, or may be automatically selected given a set of candidates.
2. Automatic selection can be done by putting a sparse prior on δ,like $\delta_j \sim Laplace(0, \tau)$, where parameter τ directly controls the flexibility of the model in altering its rate.

Automatic Changepoint Selection

Second Strategy

- Parameter estimate: S changepoints over a history of T points, each of which has a rate change $\delta_j \sim Laplace(0, \tau)$.
  - In fully Bayesian framework this could be done with a hierarchical prior on τ to obtain its posterior.
  - Maximum likelihood estimate: $\lambda = \frac{1}{S}\Sigma_{j=1}^{S}|\delta_j|$.

$$\forall j > T, \quad \begin{cases} \delta_j = 0 \text{ w.p. } \frac{T-S}{T}, \\ \delta_j \sim \text{Laplace}(0, \lambda) \text{ w.p. } \frac{S}{T}. \end{cases}$$

- Strong assumption(easy overfitting): assuming that the future will see the same average frequency and magnitude of rate changes that were seen in the history.
  - τ increased, model has more flexibility, training error will drop and this will produce wide uncertainty intervals.

Seasonality

Periodic effects

$$s(t) = \sum_{n=1}^{N} \left(a_n \cos\left(\frac{2\pi n t}{P}\right) + b_n \sin\left(\frac{2\pi n t}{P}\right) \right)$$

- P: regular period we expect the time series to have
- Increase N allows for fitting seasonal patterns changing more quickly, albeit with increased risk of overfitting.

Example

Let $N = 10$ and $P = 365.25$, we get:

$$X(t) = \left[\cos\left(\frac{2\pi(1)t}{365.25}\right), \ldots, \sin\left(\frac{2\pi(10)t}{365.25}\right) \right]$$

- Seasonal component: $s(t) = X(t)\beta$
  - $\beta = [a_1, b_1, ..., a_N, b_N]$
  - $\beta \sim Normal(0, \sigma^2)$: a smoothing prior on the seasonality.

Holidays and Events

Holidays and events provide large, somewhat predictable shocks to many business time series and often do not follow a periodic pattern, so their effects are not well modeled by a smooth cycle.

Mathematical Form

$$Z(t) = [\mathbf{1}(t \in D_1), \ldots, \mathbf{1}(t \in D_L)]$$

$$h(t) = Z(t)\kappa, \kappa \sim Normal(0, v^2)$$

- D_i: set of past and future dates for holiday i.
- κ_i: corresponding change in the forecast for holiday i.

Important to include effects for a window of days around a particular holiday such as the weekend of Thanksgiving.

Holidays and Events

Example

Holiday	Country	Year	Date
Thanksgiving	US	2015	26 Nov 2015
Thanksgiving	US	2016	24 Nov 2016
Thanksgiving	US	2017	23 Nov 2017
Thanksgiving	US	2018	22 Nov 2018
Christmas	*	2015	25 Dec 2015
Christmas	*	2016	25 Dec 2016
Christmas	*	2017	25 Dec 2017
Christmas	*	2018	25 Dec 2018

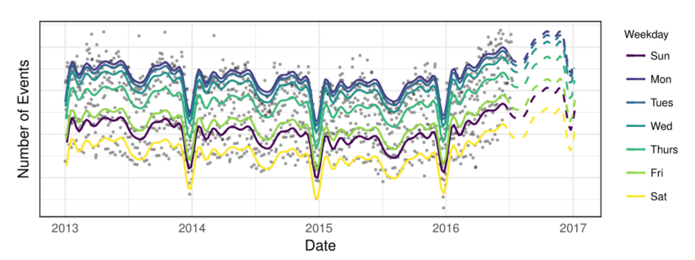
Model fitting & Performance

Model fitting

Use Stan's L-BFGS.(pystan package)

We will show some performance of Prophet.

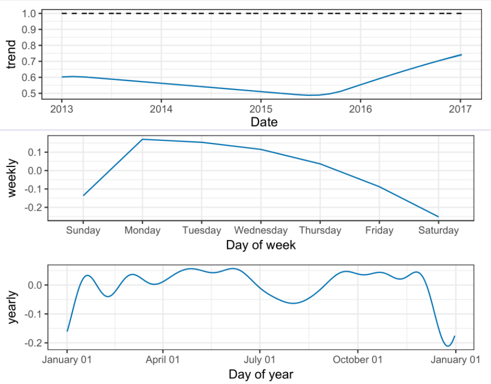
Example



Model fitting & Performance

Look at each component of the forecast separately

Trend & Seasonality



Analyst-in-the-Loop Modeling

About overfitting

Not enough historical data to select the best regularization parameters via cross-validation.

Set default values that are appropriate for most forecasting problems.

Advantages for Analysts without statistical knowledge

- Capacities: Analysts can directly specify capacities.
- Changepoints: Known dates of changepoints, such as dates of product changes.
- Holidays and seasonality: Analysts have experience with which holidays impact growth in which regions.
- Smoothing parameters: By adjusting τ and (σ, v) allow the analyst to tell the model how much of the historical seasonal variation is expected in the future.

Analyst-in-the-Loop Modeling

Motivation for Analyst-in-the-Loop Modeling

Analyst-in-the-loop modeling approach is an alternative approach that attempts to blend the advantages of statistical and judgmental forecasts by focusing analyst effort on improving the model when necessary rather that directly producing forecasts through some unstated procedure.

Automating Evaluation of Forecasts

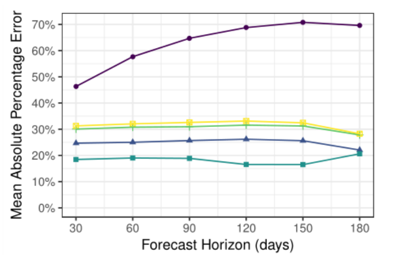
Use of Baseline Forecasts

- Compare simplistic models (last value and sample mean) as well as the automated forecasting procedures.

Modeling Forecast Accuracy

- Mean absolute percentage error (MAPE)

Simulated Historical Forecasts



Modeling Forecast Accuracy

Necessity for Automating Modeling Forecast Accuracy

When there are too many forecasts for analysts to manually check each of them, it is important to be able to automatically identify forecasts that may be problematic.

Advantages of Automating Modeling Forecast Accuracy

- When the forecast has large errors relative to the baselines, the model may be misspecified.
- Large errors for all methods on a particular date are suggestive of outliers
- 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.

Outline

Conclusion

Conclusion

In this paper, the authors:

1. Use a simple, modular regression model that often works well with default parameters and allows analysts to select the components that are relevant to their forecasting problem and easily make adjustments as needed.
2. Develop a system for measuring and tracking forecast accuracy, and flagging forecasts that should be checked manually to help analysts make incremental improvements.
3. Adjustable models and scalable performance monitoring in combination allow a large number of analysts to forecast a large number and a variety of time series – what we consider forecasting at scale.