Prophet: an Additive, Popular Approach to Time Series

Isaac Laughlin • isaac.laughlin@gmail.com • linkedin.com/ilaughlin

Managing Lead Instructor & Principal Data Scientist, Galvanize

Outline

Preliminaries

This talk might be for you if:

- You're familiar with supervised learning
- You're interested in doing predictions on time series
- You use Python or R

OR

- You're an unusually open-minded econometrician

Creators

Released in February, 2017 by

Sean J. Taylor, Facebook, Benjamin Letham, Facebook

Prior art

A promise to you (and me)

I promise not to compare Prophet to every other time series approach out there.

Classical Time Series Methods

- Based on formal structural assumptions AR, MA
- Correct model/parameter choice generally non-trivial (but important)
- Some modern flexible approaches like TBATS, Exponential Smoothing State Space Model With Box-Cox Transformation, ARMA Errors, Trend And Seasonal Components

Machine Learning methods

- Avoids structural assumptions
- No decomposition of components
- Uncertainty estimates may be hard to get
- How hyperparameters will affect specific weaknesses not clear

What Prophet brings

- Scale
- Decomposible additive components
- Straightforward parameters
- Moderate structural formalism
- Uncertainty

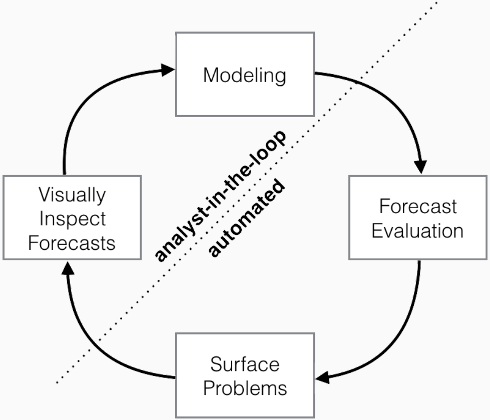
About Prophet

How it works

Technology

- Bayesian estimation of additive model
- Fit with Stan
- MAP optimization
- Interface with Python or R

Workflow



Modeling
Forecast Evaluation
Surface Problems
Visually Inspect Forecasts
analyst-in-the-loop
automated

The model

Components

- Trend, saturating growth or piecewise linear
- Yearly seasonality
- Weekly seasonality
- User-provided list of holidays
- External regressors

Generalized Additive Model

$$y(t) = g(t) + s(t) + h(t) + \epsilon_t \tag{1}$$

e.g. trend + seasonality + holiday

Simplified saturating growth

$$g(t) = \frac{C}{1 + exp(-k(t - m))} \tag{2}$$

Notes

- C is a carrying capacity
- k is a growth rate
- m an offset
- Modeled with finite number of changepoints.

Actual saturating growth

$$g(t) = \frac{C(t)}{1 + exp(-(k + \boldsymbol{a}(t)^T\delta)(t - (m + \boldsymbol{a}(t)^T\gamma)))} \tag{3}$$

Notes

- $\mathrm{a(t)} = \begin{cases} 1, \text{if } t \geq \mathrm{s_j}, \\ 0, \text{otherwise} \end{cases}$
- s_j time at which a changepoint occurs.
- γ serves to make piecewise-trend continuous

Piece-wise linear trend with changepoints

$$g(t) = (k + \boldsymbol{a}(t)^T\delta)t + (m + \boldsymbol{a}(t)^T\gamma) \tag{4}$$

Changepoint selection

- Specified by user (if known changepoints, like product launches)
- Detected naturally by sparse prior on $\delta_j \sim \mathit{Laplace}(0, \tau)$
- τ will become an important parameter of our model

Trend Uncertainty

- Assumes trend will change with same frequency and magnitude as in history.

Seasonality

Fourier series

$$s(t) = \sum_{n=1}^{N} (a_n cos(\frac{2\pi nt}{P}) + b_n sin(\frac{2\pi nt}{P})) \quad (5)$$

Notes

- Used to approximate arbitrary smooth seasonal effects
- Fitting requires estimating $2n$ parameters
- Prior on coefficients on matrix of fourier features $\sim Normal(0, \sigma^2)$
- Sigma will become an important parameter of our model.

Holidays and Events

Holidays

Create a matrix of holiday indicators

$$Z(t) = [\mathbf{1}(t \in D_1), \ldots, \mathbf{1}(t \in D_L)] \tag{6}$$

So we specify our additive holiday component as:

$$h(t) = Z(t)\boldsymbol{\kappa} \tag{7}$$

Notes

- D_i is the set of past and future dates of a holiday.
- *kappa* are given a $\sim Normal(0, \nu^2)$ prior.

Practical example in R