Forecasting at Scale

Christian Barz

2020-07-22

Prophet's goals and alternative

Prophet library

The Prophet library[1] is a model and a framework.

- targets at non-experts with background business knowledge
- less knowledge about time series is required
- easy to configure
- easily interpretable parameters
- flexible for a wide range of business problems

[1] Taylor, Letham, Forecasting at scale, 2017, The American Statistician

Alternative

The `forecast` R-package contains some alternative for automated times series models.

- `auto.arima`[2], fits mutliple ARIMA models and take the best fit
- `ets`[3], fits mutliple exponential smoothing models and take the best fit
- `snaive`[4], random walk model with seasonality

[2] Hyndman, Khandakar et al. 2007, Automatic time series for forecasting: the forecast package for R

[3] Hyndman, Koehler., Snyder & Grose, 2002, 'A state space framework for automatic forecasting using exponential smoothing methods'

[4] De Livera, Hyndman. & Snyder, 2011, 'A state space framework for automatic forecasting using exponential smoothing methods'

Prophet model

Generalized additive models

Prophet uses a generalized additive model (GAM)[5]

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

- $y(t)$ target
- $g(t)$ trend function
- $s(t)$ Foruier serie for periodic changes
- $h(t)$ represents holiday effects, i.e. irregular schedules over one or more days
- ϵ_t error term (not accomodated by the model)

[5] Hastie & Tibshirani, 1987, 'Generalized additive models: some applications'

The trend model

Trend can be modeled with

- saturated growth model
- piecewise linear model

Nonlinear Saturating growth (basic)

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

- C carrying capacity (upper bound)
- k growth rate
- m offset parameter

Nonlinear Saturating growth (time dependent parameters)

- C (carrying capacity) and k (growth rate) are usually not constant
- incoporate trend changes by explicite defined change points
- change point can be set by analyst (e.g. product launches) or automatically selected

Nonlinear Saturating growth (time dependent carrying capacity)

TBA - this is not explained in the non reviewed paper

Nonlinear Saturating growth (non constant growth rate - idea)

let $s_1, \ldots, s_n$ be change points (time stamps).

define $\delta \in \mathbb{R}^n$ (vector of rate adjustments)

define growth rate at time t by

$$k(t) := k + \sum_{j: s_j < t} \delta_j$$

where k is the base growth rate

Nonlinear Saturating growth (non constant growth rate - mathematically correct)

define $a(t) \in \{0,1\}^n$ by

$$a_j(t) := \begin{cases} 1, & \text{for } t \leq s_j \\ 0, & otherwise \end{cases}$$

then rate at time t is given by

$$k(t) := a(t)^T \delta$$

Nonlinear Saturating growth (non constant growth rate - mathematically correct)

The offset parameter m is than adjusted according to the change points by a formula:

$$\gamma := f(s, m, k, \delta).$$

Putting all together we recieve :

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

Linear trend with Changepoints

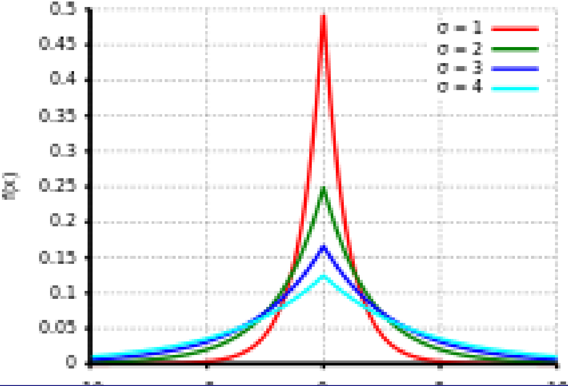
using the change points from above a piecewise linear model is given by

$$g(t) := (k + a(t)^T \delta) * t + (m + a(t)^T \gamma)$$

Automatic changepoint selection

Change points are automatically detected by putting a sparse prior on δ, i.e.

$$\delta_j \; \mathit{Laplace}(0, \tau)$$



The seasonal model

Fourier series

"Holidays and Events"

Stan model and fitting

Stan model and fitting

Stan is a platform for Bayesian inference using MCMC sampling and more.

Stan model with logistic growth trend model

```
model {
// Priors
k ~ normal(0, 5);
m ~ normal(0, 5);
epsilon ~ normal(0, 0.5);
delta ~ double_exponential(0, tau);
beta ~ normal(0, sigma);
// Logistic likelihood
y ~ normal(C ./ (1 + exp(-(k + A * delta)
                            .* (t - (m + A * gamma)))) +
              X * beta, epsilon);
}
```

Stan fitting for prophet

- Prophet uses the L-BFGS algorithm from Stan to fit the GAM
- BFGS = Broyden–Fletcher–Goldfarb–Shanno algorithm, quasi Newton method

Future Talk's

regarding Prophet

- how the time dependet carrying capacity is determined ?
- how additional regressors are included ?
- how outliers are handeled ?

theoretical background

- Stan models
- GAMS
- Baysian models
- L-BFGS
- Marcov chain Monte Carlo

Further reading

Further reading

- The textbook for forecast R-package, but with a lot of theory and practice