Bayesian Time Series Forecasting With Change Point and Anomaly Detection

Dasha Demidova, Marina Gomtsyan

Skoltech, HSE

26 october 2018

Time Series Model

Observation equaton

$y_t = \mu_t + \gamma_t + z_t^a o_t + (1 - z_t^a)\varepsilon_t$

Transition Equations

Trend:

$\mu_t = \mu_{t-1} + \delta_{t-1} + z_t^c r_t + (1 - z_t^c) u_t$

$\delta_t = \delta_{t-1} + v_t$

Seasonality:

$$\gamma_t = -\sum_{s=1}^{S-1} \gamma_{t-s} + w_t$$

Anomaly point

$\{z_t^a\}_{t=1}^n \sim \text{Ber}(p_a)$, i.i.d.

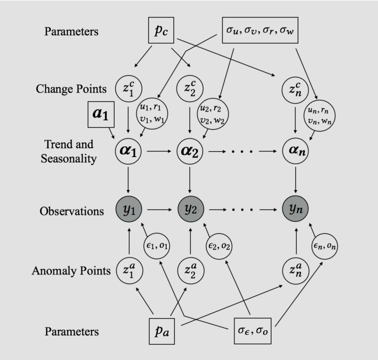
Change point

$\{z_t^c\}_{t=1}^n \sim \text{Ber}(p_c)$, i.i.d.

Noise

$o_t, \varepsilon_t, r_t, u_t, v_t, z_t$ - independent zeros mean normal noises

Graphical Model



Latent Variables

$$\{\alpha_t\}_{t=1}^n = (\mu_t, \delta_t, \gamma_t, \dots, \gamma_{t-S+2})_{t=1}^n$$

$$\{z_t\}_{t=1}^n = \{(z_t^a, z_t^c)\}_{t=1}^n$$

Parameters

$$a_1 = (\mu_o, \delta_0, \gamma_0, \dots, \gamma_{2-S})$$
$$p = (p_a, p_c)$$
$$\sigma = (\sigma_\varepsilon, \sigma_o, \sigma_u, \sigma_r, \sigma_v, \sigma_w)$$

Algorithm

Part I: Initialization

1. Initialize $\sigma_\varepsilon, \sigma_o, \sigma_u, \sigma_r, \sigma_v, \sigma_w$ with the empirical standard deviation.
2. Initialize a_1: $a_1[0] = \frac{1}{S}\sum_{t=1}^{S} y_t$, $a_1[1:] = 0$.
3. Initialize $p_a = p_c = \frac{1}{n}$, $\{z_t^a\}_{t=1}^n \sim \mathrm{Ber}(p_a)$, $\{z_t^c\}_{t=1}^n \sim \mathrm{Ber}(p_c)$

Algorithm

Part II: Inference

while $L_{a_1,p,\sigma}(y, \alpha, z)$ does not converges:

1. $\alpha \sim p_{a_1,p,\sigma}(\alpha|y, z)$ by Kalman filter, Kalman smoothing and "fake-path" trick
2. $z^a_t \sim p_{a_1,p,\sigma}(z^a_t|y, \alpha) = \text{Ber}(p^a_t)$,
   $z^c_t \sim p_{a_1,p,\sigma}(z^c_t|y, \alpha) = \text{Ber}(p^c_t)$
   $\{p^a_t\}^n_{t=1} = \mathbb{P}(z^a_t = 1|y, \alpha)$, $\{p^c_t\}^n_{t=1} = \mathbb{P}(z^c_t = 1|y, \alpha)$
3. Segment control on z_c: requirement on the length of segment among two consecutive change points
4. Using α and z, update σ by the empirical standard deviation
5. Update a_1. $a_1[:2] = \alpha_1[:2]$, $a_1[2:] = \alpha_{S+1}[2:]$
6. Calculate $L_{a_1,p,\sigma}(y, \alpha, z)$

Algorithm

The Joint Likelihood function

$$L_{a_1,p,\sigma}(y,\alpha,z) =$$

$$= \prod_{t:z_t^a=0} \mathcal{N}(y_t|\mu_t+\gamma_t,\sigma_\varepsilon) \cdot \prod_{t:z_t^a=1} \mathcal{N}(y_t|\mu_t+\gamma_t,\sigma_o) \cdot$$

$$\cdot \prod_{t:z_t^c=0} \mathcal{N}(\mu_t|\mu_{t-1}+\delta_{t-1},\sigma_u) \cdot \prod_{t:z_t^c=1} \mathcal{N}(\mu_t|\mu_{t-1}+\delta_{t-1},\sigma_r) \cdot$$

$$\cdot \prod_{t=1}^{n} \mathcal{N}(\delta_t|\delta_{t-1},\sigma_v) \cdot \prod_{t=1}^{n} \mathcal{N}(\gamma_t| -\sum_{s=1}^{S-1}\gamma_{t-s},\sigma_w) \cdot$$

$$\cdot \prod_{t=1}^{n} (p_a)^{z_t^a}(1-p_a)^{1-z_t^a}(p_c)^{z_t^c}(1-p_c)^{1-z_t^c}$$

Algorithm

State Space Equations

$y_t = Z_t\alpha_t + A_t o_t + (1 - A_t)\varepsilon_t$

$\alpha_{t+1} =$

$T_t\alpha_t + R_t C_t \eta_t + R_t(1 - C_t)\xi_t,$

where

$o_t \sim \mathcal{N}(0, H_t^o)$

$\varepsilon_t \sim \mathcal{N}(0, H_t^\varepsilon)$

$\eta_t \sim \mathcal{N}(0, H_t^\eta)$

$\xi_t \sim \mathcal{N}(0, H_t^\xi),$

$A_t \sim \text{Bernoulli}(p_a)$

$C_t \sim \text{Bernoulli}(p_c)$

Transition equations

$\alpha_t = (\mu_t, \delta_t, \gamma_t, \ldots, \gamma_{t-S+2})^T$

$\eta_t = (r_t, v_t, w_t),$

$\xi_t = (u_t, v_t, w_t)$

$Z_t = (1, 0, 1, 0, \ldots, 0)$

$T_t = \text{diag}(T_\mu, T_\gamma)$

$R = \text{diag}(R_\mu, R_\gamma)$

$H_t^o = \sigma_o^2,$

$H^\varepsilon = \sigma_\varepsilon^2$

$Q^\eta = \text{diag}(\sigma_r^2, \sigma_v^2, \sigma_w^2)$

$Q^\xi = \text{diag}(\sigma_u^2, \sigma_v^2, \sigma_w^2)$

Algorithm

Kalman Filter Recursion

$v_t = y_t - Z_t a_t$
$F_t = Z_t P_t Z_t^T + p_t^a H_t^o + (1 - p_t^a) H_t^\varepsilon$
$a_{t|t} = a_t + P_t Z_t^T F_t^{-1} v_t$
$P_{t|t} = P_t Z_t F_t^{-1} Z_t P_t$
$K_t = T_t P_t Z_t^T F_t^{-1}$

Update equations:

$a_{t+1} = T_t a_t + K_t v_t$
$P_{t+1} = T_t P_{t|t} T_t^T + R_t (p_t^c Q_t^\eta + (1 - p_t^c Q_t^\xi)) R_t^T$

Kalman Smoothing Recursion

Inputs from Kalman Filter:

a_t, P_t, v_t, F_t, K_t
$r_n = 0, N_n = 0$

Update equations:

$L_t = T_t - K_t Z_t$
$r_{t-1} = Z_t F_t^{-1} v_t + L_t r_t$
$\hat{\alpha}_t = a_t + P_t r_{t-1}$
$N_{t-1} = Z_t^T F_t^{-1} Z_t + L_t^T N_t L_t$
$V_t = P_t - P_t N_{t-1} P_t$

Algorithm

"Fake-path" Trick

The purpose is to obtain posterior distribution $\mathbb{P}_{\alpha_1,p,\sigma(\alpha|y,z)}$. All hidden variables z, p, σ are given:

1. Pick some vector $\tilde{a}_1$ and generate a sequence of time series $\tilde{y}$ from it. We also observe $\tilde{\alpha}$.
2. Obtain $\{\mathbb{E}(\tilde{\alpha}_t|\tilde{y})\}_{t=1}^n$ from $\tilde{y}$ by Kalman filter and Kalman smoothing.
3. Use $\{\tilde{\alpha}_t - \mathbb{E}(\tilde{\alpha}_t|\tilde{y}) + \mathbb{E}(\alpha_t|y)\}_{t=1}^n$ as sampling distribution from the conditional distribution.

Algorithm

Segment control on change points

Denote $t_1 < t_2 < ...$ to be all the indexes such that $z_{t_i}^c = 1$

while there exists i such that $|t_{i+1} - t_i| < l$ ***do*** **:**

1. Check if $|\mu_{t_i - 1} - \mu_{t_{t+1}+1}| \leq 2$. If so, exclude both of them from change points by setting $z_{t_i}^c = z_{t_{i+1}}^c = 0$. Otherwise, randomly exclude one of them by setting the corresponding coordinate in z^c to be 0.
2. Update all the indexes of change points in z^c.

end

Algorithm

Part III: Forecasting

1. With a_n and σ, generate future time series y_{future} with length m. Repeat generative procedure to obtain future paths $y_{future}^{(1)}, y_{future}^{(2)}, \ldots, y_{future}^{(N)}$.
2. Combine all the predictive paths and give the distribution for the future time series forecasting. Calculate the point-wise quantile intervals.

Algorithm

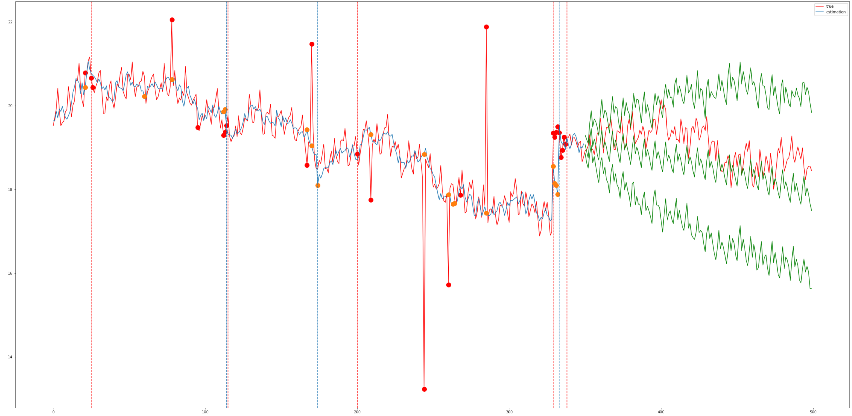
Generative Procedure

1. Generate the indexes of anomalies or change points occur

$$\{z_t^a\}_{t=1}^n \sim \mathrm{Ber}(p_a), \quad \{z_t^c\}_{t=1}^n \sim \mathrm{Ber}(p_c).$$

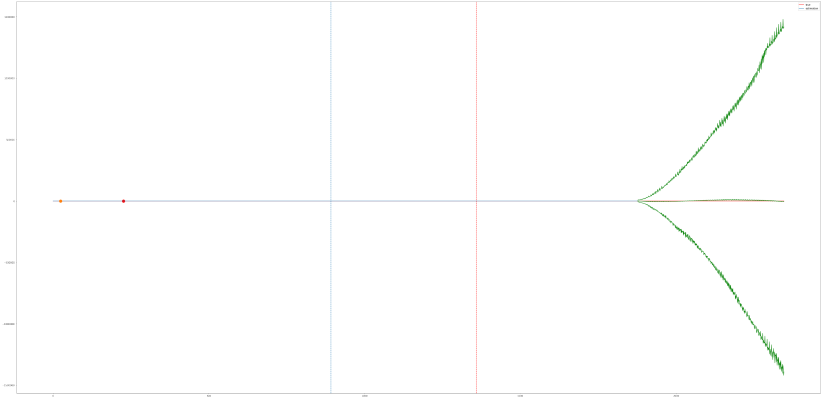
2. Gnerate $\varepsilon, o, u, r, v, w$ as independent normal r.v.'s with zero mean and standard deviations $\sigma_\varepsilon, \sigma_o, \sigma_u, \sigma_r, \sigma_v, \sigma_w$.
3. Generate ${\alpha_t}_{t=1}^m$ by transition functions.
4. Generate time series $\{y_t\}_{t=1}^m$ by the observation function.

Experiments



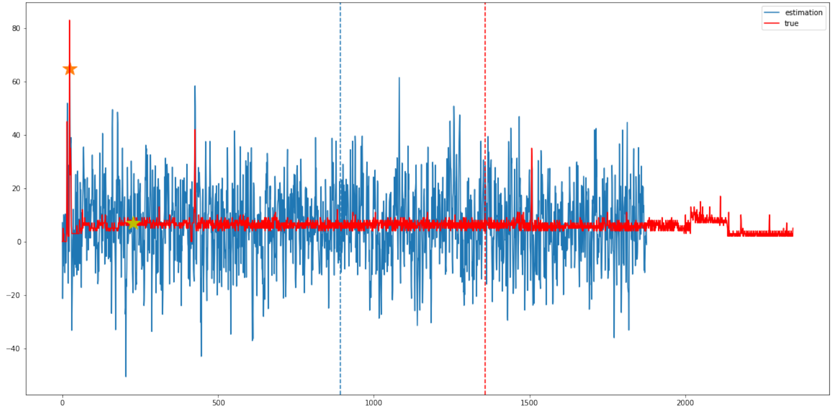
MSE on future prediction part is 18.66.

Real life Experiment



MSE on future prediction part is 50,116.

Real life Experiment



Real life Experiment

ARIMA results, with MSE 1.64