gfpop: an R Package for Univariate Graph-Constrained Change-Point Detection

Machine Learning for Time Series

Simon Queric and Vincent Herfeld

ENS Paris-Saclay, Télécom Paris

January 1, 2024

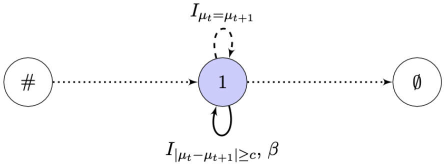
Table of content

Introduction

$$Q_n^{std}(\mu) = \sum_{t=1}^{n} (y_t - \mu_t)^2 + \beta \sum_{t=1}^{n-1} I_{\mu_t \neq \mu_{t+1}}$$

$$Q_n^{iso}(\mu) = \sum_{t=1}^{n} (y_t - \mu_t)^2 \text{ such that } \mu_t \leqslant \mu_{t+1}$$

Introduction



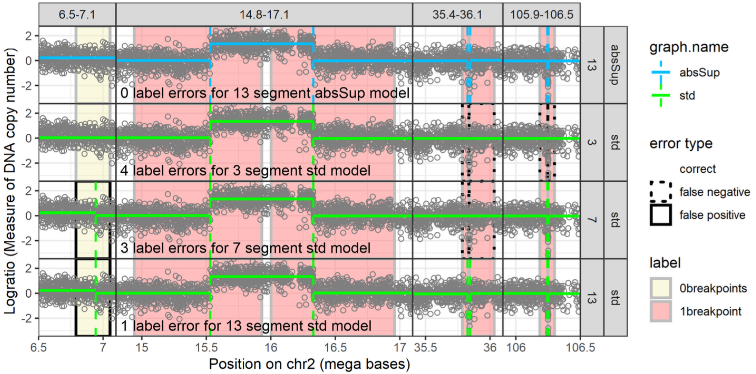


Table of content

Introduction to graph of constraints

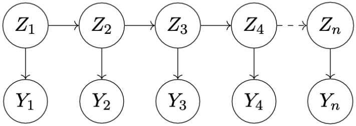


Figure: Hierarchical representation

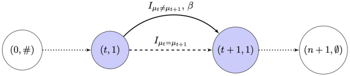


Figure: Graph of constraints

A simple example

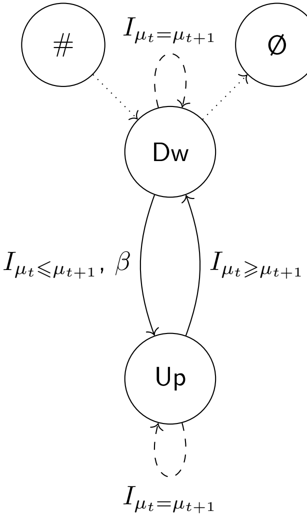


Figure: Collapsed graph for multi-modal regression

Optimization problem

Classical multiple change points detection problem :

$$(\hat{t}_1, ..., \hat{t}_K) = \underset{(t_1,...,t_k)}{\operatorname{argmin}} \sum_{k=0}^{K} c(x[t_k : t_{k+1}])$$

The article's problem :

$$Q_n^s(\theta) = \min_{\substack{p=(v,e)\in\mathcal{G}_n \\ \mu|p(\mu),\mu_n=\theta,v_n=s}} \sum_{t=1}^{n} (\gamma_{e_t}(y_t, \mu_t) + \beta_{e_t})$$

c and γ_e are always negative log-probabilities.

Cost functions

$$\text{L2 decomposition}: \ f_1 : \theta \mapsto 1, f_2 : \theta \mapsto \theta, f_3 : \theta \mapsto \theta^2$$
$$\text{Lin-log decomposition}: \ f_1 : \theta \mapsto 1, f_2 : \theta \mapsto \theta, f_3 : \theta \mapsto \log \theta$$

$$\gamma_{\ell^2}(y, \mu) = (y - \mu)^2 = y^2 - 2y\mu + \mu^2 = y^2 f_1(\mu) - 2y f_2(\mu) + f_3(\mu)$$

$$\gamma_{\text{Poisson}}(y, \mu) = -\log\left(e^{-\mu}\frac{\mu^y}{y!}\right) = \log(y!) f_1(\mu) + f_2(\mu) - y f_3(\mu)$$

Algorithm to solve the problem

Algorithm 1 Backtracking $\hat{s}$ and $\hat{\mu}$

procedure BACKTRACK$((Q_1^1, ..., Q_1^S), ..., (Q_n^1, ..., Q_n^S))$

$\hat{\mu} \leftarrow$ empty vector of size n

$\hat{s} \leftarrow$ empty vector of size n

$(\hat{s}_n, \hat{\mu}_n) = \underset{(s,\mu)}{\text{argmin}}\{Q_n^s(\mu)\}$

$\triangleright$ We can impose a subset of arrival states $\tilde{\mathcal{S}} \subset \{1, ..., S\}$ by $(\hat{s}_n, \hat{\mu}_n) \leftarrow \{\underset{(s,\mu)}{\text{argmin}}\{Q_n^s(\mu)\}, \, s \in \tilde{\mathcal{S}}\}$

for $t = n-1$ **to** $t = 1$ **do**

$\quad \hat{s}_t = \underset{s | \exists \, edge \, (s, \hat{s}_{t+1})}{\text{argmin}} \left\{ O_t^{s, \hat{s}_{t+1}}(\hat{\mu}_{t+1}) + \gamma_{(s, \hat{s}_{t+1})}(y_{t+1}, \hat{\mu}_{t+1}) + \beta_{(s, \hat{s}_{t+1})} \right\}$

$\quad \hat{\mu}_t = \underset{\mu | I_{(\hat{s}_t, \hat{s}_{t+1})}(\mu, \hat{\mu}_{t+1})}{\text{argmin}} \left\{ Q_t^{\hat{s}_t}(\mu) \right\}$ $\triangleright$ If $\hat{\mu}_t$ is such that the constraint is active, we have 'forced = TRUE' in ″gfpop() response

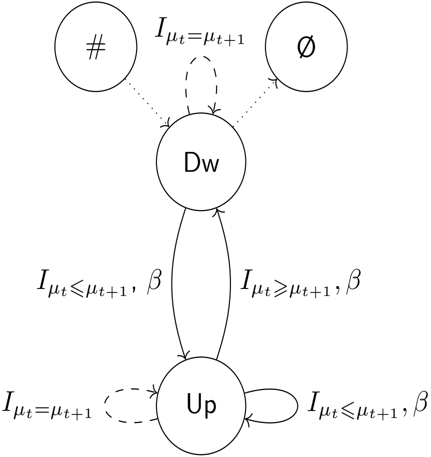
end for

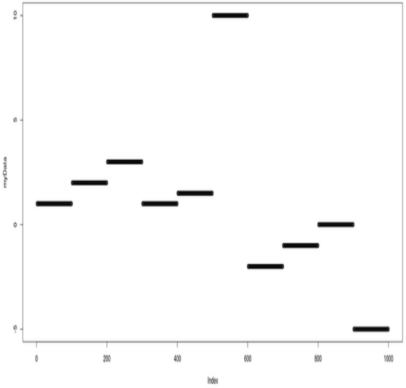
return $(\hat{s}, \hat{\mu})$

Figure: Backtrack algorithm

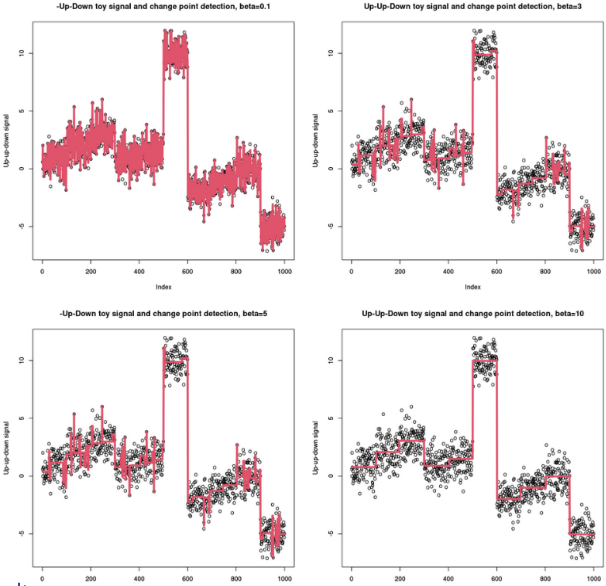
Table of content

Up-up-down signal

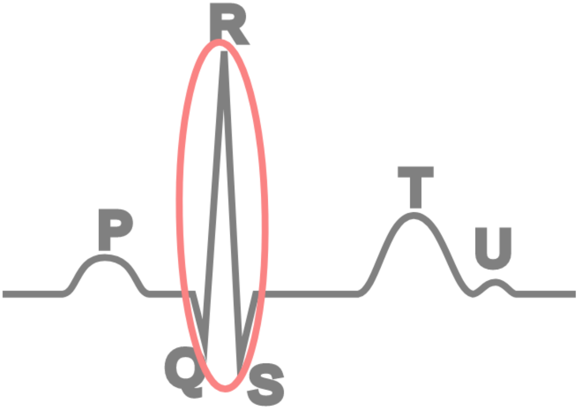




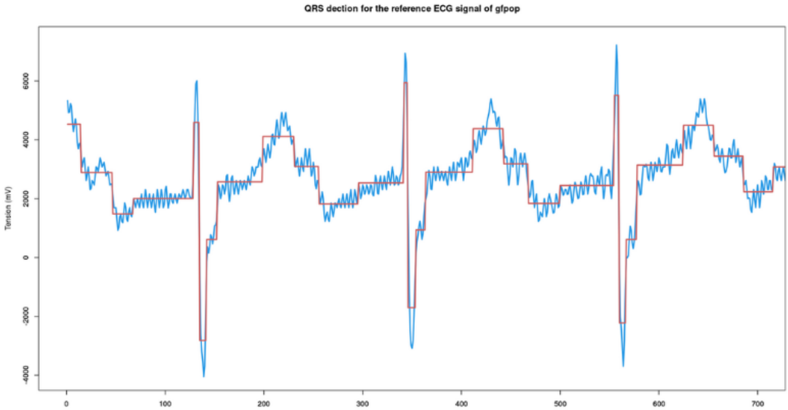
Up-up-down signal



QRS complex in ECG signals

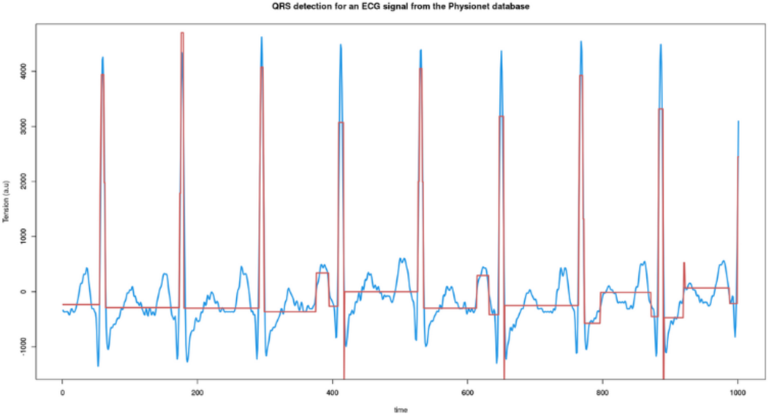


QRS complex in ECG signals



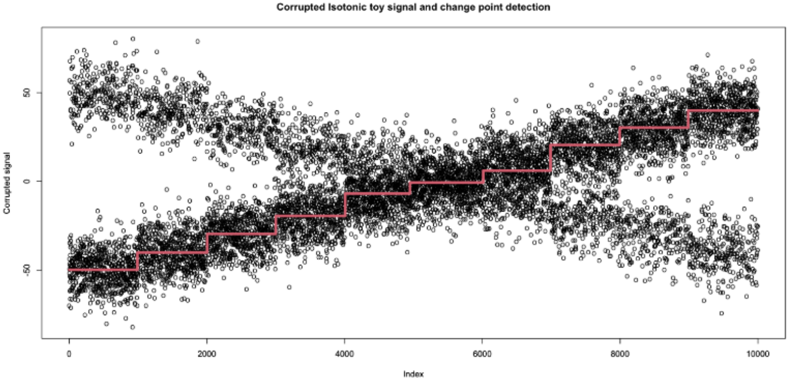
Sample rate : 250 Hz

QRS complex in ECG signals

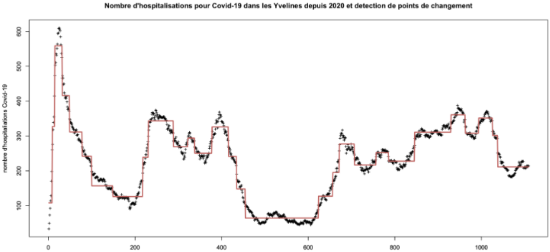


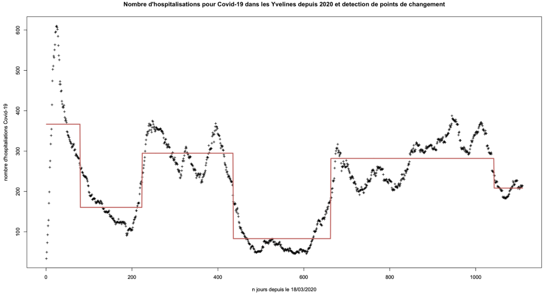
Sample rate : 250 Hz

Isotonic regression

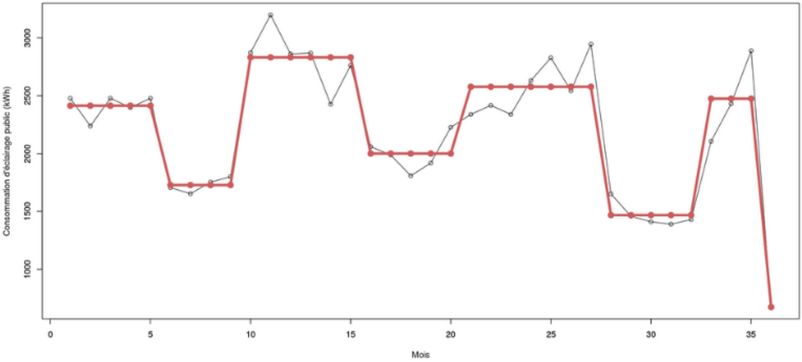


Number of Covid-19 hospitalisations





Public lighting energy consumption



We fit a simple Up-Down model.

Discussion

Strengths

1. Generalize the change point detection problem
2. Easy to use R package
3. Flexible
4. Reproducibility : github available
   `https://github.com/vrunge/gfpop/tree/master`

Weaknesses

1. Hard to tune hyperparameters and perform model selection
2. Rigid

Table of content

References

1. Toby Dylan Hocking et al. "Constrained Dynamic Programming and Supervised Penalty Learning Algorithms for Peak Detection in Genomic Data". In: Journal of Machine Learning Research 21.87 (2020), pp. 1–40. url: http://jmlr.org/papers/v21/18-843.html
2. Vincent Runge et al. gfpop: an R Package for Univariate Graph-Constrained Change- Point Detection. 2022.
3. Charles Truong, Laurent Oudre, and Nicolas Vayatis. "A review of change point detection methods". In: CoRR abs/1801.00718 (2018). url: http://arxiv.org/abs/1801.00718.