Automatic calibration of change-point detection methods

StatScale ECR Meeting (2022-12-14)

Charles Truong

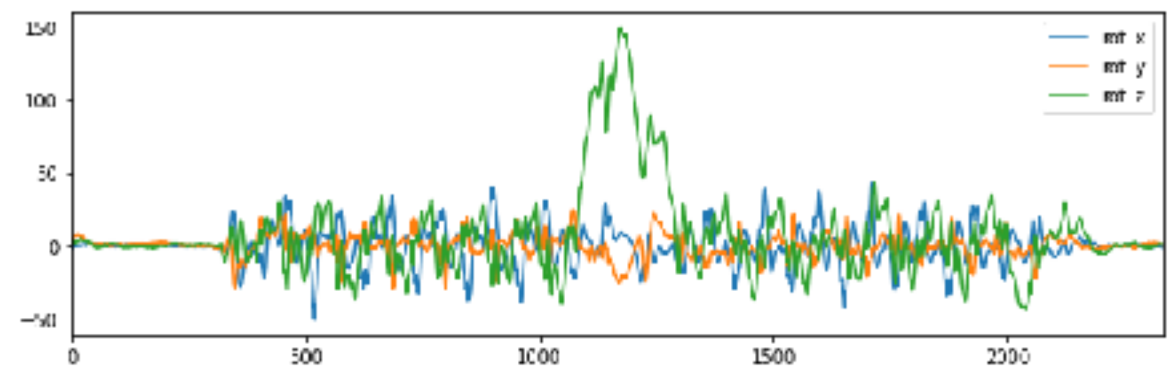
Example: study of human gait

- Human gait can be altered by a wide range of pathologies such as Parkinson's disease, arthritis, stroke,…

- Medical researchers try to objectively quantify gait characteristics [1-3].

Healthy and neurologically impaired subjects follow a fixed protocol:

- Standing still,
- Walking forward (10m),
- Turnaround,
- Walking backward (10m),
- Standing still.



[1] Heesen et al. (Multiple Sclerosis Journal, 2008).
[2] Oudre et al. (Sensors, 2018).
[3] Bois et al. (PLoS One, 2022).

Motivations

- Calibration is a hard task.
- Labels are often available.

Objective of this presentation.

- Propose a supervised approach to calibrate change-detection methods

Problem statement

- Let $\widehat{\mathcal{T}}_w$ be a family of change-point detection algorithms, indexed by a parameter vector $w \in \mathcal{W} \subset \mathbb{R}^d$.

- For a signal $\mathbf{y} = [y_1, y_2, \ldots, y_T]$, the estimated change-points are $\widehat{\mathcal{T}}_w(\mathbf{y}) = \{\hat{t}_1, \hat{t}_2, \ldots\}$

Objective. Find $w \in \mathcal{W}$ that minimizes the risk

$$L(w) := E_{(\mathbf{y},\mathcal{T})\sim\mathbb{D}}\left[\Delta\left(\mathcal{T}, \widehat{\mathcal{T}}_w(\mathbf{y})\right)\right]$$

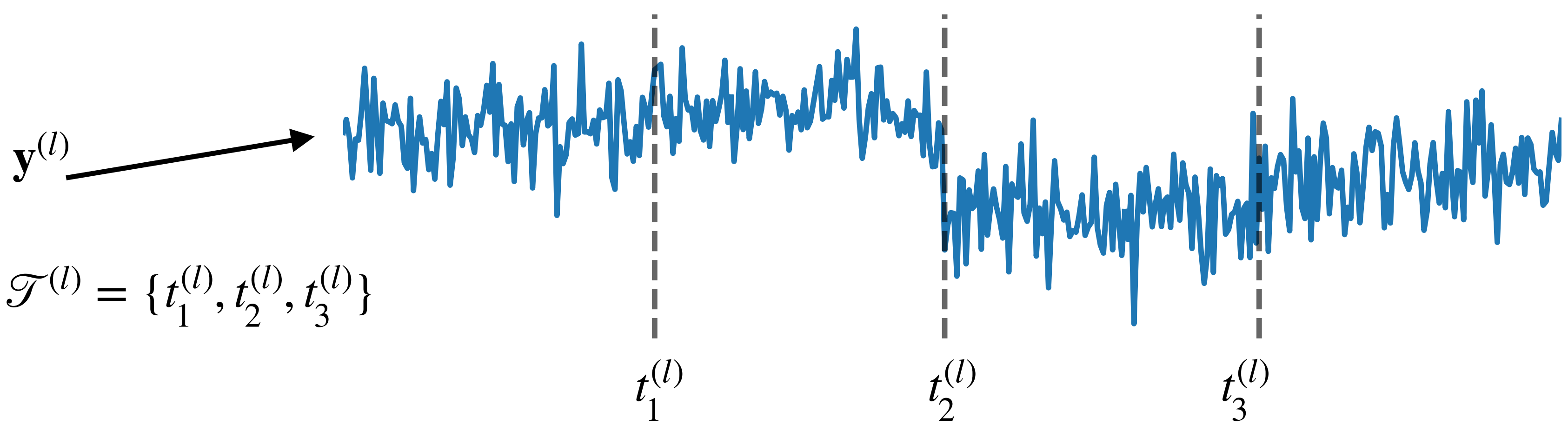
where $\Delta(\cdot,\cdot)$ is a distance on the space of segmentations and $\mathbb{D}$ is a distribution on the "(signal, segmentation)" space.

- Impossible task when $\mathbb{D}$ is unknown.

Empirical risk minimization

Supervised setting. In many situations, an expert is able to provide a few annotations, i.e.

- L signals $\mathbf{y}^{(l)}$ as well as their target segmentations $\mathcal{T}^{(l)}$ ($l = 1, \ldots, L$).



Empirical risk. Instead of the true risk, minimize the empirical risk $\bar{L}(w)$ w.r.t. w:

$$\bar{L}(w) := \sum_{l=1,\ldots,L} \Delta\left(\mathcal{T}^{(l)}, \widehat{\mathcal{T}}_w(\mathbf{y}^{(l)})\right)$$

- Still hard without assumptions on the family $(\widehat{\mathcal{T}}_w)_{w \in \mathcal{W}}$ of change-detection algorithms.

Family of change-point detection algorithms

Detect mean-shifts in multivariate signals, with a linear penalty.

$$\widehat{\mathscr{T}}_w(\mathbf{y}) := \arg \min_{\mathscr{T}=\{t_1,t_2,\ldots\}} V_w(\mathscr{T})$$

where $V_w(\mathscr{T}) := \sum_k \sum_{t=t_k}^{t_{k+1}-1} \|y_t - \bar{y}_{t_k..t_{k+1}}\|_w^2 \quad + \quad |\mathscr{T}|\log(T)$

where the norm $\|\cdot\|_w$ is defined by $\|x\|_w^2 := x^\intercal \mathrm{diag}(w) x$. Here, $\mathscr{W} = \mathbb{R}_+^d$.

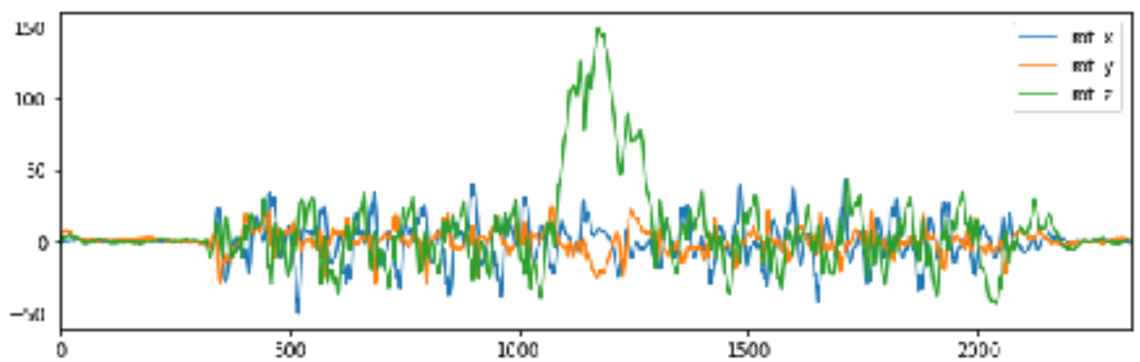
Intuition. This is a component-wise scaling [Witten and Tibshirani (JASA, 2012)]:

$$y_{\bullet,1}, \ldots, y_{\bullet,d} \xrightarrow{\text{becomes}} \sqrt{w_1}\, y_{\bullet,1}, \ldots, \sqrt{w_d}\, y_{\bullet,d}\,.$$

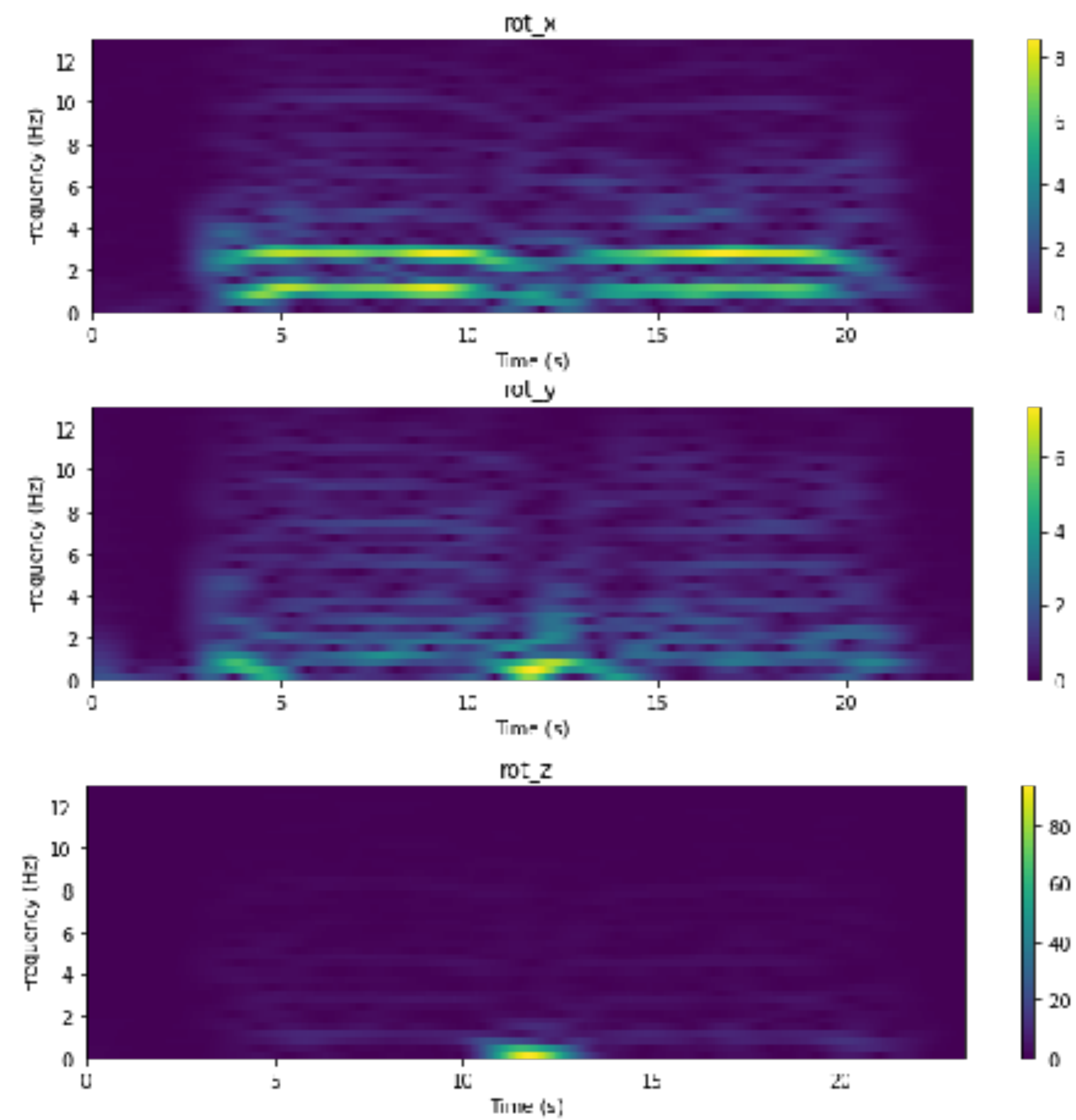
For univariate signals, this amounts to finding a “good” penalty value.

Example on a time-frequency representation

Study of human gait. We use the time-frequency representation of the raw-signal.



Each frequency is reweighed according to its usefulness to the change-point detection task.



Structured learning

Metric on segmentations. The metric Δ is defined by

$$\Delta(\mathcal{T}, \mathcal{T}') := |\mathcal{T}| + |\mathcal{T}'| - 2 \sum_k \sum_{k'} \frac{|t_k..t_{k+1} \cap t'_{k'}..t'_{k'+1}|^2}{|t_k..t_{k+1}| \times |t'_{k'}..t'_{k'+1}|}.$$

We have $0 \leq \Delta(\mathcal{T}, \mathcal{T}') \leq |\mathcal{T}| + |\mathcal{T}'|$ and $(\Delta(\mathcal{T}, \mathcal{T}') = 0) \Leftrightarrow (\mathcal{T} = \mathcal{T}')$ [Lajugie et al. (ICML, 2014)].

Surrogate loss. The empirical risk minimization is intractable, so take a surrogate loss, the structured hinge loss [Tsochantaridis et al. (JMLR, 2005)]:

$$\bar{L}_h(w) := \sum_{l=1}^{L} V_w(\mathcal{T}^{(l)}) - \min_{\mathcal{T}} \left(V_w(\mathcal{T}) - \Delta(\mathcal{T}^{(l)}, \mathcal{T}) \right)$$

Properties. The surrogate loss is **convex** w.r.t. w and is an **upper bound** on the empirical risk

$$\Delta(\mathcal{T}^{(l)}, \widehat{\mathcal{T}}_w(\mathbf{y}^{(l)})) \leq V_w(\mathcal{T}^{(l)}) - \min_{\mathcal{T}} \left(V_w(\mathcal{T}) - \Delta(\mathcal{T}^{(l)}, \mathcal{T}) \right).$$

One evaluation of $\bar{L}_h(w)$ requires a dynamic programming approach.

Final algorithm

Learning.

- Take the labels $(\mathbf{y}^{(l)}, \mathcal{T}^{(l)})$ $(l = 1,\ldots,L)$ and solve the following optimization problem:

$$\min_{w \in \mathbb{R}_+^d} [\bar{L}_h(w) + \Omega(w)]$$

where $\Omega(\cdot)$ is a regularization on the weight vector w (typically $\Omega(w) = \|w\|_1$).

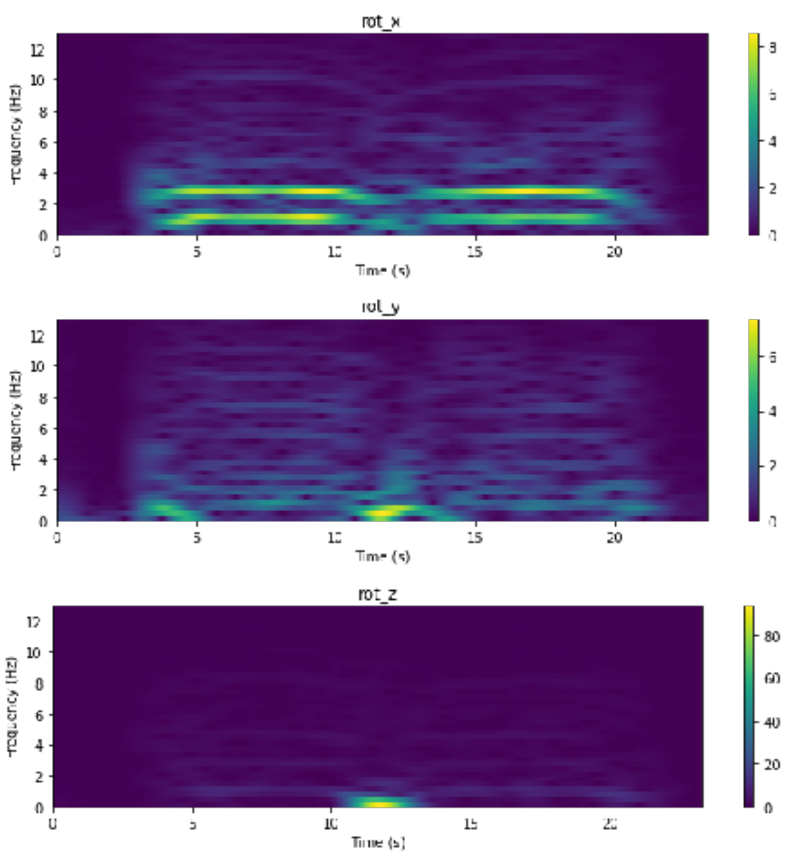
- Solved with projected sub-gradient descent as the objective function is not differentiable.

Predicting. For an out-of-sample signal $\mathbf{y}$,

- Estimate the change-points by $\widehat{\mathcal{T}}_{\hat{w}}(\mathbf{y})$, i.e. re-weight the components of $\mathbf{y}$ with the learned $\hat{w}$ and detect multivariate mean-shifts.

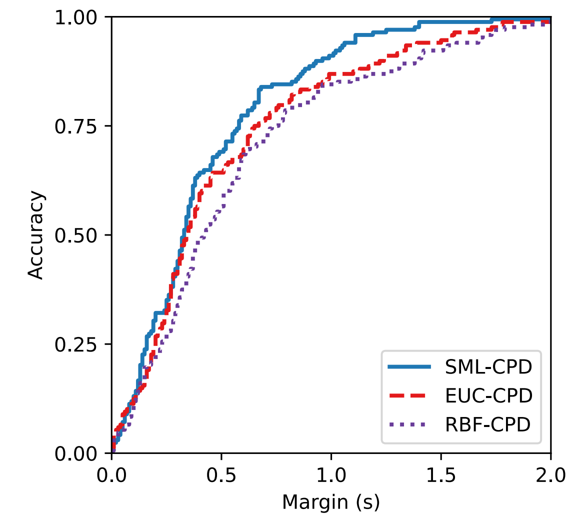
Conclusion

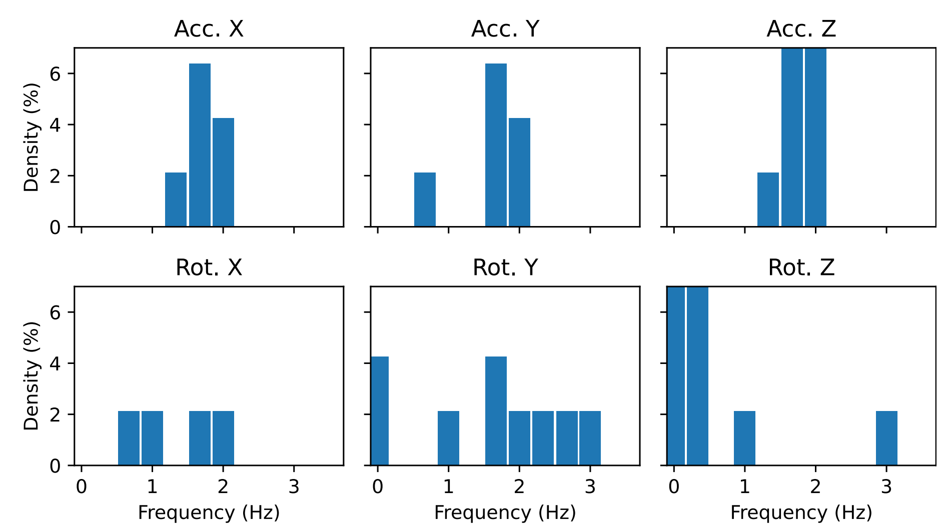
- Time-frequency representation: d = 906
- Fixed number of change-points.



Accuracy is plotted versus the allowed error margin (in seconds).

The top curve (SML-CPD) has the best accuracy for all margin levels.





Selected frequencies by SML-CPD for each dimension of the signal.
Rot. Z: low frequencies are selected.
Acc.: frequencies between 1 Hz and 2 Hz are selected.

Conclusion

- ▶ Principled way to calibrate a change-point detection method using labels
- ▶ Still rely on a user-defined family of change-point detection methods.
- ▶ Higher level objective: propose an implementation of automatic calibration procedures (supervised and unsupervised) for change-point detection: cross-validation, well-known penalties, slope heuristics, etc.