Détection de ruptures multiples – application aux signaux physiologiques

Workshop ENSPS/UM6P
Charles TRUONG (CMLA, ENS Paris-Saclay)

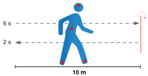
1er juillet 2019, à Ben Guerir

Introduction

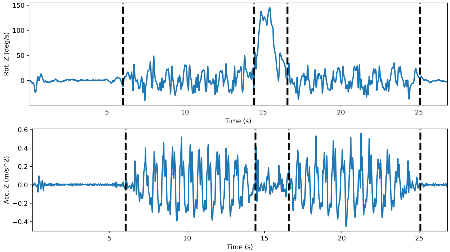
General context.

- Change point detection is a fundamental task in signal processing.
- Crucial step for the study of long non-stationary time series.
- Numerous applications exist: finance,industrial monitoring, clinical research.

Use case.



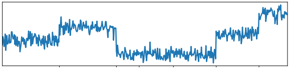
Schema of the protocol



Signal example, from a walking exercise. Rot. Z (top) and Acc. Z (bottom). Barrois-Müller, et al. (Clinical Neurophysiology, 2016)

Introduction

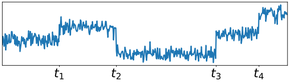
What is change point detection? Finding the segmentation that best explains the signal.



Signal $y = \{y_t\}_{t=1}^{T}$, $\mathbb{R}^d$-valued, T samples

Introduction

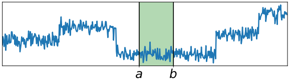
What is change point detection? Finding the segmentation that best explains the signal.



Signal	$y = \{y_t\}_{t=1}^{T}$, $\mathbb{R}^d$-valued, T samples
Segmentation, partition	Change point indexes: $\mathcal{T} = \{t_1, t_2, \ldots\}$
Number of changes	$\lvert\mathcal{T}\rvert$

Introduction

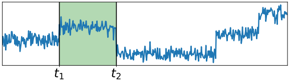
What is change point detection? Finding the segmentation that best explains the signal.



Signal	$y = \{y_t\}_{t=1}^T$, $\mathbb{R}^d$-valued, T samples
Segmentation, partition	Change point indexes: $\mathcal{T} = \{t_1, t_2, \dots\}$
Number of changes	$\lvert\mathcal{T}\rvert$
Sub-signal	$y_{a..b} = \{y_{a+1}, \dots, y_b\}$ $(0 \le a < b \le T)$

Introduction

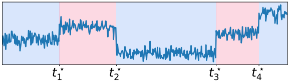
What is change point detection? Finding the segmentation that best explains the signal.



Signal	$y = \{y_t\}_{t=1}^T$, $\mathbb{R}^d$-valued, T samples
Segmentation, partition	Change point indexes: $\mathcal{T} = \{t_1, t_2, \ldots\}$
Number of changes	$\lvert\mathcal{T}\rvert$
Sub-signal	$y_{a..b} = \{y_{a+1}, \ldots, y_b\}$ $(0 \leq a < b \leq T)$
Regime, segment	- Indexes between 2 changes: $\{t_k + 1, \ldots, t_{k+1}\}$ - Samples between 2 changes: $y_{t_k..t_{k+1}}$

Introduction

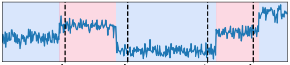
What is change point detection? Finding the segmentation that best explains the signal.



Signal	$y = \{y_t\}_{t=1}^{T}$, $\mathbb{R}^d$-valued, T samples
Segmentation, partition	Change point indexes: $\mathcal{T} = \{t_1, t_2, \ldots\}$
Number of changes	$\lvert\mathcal{T}\rvert$
Sub-signal	$y_{a..b} = \{y_{a+1}, \ldots, y_b\}$ $(0 \leq a < b \leq T)$
Regime, segment	- Indexes between 2 changes: $\{t_k + 1, \ldots, t_{k+1}\}$ - Samples between 2 changes: $y_{t_k..t_{k+1}}$
True segmentation	$\mathcal{T}^\star = \{t_1^\star, t_2^\star, \ldots\}$

Introduction

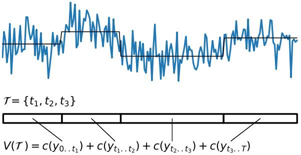
What is change point detection? Finding the segmentation that best explains the signal.



Signal	$y = \{y_t\}_{t=1}^T$, $\mathbb{R}^d$-valued, T samples
Segmentation, partition	Change point indexes: $\mathcal{T} = \{t_1, t_2, \dots\}$
Number of changes	$\lvert\mathcal{T}\rvert$
Sub-signal	$y_{a..b} = \{y_{a+1}, \dots, y_b\}$ $(0 \leq a < b \leq T)$
Regime, segment	- Indexes between 2 changes: $\{t_k + 1, \dots, t_{k+1}\}$ - Samples between 2 changes: $y_{t_k..t_{k+1}}$
True segmentation	$\mathcal{T}^\star = \{t_1^\star, t_2^\star, \dots\}$
Estimated segmentation	$\widehat{\mathcal{T}} = \{\hat{t}_1, \hat{t}_2, \dots\}$

Introduction

How to choose a segmentation?



The "best segmentation" $\widehat{\mathcal{T}}$ is the minimizer of a criterion $V(\mathcal{T})$:

$$V(\mathcal{T}) := \sum_{k=0}^{K} c(y_{t_k..t_{k+1}}).$$

Problem 1.
Fixed number K of change points:

$$\widehat{\mathcal{T}} := \arg\min_{\mathcal{T}} V(\mathcal{T}) \quad \text{s.t. } |\mathcal{T}| = K.$$

Problem 2.
Unknown number of change points:

$$\widehat{\mathcal{T}} := \arg\min_{\mathcal{T}} V(\mathcal{T}) + \text{pen}(\mathcal{T})$$

where $\text{pen}(\mathcal{T})$ measures the complexity of a segmentation $\mathcal{T}$.

Outline

Selective review

Framework

Detection methods are the combination of three elements.

Cost function | Search method | Constraint

Criterion $V(\mathcal{T})$ to minimize: $V(\mathcal{T}) := \sum_{k=0}^{K} c(y_{t_k..t_{k+1}})$.

Problem 1.
Fixed number K of change points:

$$\widehat{\mathcal{T}} := \underset{\mathcal{T}}{\arg\min}\ V(\mathcal{T}) \quad \text{s.t.} \quad |\mathcal{T}| = K.$$

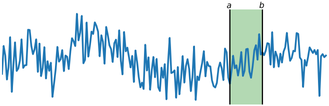
Problem 2.
Unknown number of change points:

$$\widehat{\mathcal{T}} := \underset{\mathcal{T}}{\arg\min}\ V(\mathcal{T}) + \mathrm{pen}(\mathcal{T})$$

where $\mathrm{pen}(\mathcal{T})$ measures the complexity of a segmentation $\mathcal{T}$.

Selective review

Cost function | Search method | Constraint



The cost function:

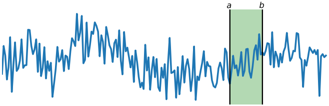
- measures homogeneity,
- encodes the type of change to detect.

(Changing parameters are highlighted)

- Change in distribution (maximum likelihood estimation) (Frick et al., 2014).
  - Multivariate Gaussian (Basseville and Nikiforov, 1993; Lavielle and Moulines, 2000; Birgé and Massart, 2007; etc.)
  - Poisson (Ko et al., 2015)
- Multiple linear model (Bai and Perron, 2003; Bai and Perron, 2006; etc.)
  - Change in an autoregressive process (Chakar et al., 2017; etc.)
- Change in distribution (non-parametric)
  - Estimation of the c.d.f. (Zou et al., 2014)
  - Rank-based statistics (Lévy-Leduc and Roueff, 2009; Lung-Yut-Fong et al., 2015)
  - Kernel-based detection (Harchaoui et al., 2008; Garreau and Arlot, 2017; Celisse et al., 2017, etc.)

Selective review

Cost function | Search method | Constraint



The cost function:

- measures homogeneity,
- encodes the type of change to detect.

(Changing parameters are highlighted)

Change in mean: $\{y_t\}_t$ such that $\mathbb{E}[y_t] = \boxed{\mu}$ (Chernoff and Zacks, 1964; Sen and Srivastava, 1975; Basseville and Nikiforov, 1993, etc.)

Cost: $c_{L_2} := \sum_{t=a+1}^{b} \|y_t - \bar{y}_{a..b}\|^2$. where $\bar{y}_{a..b}$ is the mean of $y_{a..b}$.

Change in distribution (non-parametric): $y_t \sim \boxed{F(\cdot)}$ with F a c.d.f. (Harchaoui et al., 2008; Garreau and Arlot, 2017; Celisse et al., 2017, etc.)

(Kernel setting) Denote by $k(\cdot,\cdot)$ a kernel function, by $\mathcal{H}$, the associated RKHS, by $\phi : \mathbb{R}^d \to \mathcal{H}$ the mapping function such that $\phi(y_t) = k(y_t,\cdot)$ and $\langle \phi(y_s) | \phi(y_t) \rangle_{\mathcal{H}} = k(y_s, y_t)$.

Cost: $c_{\text{kernel}} := \sum_{t=a+1}^{b} \|\phi(y_t) - \bar{\mu}_{a..b}\|^2_{\mathcal{H}}$ where $\bar{\mu}_{a..b}$ is the mean of $\{\phi(y_t)\}_{a..b}$.

Selective review

Cost function | **Search method** | Constraint

How to minimize over the space of segmentations? Exhaustive enumeration is impractical: $\mathcal{O}(T^K)$ segmentations with K changes.

Optimal resolution

For Problem 1 (fixed K) and Problem 2 (unknown K), based on **dynamic programming**, using

$$\min_{|\mathcal{T}|=K} V(\mathcal{T}) = \min_{1\leq t<T} \left[c(y_{0..t}) + V(\widehat{\mathcal{T}}^{K-1}_{t..T}) \right]$$

where $\widehat{\mathcal{T}}^{K-1}_{t..T}$ is the optimal segmentation of $y_{t..T}$ with $K-1$.

Complexity.

- Problem 1: $\mathcal{O}(T^2)$ (Bellman, 1955).
- Problem 2: $\mathcal{O}(T)$ (Killick et al., 2012).

Selective review

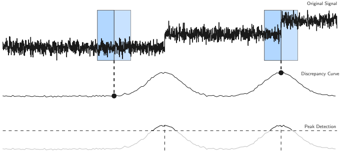
Cost function | Search method | Constraint

Approximate resolution

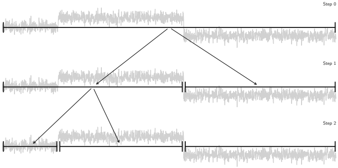
Based on *single* change point detection on sub-signals $y_{a..b}$:

$$\min_{a<t\leq b} c(y_{a..t}) + c(y_{t..b}).$$

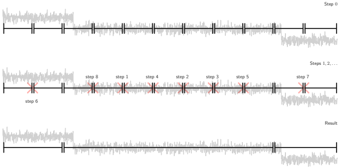
Window sliding



Binary segmentation



Bottom-up segmentation



Selective review

Cost function | Search method | **Constraint**

The correct number of changes is a trade-off between data fitting and complexity:

$$\min_{\mathcal{T}} \; V(\mathcal{T}) + \text{pen}(\mathcal{T})$$

Linear penalty. (Killick et al., 2012) Also constraint l_0.

$$\text{pen}_{l_0}(\mathcal{T}) := \beta|\mathcal{T}| \quad \text{with } \beta > 0 \text{ the smoothing parameter.}$$

Bayesian information criterion (BIC). (Yao, 1988) For instance, for c_{L_2}:

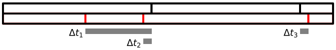
$$\text{pen}_{\text{BIC}}(\mathcal{T}) := \sigma^2 \log T \, |\mathcal{T}|.$$

Model selection. Active research under the Gaussian assumption ((Birgé and Massart, 2001; Lebarbier, 2005) or with a kernel (Arlot et al., 2012). Computationally intensive.

Evaluation framework

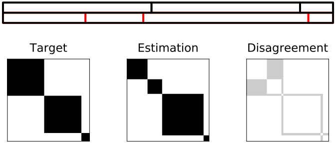
Metrics

HAUSDORFF is the worst estimation error.
Expressed in number of samples.
(Harchaoui and Lévy-Leduc, 2010)

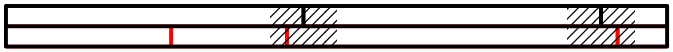


$$\text{HAUSDORFF}(\mathcal{T}^\star, \widehat{\mathcal{T}}) = \max(\Delta t_1, \Delta t_2, \Delta t_3)$$

RANDINDEX measures agreement.
Between 0 (disagreement) and +1 (total agreement).
(Da Silva et al., 2009)



F1 SCORE combines precision and recall.
A change is detected up to a user-defined margin.

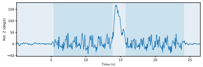


Precision is $2/3$, recall is $2/2$, F1 SCORE is $4/5$.

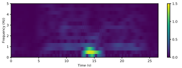
Evaluation framework

Data sets

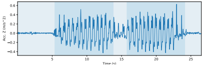
The *Gait* data set
- 262 signals
- 17-40 sec. ($f_s = 100$ Hz)
- dimension $d = 2$ (raw) or $d = 32$ (Spectrogram)
- Stand, Walk, Turnaround, Walk, Stop



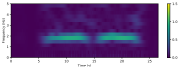
Rot. Z



Rot. Z (Spectrogram)

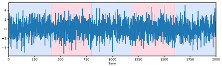


Acc. Z

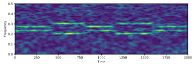


Acc. Z (Spectrogram)

The *FreqShift* data set
- piecewise periodic with noise ($d = 1$)
- 4×100 signals

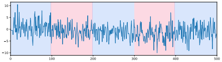


SNR $= -1$ dB



SNR $= -1$ dB (Spectrogram)

The *MeanShift* data set
- piecewise constant with noise ($d = 20$)
- 4×100 signals



Scenario 2

Scenario 1: $T = 500, \quad \sigma = 1$
Scenario 2: $T = 500, \quad \sigma = 3$
Scenario 3: $T = 2000, \quad \sigma = 1$
Scenario 4: $T = 2000, \quad \sigma = 3$

Change point detection for physiological data

Open questions.

Computational cost	In routine clinical practice, computational ressources are limited to the clinician's laptop or embedded device.
Versatility	Ability to accommodate a wide range of protocols, sensors and patient.
Automatic calibration	Translate medical expertise in statistical terms is time-consuming.

Outline

Calibrating the smoothing parameter

Unknown number of change points (Problem 2), the optimal segmentation is

$$\widehat{\mathcal{T}} := \arg\min_{\mathcal{T}} \; V(\mathcal{T}) \; + \; \beta|\mathcal{T}|.$$

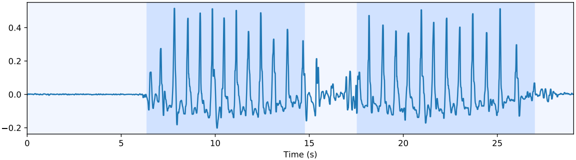
Motivations. Calibration of β:

- For c_{L_2} (mean-shifts), several closed-form expression in the literature, for instance *BIC* (Yao, 1988).
- For arbitrary cost functions, a few heuristics which consist in computing the optimal segmentation for $K = 1, \ldots, K_{\text{max}}$ (Lavielle et al., 2005).
- Supervised learning, (Hocking et al., 2013): introduced for c_{L_2}, quadratic complexity.

Calibrating the smoothing parameter

Supervised setting.

Set of signals $y^{(l)}$ and annotations (target segmentation) $\mathcal{T}^{(l)}$ ($l = 1, \ldots, L$).
Provided by an expert on a few training signals.



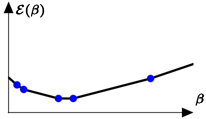
Example of an annotated signal

Calibrating the smoothing parameter

Definition. For a signal $y^{(l)}$ and its annotation $\mathcal{T}^{(l)}$, the **excess risk** $\mathcal{E}^{(l)}$ is

$$\mathcal{E}^{(l)}(\beta) := \left(V(\mathcal{T}^{(l)}) + \beta|\mathcal{T}^{(l)}|\right) - \left(\min_{\mathcal{T}} [V(\mathcal{T}) + \beta|\mathcal{T}|]\right).$$

- The excess risk is convex and $\frac{\mathrm{d}}{\mathrm{d}\beta}\mathcal{E}^{(l)}(\beta) = |\mathcal{T}^{(l)}| - |\widehat{\mathcal{T}}_\beta^{(l)}|$.
- $\mathcal{E}^{(l)}(\beta) = 0 \Rightarrow \widehat{\mathcal{T}}_\beta^{(l)} = \mathcal{T}^{(l)}$.
- Each evaluation is linear in the number of training samples (Killick et al.,2012).



`Alpin`: **Adaptive Linear Penalty Inference.** The optimal smoothing parameter $\hat{\beta}$ is

$$\hat{\beta} := \arg\min_{\beta > 0} \sum_{l=1}^{L} \mathcal{E}^{(l)}(\beta).$$

Calibrating the smoothing parameter

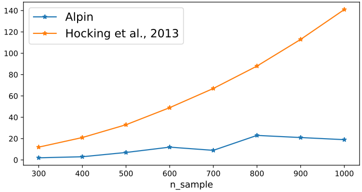
Experimental evaluation. On the *MeanShift* signals ($T = 500$, $\sigma = 3$).

	Supervised		Unsupervised	
Metric	`Alpin`	Hocking, 2013	BIC	Lavielle, 2005
Hausdorff	**8.34 (±10.61)**	9.05 (±11.34)	96.42 (±24.86)	58.06 (±18.43)
RandIndex	**0.98 (±0.01)**	**0.98 (±0.01)**	0.87 (±0.05)	0.94 (±0.01)
F1 score	**0.96 (±0.09)**	**0.96 (±0.09)**	0.72 (±0.14)	0.65 (±0.06)
$\left\lvert \lvert\widehat{\mathcal{T}}\rvert - \lvert\mathcal{T}^\star\rvert \right\rvert$	**0.08 (±0.31)**	0.12 (±0.38)	1.52 (±0.59)	4.01 (±0.10)

Hausdorff in number of samples. The margin of F1 score is $M = 10$ samples.

Comments.

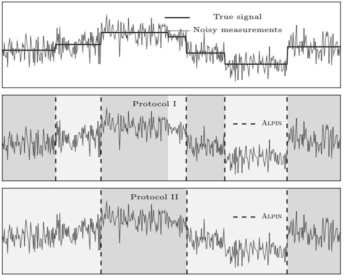
- `Alpin` performs better than unsupervised methods, and slightly better than Hocking, 2013.
- `Alpin` is the fastest supervised methods.
- `Alpin` can accommodate an arbitrary cost function.



Execution time (in minutes) for 100 signals.

Calibrating the smoothing parameter

Double label. `Alpin` adapts to the annotation.



(Top) Signal example and its noisy version ($T = 500$, $\sigma = 2$).

(Middle) Prediction and expert partition according to Protocol I.

(Bottom) Prediction and expert partition according to Protocol II.

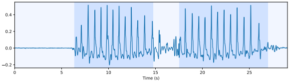
Outline

Metric learning for change point detection

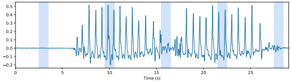
Motivations.
How to automatically choose a cost function that agrees with annotations?

Supervised setting.

Set of signals $y^{(l)}$ and annotations (target segmentation) $\mathcal{T}^{(l)}$ $(l = 1, \ldots, L)$.
Provided by an expert on a few training signals.



Example of an fully annotated signal



Example of an partially annotated signal

(Full) All change points are provided.

(Partial) Only an homogeneous portion of each regime is provided.

Metric learning for change point detection

Cost functions. Define the general class of cost functions, for a kernel k:

$$c_{\mathcal{H},M}(y_{a..b}) := \sum_{t=a+1}^{b} \|\phi(y_t) - \bar{\mu}_{a..b}\|^2_{\mathcal{H},M} \quad (0 \leq a < b \leq T)$$

where $\bar{\mu}_{a..b}$ is the mean value of $\{\phi(y_t)\}_{t=a+1}^{b}$ and $\|\cdot\|_{\mathcal{H},M}$ is the Mahalanobis-type (pseudo-)norm given by

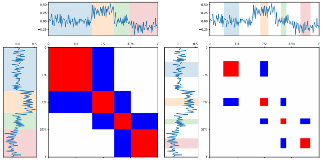
$$\|\phi(y_t)\|^2_{\mathcal{H},M} := \phi(y_t)' M \phi(y_t)$$

where M is a positive semi-definite matrix.

Metric learning for change point detection

From labels into constraints.

Similarity/dissimilarity matrix A:
$A_{st} = +1$ (red) if y_s and y_t are similar,
$A_{st} = -1$ (blue) if y_s and y_t are dissimilar,
$A_{st} = 0$ (white) otherwise.



(a) Full labels. (b) Partial labels.

Kernel-based metric learning. (Jain et al., 2012) The learned metric matrix $\widehat{M}$ is the solution to the following constrained optimization problem:

$$\min_{M \succeq 0, M = M^t} \quad D(M, I) \quad \text{s.t.} \quad \left\| \phi(y_s^{(l)}) - \phi(y_t) \right\|^2_{\mathcal{H},M} \leq u, \; y_s^{(l)} \text{ and } y_t^{(l)} \text{ similar}$$

$$\left\| \phi(y_s^{(l)}) - \phi(y_t) \right\|^2_{\mathcal{H},M} \geq v, \; y_s^{(l)} \text{ and } y_t^{(l)} \text{ dissimilar}$$

where $D(M, M_0) := \mathrm{tr}(MM_0^{-1}) - \log\det(MM_0^{-1})$ and $u, v > 0$.

Metric learning for change point detection

Computing the metric on new samples. The metric matrix is not explicitly available, but the inner-product matrix is:

$$\widehat{G} := \phi(y_s^{(l)})' \widehat{M} \phi(y_t^{(l)})$$

This is enough to compute the inner-product for any new samples z_s and z_t:

$$\phi(z_s)' \widehat{M} \phi(z_t) = k(z_s, z_t) + k_s' \, G^{-1} (\widehat{G} - G) G^{-1} \, k_t$$

where $k_\bullet := [k(z_\bullet, y_1), \; k(z_\bullet, y_2), \dots]'$ with $\{y_1, y_2, \dots\}$ the set of training samples and $G_{st} := k(y_s, y_t)$.

Metric learning for change point detection

Experimental evaluation. On the *Gait* signals (time-domain).

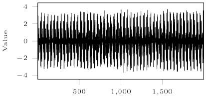
	♡`gkCPD`	`gkCPD`	♡`Win` (c_{rbf})	`Win` (c_{rbf})
Hausdorff	**0.94 (±0.58)**	1.35 (±1.35)	5.93 (±2.65)	6.01 (±2.24)
RandIndex	**0.93 (±0.03)**	0.92 (±0.05)	0.84 (±0.05)	0.83 (±0.04)
F1 score	**0.91 (±0.15)**	0.86 (±0.18)	0.62 (±0.16)	0.60 (±0.15)

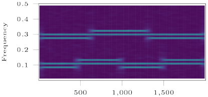
Comments.

- Supervision always improves accuracy.
- ♡`gkCPD` is the most efficient method of all tested methods.
- Raw signals as input.

Metric learning for change point detection

Double label. Partial annotations and linear kernel.

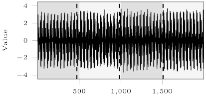


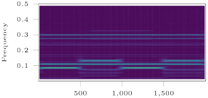


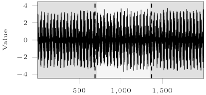
Two types of change:

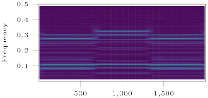
- Low frequency
- High frequency.

Comment. The detected changes depend on the input annotations.









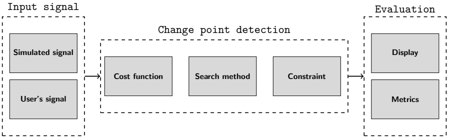
Using $M = U'U$ yields

$$\|\phi(y_t)\|_{\mathcal{H},M} = \|U\phi(y_t)\|_{\mathcal{H}}$$

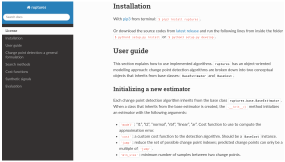
where ϕ is unsupervised (i.e. not task-specific), and U is linear and task-specific.

Outline

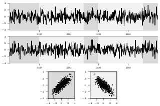
ruptures : change point detection in Python



Schematic view of the ruptures package.



Online documentation



Simulated 2D signal

```
import ruptures as rpt

# signal generation
signal, bkps = rpt.pw_normal(n_samples=500, n_bkps=4)

# change point detection
algo = rpt.Dynp(model="rbf").fit(signal)
result = algo.predict(n_bkps=4)
```

Python code.

Conclusion

Summary of contributions.

- Supervised calibration procedures which build on the clinician's expertise.
  - For the smoothing parameter (`Alpin`).
  - For the cost function (metric learning)
- A general framework for the study and comparaison of detection methods.
  - Literature review.
  - Companion Python package (`github.com/deepcharles/ruptures`).
- Data are available online (`github.com/deepcharles/gait-data`).

Summary

PyPI link	https://pypi.org/project/ruptures
Total downloads	39607
Total downloads - 30 days	5717
Total downloads - 7 days	1237

Download summary (from `pepy.tech`)

for the period January 2018 - June 2019.

Python at Netflix

Netflix Technology Blog Follow

Apr 29 · 8 min read

Mentionned by Netflix as part of their stack.

References

Bai, J. (1996). Testing for parameter constancy in linear regressions: an empirical distribution function approach. Econometrica, 64(3), 597–622.

Basseville, M., & Nikiforov, I. (1993). Detection of abrupt changes: theory and application (Vol. 104). Prentice Hall Englewood Cliffs.

Bellman, R. (1955). On a routing problem. Quaterly of Applied Mathematics, 16(1), 87–90.

Chakar, S., Lebarbier, É., Levy-Leduc, C., & Robin, S. (2017). A robust approach for estimating change-points in the mean of an AR(1) process. Bernouilli Society for Mathematical Statistics and Probability, 23(2), 1408–1447.

Da Silva, A., Lechevallier, Y., Rossi F., de A.T. De Carvahlo, F. Clustering Dynamic Web Usage Data. Innovative Applications in Data Mining, Springer Berlin Heidelberg, pp.71-82, 2009, Studies in Computational Intelligence.

Frick, K., Munk, A., & Sieling, H. (2014). Multiscale change point inference. Journal of the Royal Statistical Society. Series B: Statistical Methodology, 76(3), 495–580.

Harchaoui, Z., Bach, F., & Moulines, É. (2008). Kernel change-point analysis. In Advances in Neural Information Processing Systems 21 (NIPS 2008) (pp. 609–616). Vancouver, Canada.

Harchaoui, Z., & Lévy-Leduc, C. (2010). Multiple Change-Point Estimation With a Total Variation Penalty. Journal of the American Statistical Association, 105(492), 1480–1493.

Hocking, T., Rigaill, G., Vert, J.-P., & Bach, F. (2013). Learning sparse penalties for change-point detection using max margin interval regression. In Proceedings of the International Conference on Machine Learning (ICML) (pp. 172–180). Atlanta, USA.

Jain, P., Kulis, B., Davis, J. V., & Dhillon, I. S. (2012). Metric and kernel learning using a linear transformation. Journal of Machine Learning Research (JMLR), 13, 519–547.

Killick, R., Fearnhead, P., & Eckley, I. A. (2012). Optimal detection of changepoints with a linear computational cost. Journal of the American Statistical Association, 107(500), 1590–1598.

Ko, S. I. M., Chong, T. T. L., & Ghosh, P. (2015). Dirichlet process hidden Markov multiple change-point model. Bayesian Analysis, 10(2), 275–296.

Lavielle, M. (1999). Detection of multiples changes in a sequence of dependant variables. Stochastic Processes and Their Applications, 83(1), 79–102.

Lavielle, M. (2005). Using penalized contrasts for the change-point problem. Signal Processing, 85(8), 1501–1510.

Yao, Y.-C. (1988). Estimating the number of change-points via Schwarz' criterion. Statistics and Probability Letters, 6(3), 181–189.

Zou, C., Yin, G., Long, F., & Wang, Z. (2014). Nonparametric maximum likelihood approach to multiple change-point problems. The Annals of Statistics, 42(3), 970–1002.