Supervised change-point detection with dimension reduction, applied to physiological signals

Charles Truong
Université Paris-Saclay,
ENS Paris-Saclay, CNRS,
Centre Borelli,
F-91190 Gif-sur-Yvette, France
charles.truong@ens-paris-saclay.fr

Laurent Oudre
Université Paris-Saclay,
ENS Paris-Saclay, CNRS,
Centre Borelli,
F-91190 Gif-sur-Yvette, France
laurent.oudre@ens-paris-saclay.fr

Abstract

This paper proposes an automatic method to calibrate change point detection algorithms for high-dimensional time series. Our procedure builds on the ability of an expert (e.g. a medical researcher) to produce approximate segmentation estimates, called partial annotations, for a small number of signal examples. This contribution is a supervised approach to learn a diagonal Mahalanobis metric, which, once combined with a detection algorithm, is able to reproduce the expert's segmentation strategy on out-of-sample signals. Unlike previous works for change detection, our method includes a sparsity-inducing regularization which perform supervised dimension selection, and adapts to partial annotations. Experiments on activity signals collected from healthy and neurologically impaired patients support the fact that supervision markedly ameliorate detection accuracy

1 Introduction

The task of change-point detection, or signal segmentation, is a crucial step in numerous machine learning pipelines that handle time series. Roughly, it consists in finding the temporal boundaries of the successive regimes of a multivariate signal. There are a great deal of applications, from sleep monitoring [8], DNA sequences [4], study of neurological disorders [2], etc. Practically, the expert (e.g. an medical researcher or a biologist) must choose by themself the most suitable change-point detection procedure from the vast associated literature [12]. One particularly important parameter to select is the kind of change to detect, which is related to the signal representation or, similarly, the metric to measure the distance between samples. This calibration step is complex, time-consuming and often achieved by a trial and error. However, more often than not, the expert is able to manually segment a few signals, at least partially (i.e. give approximate change locations). For instance, Figure 1 shows the partial annotation of an expert: on a signal collected by monitoring, with an inertial sensor, a subject performing a sequence of simple activities (stand, walk, turn around, walk, stop) [2], a medical researcher has indicated a rough estimation for the activity changes. The objective of this work is to formulate a procedure to automatically learn from segmentation examples (i.e. signals and their partial annotations) an appropriate metric. Combining the learned metric with a change-point detection algorithm could then replicate the expert's segmentation strategy.

The change point detection problem. Consider a $\mathbb{R}^d$-valued signal $y = [y_1, y_2, \dots, y_T]$ with T samples. Formally, change-point detection with an fixed number K of changes consists in solving the following discrete optimization problem

$$\{\hat{t}_1, \hat{t}_2, \dots, \hat{t}_{\hat{K}}\} = \underset{\{t_1,t_2,\dots,t_K\}}{\arg\min} \left[\sum_{k=0}^{K} \sum_{t=t_k}^{t_{k+1}-1} \left\| y_t - \bar{y}_{t_k..t_{k+1}} \right\|^2 \right] \tag{1}$$

36th Conference on Neural Information Processing Systems (NeurIPS 2022).

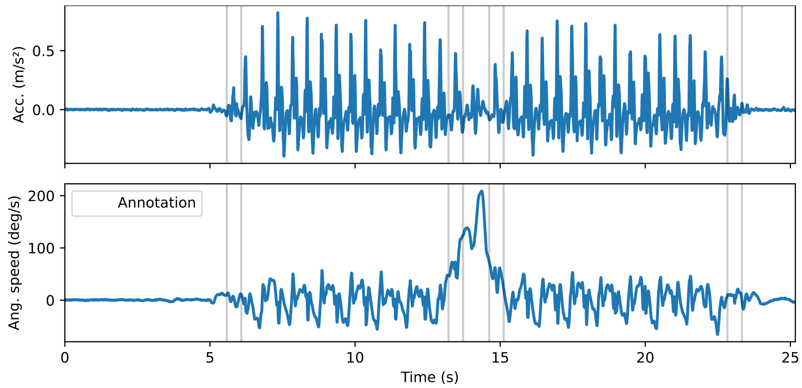


Figure 1: Signal example with partial annotation. Two (out of 6) dimensions of a gait signal (acceleration and angular speed along one axis) are shown (see Section 3 for details). The annotation here is partial (the exact location of the change is not provided); annotated portions are 0.5 second long.

where $y_{a..b}$ is the empirical mean of the sub-signal $\{y_t\}_{t=a}^{b-1}$ and $t_0 := 1$ and $t_{K+1} = T+1$ are dummy indexes and $\|\cdot\|$ is a user-defined norm on $\mathbb{R}^d$ (e.g. the Euclidean norm). The indexes $\{\hat{t}_1, \hat{t}_2, \dots, \hat{t}_{\hat{K}}\}$ are the instants when the signal has the most significant mean-shifts. A number of methods have been developed to optimize this sum of residuals (see [12] for a review). Algorithms based on dynamic programming solves Problem 1 exactly with a complexity of $\mathcal{O}(dKT^2)$. This is the method that will be adopted. Any faster but approximate methods such as window-based procedures and binary segmentation could be used instead, depending on the operational constraints.

Learning a relevant metric. This article focuses on the calibration of the norm $\|\cdot\|$, which is related to the type of change that can be detected. The chosen norm is a Mahalanobis-type (pseudo-)norm $\|\cdot\|_M$ such that $\|x\|_M^2 := x^\intercal M x$ ($\forall x \in \mathbb{R}^d$) where $M \in \mathbb{R}^{d\times d}$ a positive semi-definite (psd) matrix. In this work, only diagonal matrix are considered, i.e. $M = \operatorname{diag}(w)$ for a vector $w \in \mathbb{R}_+^d$ of positive weights. Calibration reduces to finding an appropriate w, which can be seen as a scaling of each dimension p by w_p. If w is sparse, the learn metric even performs variable selection. Replacing in (1), the norm $\|\cdot\|$ by $\|\cdot\|_{\operatorname{diag}(w)}$ with a properly calibrated w is particularly important for high-dimensional signals which might contain noisy components that alter change detection algorithms. While there are many articles focused on calibrating segmentation methods in a unsupervised way [12], only a few works have tackle this problem from a supervised standpoint [7, 11] and none, to the best of our knowledge, have added a sparsity constraints to the weight vector w.

Contributions. We propose a procedure to learn from expert labels an appropriate norm that can both replicate the segmentation strategy of the expert and and select important signal dimensions for this task, thanks to a sparsity regularization. This procedure is applied on physiological signals, collected from healthy and neurologically impaired patients to asses their gait.

2 Method

The procedure consists in two steps: (i) a learning step during which, a weight vector $\hat{w}$ is learned using labelled signals, and (ii) a prediction step during which a change-point detection procedure is applied to out-of-sample signals to segment them. This section formally introduces the nature of the labels provided by the expert and the metric learning approach.

2.1 From annotations to constraints.

Annotations are provided by an expert and transformed into triplet constraints, which are then fed to the sparse metric learning algorithm. Annotations can either be full or partial.

Full annotations. For each training signal $y^{(l)}$, a full label consists in the set of change points $\mathcal{T}^{(l)} = \{t_1^{(l)}, t_2^{(l)}, \dots\}$. The set $\mathcal{T}^{(l)}$ includes all the changes contained in the signal $y^{(l)}$, according to the expert.

Partial labels. For each training signal $y^{(l)}$, a partial label consists in the set of intervals $\mathcal{S}^{(l)} = \{[s_1^{(l)}, e_1^{(l)}], [s_2^{(l)}, e_2^{(l)}], \dots\}$ that contain a change point. Instead of giving the exact position of a change $t_k^{(l)}$, the expert only provides an approximate position $[s_k^{(l)}, e_k^{(l)}]$ such that $t_k^{(l)} \in [s_k^{(l)}, e_k^{(l)}]$. All the changes contained in the signal $y^{(l)}$, according to the expert, are in one of the intervals of the set $\mathcal{S}^{(l)}$. Each interval $[s_k^{(l)}, e_k^{(l)}]$ contains only one change and the intervals do not overlap. An example of partial annotations is shown on Figure 1.

From label to triplets of samples. The proposed metric learning procedure relies on triplets of samples (anchor sample, positive sample, negative sample). Using a full label $\mathcal{T}^{(l)}$, a triplet can be created as follows: for any anchor sample y_t in a certain segment $[t_k^{(l)}, t_{k+1}^{(l)}[$, a positive sample is any element of the same segment y_{t^+} of $[t_k^{(l)}, t_{k+1}^{(l)}[$ (except the anchor sample) and a negative sample y_{t^-} is any element of the previous segment $[t_{k-1}^{(l)}, t_k^{(l)}[$ or the following segment $[t_{k+1}^{(l)}, t_{k+2}^{(l)}[$. Using a partial label $\mathcal{S}^{(l)}$, a triplet can be created similarly: for any anchor sample y_t in a certain segment $[e_k^{(l)}, s_{k+1}^{(l)}]$, a positive sample is any element of the same segment y_{t^+} of $[e_k^{(l)}, s_{k+1}^{(l)}]$ (except the anchor sample) and a negative sample y_{t^-} is any element of the previous segment $[e_{k-1}^{(l)}, s_k^{(l)}]$ or the following segment $[e_{k+1}^{(l)}, s_{k+2}^{(l)}]$. Intuitively, two samples that belong to the same homogeneous segment (i.e. without change) are from the same class, while two samples that belong to two consecutive segments (i.e. separated by a change point) are from different classes.

2.2 Sparse metric learning

Let $\mathcal{D}^{(l)}$ be the set of triplets generated from the labels (full $\mathcal{T}^{(l)}$ or partial $\mathcal{S}^{(l)}$). The sparse metric learning procedure for change-point detection consists in solving the following optimization problem

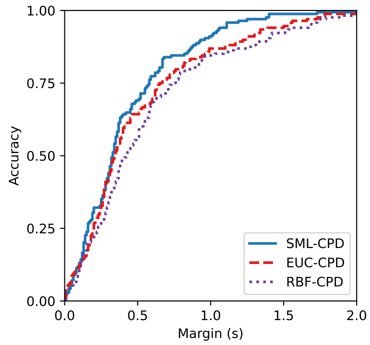
$$\min_{w \in \mathbb{R}_+^d} \left[\left(\sum_l \frac{1}{|\mathcal{D}^{(l)}|} \sum_{(y_t, y_{t^+}, y_{t^-}) \in \mathcal{D}^{(l)}} \left(1 + \|y_t - y_{t^+}\|_w^2 - \|y_t - y_{t^-}\|_w^2 \right)_+ \right) + \lambda \|w\|_1 \right] \quad (2)$$

where $[\cdot]_+ = \max(0, \cdot)$ and $\lambda > 0$ controls the trade-off between the sparsity of w and the triplet constraints. This is simply the sum over the training set of the margin-based hinge loss and a sparsity inducing regularization. The learned $\hat{w}$, which is the solution of Problem 2, is then such that the distance between samples from the same segment is smaller than the distance between samples from consecutive regimes (separated by a change-point). Because there can be a large number of possible triplets in $\mathcal{D}^{(l)}$, learning a weight vector w can be computationally costly. A sampling strategy is frequently used to focus to reduce the computational burden; such a strategy is often called triplet mining [6]. In this work, a fixed number of triplets is simply sampled at random from each set $\mathcal{D}^{(l)}$. Also, Problem 2 includes a non-smooth regularization and large number of triplet constraints, stochastic composite optimization has been proposed [9]. This is an iterative minimization algorithm where each step is a stochastic gradient step followed by the application of a proximal operator (for the ℓ_1 norm). This work uses the implementation of [3].

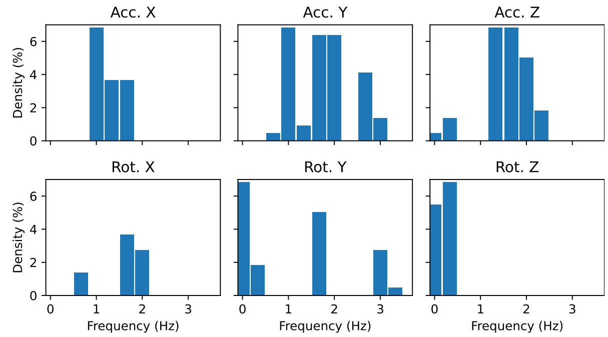
3 Applications on physiological signals

In the following, our method is denoted *SML-CPD* for Sparse Metric Learning for Change-Point Detection.

Data. The *Gait* data set contains 42 labelled time series (sampling frequency: 100 Hz) from an inertial sensor placed at the lower back of a subject performing a fixed sequence of simple activities: "Stand", "Walk", "Turnaround", "Walk", "Stop". The objective is to detect the time indexes at which the activity of the subject changes (each signal has 4 change-points). The time series have $d = 6$ dimensions: the accelerations (m/s^2) along three axes (X, Y and Z) and the angular velocities (deg/s) around the same three axes. Figure 1 shows an example (only two dimensions are displayed). The time-frequency representation is the short-term Fourier transform (STFT), computed with 300 samples per segment and an overlap of 299 samples, of each dimension; the concatenation of all STFT yields a $d = 906$-dimensional signal. As for the partial annotations, a medical researcher used an annotation tool to provide portions of 50 samples (0.5 s) around activity's changes.



(a) Accuracy vs error margin



(b) Distribution of selected frequencies

Figure 2: (a) Accuracy is plotted versus the allowed error margin (in seconds). The top curve (SML-CPD) has the best accuracy for all margin levels. (b) Selected frequencies by SML-CPD for each dimension of the signal.

Detection algorithms. Our method *SML-CPD* is compared to two common change-point detection algorithms: *EUC-CPD* which is equivalent to *SML-CPD* without the sparse metric learning step (i.e. the norm $\|\cdot\|_w$ reduces to the Euclidean norm $\|\cdot\|$) and *RBF-CPD* which is a kernel-based segmentation procedure that can detect general changes in the distribution of the samples [1]. The chosen kernel is the radial basis function (RBF). Note that both *SML-CPD* and *EUC-CPD* are applied on the time-frequency representation of the signal, while *RBF-CPD* is applied on the original data[1].

Evaluation metrics. The detection power is evaluated with the accuracy which is the proportion of correctly detected changes. For a given margin $M > 0$, a true change t is considered detected if the estimated change-point $\hat{t}$ is such that $|t - \hat{t}| < M$. All scores are computed with a 5-fold cross-validation.

Results A number of observations can be made from the results:

- *Supervision improves detection accuracy.* The cross-validated accuracy is shown on Figure 2-a. The accuracy curve can be read like a ROC curve: here, *SML-CPD* has the highest curve and outperforms other methods, meaning that supervision markedly improves the detection at all margins. For a reasonable margin $M = 1$ s, accuracies are 91.1% for *SML-CPD*, 86.9% for *EUC-CPD*, 83.9% for *RBF-CPD*.
- *Our method projects the signals into a low dimension space.* The number of non-zero coefficients in the learned w of *SML-CPD* is around 15 in the different folds of the cross-validation, meaning that only 15 dimensions are kept to perform the segmentation, compared to the 906 dimensions of the original STFT.
- SML-CPD *provides useful insights on the segmentation.* The learned weight vector $\hat{w}$ in *SML-CPD* helps the expert understand the important dimensions to segment their signals. The distribution of the selected dimensions/frequency in the STFT is displayed on Figure 2-b First, even though possible frequencies range from 0 Hz to 50 Hz, no frequency above 3 Hz was ever chosen. Second, for the accelerations, most frequencies are picked from the [1 Hz - 2.5 Hz] band; this corresponds to the frequency of the main phenomenon during the walk: the repetitions of footsteps. A footstep lasts about 0.8 second for healthy subjects and less for neurological impaired patients (both are present in the *Gait* data set). Third, for the angular velocities, most of the selected frequencies are below 0.5 Hz. This is consistent with the behaviour of the signal during the turnaround: there is a relatively smooth peak in the angular velocity (see Figure 1) which is visible at frequencies below 0.5 Hz.

4 Conclusion and future work

In this paper, we have presented a procedure to integrate expert's annotations to improve change-point detection algorithms, without resorting to a time-consuming calibration by trial and error. In addition, the learned metric performs an informative dimension selection. In future works, we will tackle the situation where the number of changes is unknown, by combining existing approaches [5, 10] with our own. Also, more complex transformations of the signal could be considered, e.g. neural networks.

[1] We use the Python package "ruptures" [12] for the segmentation algorithms.

Acknowledgments and Disclosure of Funding

Charles Truong is funded by the Industrial Data Analytics and Machine Learning (IDAML) chair of ENS Paris-Saclay. Part of this work has been funded by Region Ile-de-France. Part of the computations has been executed on Atos Edge computer.

References

[1] S. Arlot, A. Celisse, and Z. Harchaoui. A kernel multiple change-point algorithm via model selection. *Journal of Machine Learning Research*, 20(162):1–56, 2019.

[2] R. Barrois, Th. Gregory, L. Oudre, Th. Moreau, Ch. Truong, A.A. Pulini, A. Vienne, Ch. Labourdette, N. Vayatis, S. Buffat, A. Yelnik, C. De Waele, S. Laporte, P.P. Vidal, and D. Ricard. An automated recording method in clinical consultation to rate the limp in lower limb osteoarthritis. *PLoS ONE*, 11(10), 2016.

[3] W. De Vazelhes, C. J. Carey, Y. Tang, N. Vauquier, and A. Bellet. metric-learn: metric mearning algorithms in Python. *Journal of Machine Learning Research (JMLR)*, 21:1–6, 2020.

[4] T. D. Hocking, G. Rigaill, P. Fearnhead, and G. Bourque. Constrained dynamic programming and supervised penalty learning algorithms for peak detection in genomic data. *Journal of Machine Learning Research (JMLR)*, 21:1–40, 2020.

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

[6] M. Kaya and H. S. Bilge. Deep metric learning : a survey. *Symmetry*, 11(9):1066, 2019.

[7] R. Lajugie, F. Bach, and S. Arlot. Large-margin metric learning for constrained partitioning problems. In *Proceedings of the 31st International Conference on Machine Learning (ICML)*, pages 297–395, Beijing, China, 2014.

[8] Mathias Perslev, Sune Darkner, Lykke Kempfner, Miki Nikolic, Poul Jørgen Jennum, and Christian Igel. U-Sleep: resilient high-frequency sleep staging. *npj Digital Medicine*, 4(1):1–12, 2021.

[9] Y. Shi, A. Bellet, and F. Sha. Sparse compositional local metric learning. In *Proceedings of the AAAI Conference on Artificial Intelligence (AAAI)*, volume 28, pages 2078–2084, Québec City, Québec, Canada, 2014.

[10] C. Truong, L. Oudre, and N. Vayatis. Penalty learning for changepoint detection. In *Proceedings of the European Signal Processing Conference (EUSIPCO)*, pages 1569–1573, Kos, Greece, 2017.

[11] C. Truong, L. Oudre, and N. Vayatis. Supervised kernel change point detection with partial annotations. In *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 1–5, Brighton, UK, 2019.

[12] C. Truong, L. Oudre, and N. Vayatis. Selective review of offline change point detection methods. *Signal Processing*, 167, 2020.