Penalty learning for changepoint detection

Conference Paper · August 2017

DOI: 10.23919/EUSIPCO.2017.8081473

CITATIONS
4

READS
318

3 authors:

Charles Truong

11 PUBLICATIONS **391** CITATIONS

SEE PROFILE

Laurent Oudre

Ecole Normale Supérieure Paris-Saclay

79 PUBLICATIONS **784** CITATIONS

SEE PROFILE

Nicolas Vayatis

Ecole normale supérieure Paris-Saclay

191 PUBLICATIONS **2,731** CITATIONS

SEE PROFILE

Penalty Learning for Changepoint Detection

Charles Truong*†, Laurent Oudre‡†, Nicolas Vayatis*†
*CMLA, ENS Cachan, CNRS, Université Paris-Saclay, 94235, Cachan, France
†COGNAC G, University Paris Descartes, CNRS, 75006 Paris, France
‡L2TI, University Paris 13, 93430 Villetaneuse, France

***Abstract*—We consider the problem of signal segmentation in the setup of supervised learning. The supervision lies here in the existence of labelled changepoints in a historical database of similar signals. Typical segmentation techniques rely on a penalized least square procedure where the smoothing parameter is fixed arbitrarily. We introduce the ALPIN (Adaptive Linear Penalty INference) algorithm to tune automatically the smoothing parameter. ALPIN has linear complexity with respect to the sample size and turns out to be robust with respect to noise and diverse annotation strategies. Numerical experiments reveal the efficiency of ALPIN compared to state-of-the-art methods.**

I. Introduction

The task of changepoint detection consists in retrieving the time stamps where a signal undergoes abrupt changes or, equivalently, in finding the different regimes a signal is composed of [1], [2]. Signal segmentation is a standard preprocessing step in numerous signal processing tasks such as indexation or feature extraction, which are often built on the assumption that only one phenomenon is observable in the signal. In several fields such as geology, bioengineering, or biology the segmentation is still performed manually by experts, who are able to pinpoint where changepoints occur in a signal thanks to experience and a high level understanding of the underlying phenomenon which causes them. Nevertheless, this task is fastidious and time-consuming and several approaches for automatic segmentation based on statistical tests, global optimization or data modelling have been introduced to address this problem.

The main limitations to the use of these approaches in real-life situations is that those methods are seldom fully automatic and rely on several parameters (like the statistical confidence level [1], [3], the number of regimes [4], [5], etc.). The manual tuning of these parameters is more art than science, especially when the data are noisy, does not fit a standard model or when the end user has particular expectations (for example detecting only changepoints of a certain magnitude). Furthermore, this manual tuning by trial-and-error is often suboptimal as it only explores a fraction of the parameter space and a limited number of train signals. Finding a quantitative means to calibrate segmentation algorithms to best match the user's expectations is a subject of active research [6]–[10]. Some heuristics have been introduced that can provide a rough estimation of the parameters to be tuned. Recent methods have used supervised machine learning techniques to adapt parameters to the data by using annotated databases [6], [11] with promising results for DNA fragmentation.

This article follows this approach and proposes a novel method to fully automatize a signal segmentation algorithm. In the setup of supervised learning, the smoothing parameter of a penalized least square procedure is learned by optimizing a convex loss function measuring the differences between the labels and the detected changepoints. We show that this low-complexity procedure adapts to the annotation strategy of the expert. This information can then be used on new unlabeled data to provide segmentation results that are coherent with the expert specifications.

In Section II, the framework of signal segmentation is described, as well as existing methods to calibrate the penalty parameter. Then in Section III we propose a scheme to learn it from a data set of annotated signals. The experimental methodology is described with detail in Section IV and the results are commented in Section V.

II. Background

In this section, we present the problem of penalized changepoint detection and the existing penalties that can be found in the literature.

A. Penalized changepoint detection

Let $y \in \mathbb{R}^n$ be a signal of n samples, corrupted by an additive noise $\epsilon \sim \mathcal{N}(\mathbf{0}, \sigma^2 \mathbf{I}_n)$. We assume that y is not stationary and is composed of several successive regimes. Let $\mathcal{P}$ be the set of partitions of $\{1, \dots, n\}$ which consist solely of integer intervals. For a partition $A = \left\{\underline{a}_1, \dots, \underline{a}_{|A|}\right\} \in \mathcal{P}$, $\underline{a}_1, \dots, \underline{a}_{|A|}$ represent different regimes. The number of regimes in A is $|A|$ and $|A| - 1$ the number of changepoints. The aim of changepoint detection is to retrieve the different regimes present in the signal and the times where the signal switches from one regime to another.

For a given partition $A \in \mathcal{P}$, a natural measure of the approximation quality is the empirical quadratic risk [7]–[9]:

$$R(y, A) := \sum_{\underline{a} \in A} \sum_{i \in \underline{a}} (y_i - \bar{y}_{\underline{a}})^2 \tag{1}$$

where y_i is the i-th sample and $\bar{y}_{\underline{a}}$ is the mean value of y on segment $\underline{a}$. The relationship between a signal and its approximation on a partition is illustrated on Figure 1. This quantity can be made arbitrarily close to zero by choosing a partition with a sufficient number of regimes. Therefore, it can only be minimized if the number of regimes $|A|$ is known, which is often not the case in real-life scenario.

ISBN 978-0-9928626-7-1

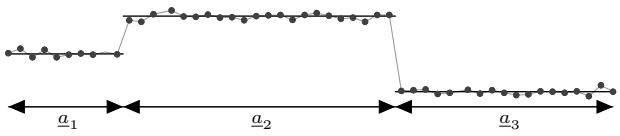


Fig. 1. Signal example. Dotted line: signal example $y \in \mathbb{R}^{100}$. Thick line: best piecewise approximation with underlying partition $A = \{\underline{a}_1, \underline{a}_2, \underline{a}_3\}$.

When $|A|$ is unknown, a penalized empirical quadratic risk can be used instead

$$\arg\min_{A \in \mathcal{P}} R(y, A) + \text{pen}(A), \tag{2}$$

where $\text{pen}(\cdot)$ denotes a suitable nonnegative function defined on $\mathcal{P}$. The penalty term allows to control the balance between signal approximation and model complexity. The function $\text{pen}(\cdot)$ is increasing when the partition A is more complex (e.g. $|A|$ increases). The objective of the method presented in this article is the learn a suitable function $A \mapsto \text{pen}(A)$ such that the result of (2) agrees with the expert's annotation.

B. Examples of penalties

Several penalty terms have been used in the literature, either justified from theoretical assumptions or inferred from data.

- **BIC** ([12]). Also known as the Schwarz's criterion, the BIC penalty has been thoroughly used for model selection.

$$\text{pen}(A) = \sigma^2 \log(n)|A| \tag{3}$$

  This penalty depends on the variance of the noise (which has to be estimated beforehand), the number of changepoints and the signal length. By design it maximizes an asymptotic approximation of the posterior probability with a uniform prior. The same parameters are used in various penalties [13], [7].
- **Hocking, 2013** ([6], [11]). The main originality of this penalty is that the constants w_1, w_2, w_3 are learned from annotated data.

$$\text{pen}(A) = e^{w_1 \log(n) + w_2 \log(\sigma) + w_3}|A| \tag{4}$$

  More precisely, the objective is to find the parameters w_1, w_2, w_3 that minimize the annotation error, which is the difference between the number of regimes of the solution of (2) and the number of regimes given by the expert. Since the annotation error is not convex and cannot be optimized as such, the authors propose to use a convex relaxation of the annotation error, which requires the computation of a large number of segmentations. The method presented in this work differs essentially on the mapping and skips the initial calculation of all segmentations.
- **Lavielle, 2005** ([9]). This penalty does not have a closed form but consists in a heuristic. The best empirical quadratic risk is computed for an increasing number of regimes As soon as adding a changepoint does not significantly reduce this quantity, the method outputs the current number of regimes as the final segmentation.

III. Method

In this section we introduce the ALPIN algorithm, which aims at efficient supervised segmentation in the case of linear penalties thanks to the empirical estimation of the smoothing parameter.

A. Penalized changepoint detection with linear penalty

We consider here the case of linear penalties, which is the most common in the literature [9], [12], [14]. The penalty term is assumed to be linear in the number of regimes, *i.e.*

$$\text{pen}(A) := \beta|A| \quad \text{with } \beta > 0. \tag{5}$$

The smoothing parameter β controls the trade-off between model complexity and goodness of fit. Low values of β favour partitions with many regimes and high values of β discard most changepoints.
Considering a signal $y \in \mathbb{R}^n$, the penalized changepoint detection problem of (2) now consists in finding the β-optimal partition $\hat{A}_\beta(y)$ defined as :

$$\begin{aligned} \hat{A}_\beta(y) &:= \arg\min_{A \in \mathcal{P}} R_\beta(y, A) \\ \text{with } R_\beta(y, A) &:= R(y, A) + \beta|A|. \end{aligned} \tag{6}$$

Interestingly, there exists an efficient algorithm to recover the β-optimal partition, whose complexity is in $\mathcal{O}(n)$ ([15]).

B. Excess penalized risk

Our goal is to learn the value for β to use in (6) in order to output a partition that agrees with the expert view. The approach described in this paper assumes that expert annotation on past signals are available. For an annotated signal $y \in \mathbb{R}^n$, the expert annotation is a partition $A^{\text{lab}}(y) \in \mathcal{P}$. The objective is to find a value of $\beta \in]0, +\infty]$ such that the β-optimal partition $\hat{A}_\beta(y)$ coincides with the expert partition $A^{\text{lab}}(y)$. We introduce the excess penalized risk

$$\mathcal{E}(y, \beta) := R_\beta(y, A^{\text{lab}}(y)) - \min_{A \in \mathcal{P}} R_\beta(y, A). \tag{7}$$

For illustration purposes, a view of the excess penalized risk of a signal is shown on Figure 2. This quantity will serve as a loss function in the supervised learning strategy described in the following section. It is the difference between the empirical quadratic risk of the expert partition and the one of the β-optimal partition. The excess penalized risk is always nonnegative and the function $\beta \mapsto \mathcal{E}(y, \beta)$ is convex for any signal $y \in \mathbb{R}^n$. Indeed, for a fixed partition $A \in \mathcal{P}$ the function $\beta \mapsto R_\beta(y, A)$ is affine, so the function $\beta \mapsto \min_{A \in \mathcal{P}} R_\beta(y, A)$, which is the pointwise minimum of a finite set of affine functions, is concave. The excess penalized risk is by definition an affine function minus a concave function, thus it is convex. As a result, any method for convex optimization can be used to compute the true minimum of the excess penalized risk way relative to β. In this article we use a limited memory quasi-Newton algorithm [16], which requires the computation of the excess penalized risk at each iteration. As stated in Section III-A, the β-optimal partition can be computed in

ISBN 978-0-9928626-7-1 © EURASIP 2017

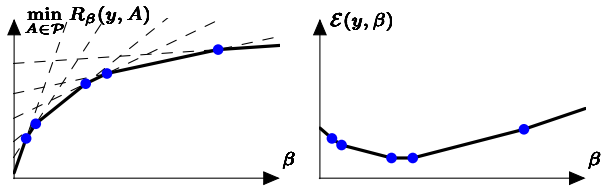


Fig. 2. (Top) For a given signal $y \in \mathbb{R}^n$ the minimum empirical risk over all partitions $A \in \mathcal{P}$ is plotted versus the penalty level. Dashed lines are empirical risks for a few partitions. (Bottom) The corresponding excess penalized risk is plotted versus the penalty level.

$\mathcal{O}(n)$ operations thus so does the excess penalized risk. This learning strategy is more efficient than "Hocking, 2013" which requires $\mathcal{O}(n^2)$ operations [6]. When predicting changepoints, BIC, "Hocking, 2013" and ALPIN achieve this task in linear time, contrary to "Lavielle, 2005" which has quadratic complexity [9].

C. Estimation of the smoothing parameter

Given N annotated signals $y^{(1)}, \dots, y^{(N)}$ and their associated expert partitions $A^{\text{lab}}(y^{(1)}), \dots, A^{\text{lab}}(y^{(N)})$, we infer the penalty β which yields the lowest average excess penalized risk:

$$\beta_{\text{opt}} := \arg\min_{\beta>0} \frac{1}{N} \sum_{i=1}^{N} \mathcal{E}(y^{(i)}, \beta). \tag{8}$$

Since the contributions of each signal $y^{(i)}$ to the average excess penalized risk is independent, all calculations can be done in parallel. In order to improve the performances of the optimization procedure, we initialize the algorithm by randomly picking a signal from the training database. The penalty that minimizes its excess penalized risk is found and this value is used as a warm start for the optimization of the global average excess penalized risk.

IV. EXPERIMENTS

This section presents the experimental setting used for analysing the performance of the ALPIN algorithm.

A. Data set

For testing, we use a synthetic data set constructed as follows. A set of 100 piecewise constant functions from $[0, 1]$ to $\mathbb{R}$ are simulated, with a number of changepoints randomly chosen between 3 and 7. The length of each regime is drawn uniformly between 0.05 and 0.3 and the jumps between regimes have random amplitudes between 1 and 5. Those functions are then sampled on an equispaced grid of n points and corrupted by a Gaussian noise of variance σ^2. The expert partitions are defined as the true partitions used to simulate the data.

B. Performance metrics

The performance on the data set is assessed with several metrics. Let $A \in \mathcal{P}$ and A^{lab} be respectively the partition output by the algorithm and the expert partition. Let $a_1, \dots, a_{|A|-1}$ and $a_1^{\text{lab}}, \dots, a_{|A^{\text{lab}}|-1}^{\text{lab}}$ their respective associated changepoints.

- HAUSDORFF metric is large when a changepoint from either A or A^{lab} is far from every changepoint of A^{lab} or A respectively [5]. Oversegmentation as well as undersegmentation is penalized.
- PRECISION and RECALL measure the ability of the method to find the correct change times (with a tolerance of 10 samples). Oversegmentation of a signal causes the precision to be close to zero and the recall close to one. Undersegmentation has the opposite effect.
- ANNOTATIONERROR measures the error made in estimating the number of changepoints [6].
- RANDINDEX has been introduced to evaluate clustering methods [5]. It measures the ability of the method to assign the samples to the same regime than in the expert partition.

The formulas of all metrics are summarized in Table I.

TABLE I
METRIC FORMULAS.

Name	Value
HAUSDORFF	$\max\{\max_i \min_j \lvert a_i - a_j^{\text{lab}}\rvert, \max_j \min_i \lvert a_i - a_j^{\text{lab}}\rvert\}$
PRECISION	$\frac{\text{card}\{a_i^{\text{lab}} \text{ s.t. } \min_j \lvert a_i^{\text{lab}} - a_j\rvert < 10\}}{\lvert A\rvert - 1}$
RECALL	$\frac{\text{card}\{a_i^{\text{lab}} \text{ s.t. } \min_j \lvert a_i^{\text{lab}} - a_j\rvert < 10\}}{\lvert A^{\text{lab}}\rvert - 1}$
ANNOTATIONERROR	$\big\lvert \lvert A\rvert - \lvert A^{\text{lab}}\rvert \big\rvert$
RANDINDEX	$\frac{\text{card}\{(s,t) \lvert A^{\text{lab}} \text{ and } A \text{ agree on } (s,t)\}}{n(n-1)}$

V. RESULTS

In this section, we first compare the ALPIN algorithm to a standard changepoint detection method using sliding windows. Then we compare the penalty learned with ALPIN to the state-of-the-art penalties described in Section II-B. Finally, we discuss the adaptivity of the method to the expectations of the expert.

A. Comparison with a standard procedure

The most common changepoint detection method does not use learning nor optimization, but a statistical test computed on a sliding window. This method, has been extensively used for online segmentation in various application fields [1], [17], [18]. For each sliding window, a Student-s T-test is performed between the samples in the first half of the window and those in the last half. If the test is positive, *i.e.* if a significant difference is detected between the two sets of samples, a changepoint is detected. In our implementation of this method, the confidence level is set at 95%, the real value of the noise variance σ^2 is supposed to be known and the window length is equal to the smallest segment found in the database (i.e. $0.05 \times n$ samples).

The average PRECISION and RECALL on the synthetic data set are computed for this method as well as for the ALPIN method, for different noise level σ and different signal lengths n. For the ALPIN method these metrics are computed using a 10-fold cross-validation. Results shown on Figure 3 show that ALPIN displays better performances the noise level, which illustrates the relevance of the learning strategy. It is also noticeable that the t-test method is sensitive to the noise level :

ISBN 978-0-9928626-7-1 © EURASIP 2017

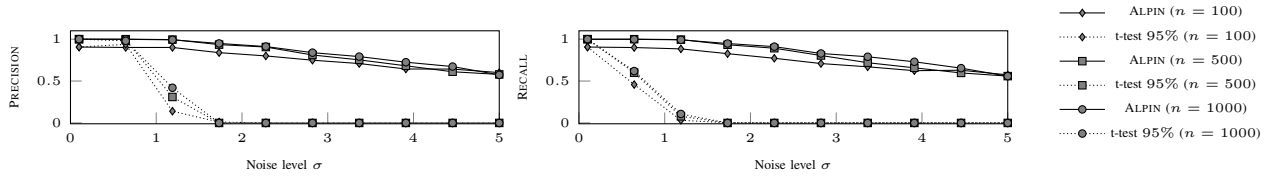
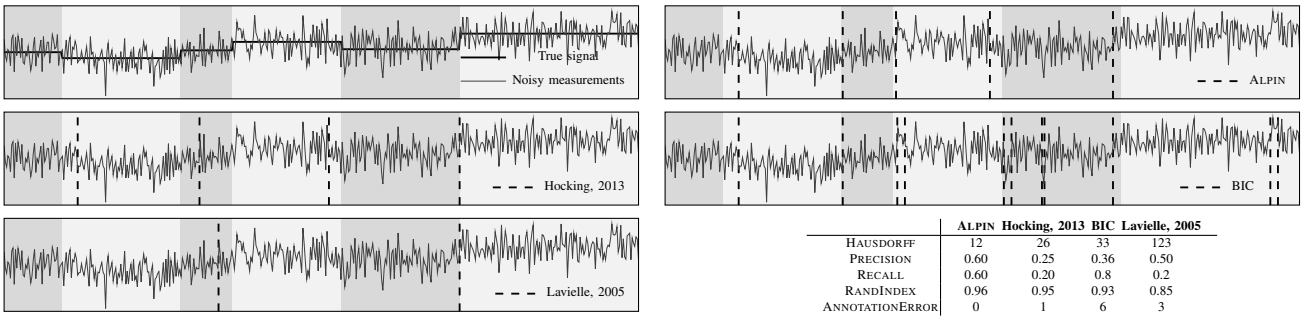


Fig. 3. Performances of ALPIN and the Student's t-test for several combinations of (n, σ).



	ALPIN	Hocking, 2013	BIC	Lavielle, 2005
HAUSDORFF	12	26	33	123
PRECISION	0.60	0.25	0.36	0.50
RECALL	0.60	0.20	0.8	0.2
RANDINDEX	0.96	0.95	0.93	0.85
ANNOTATIONERROR	0	1	6	3

Fig. 4. Signal example ($n = 500, \sigma = 2$) and the computed segmentation for ALPIN, "Hocking, 2013", BIC, "Lavielle, 2005". Consecutive regimes are in different shades of grey. The metric values are given in the table.

TABLE II
SCORES. MEANS AND STANDARD DEVIATIONS ARE SHOWN. FOR EACH METRIC, THE BEST SCORE IS IN BOLD TYPE.

$n = 500, \sigma = 1$	ALPIN	Hocking, 2013	BIC	Lavielle, 2005
HAUSDORFF	2.1 (3.8)	**1.9** (3.2)	43.7 (34.6)	78.7 (78.6)
PRECISION	**0.99** (0.05)	**0.99** (0.04)	0.64 (0.20)	**0.99** (0.07)
RECALL	**0.99** (0.04)	**0.99** (0.04)	**0.99** (0.04)	0.71 (0.26)
RANDINDEX	**0.997** (0.006)	**0.997** (0.006)	0.970 (0.022)	0.904 (0.109)
ANNOTATIONERROR	**0.0** (0.1)	**0.0** (0.0)	3.1 (2.3)	1.4 (1.4)

$n = 500, \sigma = 2$	ALPIN	Hocking, 2013	BIC	Lavielle, 2005
HAUSDORFF	20.6 (31.0)	**19.4** (29.2)	46.3 (35.5)	90.8 (74.7)
PRECISION	0.92 (0.12)	0.94 (0.12)	0.63 (0.23)	**0.97** (0.10)
RECALL	0.91 (0.13)	0.90 (0.15)	**0.93** (0.13)	0.64 (0.25)
RANDINDEX	**0.980** (0.023)	0.976 (0.035)	0.959 (0.029)	0.884 (0.103)
ANNOTATIONERROR	**0.27** (0.49)	0.31 (0.56)	2.8 (2.5)	1.6 (1.3)

when σ is larger that 2, both metrics go to zero, meaning that the standard method is not able to detect any changepoint. The ALPIN algorithm, thanks to the learning process, adapts to the considered problem and maintain acceptable performances even when the noise level is large.

B. Comparison with existing methods

In this section, the penalty learned by ALPIN is compared to existing penalties described in Section II-B. Two scenarii are defined that correspond to two different noise levels : moderate ($\sigma = 1$) and difficult ($\sigma = 2$). The signal length is set to $n = 500$. The true value of σ is fed to the penalties relying on it and the learning for "Hocking, 2013" and ALPIN is performed using a 10-fold cross-validation. Table II presents the metrics values (mean and standard deviation) for both values of σ. The first observation is that the learning step of "Hocking, 2013" and ALPIN clearly improves the performances. On average, when $\sigma = 1$ (resp. $\sigma = 2$), the maximum error when predicting a changepoint is around 2 samples (resp. 20 samples) for "Hocking, 2013" and ALPIN. (See HAUSDORFF in Table II.) In particular, BIC and "Lavielle, 2005" respectively overestimate and underestimate the number of regimes for both scenarii. As a result each of these techniques can optimize either PRECISION or RECALL but not both. Conversely, "Hocking, 2013" and ALPIN keep both PRECISION and RECALL over 90%. Penalty learning approaches have almost perfect reconstruction for signals with a moderate noise level ($\sigma = 1$), but in the difficult scenario ($\sigma = 2$), accuracy decreases as noise blurs small jumps. Figure 4 presents an example of the results obtained with the different penalties for the difficult scenario. Interestingly, although some changepoints are not really visible, they are still recovered almost perfectly with the ALPIN algorithm. The first and fourth predictions of ALPIN are detected with an error larger than 10 signal samples, but this phenomenon also occurs for "Hocking, 2013". However "Hocking, 2013" fails to find the changepoint between the second and third regimes, which is characterized by a small amplitude jump and a short duration. As previously described, the BIC method tends to oversegment the signal, while the "Lavielle, 2005" undersegments it.

Although the performances of "Hocking, 2013" and ALPIN are not significantly different on this data set, the computation time to run both techniques is. Indeed, the execution time for processing 100 signals of length $n = 500$ (learning step) is 33 minutes for "Hocking, 2013" but only 7 minutes for ALPIN[1]. The convex excess penalized risk introduced in Section III-B and the possibility to directly minimize it with standard optimization methods allows to keep a reasonable computing time, which makes it suitable for real-life situations.

C. Adaptiveness to expert labels

One interesting feature of learning the penalty from annotated data is that ALPIN algorithm is able to adapt to different expert annotations. More precisely, depending of the changepoints that have been annotated by the expert, the final

[1]All times refer to running a Python implementation of ALPIN and [6] on a Linux computer with 24 Intel processors running at 2.80 GHz (CPU)

ISBN 978-0-9928626-7-1 © EURASIP 2017

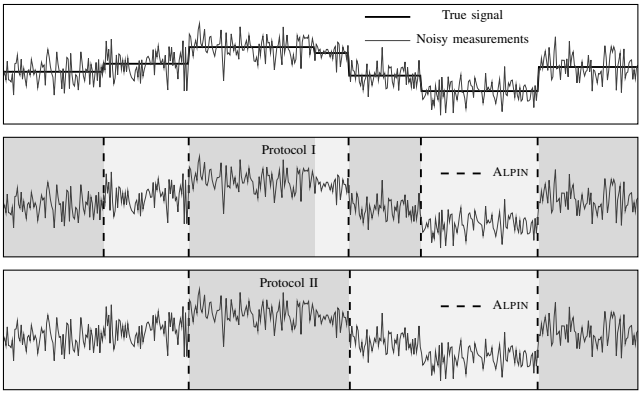


Fig. 5. (Top) Signal example and its noisy version ($n = 500, \sigma = 2$). (Middle) Prediction and expert partition according to Protocol I. (Bottom) Prediction and expert partition according to Protocol II.

segmentation algorithm can be different. To investigate this idea, we propose to consider two annotation protocols for the database described in Section IV-A.

- In Protocol I, all changepoints from the true underlying function are considered.
- In Protocol II, only the biggest and most visible changepoints are regarded as real changepoints. More precisely the changes with amplitude below 3 are discarded.

The idea behind this experiment is to see if the algorithm can learn the differences between these two protocols and especially if the final segmentation algorithms with the two learned penalties yield to different results. Figure 5 presents an example of the results obtained with ALPIN by using only train data from Protocol I or only train data from Protocol II. It is visible that the number of changepoints detected for both protocols are different, although the input signal is exactly the same. This proves that the algorithm was able to learn that Protocol II only considers the largest changes. In Protocol I, one changepoint is missed (middle plot) which is due to the fact that the mean shift is small, a situation that is rarely found in the training database.

VI. CONCLUSION

In this article, we have introduced a learning strategy to customize segmentation algorithms thanks to an annotated database. By using a convex excess penalized risk, this procedure can be performed in acceptable runtime and is suitable for real-life applications. Results show that the method adapts well to a wide range of situations (noisy signals, small and large jumps) and is able to learn the magnitude and number of changepoints expected by the expert.

ACKNOWLEDGMENT

This work was supported by a public grant as part of the Investissement d'avenir project, reference ANR-11-LABX-0056-LMH, LabEx LMH.

REFERENCES

[1] M. Basseville and I. Nikiforov, *Detection of Abrupt Changes: Theory and Application*. Prentice Hall Englewood Cliffs, 1993, vol. 104.
[2] L. Oudre, A. Lung-Yut-Fong, and P. Bianchi, "Segmentation of accelerometer signals recorded during continuous treadmill walking," in *Proceedings of the 19th European Signal Processing Conference (EUSIPCO)*, 2011, pp. 1564–1568.
[3] F. Desobry, M. Davy, and C. Doncarli, "An online kernel change detection algorithm," *IEEE Transactions on Signal Processing*, vol. 53, no. 8, pp. 2961–2974, 2005.
[4] J. Bai and P. Perron, "Critical values for multiple structural change tests," *Econometrics Journal*, vol. 6, no. 1, pp. 72–78, 2003.
[5] L. Boysen, A. Kempe, V. Liebscher, A. Munk, and O. Wittich, "Consistencies and rates of convergence of jump-penalized least squares estimators," *The Annals of Statistics*, vol. 37, no. 1, pp. 157–183, 2009.
[6] T. 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 30th International Conference on Machine Learning (ICML)*, Atlanta, USA, 2013, pp. 172–180.
[7] É. Lebarbier, "Detecting multiple change-points in the mean of gaussian process by model selection," *Signal Processing*, vol. 85, no. 4, pp. 717–736, 2005.
[8] M. Lavielle and E. Moulines, "Least-squares estimation of an unknown number of shifts in a time series," *Journal of Time Series Analysis*, vol. 21, no. 1, pp. 33–59, 2000.
[9] M. Lavielle, "Using penalized contrasts for the change-point problem," *Signal Processing*, vol. 85, no. 8, pp. 1501–1510, 2005.
[10] S. Arlot and A. Celisse, "A survey of cross-validation procedures for model selection," *Statistical Surveys*, vol. 4, pp. 40–79, 2010.
[11] T. Hocking, G. Schleiermacher, I. Janoueix-Lerosey, V. Boeva, J. Cappo, O. Delattre, F. Bach, and J.-P. Vert, "Learning smoothing models of copy number profiles using breakpoint annotations," *BMC Bioinformatics*, vol. 14, no. 1, p. 164, 2013.
[12] Y.-C. Yao, "Estimating the number of change-points via Schwarz' criterion," *Statistics and Probability Letters*, vol. 6, no. 3, pp. 181–189, 1988.
[13] Z. Harchaoui, F. Vallet, A. Lung-Yut-Fong, and O. Cappé, "A regularized kernel-based approach to unsupervised audio segmentation," in *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, Taipei, Taiwan, 2009, pp. 1665–1668.
[14] Z. Harchaoui, E. Moulines, and F. Bach, "Kernel Change-point Analysis," in *Advances in Neural Information Processing Systems 21 (NIPS 2008)*, Vancouver, Canada, 2008, pp. 609–616.
[15] R. Killick, P. Fearnhead, and I. Eckley, "Optimal detection of changepoints with a linear computational cost," *Journal of the American Statistical Association*, vol. 107, no. 500, pp. 1590–1598, 2012.
[16] R. Byrd, P. Lu, J. Nocedal, and C. Zhu, "A Limited Memory Algorithm for Bound Constrained Optimization," *SIAM Journal on Scientific Computing*, vol. 16, no. 5, pp. 1190–1208, 1995.
[17] S. Kay, *Fundamentals of Statistical Signal Processing, Volume II: Detection Theory*, 1st ed. Prentice Hall, 1998.
[18] E. Keogh, S. Chu, D. Hart, and M. Pazzani, "An online algorithm for segmenting time series," in *Proceedings of the 2001 IEEE International Conference on Data Mining (ICDM)*, 2001, pp. 289–296.

ISBN 978-0-9928626-7-1 © EURASIP 2017