ESANN 2016 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Bruges (Belgium), 27-29 April 2016, i6doc.com publ., ISBN 978-287587027-8. Available from http://www.i6doc.com/en/.

Comparison of three algorithms for parametric change-point detection

Cynthia Faure[1], Jean-Marc Bardet[1], Madalina Olteanu[1] and Jérôme Lacaille[2]

1- SAMM EA 4543, Université Panthéon Sorbonne, Paris, France

2- Snecma (Safran Group), France

Abstract. Numerous sensors placed on aircraft engines capture a considerable amount of data during tests or flights. In order to detect potential crucial changes of characteristic features, it is relevant to develop powerful statistical algorithms. This manuscript aims at detecting change-points, in an off-line framework, in piecewise-linear models and with an unknown number of change-points. In this context, three recent algorithms are considered, implemented and compared on simulated and real data.

1 Introduction

Numerous sensors placed on aircraft engines capture a considerable amount of data and monitor these engines in order to ensure the operational reliability during flights. The present manuscript is concerned with the study of several time series issued from flights on aircrafts equipped with SNECMA engines, and representing different recorded features : exogenous (lever, altitude, ...) and endogenous (temperature, ...). Detecting change-points in these data may be seen as a first step in the identification of a potential abnormality of the engine. Hence, the aim of this paper is to search for an efficient change-point detection method, with a low computational cost and a high performance. The framework chosen here is the off-line parametric change-point detection, with an unknown number of change-points. Training a parametric model that fits all available recorded features is not an easy task, mainly because of the heterogeneity of the data. To begin, we focused on data having a piecewise linear behavior. Hence, the algorithms to be further introduced will search for change-points in the slope. We will see further that this framework is not too constraining, since the algorithms to be presented may be applied on any time-series, as long as an appropriate cost-function is defined.

Detecting change-points in the slope of a time-series can be achieved, for instance, by minimizing the least-squared residuals contrast, as described in the seminal paper [1]. Since the number of change-points is unknown, a penalty term is usually added to the contrast function, as proposed in [2]. Then, the penalized contrast function may be minimized using various approaches based on dynamic programming. Three of these algorithms are investigated here : optimal partitioning [OP] (see, for instance, [3]), pruned exact linear time method [PELT] (see [4]) and slope heuristics using the slope estimation procedure [SEP] (see [5]). OP and PELT algorithms are both searching for the optimal partition minimizing the penalized contrast function, but the interest of PELT is a reduced

ESANN 2016 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Bruges (Belgium), 27-29 April 2016, i6doc.com publ., ISBN 978-287587027-8. Available from http://www.i6doc.com/en/.

computational time : the complexity of OP is quadratic, while the complexity of PELT is linear. Both methods require the a priori choice of a penalty term, usually the AIC or the BIC penalties. The interest of SEP comes from choice of the penalty, which is data driven. The slope heuristics in the framework of model selection was introduced in [5]. [6] further developed slope heuristics in a more general context, and then the SEP was adapted to the change-point detection issue by [7] and [8].

The rest of paper is organized as follows : section 2 summarizes the basic knowledge in off-line parametric change-point detection, and introduces the OP, PELT and SEP algorithms. Numerical applications are presented in Section 3.

2 Background

Let $Y = (Y_1, \ldots, Y_N)$ be a sequence of random variables. For $t \in \{1, \ldots, N\}$, suppose that Y_t is a function of $X_t \in \mathbb{R}^p$, where X_t is a random or deterministic vector. Also, assume there exists K^* (unknown) parametric changes in the relationship between (Y_t) and (X_t) : there exists an unknown vector $(\tau_1^*, \ldots, \tau_{K^*}^*) \in \mathbb{N}^{K^*}$ such that $\tau_1^* < \tau_2^* < \cdots < \tau_{K^*}^*$, and $K^* + 1$ unknown vectors $\theta_i^* \in \mathbb{R}^p$ satisfying $Y_t = g_{\theta_i^*}(X_t, \varepsilon_t)$ when $t \in \{\tau_i^* + 1, \tau_i^* + 2, \ldots, \tau_{i+1}^*\}$, where by convention $\tau_0^* = 0$ and $\tau_{K^*+1}^* = N$. $(\varepsilon_1, \ldots, \varepsilon_N)$ is a family of random vectors (unobserved) and $(\theta, x, e) \in \mathbb{R}^p \times \mathbb{R}^p \times \mathbb{R} \mapsto g_\theta(x, e)$ is a known or chosen function. The paper aims at estimating K^*, $(\tau_1^*, \ldots, \tau_{K^*}^*)$ and $(\theta_i^*)_i$, the parameters of the "true" model to be retrieved from an observed sample. The off-line change-point detection strategy chosen here consists in minimizing in $(K, (\tau_i), (\theta_i))$ a penalized contrast defined by :

$$\sum_{i=0}^{K} \sum_{t=\tau_i+1}^{\tau_{i+1}} C\big(Y_t, X_t, \theta_i\big) + \beta\, f(K) \ , \tag{1}$$

where the cost function C may be defined, for example, as a quadratic loss or a -log-likelihood (see for instance [9]). The term $\beta f(K)$ is the penalty which prevents from overfitting. The choice of the penalty term is usually linear (for instance, AIC or BIC) in the number of break points. The minimization of contrast (1) can be simplified by plugging the estimates of θ_i. Let $\hat\theta_{u,v} = \arg\min_{\theta \in \mathbb{R}^p} \sum_{t=u+1}^{v} C\big(Y_t, X_t, \theta\big)$ be the estimate of θ computed in the time-interval $\{u+1, ..., v\}$. Then, the change-point detection problem becomes:

$$(\hat K, \hat\tau_1, \ldots, \hat\tau_{\hat K}) = \underset{K; \tau_1 < \tau_2 < \cdots < \tau_K}{\arg\min} \left\{ \sum_{i=0}^{K} \sum_{t=\tau_i+1}^{\tau_{i+1}} C\big(Y_t, X_t, \hat\theta_{\tau_i, \tau_{i+1}}\big) + \beta\, f(K) \right\} \tag{2}$$

Remark: In Section 3, devoted to numerical applications, the case of a linear model depending on the time only is considered. In this case, $X_t = (t, 1)$, $\theta_i^* = (\theta_i^{*(1)}, \theta_i^{*(2)})$ and $g_\theta(x, e) =< x, \theta > +e$, with $< \cdot, \cdot >$ the inner product. Classical least squares estimates are chosen for θ and the cost function to be minimized is the MDL (minimum description length), [10] : $C(Y_t, X_t, \hat\theta_{u,v}) = 3 \ln(v-u) + (v-u) \log(2\pi\hat\sigma^2)$, where $\hat\sigma^2 = \frac{1}{v-u} \sum_{t=u+1}^{v} \big(Y_t - \hat\theta_{u,v}^{(1)}\, t - \hat\theta_{u,v}^{(2)}\big)^2$.

ESANN 2016 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Bruges (Belgium), 27-29 April 2016, i6doc.com publ., ISBN 978-287587027-8. Available from http://www.i6doc.com/en/.

2.1 Optimal partitioning [OP]

The optimal partitioning [OP] algorithm used here for minimizing the penalized contrast in (2) was first described in [3]. The idea is to use dynamical programming in order to reduce the exponential complexity of an exhaustive search to a quadratic one. With the previous notations, a change-point occurs between the instants u and v if there exists some instant $u < l < v$ such that

$$\sum_{t=u+1}^{l} C(Y_t, X_t, \hat{\theta}_{u,l}) + \sum_{t=l+1}^{v} C(Y_t, X_t, \hat{\theta}_{l,v}) + \beta < \sum_{t=u+1}^{v} C(Y_t, X_t, \hat{\theta}_{u,v}) \qquad (3)$$

Using criterion (3), the OP algorithm scans the data iteratively and associates the optimal value of the penalized contrast, $F(Y_{1:u})$, to each subsample $Y_{1:u} = (Y_1, ..., Y_u)$. The proof that $F(Y_{1:u})$ contains the minimum of the penalized contrast for each $u = 1, ..., N$, may be sketched by the following recursion : $F(Y_{1:N}) = \min_{u=1,...,N} \left\{ F(Y_{1:u}) + \sum_{t=u+1}^{N} C(Y_t, X_t, \hat{\theta}_{u,N}) + \beta \right\}$
One may see immediately that the time instant τ which gives the optimal $F(Y_{1:N})$ corresponds to the last change-point before N. The algorithm computes iteratively $F_{1:N} = (F(Y_{1:1}), ..., F(Y_{1:N}))$ and $CP_{1:N} = (CP(Y_{1:1}), ..., CP(Y_{1:N}))$, where $F(Y_{1:u})$ is the minimum value of the penalized contrast and $CP(Y_{1:u})$ represents the last change point before u, for each subsample $Y_{1:u} = (Y_1, ..., Y_u)$, $u = 1, ..., N$. The optimal partition may be retrieved by scanning backwards $CP_{1:N}$. The OP algorithm is summarized in *Procedure 1* hereafter.

2.2 Pruned Exact Linear Time [PELT]

Introduced in [4], the PELT algorithm aims at reducing the computational complexity of the OP algorithm, while still retrieving the optimal solution. This is achieved by pruning the set of possible solutions when minimizing $F(Y_{1:u})$. Pruning is justified by the following property of change-points, proven in [4]: if equation (3) holds for some $u < l < v$ and if $F(Y_{1:u}) + \sum_{t=u+1}^{l} C(Y_t, X_t, \hat{\theta}_{u,l}) + \beta \geq F(Y_{1:l})$, then u can never be the last change-point before v. Pruning the set of possible change-points reduces the quadratic complexity of OP to a linear one. The algorithm, which consists in adding a supplementary step to the OP procedure, is summarized in *Procedure 2*.

2.3 Slope estimation procedure [SEP]

In both algorithms previously described, the penalty term has to be a priori chosen. Usually, AIC or BIC penalty terms are used in practice, with a slight preference for the BIC, justified by its parsimony and consistency properties. However, in some cases, a priori choices for the penalty term may not be exactly appropriate for the data and the problem to be solved. Recently, a data-driven model for calibrating the penalty term, called "slope heuristics", was introduced in [5]. In this paper, the optimal penalty is selected using the slope estimation procedure described in [6]. The contrast function C is minimized for a fixed

ESANN 2016 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Bruges (Belgium), 27-29 April 2016, i6doc.com publ., ISBN 978-287587027-8. Available from http://www.i6doc.com/en/.

number of change-points, K, with K varying from 1 up to some fixed bound, K_{max}. Then, the optimal values of C for $K = 1, ..., K_{max}$ are plotted. The resulting scatterplot has a linear behavior once K becomes "big enough", as shown in Figure 1. The optimal penalty is then chosen as twice the slope of the linear model. The algorithm is briefly described in *Procedure 3*. While the slope heuristics allows to have a data-driven penalty, its main drawback is the computational complexity. Indeed, the SEP algorithm requires to repeat K_{max} times a non-penalized change-point detection method (using dynamical programming with complexity $O(N^2)$ at each run) in order to estimate the slope α.

Procedure 1: Optimal Partitioning [OP]

- Initialize $F(Y_{1:1}) = 0$ and $CP = NULL$.
- For $u = 1, ..., N$, compute iteratively (forward) $F(Y_{1:u})$ and the corresponding break point $\tau = \arg\min_{\tau=1,...,u} F(Y_{1:u})$, $CP = CP \cup \tau$.
- Compute iteratively (backward) the optimal change points : $\hat{\tau}_K \leftarrow \arg\min_{1,...,N} F(Y_{1:N})$, $\hat{\tau}_{K-1} \leftarrow \arg\min_{1,...,\hat{\tau}_K} F(Y_{1:\hat{\tau}_K})$, ...

Procedure 2: Pruned Exact Linear Time [PELT]

- Initialize $F(Y_{1:1}) = -\beta$, $R = \{1\}$ and $CP = NULL$
- For $u = 1, ..., N$, compute iteratively (forward) $F(Y_{1:u})$ and the corresponding break point : $\tau = \arg\min_{\tau \in R} F(Y_{1:u})$, $CP = CP \cup \tau$.

  Update the set of plausible change-points, R :
  $R \leftarrow \{v \in R \cup \tau \; : \; F(Y_{1:v}) + \sum_{t=v+1}^{\tau} C(Y_t, X_t, \hat{\theta}_{v,\tau}) + \beta \leq F(Y_{1:\tau})\}$
- Compute iteratively (backward) the optimal change points : $\hat{\tau}_K \leftarrow \arg\min_{1,...,N} F(Y_{1:N})$, $\hat{\tau}_{K-1} \leftarrow \arg\min_{1,...,\hat{\tau}_K} F(Y_{1:\hat{\tau}_K})$, ...

Procedure 3: Slope Estimation Procedure [SEP]

- Compute, for $K = 1, ..., K_{max}$:
  $\mathbb{C}(K) = \min_{\tau_0<\tau_1<...<\tau_K<\tau_{K+1}} \sum_{i=0}^{K} \sum_{t=\tau_i+1}^{\tau_{i+1}} C(Y_t, X_t, \hat{\theta}_{\tau_i,\tau_{i+1}})$
- Draw the scatterplot $(K, \mathbb{C}(K))_{1 \leq K \leq K_{max}}$
- Compute the slope α of the linear model for K "big enough"
- The optimal penalty term will be $\beta = -2\alpha$

3 Numerical applications

3.1 Simulated data

Two scenarios were considered for simulations: (**A**) a time series of length $N = 300$ and three random change points, and (**B**) a time series of length $N =$

ESANN 2016 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Bruges (Belgium), 27-29 April 2016, i6doc.com publ., ISBN 978-287587027-8.
Available from http://www.i6doc.com/en/.

$\mathbf{A} : N = 300, K^\star = 3$	OP	PELT	SEP($K_{max} = 10$)
Exact number of cp	**90.0%**	89.0%	**90.0%**
Performance ($\lvert\tau_i - \hat{\tau}_i\rvert \leq 5$)	85.2%[17.6]	90.0% [5.4]	**93.0% [14.0]**
Time (seconds)	2.1 [0.1]	**0.1 [0.0]**	32.3 [0.8]
$\mathbf{B} : N = 1000, K^\star = 6$	OP	PELT	SEP($K_{max} = 15$)
Exact number of cp	90.6%	82.7%	**95.0%**
Performance ($\lvert\tau_i - \hat{\tau}_i\rvert \leq 10$)	97.0% [7.2]	99.0% [1.0]	**99.8% [2.0]**
Time (seconds)	32.6 [5.1]	**0.4 [0.1]**	670.0 [30.0]

Table 1: Performances on simulated data (1000 samples for each scenario)

1000 and six random change points. For each scenario, the time series was replicated 1000 times. A condition on the slopes was added in order to ensure the continuity of the data. The BIC penalty was used both in the OP and PELT algorithms. Three performance criteria were considered: the proportion (among the 1000 replications) of correctly identified change-points (3 for scenario **A** and 6 for **B**); the precision: among the simulations with a well identified number of change-points, the precision of these change-points, measured by the absolute difference $|\tau_i - \hat{\tau}_i|$; the computational time. The results summarizing the experiments (mean values with standard-errors in brackets) are available in Table 1. Although the SEP procedure achieves the best results in terms of accuracy, its computational complexity becomes very heavy for a relatively long series (1000 data).

3.2 Real data

According to the previous section, PELT achieves the best trade-off between accuracy of results and computational time. Hence, this method was selected to be trained and tested on large real data. The database we used was provided by SNECMA and contains high-frequency records (from 1Hz to 100 Hz), for various features captured by the sensors and with various lengths (from a few hundreds to some tens of thousands points). For illustration, the change-points detected on two features (shaft speed and engine temperature) are represented in Figure 2. The behavior of these two features is mainly piecewise linear, hence the PELT algorithm previously described and used for simulations was used as such, but for more complex time-series, it is sufficient to select a more appropriate cost-function C and train exactly the same algorithm described in *Procedure 2*. Globally, the change-points are well detected, even the small ones. The meaning of each change-point is settled afterwards with the help of experts. A change-point may be observed, for instance, after an action of the pilot (pulling the lever). Detecting these change-points allows to assume causal relations between different features with some delay (a raise of the lever implies a raise of the shaft speed). Eventually, machine-learning techniques which will learn the "normal" change-points and set alarms for "abnormal" change-points may be applied afterwards in an operational context.

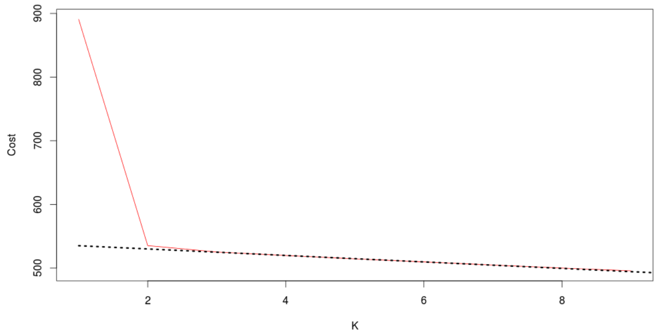

Fig. 1: Representation of the cost function (not penalized) for $1 \leq K \leq K_{max}$

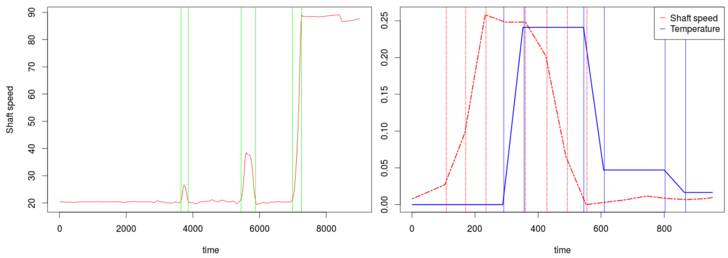

Fig. 2: Change-points (PELT) on the shaft speed and on the engine temperature

References

[1] J. Bai and P. Perron. Estimating and Testing Linear Models with Multiple Structural Changes. *Econometrica*, 66(1):47–78, 1998.

[2] M. Lavielle and E. Moulines. Least-squares estimation of an unknown number of shifts in a time series. *Journal of Time Series Analysis*, 21:33–59, 2000.

[3] B. et. al. Jackson. An algorithm for optimal partitioning of data on an interval. *IEEE Signal Processing Letters*, 12(2):105–108, 2005.

[4] Fearnhead P. Killick R. and I. Eckley. Optimal detection of changepoints with a linear computational cost. *JASA*, 107(500):1590–1598, 2012.

[5] S. Arlot and P. Massart. Data-driven calibration of penalties for least-squares regression. *The Journal of Machine Learning Research*, 10:245–279, 2009.

[6] Maugis C. Baudry J-P. and Michel B. Slope heuristics: overview and implementation. *RR-INRIA*, (7223), 2010.

[7] E. Lebarbier. Detecting multiple change-points in the mean of Gaussian process by model selection. *Signal Processing*, 85:717–736, 2005.

[8] J.-M. Bardet, W. Kengne, and O. Wintenberger. Multiple breaks detection in general causal time series using penalized quasi-likelihood. *Electronic Journal of Statistics*, 6:435–477, 2012.

[9] L Horvath. The Maximum-Likelihood Method for Testing Changes in the Parameters of Normal Observations. *Annals of Statistics*, 21(2):671–680, 1993.

[10] R. A. Davis, T. C.M. Lee, and G. A. Rodriguez-Yam. Structural break estimation for non-stationary time series models. *Journal of the American Statistical Association*, 101:223–239, 2006.