ESANN 2012 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Bruges (Belgium), 25-27 April 2012, i6doc.com publ., ISBN 978-2-87419-049-0. Available from http://www.i6doc.com/en/livre/?GCOI=28001100967420.

A CUSUM approach for online change-point detection on curve sequences

Nicolas Cheifetz[1,2], Allou Samé[1], Patrice Aknin[1] and Emmanuel de Verdalle[2]

1- Université Paris-Est, IFSTTAR, GRETTIA
2- Veolia Environnement Recherche & Innovation (VERI)

Abstract. Anomaly detection on sequential data is common in many domains such as fraud detection for credit cards, intrusion detection for cyber-security or military surveillance. This paper addresses a new CUSUM-like method for change point detection on curves sequences in a context of preventive maintenance of transit buses door systems. The proposed approach is derived from a specific generative modeling of curves. The system is considered out of control when the parameters of the curves density change. Experimental studies performed on realistic world data demonstrate the promising behavior of the proposed method.

1 Introduction

This study is motivated by the predictive maintenance of pneumatic doors in transit buses. Such a task has led us to detect changes on a sequence of complex observations. In this context, each observation is a multidimensional trajectory (or curve) representing the couple (air pressure of actuators, door position) during an opening/closing cycle of a door.

Change point detection on a sequence of observations is generally formulated as a sequential hypothesis testing problem consisting in detecting earlier the occurrence of a change with a low false alarm rate [1]. A large amount of detection rules has been proposed in the literature [1], [2] to address this issue but these approaches are very often based on multivariate sequential data. This paper deals with the problem of on-line change detection on a sequential data where each observation consists in a multivariate curve. For this purpose, a generative model inspired from the Hidden Process Regression Model [3] has been used to represent multivariate curves. Based on the resulting curves density, an on-line CUSUM-like detection approach is derived.

The next section gives a brief review of the CUSUM and Generalized Likelihood Ratio (GLR) detection algorithms. In the third section, we describe the generative model used for the modeling of multivariate curves and the sequential strategy proposed to detect changes from curves sequences. An experiment on both synthetic and real data related to the monitoring of pneumatic door systems is detailed in section 4.

2 Review of CUSUM-like algorithms for multivariate data

Let us suppose that data are multivariate observations $\mathbf{x}_1, \ldots, \mathbf{x}_t, \ldots$ sequentially received. The classical CUSUM algorithm [2] consists in deciding about which

ESANN 2012 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Bruges (Belgium), 25-27 April 2012, i6doc.com publ., ISBN 978-2-87419-049-0. Available from http://www.i6doc.com/en/livre/?GCOI=28001100967420.

of the following hypotheses to choose when the parameters θ_0 and θ_1 are known:

$$\begin{cases} (H_0) & \mathbf{x}_i \overset{iid}{\sim} p(\mathbf{x}_i;\theta_0) \qquad \forall i = 1,\dots,t \\ (H_1) & \mathbf{x}_i \overset{iid}{\sim} p(\mathbf{x}_i;\theta_0) \qquad \forall i = 1,\dots,p-1 \\ & \mathbf{x}_i \overset{iid}{\sim} p(\mathbf{x}_i;\theta_1) \qquad \forall i = p,\dots,t \end{cases} \tag{1}$$

where p is the change time point. The detection statistic is then written as:

$$g_t = \max_{2\leq p\leq t} \log\left(\frac{p(\mathbf{x}_1,\dots,\mathbf{x}_{p-1};\theta_0)\times p(\mathbf{x}_p,\dots,\mathbf{x}_t;\theta_1)}{p(\mathbf{x}_1,\dots,\mathbf{x}_t;\theta_0)}\right) = \max_{2\leq p\leq t}\sum_{i=p}^{t}\log\frac{p(\mathbf{x}_i;\theta_1)}{p(\mathbf{x}_i;\theta_0)}. \tag{2}$$

The change time t_A (or alarm time) is defined by: $t_A = \inf\big\{t : g_t \geq h\big\}$. The optimality of this detection rule has been proved in both the asymptotic case [4] and the non-asymptotic case [5].

The Generalized Likelihood Ratio (GLR) test [1] can be considered when the parameter θ_1, after change, is unknown. In contrast to the CUSUM detection statistic (eq. (2)) that can be recursively written, the GLR rule does not have a recursive formulation [1].

This paper considers that θ_0 and θ_1 are unknown. In a perspective of online change-point detection from a curve sequence, the next section begins with the specification of a probability density function for multivariate curves.

3 Sequential change point detection approach for multidimensional curves

3.1 Generic regression model for multivariate curves

The curve modeling approach described here is inspired from the Hidden Process Regression Model initiated in [3]. Originally, this model was dedicated to the description of mono-dimensional curves presenting some changes in regimes. As the multidimensional curves (related to the door opening/closing cycles) studied in this paper are themselves subject to changes in their regimes, we propose an extension of the latter model to deal with multidimensional curves. Let $\big(\mathbf{x}_1,\dots,\mathbf{x}_t\big)$ be a curves sequence, where $\mathbf{x}_i = \big(x_{i1},\dots,x_{im}\big)$ and $x_{ij}\in\mathbb{R}^d, d\geq 1$. Each trajectory $\mathbf{x}_i$ is associated with a time vector $\mathbf{t} = \big(t_1,\dots,t_m\big)$, where $t_1 < \dots < t_m$. Moreover, we assume that at each time point t_j, the variable x_{ij} follows one of K polynomial regression models of order r. That is to say x_{ij} is generated by:

$$x_{ij} = \boldsymbol{\beta}_{z_{ij}}\cdot\mathbf{T}_j + \epsilon_{ij} \tag{3}$$

where $\epsilon_{ij}\sim\mathcal{N}(0,\mathbf{\Sigma}_{z_{ij}})$ is a d-dimensional Gaussian noise with covariance matrix $\mathbf{\Sigma}_{z_{ij}}$, matrix $\boldsymbol{\beta}_{z_{ij}}$ is the $d\times(r+1)$ polynomial coefficients and $\mathbf{T}_j = \big(1,t_j,\dots,(t_j)^r\big)^T$ is the covariate vector. The difference between this extension and the original model for monodimensional curves [3] remains in the specification of the parameter $\boldsymbol{\beta}_{z_{ij}}$ which is a matrix in our situation.

The hidden variable z_{ij} ($\forall i = 1,\dots,t$ and $\forall j = 1,\dots,m$) is assumed to be generated independently by a multinomial distribution $\mathcal{M}(1,\pi_1(t_j;\boldsymbol{a}),\dots,\pi_K(t_j;\boldsymbol{a}))$ where $\pi_k(t;\boldsymbol{a})$ is a logistic function of time. The logistic transformation allows to model dynamical changes between segments with flexibility.

ESANN 2012 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Bruges (Belgium), 25-27 April 2012, i6doc.com publ., ISBN 978-2-87419-049-0. Available from http://www.i6doc.com/en/livre/?GCOI=28001100967420.

The parameter $\boldsymbol{\theta} = \left(\boldsymbol{a}, \boldsymbol{\beta}_1, ..., \boldsymbol{\beta}_K, \boldsymbol{\Sigma}_1, ..., \boldsymbol{\Sigma}_K\right)$ of this model is estimated by maximizing the log-likelihood through the Expectation-Maximization algorithm [6]. The pseudo-code of this method is provided in Algorithm 1. As in the univariate case, parameter $\boldsymbol{a}$ is updated through a multi-class Iterative Reweighted Least Squares (IRLS) algorithm [7]. Therefore, the algorithm 1 is the multidimensional version of the EM algorithm introduced by Chamroukhi [3].

Algorithm 1: Pseudo-code for Multivariate RHLP. EM($\mathbf{x}$, r, K)

Input: Observation matrix $\mathbf{x}$ of size t, degree r of the polynomial, number K of segments for each curve and eventually $\boldsymbol{\theta}^{(0)}$

$q \leftarrow 0$ `// Random initialization of` $\boldsymbol{\theta}^{(0)}$ `if not provided`

while *Non convergence test* **do**

  for $(i, j, k) \in \{1, \dots, t\} \times \{1, \dots, m\} \times \{1, \dots, K\}$ **do**

$$\tau_{ijk}{}^{(q)} \leftarrow \frac{\pi_k\left(t_j; \boldsymbol{a}^{(q)}\right) \cdot \mathcal{N}\left(x_{ij}; \beta_k{}^{(q)} \cdot \mathbf{T}_j, \boldsymbol{\Sigma}_k{}^{(q)}\right)}{\sum_{l=1}^{K} \pi_l\left(t_j; \boldsymbol{a}^{(q)}\right) \cdot \mathcal{N}\left(x_{ij}; \beta_l{}^{(q)} \cdot \mathbf{T}_j, \boldsymbol{\Sigma}_l{}^{(q)}\right)}$$ `// E-step`

$$\boldsymbol{a}^{(q+1)} \leftarrow \arg\max_{\boldsymbol{a}} \sum_{j=1}^{m} \sum_{k=1}^{K} \left(\sum_{i=1}^{t} \tau_{ijk}{}^{(q)} \right) \cdot \log \pi_k\left(t_j; \boldsymbol{a}\right)$$ `// M-step` `// IRLS`

  for $k \in \{1, \dots, K\}$ **do**

$$\left(\beta_k{}^{(q+1)}\right)^T \leftarrow \left[\sum_{j=1}^{m} \left(\sum_{i=1}^{t} \tau_{ijk}{}^{(q)} \right) \mathbf{T}_j \, \mathbf{T}_j^T \right]^{-1} \left[\sum_{j=1}^{m} \mathbf{T}_j \left(\sum_{i=1}^{t} \tau_{ijk}{}^{(q)} x_{ij} \right) \right]$$

$$\boldsymbol{\Sigma}_k{}^{(q+1)} \leftarrow \frac{\sum_{i=1}^{t} \sum_{j=1}^{m} \tau_{ijk}{}^{(q)} \left[x_{ij} - \beta_k{}^{(q)} \mathbf{T}_j\right] \left[x_{ij} - \boldsymbol{\beta}_k{}^{(q)} \mathbf{T}_j\right]^T}{\sum_{i=1}^{t} \sum_{j=1}^{m} \tau_{ijk}{}^{(q)}}$$

  $q \leftarrow q + 1$

Output: Estimated model parameter $\widehat{\boldsymbol{\theta}}$

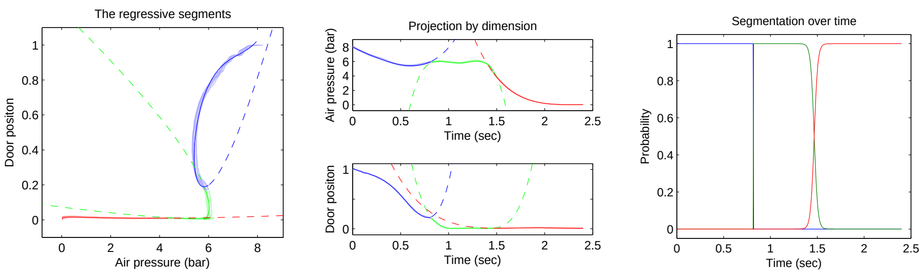

Fig. 1: Density estimation of ten defect-free bivariate curves (left), projection by dimension (middle) and logistic probabilities (right) during a closing motion. The degree of polynomial regression is set to 4 and the number of regimes is set to 3, corresponding to 3 physical operating steps.

ESANN 2012 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Bruges (Belgium), 25-27 April 2012, i6doc.com publ., ISBN 978-2-87419-049-0. Available from http://www.i6doc.com/en/livre/?GCOI=28001100967420.

3.2 On-line change-point detection strategy

Applying the CUSUM and GLR off-line rule (section 2) becomes rapidly inefficient regarding computational and memory requirements. To overcome these limitations, we propose a new on-line detection strategy which sequentially uses windows of size W [2]. Let us assume that the last change point has been detected at time point t_A and that, up to time $t > t_A$, no change has been detected. The proposed strategy then consists in detecting the change point into the time interval $t < p \leq t + W$ instead of considering the overall time interval $1 < p \leq t+W$. This particular strategy only requires the computation of $O(W)$ likelihood ratios instead of $O(t+W)$ ratios for the off-line version. The detection is performed by computing:

$$g_{t+W} = \max_{t<p\leq t+W} \log\left(\frac{p(\mathbf{x}_{t_A},\ldots,\mathbf{x}_{p-1};\widehat{\boldsymbol{\theta}}_0)\times p(\mathbf{x}_p,\ldots,\mathbf{x}_{t+W};\widehat{\boldsymbol{\theta}}_1)}{p(\mathbf{x}_{t_A},\ldots,\mathbf{x}_{t+W};\widehat{\boldsymbol{\theta}}_0)}\right).$$

The computation of this statistic requires the evaluation, for each $t < p \leq t+W$, of the log likelihood ratio where the ML estimates $\widehat{\theta}_0$ and $\widehat{\theta}_1$ are computed by running the EM algorithm respectively on the data sets $\{\mathbf{x}_{t_A},\ldots,\mathbf{x}_{p-1}\}$ and $\{\mathbf{x}_p,\ldots,\mathbf{x}_{t+W}\}$. These estimations are accelerated by initializing the EM algorithm with the parameters $\widehat{\boldsymbol{\theta}}_0$ and $\widehat{\boldsymbol{\theta}}_1$ computed in the previous window step for $p = t$. It should be noticed that the initial parameter $\widehat{\boldsymbol{\theta}}_0^{init}$ is computed by running EM on t_0 healthy curves.This procedure is described in algorithm 2.

Algorithm 2: Pseudo-code for the proposed detection method.

Input: Sequential curves $(\mathbf{x}_1,\mathbf{x}_2,\ldots)$, number t_0 of healthy curves, local window size W, degree r of polynomial, number K of segments for each curve, detection threshold h_α where α is a FA rate given by the user

$t_A \leftarrow 1$

$t \leftarrow t_A + t_0 - 1$ `// Initialization`

$\widehat{\boldsymbol{\theta}}_0^{init} \leftarrow \mathrm{EM}\big((\mathbf{x}_{t_A},\ldots,\mathbf{x}_t), r, K\big)$ `// Null hypothesis` (H_0)

while *there are new data of size* W **do**

- $g_{t+W} \leftarrow \max_{t<p\leq t+W} \sum_{i=p}^{t+W} \log p(\mathbf{x}_i;\widehat{\boldsymbol{\theta}}_1) - \log p(\mathbf{x}_i;\widehat{\boldsymbol{\theta}}_0)$
- $t^* \leftarrow \arg\max_{t<p\leq t+W} \sum_{i=p}^{t+W} \log p(\mathbf{x}_i;\widehat{\boldsymbol{\theta}}_1) - \log p(\mathbf{x}_i;\widehat{\boldsymbol{\theta}}_0)$
- where $\begin{cases} \widehat{\boldsymbol{\theta}}_0 \leftarrow \mathrm{EM}\Big((\mathbf{x}_{t_A},\ldots,\mathbf{x}_{p-1}),\ r,\ K,\ \widehat{\boldsymbol{\theta}}_0^{init}\Big) \\ \widehat{\boldsymbol{\theta}}_1 \leftarrow \mathrm{EM}\Big((\mathbf{x}_p,\ \ldots\ ,\mathbf{x}_{t+W}),\ r,\ K,\ \widehat{\boldsymbol{\theta}}_1^{init}\Big) \end{cases}$
- **if** $g_{t+W} < h_\alpha$ **then**
  - $(\widehat{\boldsymbol{\theta}}_0^{init}, \widehat{\boldsymbol{\theta}}_1^{init}) \leftarrow (\widehat{\boldsymbol{\theta}}_0, \widehat{\boldsymbol{\theta}}_1)$
  - $t \leftarrow t + W$
- **else**
  - $t_A \leftarrow t^*$
  - *Restart the initialization*

Output: Estimated change time(s) $\widehat{t}_A$

ESANN 2012 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Bruges (Belgium), 25-27 April 2012, i6doc.com publ., ISBN 978-2-87419-049-0. Available from http://www.i6doc.com/en/livre/?GCOI=28001100967420.

4 Experimental studies

Opening and closing operations have a similar behavior ; in this article, we only report results related to door closing operations. The dataset has been recorded in July 2011 on a 18-meters articulated bus. Two variables were recorded: the pressure inside door pneumatic actuators and the door position. A sampling frequency of 100Hz was adopted. The resulting curves length was $m = 204$.

The number of curves needed to learn $\widehat{\boldsymbol{\theta}}_0^{init}$ was set to 10 and the window size W was set to 5 curves. We have observed that the window size does not affect the detection quality but can slow down the process if too large.

Relevant hyperparameters embedded in the density estimation model, are determined using a physical prior i.e. $K = 3$, which corresponds to the three steps during the door operation (Fig. 1); the degree r of polynomial regression model was experimentally set to 4.

The detection threshold, based on α (the expected false alarm rate provided by the user), is estimated as follows: first, the multivariate RHLP is used to learn a parameter vector $\boldsymbol{\theta}_0$ on a set of healthy closing operation curves ; then, curves sequences of size 15 000 are generated from distribution $p(\mathbf{x}_i; \boldsymbol{\theta}_0)$; finally, the threshold is formed by the $(1 - \alpha)$ confidence interval of the test statistic. In this article, α was set to 0.1, 0.01 and 0.001.

Two performance metrics were used to evaluate the proposed method: the false alarm rate (FAR) and the average detection delay (ADD) that is to say the delay to detect an effective degraded curve after the change point. Note that FAR is computed on the healthy curves. Each curves sequence consists of 10 250 curves including 10 000 defect-free curves and 250 curves corresponding to a door blocking damage.

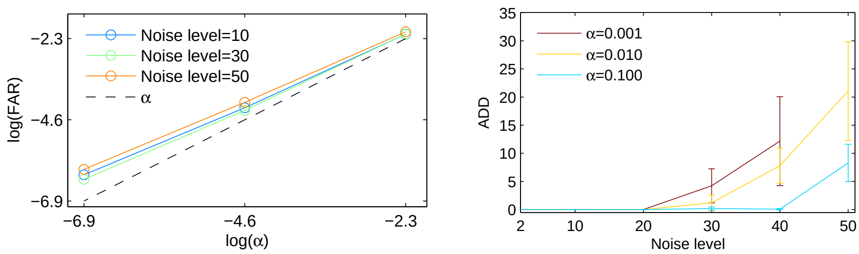

Fig. 2: $\log(\alpha)$ VS log(FAR) (left). And impact of the noise level on detection delay (right). Each value of log(FAR) and ADD is an average over 10 different sequences.

Figure 2 (left) displays the logarithm of FAR as a function of the logarithm of α. We observe that FAR indicator is stable for the different values of noise level. In fact, similar values between FAR and α means that the detection threshold has been correctly estimated. Figure 2 (right) shows the behavior of ADD in relation with the noise level which is a multiplicative factor of the variance $\boldsymbol{\Sigma}_k$. It can be seen that ADD increases with the noise level and that no change is detected when the noise level is too large (distribution before change and distribution after change are overlapping).

ESANN 2012 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence
and Machine Learning. Bruges (Belgium), 25-27 April 2012, i6doc.com publ., ISBN 978-2-87419-049-0.
Available from http://www.i6doc.com/en/livre/?GCOI=28001100967420.

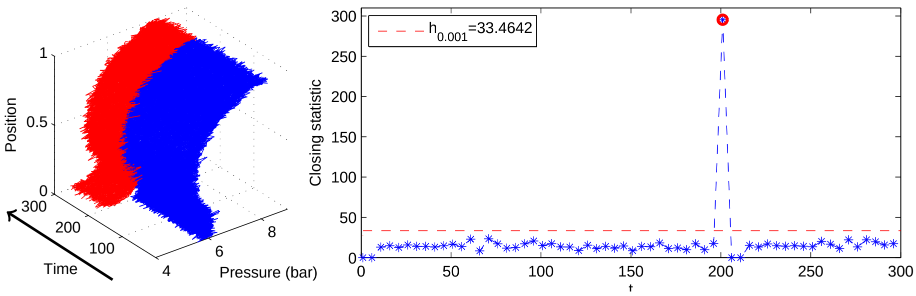

Fig. 3: Example of detection on closing curves (α is set to 0.001 and noise level is set to 2). 200 in-control observations are in blue and 100 anomalies in red (left); statistic g_t is in blue, estimated detection threshold is in red and alarm is the red circle (right).

5 Conclusion

In this paper, we have presented a sequential method to detect anomalies in a curves sequence. The proposed method uses a CUSUM-like test based on densities defined on the curves space. This approach is suitable for sudden changes, like in operating system breakdown. A generative model has been defined for density estimation which is a multivariate extension of the Hidden Process Regression Model [3]. This strategy is applied to monitor pneumatic doors. The experimental results showed a certain practicability of the approach. We believe this strategy could be used to an application when sustained special causes are observed, those that continue until they are identified and fixed.

References

[1] M. Basseville and I. Nikiforov. *Detection of Abrupt Changes: Theory and Applications*. Prentice-Hall, Englewood Cliffs, NJ, USA, 1993.

[2] T. Lai. Sequential analysis: some classical problems and new challenges. *Statist. Sinica*, 11(2):303–408, 2001.

[3] F. Chamroukhi, A. Samé, G. Govaert, and P. Aknin. A hidden process regression model for functional data description. Application to curve discrimination. *Neurocomputing*, 73(7-9):1210–1221, March 2010.

[4] G. Lorden. Procedures for Reacting to a Change in Distribution. *The Annals of Mathematical Statistics*, 42(6), 1971.

[5] G. Moustakides. Optimal Stopping Times for Detecting Changes in Distributions. *The Annals of Statistics*, 14(4):1379–1387, 1986.

[6] A. P. Dempster, N. M. Laird, and D. B. Rubin. Maximum Likelihood from Incomplete Data via the EM Algorithm. *Journal of the Royal Statistical Society. Series B (Methodological)*, 39(1):1–38, 1976.

[7] P. Green. Iteratively reweighted least squares for maximum likelihood estimation,and some robust and resistant alternatives. *Journal of the Royal Statistical Society B*, 46:149–192, 2003.