Bayesian Learn-Predict-Adjust Method for Online Detection of Recurrent Changepoints

Alexander Maslov, Mykola Pechenizkiy, Yulong Pei, Indrė Žliobaitė, Alexander Shklyaev, Tommi Kärkkäinen, Jaakko Hollmén
IJCNN’17, Anchorage, Alaska, USA (May 14-19)

Eindhoven University of Technology (The Netherlands)
University of Jyväskylä (Finland)
Aalto University, University of Helsinki (Finland)
Lomonosov Moscow State University (Russia)
BLPA is a novel online detection method which extends the Bayesian Online Changepoint Detector (BD) [1] by embedding into it a Predictive Change Confidence Function (PCCF) [2], in order to predict future changepoints in the input data stream, adjust detector’s settings dynamically and to reduce FP rate.

Outline

- Online change detection background
- Part I: Predicting recurrent changes using PCCF function
- Part II: Bayesian Online Change Detector (BD)
- Part III: BLPA. Embedding PCCF into BD
- Experimental results and conclusion
Change detection problem

‘Changepoint’ is a time moment when statistical properties of the data stream change significantly according to the predefined criteria.

The are many practical applications:

- Fault detection and monitoring
- Vibration monitoring of mechanical systems
- Automatic depth of anesthesia control
- DM: Predictive modeling under concept drift
A good *online* change detector should:

- detect all changes within an acceptable time lag (high TPR),
- be robust to noise by not raising false alarms (low FPR).

Detector has a sensitivity parameter $s$:

- If $s$ is (too) high: all changes are detected but some outliers misclassified as changes.
- If $s$ is (too) small: we skip outliers but change detection lag increases.
Part I: PCCF. Current state-of-the-art vs. our approach

• Existing approaches are *reactive*:
  • assume that change points are not predictable.

• Our approach is *proactive*:
  • model recurrence to anticipate changes and integrate this information with existing online change detectors.
Motivating example

From a sequence of previously observed changes we estimate that $t_i - t_{i-1} \sim 12.8 \pm 0.85$ hours. Can we use this information to improve performance of the change detector?

Outlier caused change detection event (FP) might be skipped.
Basic intuition

We can improve performance of the change detectors by dynamically adjusting their sensitivity parameter. In the illustrating example below the optimal dynamic threshold value $h \sim 1/s$ is depicted by the dashed line.

How can we learn to dynamically adjust the sensitivity of a detector?
Two cases: periodic and recurrent changes

**Periodic:** \( t_i = (i - 1) \times w + \delta t \), where \( w \) is the period width, e.g. changes are caused by events which happen daily 18.00 - 21.00

**Recurrent:** \( t_i = t_{i-1} + \delta t \), i.e. every next event happens after ‘approximately’ equal time intervals from the past change
Predictive change confidence function (PCCF) is the probability to observe any recurrent change out of sequence of all possible changes $c \in \langle c_i \rangle_{i=1}^{k}$ at any given time moment $t$:

$$\mathcal{P}(t \mid \theta) = \sum_{i=1}^{k} p(c_i = t \mid \theta).$$

In case of periodic changes:

$$\mathcal{P}(t \mid \theta) = \sum_{j=1}^{k=\text{div}(t,w)} p(c_1 = t - (j - 1)w \mid \theta).$$
Numerical computation of the PCCF for recurrent events

\[ p(c_1) = [p_1, p_2, p_3, p_4, p_5] \]

\[ p(c_2) = \sum_{\tau=1}^{4} p(c_2 = t | c_1 = \tau) p(c_1 = \tau) \]

\[ = \sum_{\tau=1}^{4} p(c_1 = t - \tau) p(c_1 = \tau) \]
Recurrent PCCF examples

Two Gaussian PCCF functions in which
\[ p(c_1 = t|\theta) = \mathcal{N}(t|\theta = \mu, \sigma), \]
\[ p(c_2 = t|\theta) = \mathcal{N}(t - c_1|\theta = \mu, \sigma), \] etc.

PCCF oscillates having local maximums at time moments \( k\mu \) and converging to the limit \( \frac{1}{\mu} \).
PCCF: the limit

PCCF for the moment $t$ is a sum

$$P(t|\theta) = \sum_{l=1}^{t} \frac{1}{\sigma \sqrt{2\pi l}} \exp \left( \frac{-(t - l\mu)^2}{2l\sigma^2} \right)$$

The limit $L$ of $P(t)$ when $t \to \infty$

$$L = \lim_{t \to \infty} \sum_{l=1}^{\infty} \frac{1}{\sigma \sqrt{2\pi l}} \exp \left( \frac{-(t - \mu l)^2}{2l\sigma^2} \right) = \frac{1}{\mu}.$$

The limit of the PCCF function is $\frac{1}{E(X)}$ where $X \sim \mathcal{P}(X)$. 
Part II: BD detector: Run lengths

In the Bayesian Online Changepoint detector (BD) [1] time occurrences of the changes are modelled by a latent variable run length $r_t$.

- Run length $r_t$ is the number of time steps since the most recent change.
- On each time step there are two possibilities: either the run length increases $r_t = r_{t-1} + 1$ or changepoint occurs $r_t = 0$.
- The conditional prior $P(r_t|r_{t-1})$ of the change is given by a constant-value hazard rate $h$.

\[
p(r_t|r_{t-1}) = \begin{cases} 
1 - h & \text{if } r_t = r_{t-1} + 1 \\
h & \text{if } r_t = 0 
\end{cases} \tag{1}
\]
BD detector [1] works by recursively estimating posterior probability distribution $P(r_t|x_{1:t}, \theta)$ of the run length $r_t$. Changepoint is an event when

$$\arg \max_{r_t} P(r_t|x_{1:t}, \theta) = 0$$  \hspace{1cm} (2)$$

Every time a new measurement $x_t$ is observed the posterior distribution is recalculated using the Bayes' theorem to update parameters of the distributions used to model the data

$$P(r_t|\cdot) = \sum_{r_{t-1}} P(r_t| r_{t-1}, \cdot) P(r_{t-1}|\cdot)$$  \hspace{1cm} (3)$$
The conditional prior probability of the change $P(r_t = 0|t)$ in the BD detector is a constant-value hazard rate $h$. This is a non-informative prior which does not hold enough information to distinguish outliers and noisy changes from the changepoints.

We improve performance of the BD detector by using an informative prior distribution for run lengths $r_t$ in a form of the PCCF function which is a probability estimator for the future run length values $r_t$ in the BD detector.
The approach is two-layered. The stream of input observations $\langle x_t \rangle$ and the stream of time intervals between changepoints $\langle c_i - c_{i-1} \rangle$ are modelled ‘uniformly’ - using the same data model.
BLPA: Data model

Common data model both for $\langle x_t \rangle$ and $\langle c_i - c_{i-1} \rangle$

- $\tilde{\mu}$ - estimated mean value
- $\tilde{\sigma}$ - estimated standard deviation

Following the notations in [3]

- **The prior** for $\tilde{\mu}$ and $\tilde{\sigma}$ is the normal-gamma distribution

$$P(\tilde{\mu}, \tilde{\sigma} | \tilde{\mu}_0, \tilde{\kappa}_0, \tilde{\alpha}_0, \tilde{\beta}_0) = N(\tilde{\mu}_0, \tilde{\kappa}_0 \tilde{\tau}) \Gamma(\tilde{\alpha}_0, \tilde{\beta}_0)$$ (4)

- **The posterior** is also the normal-gamma distribution (conjugacy property)

$$P(\tilde{\mu}, \tilde{\sigma} | D) \propto P(\tilde{\mu}, \tilde{\sigma} | \tilde{\mu}_0, \tilde{\kappa}_0, \tilde{\alpha}_0, \tilde{\beta}_0) P(D | \tilde{\mu}, \tilde{\tau})$$ (5)

$$\propto N(\tilde{\mu}_n, \tilde{\kappa}_n \tilde{\tau}) \Gamma(\tilde{\alpha}_0 + n/2, \tilde{\beta}_n)$$ (6)
BLPA: Pseudocode

1: $\theta \leftarrow (\mu_0, \kappa_0, \alpha_0, \beta_0)$  
2: $\theta^c \leftarrow (\mu_0^c, \kappa_0^c, \alpha_0^c, \beta_0^c)$  
3: $\theta = \theta_0$  
4: $\theta^c = \theta_0^c$  
5: $\langle H_j \rangle_{j=1}^T = \text{Pccf}(\theta_0^c)$  
6: \textbf{for} $t=1:T$ \textbf{do}  
7: \hspace{1em} $x \leftarrow [x, x_t]$  
8: \hspace{1em} $\pi_t = P(x_t|\theta)$  
9: \hspace{1em} $P(r_t = r_{t-1} + 1, x) = P(r_{t-1}, x_{1:t-1})\pi_t(1 - H_{t-1})$  
10: \hspace{1em} $P(r_t = 0, x) = H_{t-1} \sum_{r_{t-1}} P(r_{t-1}, x_{1:t-1})\pi_t$  
11: \hspace{1em} $P(r_t|x) = P(r_t, x)/P(x)$  
12: \hspace{1em} $\theta \leftarrow \text{Update}(\theta)$  
13: \hspace{1em} \textbf{if} (arg max$_{r_t} p(r_t|x, \theta) = 0$) \textbf{then}  
14: \hspace{2em} $\theta^c \leftarrow \text{Update}(\theta^c)$  
15: \hspace{2em} $\langle H_j \rangle_{j=t}^T = \text{Pccf}(\theta^c)$  
16: \hspace{1em} \textbf{end if}  
17: \textbf{end for}
The change detector can be considered as a binary classifier assigning labels ‘change’/‘not change’ to the incoming observations $x_t$.

- $e_t^+$ is the ‘change’ label assigned at the moment $t$ to $x_t$
- $e_t^-$ is the label ‘not change’ assigned at the moment $t$ to $x_t$

True Positive (TP), False Positives (FP), True Negatives (TN) and False Negatives (FN) events can be defined as follows:

- $e_t^+$ is TP if $\exists c_i : t - c_i < \delta$, and FP if $\nexists c_i : t - c_i < \delta$
- $e_t^-$ is FN if $\exists c_i : t - c_i < \delta$, and TN if $\nexists c_i : t - c_i < \delta$

The performance of the change detector is defined by TP/FP rates and by the average delay $\delta$ of the detection.
Experiments: artificial data streams

- 200 signals for each \( h \in [50 : 300] \) (\( \lambda = 1/h \) is a frequency of the changes) with 10 recurrent changes in the mean value.

**Figure 1:** Generated signal and corresponding PCCF. Dashed lines on the top plot depict moments when detector **without** PCCF alarmed changes.
Figure 2: ROC curve obtained by changing the $h = 1/\lambda$ value. FP rate is decreased while not reducing TP rate. In the worst cases the performance of both detectors is similar.
Experiments: Human Activity signal

Human activity data set [4] - sensor measurements from people performed static (standing, sitting, lying) and dynamic (walking, walking up/downstairs) activities. Changes are caused by transitions from one set of activities to another.

Figure 3: Dashed lines on the top plot are detections without PCCF.
Figure 4: ROC curve obtained by changing the $h = 1/\lambda$ value. FP rate is decreased when BD detector is used with the PCCF function.
Conclusions and future work

• We proposed **BLPA** - a new Bayesian method to improve accuracy of the Bayesian Online Changepoint detector for the data streams with recurrent changes.

• In the current version user have to
  1. test the hypothesis if changepoints in the input data stream are recurrent
  2. perform initial estimation of the prior probability distribution of the PCCF’s parameters.

• We plan to consider more real life data streams in order to develop procedures to automate this process.
Thanks!

- Please ask questions!
References

- Ryan Prescott Adams and David J.C. MacKay, “Bayesian Online Changepoint Detection”
- Michael I. Jordan “Chapter 9. The exponential family: Conjugate priors”
- J.L. Reyes-Ortiz, L. Oneto, A.Samà, X.Parra, D.Anguita, Transition-aware human activity recognition using smartphones