CHANGE DETECTION IN MULTIVARIATE DATASTREAMS CONTROLLING FALSE ALARMS

Luca Frittoli, Diego Carrera, Giacomo Boracchi

ECML PKDD 2021
VIRTUAL
13-17 September

QuantTree Exponentially Weighted Moving Average (QT-EWMA) is a novel online and nonparametric change-detection algorithm for multivariate datastreams that can be configured to yield a target Average Run Length (ARL_0), thus controlling the expected time before a false alarm. Our experiments on synthetic and real-world data demonstrate that QT-EWMA controls the ARL_0 better than state-of-the-art methods, achieving comparable detection delays.

CHANGE DETECTION

Identify distribution changes in streaming data:

$$x_1, x_2, \ldots \in \mathbb{R}^d \qquad x_t \sim \begin{cases} \phi_0 & \text{if } t < \tau \\ \phi_1 & \text{if } t \geq \tau \end{cases}$$

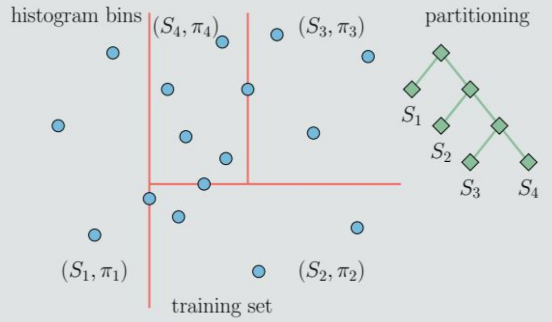
Goal: online monitoring at a **controlled Average Run Length** (ARL_0), i.e., the expected time before a **false alarm**

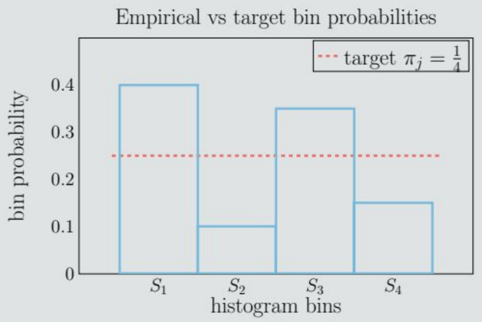
$$\mathrm{ARL}_0 = \mathbb{E}[t^*], \quad t^* = \text{detection time}$$

Applications: industrial monitoring, security, finance...

PROPOSED SOLUTION

We model ϕ_0 by QuantTree histogram (QT) [1] and monitor the bin probabilities of the datastream by EWMA statistics

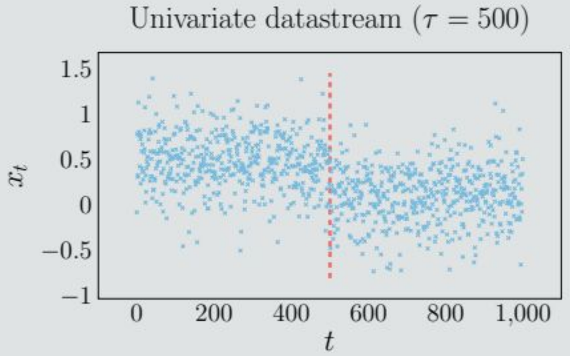


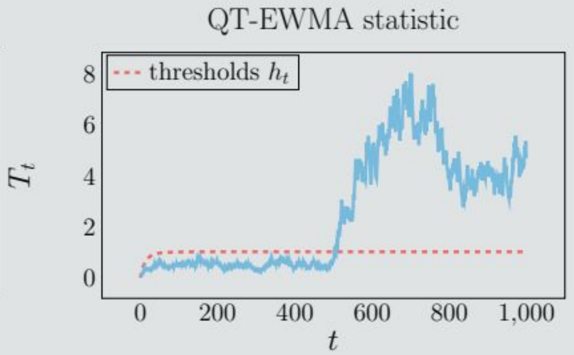


$$y_{j,t} = 1(x_t \in S_j) \quad Z_{j,t} = (1-\lambda)Z_{j,t-1} + \lambda y_{j,t}, \quad Z_{j,0} = \pi_j$$

We measure the deviation of $Z_{j,t}$ from target probabilities:

$$T_t = \sum_{j=1}^{K} \frac{(Z_{j,t} - \pi_j)^2}{\pi_j}, \text{ detection at } t^* = \min\{t : T_t > h_t\}$$





THRESHOLDS

QuantTree properties [1] imply that QT-EWMA is **non-parametric:** the distribution of T_t is **independent** from ϕ_0. We compute **thresholds** maintaining the **target ARL_0** via **Monte Carlo simulations** on univariate Gaussian data [1] by setting a constant **false alarm probability** [2]:

$$\mathbb{P}(T_t > h_t | T_k \leq h_t \forall k < t) = \alpha = \frac{1}{\mathrm{ARL}_0}$$

Then, the probability of a **false alarm** before time t is:

$$\mathbb{P}(t^* \leq t) = \sum_{k=1}^{t} \alpha(1-\alpha)^{k-1} = 1 - (1-\alpha)^t$$

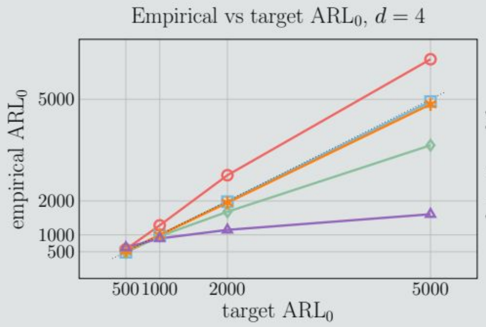
COMPUTATIONAL COST

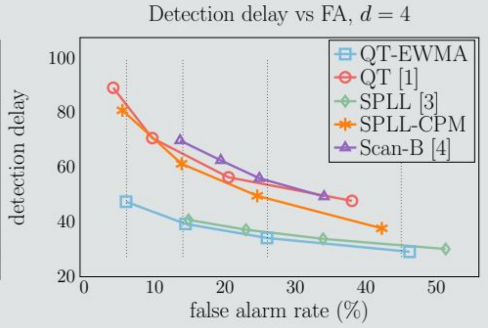
algorithm	QT-EWMA	QT [1]	SPLL [3]	SPLL-CPM	Scan-B [4]
complexity	$\mathcal{O}(K)$	$\mathcal{O}(K)$	$\mathcal{O}(md)$	$\mathcal{O}(md + w\log w)$	$\mathcal{O}(nBd)$
memory	K	K	1	w	$(n+1)Bd$

QT-EWMA has constant computational and memory costs that are independent from the data dimension d

EXPERIMENTS

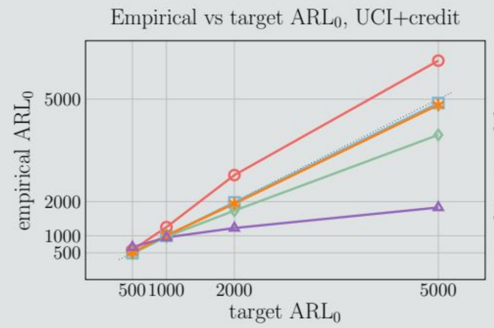
Synthetic Gaussian datastreams

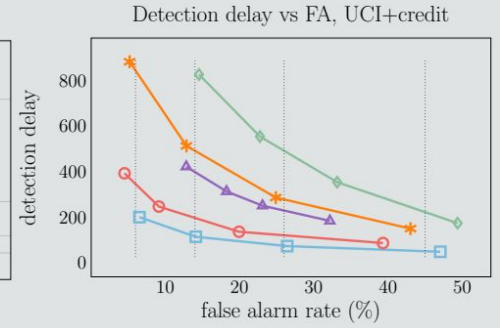




QT-EWMA achieves the target ARL_0, has low detection delays and yields the target false alarm rates

Real-world datastreams





QT-EWMA maintains the control over both ARL_0 and false alarm rates, and yields excellent detection delays

empirical ARL_0 = average stopping time t* on stationary datastreams
detection delay = average of t* - τ on datastreams with a change point at τ
false alarm rate = % of datastreams raising false alarms, i.e. t* < τ

CONCLUSIONS

- **QT-EWMA** extends QuantTree [1] to **nonparametric online change detection** controlling the ARL_0
- Maintains the **target ARL_0** on any datastream
- Effectively controls **false alarm rates**
- Achieves state-of-the-art **detection delays**

References:

[1] Boracchi, Carrera, Cervellera, Macciò "QuantTree: histograms for change detection in multivariate data streams" ICML 2018
[2] Margavio, Conerly, Woodall, Drake "Alarm rates for quality control charts" Statistics & Probability Letters 1995
[3] Kuncheva "Change detection in streaming multivariate data using likelihood detectors" TKDE 2011
[4] Li, Xie, Dai, Song "M-statistic for change-point detection" NIPS 2015

Corresponding author: Luca Frittoli (luca.frittoli@polimi.it)

code available at: https://github.com/diegocarrera89/quantTree