Scalable SUM-Shrinkage Schemes for Distributed Monitoring Large-Scale Data Streams

Kun Liu (kliu80@gatech.edu)

Collaborators: Ruizhi Zhang and Yajun Mei
H. Milton Stewart School of Industrial and Systems Engineering
Georgia Institute of Technology

Abstract

We investigate the problem of distributed monitoring large-scale data streams where an undesired event may occur at some unknown time and affect only a few unknown data streams.
We propose to develop scalable global monitoring schemes by parallel running local detection procedures and by combining these local procedures together to make a global decision based on SUM-shrinkage techniques.

Problem Formulation and Existing Research

Problem formulation

Online Monitoring independent large-scale data streams:

$$\begin{aligned} \text{Data Stream } 1: &\quad X_{1,1}, X_{1,2}, \cdots \\ \text{Data Stream } 2: &\quad X_{2,1}, X_{2,2}, \cdots \\ \cdots &\quad \cdots \\ \text{Data Stream } K: &\quad X_{K,1}, X_{K,2}, \cdots . \end{aligned}$$

At some *unknown* time ν, an event occurs and affects a few *unknown* data streams in the sense of changing the distributions of $X_{k,n}$'s from $N(0,1)$ to $N(\mu_k, 1)$, while μ_k may or may not be known.

Objective: Detect the true change time ν as soon as possible.
Mathematically, find a stopping time T to minimize the "worst case" detection delay proposed by Lorden (1971):

$$\overline{\mathbf{E}}_{\delta_1,\cdots,\delta_K}(T) = \sup_{\nu \geq 1} \operatorname{ess\,sup} \mathbf{E}^{(\nu)}\Big((T-\nu+1)^+ \Big| \mathcal{F}_{\nu-1}\Big),$$

subject to the global false alarm constraint:

$$\mathbf{E}^{(\infty)}(T) \geq \gamma. \tag{1}$$

Applications: Industrial quality control, signal detection, biosurveillance (CDC Biosense) etc.

Challenges:

- *Time domain*: Repeatedly test hypotheses of $H_0: \nu = \infty$ (no change) against $H_1: \nu = 1, 2, \ldots,$ (a change occurs) at each and every time step n when new data arrives.
- *Spatial domain*: We do not know which subset of data streams is affected, and the post-change parameter μ_k's might also be unknown. If r out of K data streams are affected, there are $\binom{K}{r}$ possible combinations.

Existing Research

- Tartakovsky and Veeravalli (2008) and Mei (2010): Assume the post-change parameter μ_k's are completely specified if affected.
- Xie and Siegmund (2013) proposed a semi-Bayesian scheme by assuming the proportions :

$$T_{XS}(a, p_0) = \inf\Bigg\{ n \geq 1 : \max_{0 \leq i < n} \sum_{k=1}^{K} \log(1 - p_0 + p_0 \exp\Big[\Big(\max\Big(0, \frac{1}{\sqrt{n-i}} \sum_{j=i+1}^{n} X_{k,j}\Big)\Big)^2 / 2\Big] \geq a \Bigg\}, \tag{2}$$

where p_0 is the fraction of affected data streams.

Our Proposed Methodology

Our proposed research has two components:

- Local detection statistics $W_{k,n}$'s that can efficiently detect local change at kth local sensor up to time n.
- A SUM-shrinkage method that combines the local detection statistics $W_{k,n}$'s suitably.

Let us postpone the discussion of $W_{k,n}$'s and focus on the SUM-shrinkage method first, which is motivated by parallel computing in the censoring sensor networks.

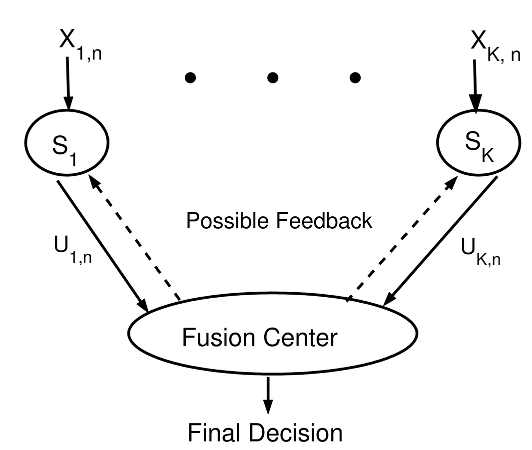

Figure 1: A widely used configuration of censoring sensor networks [8].

At time n, each local data stream does the dimension reduction by automatically filtering out non-changing streams.

$$U_{k,n} = h_k(W_{k,n}) =^{e.g.} \begin{cases} W_{k,n}, & \text{if } W_{k,n} \geq b_k \\ \text{NULL}, & \text{if } W_{k,n} < b_k \end{cases},$$

where $b_k \geq 0$ is the kth local censoring parameter.
At the global level, we use the local data streams that appear to be affected by the occurring event to make the decision. The general "SUM-shrinkage" form:

$$G_n = \sum_{k=1}^{K} U_{k,n} = \sum_{k=1}^{K} h_k(W_{k,n}). \tag{3}$$

Raise a global alarm at the time:

$$N_G(a) = \inf\{n \geq 1 : G_n \geq a\}. \tag{4}$$

Three special choices of G_n's are as follows:

- **Hard-thresholding**: Treat the "NULL" values as lower limit 0, and thus $h(x) = x\mathbf{1}\{x \geq b\}$ for some constant b,
$$G_n = \sum_{k=1}^{K} W_{k,n}\mathbf{1}\{W_{k,n} \geq b_k\}.$$
- **Soft-thresholding**: Treat the "NULL" values as upper limit b_k's, and thus $h(x) = \max\{x - b, 0\}$ for some constant b,
$$G_n = \sum_{k=1}^{K} \max(W_{k,n}, b_k).$$
- **Order-thresholding**: If (at most) r out of K data streams are affected by the occurring event, and thus $h(x) = x\mathbf{1}\{x \geq w_{(r)}\}$, $w_{(r)}$ is the r-th largest of $w_1, \ldots, w_K$,
$$G_n = \sum_{k=1}^{r} U_{(k),n}.$$

Non-Homogeneous Sensors with Known Post-Change Distributions

The $W_{k,n}$'s are chosen as the well-known local CUSUM statistics (Page 1954)

$$W_{k,n} = \max\Big(W_{k,n-1} + \mu_{k,n} - \frac{\mu_{k,n}^2}{2},\ 0\Big), \tag{5}$$

for $n \geq 1$ and $W_{k,0} = 0$ for $k = 1, \cdots, K$.

The choice of b_k's: If sensors are homogeneous, a "good" choice is $b_k = \rho_k b$, for $k = 1, \ldots, K$ and constant $b \geq 0$, where $\rho_k = \frac{I(g_k, f_k)}{\sum_{k=1}^{K} I(g_k, f_k)}$ and $I(g_k, f_k)$ is the Kullback-Leibler information number.
A choice of $b = (1/\rho_{\min}) \log \eta^{-1}$ will guarantee that on average, at most $100\eta\%$ of K sensors will transmit messages at any given time when no event occurs.

Theoretical results: Suppose that the delay effects δ_k's satisfy the following post-change hypothesis set Δ :

$$\Delta = \Big\{(\delta_1, \ldots, \delta_K) : \text{the } \delta_k\text{'s either} = \infty \text{ or satisfy } 0 \leq \delta_k << \log\gamma \text{ and } \min_{1 \leq k \leq K} \delta_k = 0\Big\},$$

where γ is the false alarm constraint in (1).

Theorem 1. *For any given post-change hypothesis $(\delta_1, \ldots, \delta_K) \in \Delta$ subject to the false alarm constraint (1), as γ goes to ∞, the hard-thresholding scheme $N_{hard}(a, b)$ asymptotically minimize $\overline{\mathbf{E}}(N_{hard}(a, b))$ (up to the first-order). The conclusion also holds for the soft thresholding scheme $N_{soft}(a, b)$ and the combined thresholding scheme $N_{comb,r}(a, b)$ when the occurring event affects at most r data streams.*

Homogeneous Sensors with Unknown Post-Change Distributions

Challenge: Determine the local detection statistics $W_{k,n}$'s properly when the post-change mean μ_k is unknown.

Motivation: The recursive register approach for one-sided problem in one-dimensional case by Lorden and Pollak (2008).

A technical assumption: $\mu \geq \rho$, where ρ is the smallest mean shift that is meaningful in practice.

Idea: Replace the unknown μ by its estimate from the past observed data in (5). At time n, the W_n can produce a candidate post-change time $\hat{\nu} \in \{0, 1, \cdots, n-1\}$, and thus μ is estimated by $X_{\hat{\nu}}, X_{\hat{\nu}+1}, \cdots, X_{n-1}$.

Procedure:

- Define $\hat{\nu}$ as the largest $0 \leq i \leq n-1$ such that $W_i = 0$, and denote by $T_n = n - \hat{\nu}$ and $S_n = \sum_{i=\hat{\nu}}^{n-1} X_i$ the total number and the summation of observations X_i's between the candidate post-change time $\hat{\nu}$ and time step $n-1$.
- A Bayes-type estimate of μ at time n:
$$\hat{\mu}_n = \max\Big(\rho, \frac{s + S_n}{t + T_n}\Big), \tag{6}$$
- Let $S_0 = T_0 = W_0 = X_0 = 0$, and $\hat{\mu}_1 = \rho$. For all $n \geq 1$,
$$W_n = \max\Big(W_{n-1} + \hat{\mu}_n X_n - \frac{1}{2}(\hat{\mu}_n)^2, 0\Big), \tag{7}$$
where the S_n and T_n in (6) has the recursive formula:

Homogeneous Sensors with Unknown Post-Change Distributions (Cont' d)

$$\begin{pmatrix} S_n \\ T_n \end{pmatrix} = \begin{cases} \begin{pmatrix} S_{n-1} + X_{n-1} \\ T_{n-1} + 1 \end{pmatrix}, & \text{if } W_{n-1} > 0, \\ \begin{pmatrix} 0 \\ 0 \end{pmatrix}, & \text{if } W_{n-1} = 0. \end{cases} \tag{8}$$

For the two-side test in multi-dimensional case: The local detection statistics for kth data stream is

$$W_{k,n} = \max(W_{k,n}^{(1)}, W_{k,n}^{(2)}),$$

where $W_{k,n}^{(1)}$ and $W_{k,n}^{(2)}$ are applied to detect positive and negative mean shifts, respectively. The estimate of μ_k's are defined in the similar form as in (6).
The three SUM-shrinkage methods can then be applied to combine this new local detection statistics $W_{k,n}$'s together.

Comparison: Xie and Siegmund's schemes are computationally heavy with large local memory requirements to store past information, so it is computationally infeasible for online monitoring large-scale data streams over long time period. However, our proposed scheme is scalable, computationally simple and fast.

Simulation Results:

Table 1: A comparison of detection delays when the change is instantaneous and the post-change mean $\mu_k = 1$ if affected. Results are based on 2500 Monte Carlo simulations.

γ		# local data streams affected								
		1	3	5	8	10	20	30	50	100
	Smallest standard error	0.19	0.08	0.06	0.04	0.03	0.02	0.01	0.01	0.00
	Largest standard error	0.40	0.14	0.08	0.05	0.04	0.03	0.02	0.02	0.01
5000	Xie and Siegmund's schemes $T_{XS}(a, p_0)$									
	$T_{XS}(a = 53.5, p_0 = 1)$	52.4	18.3	11.1	7.1	5.7	2.9	2.0	1.2	1.0
	$T_{XS}(a = 19.5, p_0 = 0.1)$	31.1	13.4	9.2	6.7	5.7	3.5	2.5	1.8	1.0
	Soft-thresholding Schemes $N_{soft}(a)$									
	$N_{soft}(a = 127.86, b_1 = 0)$	75.0	35.4	25.2	18.5	16.0	10.3	8.1	6.1	4.1
	$N_{soft}(a = 84.91, b_1 = 0.50)$	72.1	33.9	24.1	17.7	15.3	10.0	7.9	6.0	4.2
	$N_{soft}(a = 24.01, b_1 = \log(10))$	45.8	22.0	16.4	12.8	11.5	8.5	7.3	6.1	5.0
	$N_{soft}(a = 7.88, b_1 = \log(100))$	29.0	17.2	14.2	12.0	11.2	9.2	8.3	7.3	6.4
5×10^4	Soft-thresholding Schemes $N_{soft}(a)$									
	$N_{soft}(a = 136.07, b_1 = 0)$	89.0	39.9	27.9	20.2	17.4	11.1	8.7	6.5	4.4
	$N_{soft}(a = 92.79, b_1 = 0.50)$	85.7	38.2	26.8	19.4	16.7	10.7	8.4	6.3	4.4
	$N_{soft}(a = 29.05, b_1 = \log(10))$	55.1	25.3	18.4	14.1	12.6	9.1	7.8	6.5	5.2
	$N_{soft}(a = 11.11, b_1 = \log(100))$	35.5	19.7	16.0	13.4	12.4	10.0	8.9	7.9	6.8

Selected References (39 total publications)

[1] LORDEN, G. (1971). Procedures for reacting to a change in distribution. *Ann. Math. Statist.* **42** 1897–1908. MR0309251

[2] LORDEN, G. and POLLAK, M. (2008). Sequential change-point detection procedures that are nearly optimal and computationally simple. *Sequential Analysis* **27** 476-512. MR2460209

[3] MEI, Y. (2010). Efficient scalable schemes for monitoring a large number of data streams. *Biometrika* **97.2** 419-433. MR2650748

[4] PAGE, E. S. (1954). Continuous inspection schemes. *Biometrika* **41** 100–115. MR0088850

[5] ROBERTS, S. W. (1966). A comparison of some control chart procedures. *Technometrics* **8** 411–430. MR0196887

[6] SHIRYAEV, A. N. (1963). On optimum methods in quickest detection problems. *Theory Probab. Appl.* **8** 22–46.

[7] TARTAKOVSKY, A. G., and VEERAVALLI, V. V. (2008). Asymptotically optimal quickest change detection in distributed sensor systems. *Sequential Analysis,* **27(4)**, 441-475.

[8] VEERAVALLI, V. V., BASAR, T., and POOR, V. H. (1993). Decentralized sequential detection with a fusion center performing the sequential test. *Information Theory, IEEE Transactions on,* **39(2)**, 433-442.

[9] XIE, Y. and SIEGMUND, D. (2013). Sequential multi-sensor change-point detection. *Ann. Stat.*, **41** 670–692. MR3099117