Parallel Sequential Multi-Sensor Change-Point Detection

Yao Xie*, David Siegmund†

* Duke University, † Stanford University
Joint Statistical Meeting 2012

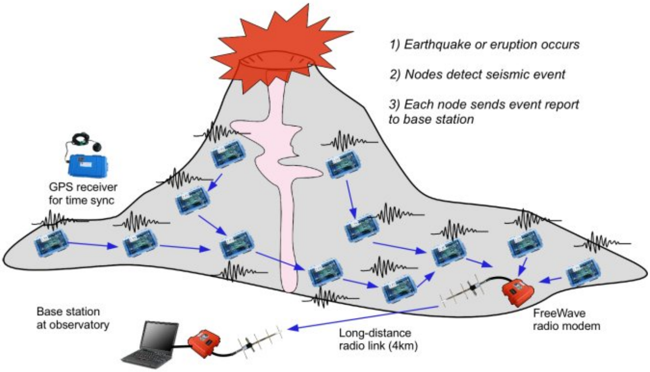
October 1, 2012

Outline

- Motivating applications
- Model
- Mixture procedure and parallel procedure
- Performance evaluation

Motivation: volcano monitoring

Change-point detection using multiple sensors

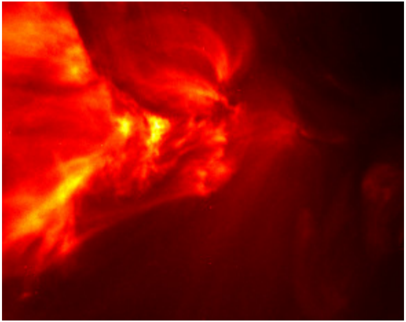


Anormaly detection
Harvard Sensor Networks Lab

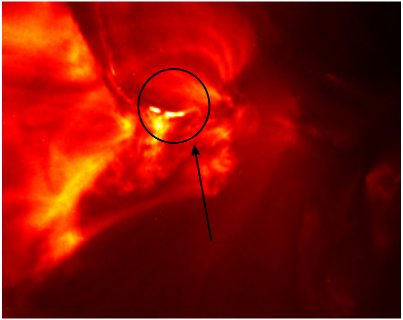
Solar flare detection

- Video sequences, each pixel is a “sensor”
- Very high-dimensional: # sensors $= 232 \times 292 = 67744$
- Goal: online detection of small and transient solar flare

t = 100



t = 226

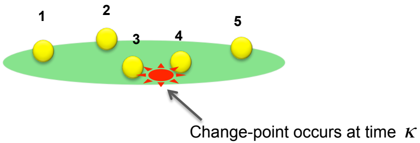


Source: NASA

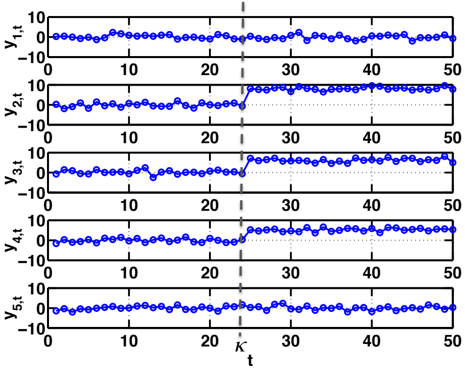
Challenges

- High-dimensional data
- Online detection
- Fundamental question:
  How to quickly detect temporal change of spatial data

Model



1
2
3
4
5
Change-point occurs at time κ



$y_{1,t}$
$y_{2,t}$
$y_{3,t}$
$y_{4,t}$
$y_{5,t}$
10
0
−10
0 10 20 30 40 50
κ
t

Formulation

Preprocessed data/model residuals:

$$\{y_{n,t}\}_{n=1,\cdot,N,t=1,2,\cdots},$$

$$\begin{aligned}
&\mathrm{H}_0: \quad y_{n,t} = w_{n,t}, \quad n = 1, \cdots, N, \quad t = 1, 2, \cdots \\
&\mathrm{H}_1: \quad \begin{cases} y_{n,t} = w_{n,t}, & n \in \mathcal{S}, \quad t = 1, \cdots, k; \\ y_{n,t} = \mu_n + w_{n,t}, & n \in \mathcal{S}^c, \quad t = k+1, \cdots \end{cases} \\
&w_{n,t} \sim \mathcal{N}(0, 1)
\end{aligned}$$

- Unknown parameters
  1. μ_n: changepoint amplitude
  2. $\mathcal{S}$: subset of sensors affected
  3. k: changepoint time

Simple procedure

- The simple approach is to use total energy:

$$\frac{1}{N}\sum_{n=1}^{N} y_{n,t}^2 \geq b$$

- Slightly more sophisticated, form a CUSUM statistic

$$T_{simple} = \inf\left\{ \max_{t-w\leq k\leq t} \sum_{i=k+1}^{t} \frac{1}{N}\sum_{n=1}^{N} y_{n,i}^2 \geq b. \right\}$$

w : window length

- These simple approaches have been widely used in engineering fields, network anomaly detection, video surveillance

Multichannel changepoint detection

- Likelihood ratio statistic

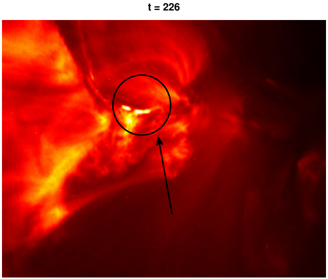
$$\max_{\mathcal{S}} \max_{1 \le k \le t} \sum_{n \in \mathcal{S}} \max_{\mu_n} \sum_{i=k+1}^{t} \{y_{n,i}\mu_n - \frac{\mu_n^2}{2}\}$$

- $\mathcal{S}$ has 2^N possibilities, $N = \#$ sensors
- One simplification:

$$T_{mc} = \inf \left\{ t : \max_{1 \le k \le t} \frac{1}{N} \sum_{n=1}^{N} \max_{\mu_n} \sum_{i=k+1}^{t} (y_{n,i}\mu_n - \frac{\mu_n^2}{2}) \right\}$$

Related to [Tartakovsky, Veeravalli 08]

Sparsity



- Number of sensors affected by the changepoint is small
- Fraction of affected sensors:

$$p = |\mathcal{S}|/N \ll 1$$

- Model this by assuming each sensor affected with probability p_0

Mixture procedure

$$T_{mix}(p_0, b)$$
$$= \inf\left\{ t : \max_{0 \le k < t} \sum_{n=1}^{N} \log\left(1 - p_0 + p_0 \exp[(U^+_{n,k,t})^2/2]\right) \ge b \right\}.$$

$$S_{n,t} = \sum_{i=1}^{t} y_{n,i},$$
$$U_{n,k,t} = (t-k)^{-1/2}(S_{n,t} - S_{n,k}).$$

Performance metrics

- Average run length (ARL)

$$\mathbb{E}^{\infty}\{T\}$$

  Average period of making false alarms when there is no changepoint.

- Expected detection delay

$$\max_{1\leq k\leq t} \mathbb{E}^{k}\{T-k|T>k\}$$

  Number of samples needed before claiming a detection when there is changepoint.

ARL of mixture procedure

Closed form expression for ARL of mixture procedure[⋆]

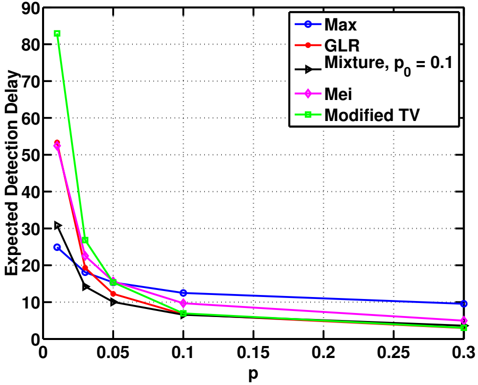
Table : Threshold for $T_{mix}(p_0, b)$, ARL ≈ 5000, $N = 100$, $m_1 = 200$.

Procedure	Monte Carlo ARL
$T_{mix}(1, 53.5)$	4978
$T_{mix}(0.1, 19.5)$	5000

⋆ "Sequential multisensor changepoint detection", Xie, Siegmund, 2012, submitted to Annuals of Stats.

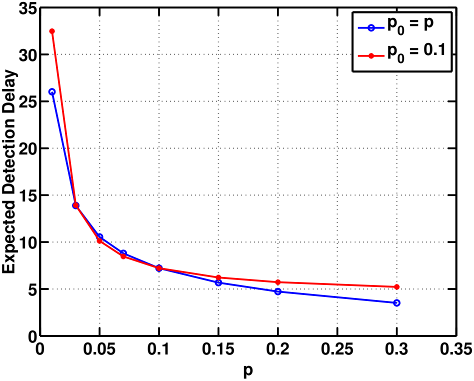
Detection delay

- Performance of mixture procedure is better



Sensitivity to p

- Mix procedure: good enough?
- We do not know true p
- When $p_0 \neq p$, N is large: $p_0 N$ very different from pN



$N = 100$

Parallel Procedure

- How to be more robust to uncertainty in p
- Use two mixture procedures

$$T_{parallel} \triangleq \min\{T_{mix}(p_1, b_1), T_{mix}(p_2, b_2)\}.$$

- Choose p_1 and p_2:

$$p \in [p_1, p_2]$$

ARL of Parellel

- No closed form ARL of Parallel
- Choose b_1 and b_2: ARL of two procedures equal
- Use a very conservative lower bound
  - $\mathbb{P}^\infty\{T_{mix}(p_i, b_i) \leq 1000\} \approx 0.05$, $i = 1, 2$
  - By Bonferroni inequality:

$$\mathbb{P}^\infty\{\min[T_{mix}(p_1, b_1), T_{mix}(p_2, b_2)] \leq 1000\} \leq 0.1,$$

  So $\mathbb{E}^\infty\{T_{parallel}\} \geq 10000$

Example

$$T_{parallel} = \min\{T_{mix}(0.02, 21.2), T_{mix}(0.33, 87.7)\}$$

Table : Thresholds $m_1 = 200$, $N = 400$.

Procedure	b	Monte Carlo ARL
T_{simple}	14.34	10000
T_{mc}	0.77	10000
$T_{mix}(0.1)$	44.7	10000
$T_{mix}(0.02)$	21.2	20000
$T_{mix}(0.33)$	87.7	20000
$T_{parallel}$		10000

Detection Delay: Parallel-1

Table : Comparison of Parallel and Single Procedures

p	$\mu_n = \mu$	FI	T_{simple}	T_{mc}	$T_{mix}(0.1)$	$T_{parallel}$
0.1	0.7	19.6	187.0	5.7	6.5	6.4
0.005	1.0	2.0	445.0	48.5	27.1	22.9
0.005	0.7	1.0	523.0	94.7	54.5	45.8
0.25	0.3	7.5	197.5	8.4	12.0	10.5

FI = Fisher Information, $\mu^2 Np/\sigma^2$

Parallel Procedure - 2

- Drawback of Parallel-1: no closed form ARL
- Parallel-2: linearly combine the statistics of $T_{mix}(p_i, b_i)$: $S(p_i)$

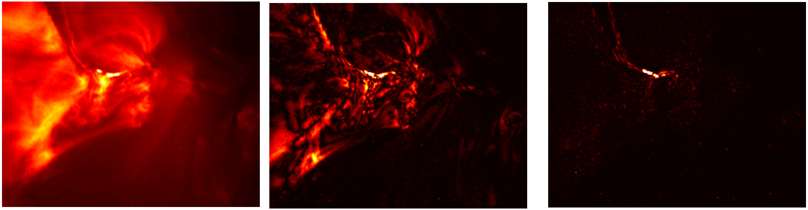
$$S_{parallel} = c_1 S(p_1) + c_2 S(p_2)$$

- Parallel-2

$$T_{parallel,2} = \inf\{t : S(parallel) > b\}$$

Summary

- Exploit sparsity in high-dimensional change-point detection
- Parallel procedure further boost performance
- Future work: how to extract detection statistic from high dimensional data
  Ongoing work:



ARL of Parallel-2

- Close-form approximation for the average run-length $N \to \infty$ and $b \to \infty$ with b/N fixed, $m_1 = o(b^r)$ for some positive integer r, and define θ by $\dot{\psi}(\theta) = b/N$

$$\mathbb{E}^{\infty}\{T_2\} \sim f(N, \theta, p_0) / \int_{[2N\gamma(\theta)/m_1]^{1/2}}^{[2N\gamma(\theta)/m_0]^{1/2}} y\nu^2(y) dy.$$

$g(x, c_1, c_2, p_1, p_2) = c_1 \log(1 - p_1 + p_1 \exp[(x^+)^2/2]) +$
$c_2 \log(1 - p_2 + p_2 \exp[(x^+)^2/2])$,
$\psi(\theta) = \log \mathbb{E}\{\exp[\theta g(U, c_1, c_2, p_1, p_2)]\}$

Table : Thresholds $N = 400$, $m_1 = 200$.

Procedure	b	Monte Carlo ARL
$T_{mix}(0.1)$	44.7	10000
$S_{parallel}$	46.24	10000

$$S_{parallel} = S(0.02) + 0.3S(0.33)$$

Table : Delay, $N = 400$, $m_1 = 200$

Procedure	p	μ	Expected Detection Delay
$T_{mix}(0.1, b)$	0.1	0.7	6.48
	0.005	1.0	27.08
	0.005	0.7	54.49
	0.25	0.3	11.96
$S_{parallel}$	0.2	0.7	4.48
	0.005	1.0	26.40
	0.005	0.7	49.79
	0.25	0.3	13.81

Improvement not as significant as Parallel procedure.