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

Anomaly 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

Source: NASA
Challenges

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

Change-point occurs at time $\kappa$
Formulation

Preprocessed data/model residuals:

\[
\{y_{n,t}\}_{n=1,\ldots,N, t=1,2,\ldots}
\]

\[H_0:\ y_{n,t} = w_{n,t}, \quad n = 1, \ldots, N, \quad t = 1, 2, \ldots\]

\[H_1:\ \begin{cases} 
  y_{n,t} = w_{n,t}, & n \in S, \quad t = 1, \ldots, k; \\
  y_{n,t} = \mu_n + w_{n,t}, & n \in S^c, \quad t = k + 1, \ldots
\end{cases}\]

\[w_{n,t} \sim \mathcal{N}(0, 1)\]

- Unknown parameters
  1. \(\mu_n\): changepoint amplitude
  2. \(S\): 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

\[
\max_S \max_{1 \leq k \leq t} \sum_{n \in S} \max_{\mu_n} \sum_{i=k+1}^{t} \{ y_{n,i} \mu_n - \frac{\mu_n^2}{2} \}
\]

- \( S \) has \( 2^N \) possibilities, \( N = \) \# sensors

- One simplification:

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

Related to [Tartakovsky, Veeravalli 08]
Sparsity

- Number of sensors affected by the changepoint is small
- Fraction of affected sensors:
  \[ p = \frac{|S|}{N} \ll 1 \]
- Model this by assuming each sensor affected with probability \( p_0 \)
Mixture procedure

\[ T_{\text{mix}}(p_0, b) \]

\[ = \inf \left\{ t : \max_{0 \leq k < t} \sum_{n=1}^{N} \log \left( 1 - p_0 + p_0 \exp\left[ (U_{n,k,t}^+)^2 / 2 \right] \right) \geq b \right\}. \]

\[ S_{n,t} = \sum_{i=1}^{t} y_{n,i}, \]

\[ U_{n,k,t} = (t - k)^{-1/2} (S_{n,t} - S_{n,k}). \]
Performance metrics

- Average run length (ARL)

\[ E^\infty \{ T \} \]

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

- Expected detection delay

\[ \max_{1 \leq k \leq t} E^k \{ T - k \mid T > k \} \]

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 $\approx 5000$, $N = 100$, $m_1 = 200$.

<table>
<thead>
<tr>
<th>Procedure</th>
<th>Monte Carlo ARL</th>
</tr>
</thead>
<tbody>
<tr>
<td>$T_{mix}(1, 53.5)$</td>
<td>4978</td>
</tr>
<tr>
<td>$T_{mix}(0.1, 19.5)$</td>
<td>5000</td>
</tr>
</tbody>
</table>

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: $p_0N$ very different from $pN$

![Graph showing expected detection delay for different values of $p$. The graph has a title indicating $N = 100$. The x-axis represents $p$, ranging from 0 to 0.35, and the y-axis represents expected detection delay in years, ranging from 0 to 35. Two curves are present, one for $p_0 = p$ and another for $p_0 = 0.1$. The graph demonstrates a downward trend as $p$ increases.]
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 $p_2$:

$$p \in [p_1, p_2]$$
No closed form ARL of Parallel

Choose $b_1$ and $b_2$: ARL of two procedures equal

Use a very conservative lower bound

1. $\mathbb{P}^\infty \{ T_{mix}(p_i, b_i) \leq 1000 \} \approx 0.05, i = 1, 2$

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

\[ T_{\text{parallel}} = \min\{ T_{\text{mix}}(0.02, 21.2), T_{\text{mix}}(0.33, 87.7) \} \]

Table: Thresholds \( m_1 = 200, \; N = 400 \).

<table>
<thead>
<tr>
<th>Procedure</th>
<th>( b )</th>
<th>Monte Carlo ARL</th>
</tr>
</thead>
<tbody>
<tr>
<td>( T_{\text{simple}} )</td>
<td>14.34</td>
<td>10000</td>
</tr>
<tr>
<td>( T_{\text{mc}} )</td>
<td>0.77</td>
<td>10000</td>
</tr>
<tr>
<td>( T_{\text{mix}}(0.1) )</td>
<td>44.7</td>
<td>10000</td>
</tr>
<tr>
<td>( T_{\text{mix}}(0.02) )</td>
<td>21.2</td>
<td>20000</td>
</tr>
<tr>
<td>( T_{\text{mix}}(0.33) )</td>
<td>87.7</td>
<td>20000</td>
</tr>
<tr>
<td>( T_{\text{parallel}} )</td>
<td></td>
<td>10000</td>
</tr>
</tbody>
</table>
### Table: Comparison of Parallel and Single Procedures

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

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 $T_{mix}(p_i, b_i)$: $S(p_i)$
  
  $S_{\text{parallel}} = c_1 S(p_1) + c_2 S(p_2)$

- Parallel-2

  $T_{\text{parallel,2}} = \inf\{t : S(\text{parallel}) > b\}$
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:
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 $\theta$ by $\psi(\theta) = b/N$

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

$$
g(x, c_1, c_2, p_1, p_2) = c_1 \log(1 - p_1 + p_1 \exp[(x^+)^2/2]) + c_2 \log(1 - p_2 + p_2 \exp[(x^+)^2/2]),
$$

$$
\psi(\theta) = \log E\{ \exp[\theta g(U, c_1, c_2, p_1, p_2)] \}.
$$
Table: Thresholds $N = 400, m_1 = 200.$

<table>
<thead>
<tr>
<th>Procedure</th>
<th>$b$</th>
<th>Monte Carlo ARL</th>
</tr>
</thead>
<tbody>
<tr>
<td>$T_{mix}(0.1)$</td>
<td>44.7</td>
<td>10000</td>
</tr>
<tr>
<td>$S_{parallel}$</td>
<td>46.24</td>
<td>10000</td>
</tr>
</tbody>
</table>

$$S_{parallel} = S(0.02) + 0.3S(0.33)$$
Table: Delay, $N = 400$, $m_1 = 200$

<table>
<thead>
<tr>
<th>Procedure</th>
<th>$p$</th>
<th>$\mu$</th>
<th>Expected Detection Delay</th>
</tr>
</thead>
<tbody>
<tr>
<td>$T_{mix}(0.1, b)$</td>
<td>0.1</td>
<td>0.7</td>
<td>6.48</td>
</tr>
<tr>
<td></td>
<td>0.005</td>
<td>1.0</td>
<td>27.08</td>
</tr>
<tr>
<td></td>
<td>0.005</td>
<td>0.7</td>
<td>54.49</td>
</tr>
<tr>
<td></td>
<td>0.25</td>
<td>0.3</td>
<td>11.96</td>
</tr>
<tr>
<td>$S_{parallel}$</td>
<td>0.2</td>
<td>0.7</td>
<td>4.48</td>
</tr>
<tr>
<td></td>
<td>0.005</td>
<td>1.0</td>
<td>26.40</td>
</tr>
<tr>
<td></td>
<td>0.005</td>
<td>0.7</td>
<td>49.79</td>
</tr>
<tr>
<td></td>
<td>0.25</td>
<td>0.3</td>
<td>13.81</td>
</tr>
</tbody>
</table>

Improvement not as significant as Parallel procedure.