Anomaly Detection in Cyber Physical Systems

Maggie Cheng

Illinois Institute of Technology

IEEE Big Data Conference
December 11, 2018
Seattle, WA
Outline

Introduction

Outlier Detection

Sequential Change Point Detection

High Dimensional Data

Summary and Outlook
Anomaly Analysis

- Unusual and significant changes in the network or CPS
- Necessary to detect, and remove/mitigate
- Analyzing anomalies from data is a big data analytics problem:
  - large amount of data
  - high speed
  - high dimensional
  - heterogeneous types
  - noisy
- Involves extracting and interpreting anomalous patterns from data
  - difficult to define a baseline for normal pattern
Anomaly Analysis

The steps in diagnosing an anomaly

1. Detection: binary, or real-valued anomaly score
   - The time at which the anomaly is observed
2. Identification: selecting the anomaly type from a set of possible candidate anomalies
   - Zero-day attack?
3. Localization
   - Which link/node, component?
4. Quantification of impact
   - A measure of the importance of the anomaly

The scope of this tutorial

▶ Detection
▶ Cyber + physical anomalies
Types of Anomalies

Outlier

- Aberrant observations that are considerably different from the majority are outliers.
- "An outlier is an observation that deviates so much from other observations as to arouse suspicion that it was generated by a different mechanism" (Hawkins, 1980)
- Indicator of problems
- With or w/o time dimension

Change Point on Time Series

- Abrupt change in time series
- Indicator of change of state in the underlying process
Types of Anomalies

Outlier
- Location outliers
- Scatter outliers
- Combination of both

Change Point
- Change in mean
- Change in variance
- Combination of both
Outlier vs. Change Point on Time Series

Change point may not be an outlier; Outlier may not be a change point

Introduction
Outline

Introduction

Outlier Detection

Sequential Change Point Detection

High Dimensional Data

Summary and Outlook
Outlier Detection

Applications of Outlier Analysis

- Credit card fraud detection
- Intrusion detection in networks
- Bad data detection in power grid
  - Observing abnormal values indicates measurement errors or errors in the generating process
- Many other applications ...

Evolution of Outlier Detection Techniques

- From univariate to multivariate (multi-dimensional)
- From one normal class to multiple normal classes (more than one generating mechanism underlying the data)
Outlier Detection

Detection Techniques

- Statistical outlier detection (Barnett and Lewis, 1994)
  - Based on distribution
- Distance-based (Knorr and Ng, 1998)
  - Based on notion of proximity
- Density-based (Breunig et al., 2000)
  - Based on local outlier factor
Example I: Bad Data Detection in Power Systems

Weighted Least Square Based Method

\[ k = 0, \text{ initialize state variable } x^0 \]

Newton-Raphson

\[ \Delta x^k = x^{k+1} - x^k \]

\[ \max |\Delta x^k| < \epsilon? \]

Yes

No

\[ k \leftarrow k + 1 \]

\[ \chi^2 - \text{test} \]

Bad data?
Example I: Bad Data Detection in Power Systems

**WLS-Based Error Detection in Power Systems**

- **Linear WLS**

  \[
  \text{minimize} \quad f = \|e\|^2 = e^T \cdot e = \sum_{i=1}^{m} w_i \left[ z_i - \sum_{j=1}^{n} a_{ij} x_j \right]^2
  \]

- **Non-Linear WLS**

  \[
  \text{minimize} \quad f = \|e\|^2 = e^T \cdot e = \sum_{i=1}^{m} \frac{1}{\sigma_i^2} \left[ z_i - h_i(x) \right]^2
  \]

  - \(x\): state variables
  - \(z\): measurements
  - \(h(x)\): non-linear measurement functions
  - \(z_i - h_i(x)\): residual of the \(i^{th}\) measurement
Example 1: Bad Data Detection in Power Systems

Weighted Least Square Based Method: Using $\chi^2$-Test

- Let $z_i$ be measurement, $\hat{z}_i$ be anticipated value for $z_i$.
- Assume residuals $X_i = z_i - \hat{z}_i$ are Gaussian and independent.
- Therefore, $\chi^2 = \sum_{i=1}^{n} (X_i/\sigma_i)^2$ follows $\chi^2$-squared distribution with DOF $\nu = n$.
- Given a significance level $\alpha$, if $\chi^2 > \chi_{\alpha}^2 \rightarrow$ bad data detected.

Assumptions in WLS

- Errors in measurements are Gaussian and independent.
- System topology model is correct.
Example II: Line Outage Detection in Power Systems

Topology Error: Discrepancy Between Assumed Model and Actual Model

(h) Assumed model

(i) Actual model

IEEE 9-Bus Test System
Example II: Line Outage Detection in Power Systems

System Matrix Known

- Hypothesis testing concerning the mean vector of the residuals
- Use the asymptotic distribution theory of Maxima

System Matrix Unknown

- Perform change point detection on time series → anomaly detected
- Identify and locate specific line outage
Example II: Line Outage Detection in Power Systems

A New Topology Error Detection Approach

\(x\): state variables
\(z\): measurements
\(h(x)\): non-linear measurement functions
\(\hat{z}_i = h_i(x)\): the anticipated value for the \(i^{th}\) measurement
\(X_i = z_i - \hat{z}_i\): the residual of the \(i^{th}\) measurement

- Without topology errors, it is expected that the residuals \(X_i\) are normally distributed with zero means:

\[ X_i \sim N(\mu_i, \sigma_i^2) \text{ and } \mu_i = 0, \forall i \]

- The problem of topology error detection is cast as a problem of testing the hypothesis of whether the mean vector of a stochastic process is zero.
Example II: Line Outage Detection in Power Systems

Hypothesis Testing Approach

- **Input:** the residuals for the \( n \) redundant measurements:
  \[ X_i = z_i - \hat{z_i}, \quad i = 1 \ldots n. \]
- **Hypotheses:**
  \[ H_0 : \quad \mu_1 = \ldots = \mu_n = 0 \]
  \[ H_1 : \quad \exists i, \text{ such that } \mu_i \neq 0 \]

New Test Statistics

Define \( M_n = \max_{i \leq n} |X_i|/\sigma_i \)

- \( M_n \) is the maximum standardized residual
- Hypothesis testing based on \( M_n \): Compute \( \bar{M}_n \) from data and compare \( \bar{M}_n \) with what is expected under the null hypothesis.
Hypothesis Testing

Hypothesis Testing Framework
Given a significance level \( \alpha \in (0, 1) \), find a cutoff value \( t_n \) such that under the null hypothesis \( H_0 \),

\[
\mathbb{P}(M_n > t_n) = \alpha
\]

▶ If \( \bar{M}_n > t_n \), reject \( H_0 \)

▶ How to find the cutoff value \( t_n \)?
Detection Threshold $t_n$: When $X_i$ Are Independent

From $M_n = \max_{i \leq n} |X_i|/\sigma_i$ and $\mathbb{P}(M_n > t_n) = \alpha$, we have

$$1 - \alpha = \mathbb{P}(M_n \leq t_n)$$

$$= \prod_{i=1}^{n} \mathbb{P}(|X_i|/\sigma_i \leq t_n)$$

$$= [1 - 2\Phi(-t_n)]^n$$

where $\Phi(x) = \int_{-\infty}^{x} \phi(u)\,du$, and $\phi(x) = (2\pi)^{-1/2}e^{-x^2/2}$.

Let $\Phi^{-1}$ be the inverse function of $\Phi$. Hence

$$t_n = -\Phi^{-1}\{[1 - (1 - \alpha)^{1/n}] / 2\}$$  \hspace{1cm} (1)
Asymptotic Theory for $M_n$

When $X_i$ are dependent

- Need to know the asymptotic distribution for $M_n$ under dependence

**Definition (Measure of Dependence)**

- Let $X_1, X_2, \ldots, X_n$ be a non-stationary Gaussian process; $r_{ij}$ be the correlation between $X_i$ and $X_j$, for $1 \leq i, j \leq n$.
- Define a skeleton index set

  $$S(\delta) = \{(i, j) : |r_{ij}| > \delta, \ 1 \leq i, j \leq n\}$$

- The cardinality of the set $S(\delta)$ is a measure of dependence.
- Large $|S(\delta)|$ implies strong overall dependence.
Asymptotic Theory for $M_n$

Theorem (Weak Dependence Condition (Wu et al., 2016))

*If there exists $\lambda \in (0, 1)$ and constant $C, \alpha > 0$, such that*

$$\max_{ij} |r_{ij}| < \lambda$$

*and for every $\delta \in (0, 1)$, the cardinality of the skeleton index set $S(\delta)$ satisfies*

$$|S(\delta)| \leq Cn\delta^{-\alpha}$$

*Then for every $0 < \alpha < 1$, we have*

$$\lim_{n \to \infty} \mathbb{P}(M_n \leq t_n) = 1 - \alpha$$

Remarks

1. Under conditions (2) and (3), the maxima $M_n$ has the same asymptotic distribution as the one obtained under independence.

2. When power grids are sparsely connected, the condition is easily satisfied.
Detecting a Topology Error in IEEE 118-Bus System

Setup

▶ 118 state variables, 301 measurements: 183 redundant measurements.
▶ Topology change: remove the transmission line between bus 37 and 38.

Procedure

▶ Obtain the residuals \((X_i)_{i=1}^{n}\) by using non-WLS state estimation.
▶ Estimate the standard error \(\hat{\sigma}\).
▶ Compute the standardized residual \(X_i/\hat{\sigma}_i, \ i = 1, \ldots, n\).
▶ Compute the cutoff value \(t_n\) with significance level \(\alpha = 0.001\) and \(n = 183\).
▶ By hypothesis testing approach, if there is an index \(i\) such that \(|X_i|/\hat{\sigma}_i > t_n\), reject \(H_0\), and report alarm.
Detecting a Topology Error in IEEE 118-Bus System

Results

- $t_n = 4.546012$
- Identify the index set $J = \{ j : |X_j| / \hat{\sigma}_j > t_n \}$

$$J = \{112, 113, 114, 115, 116, 124, 125, 127, 129, 130, 154, 155, 162, 165\}$$

(j) Residuals $X_i$

(k) Standard errors $\hat{\sigma}_i$

(l) $X_i / \hat{\sigma}_i$
Example III: PCA-Based Network Traffic Anomaly Detection

(Lakhina et al., 2004a; Huang et al., 2006)

- Outlier detection from time series
- Use Principal Component Analysis (PCA) to detect anomalies
- Detect volume anomaly in network traffic
  - Use O-D flow information
  - Each link has aggregated traffic from all O-D flows
Example III: PCA-Based Network Traffic Anomaly Detection

PCA Analysis on Network Data

- Form link data matrix $Y_{m \times n}$
  - $m$: last $m$ data points, $n$: number of links
- Perform PCA on $Y$
  - Find the principle components:
  1. The first principle component $v_1$:

$$ v_1 = \arg \max_{\|v\|=1} \|Yv\| $$

$$ \|Yv\|^2 $$ is proportional to the variance of the data measured along $v$. 

Outlier Detection from Time Series
Example III: PCA-Based Network Traffic Anomaly Detection II

2 The $k$-th principle component $v_k$ is:

$$v_k = \arg \max_{\|v\|=1} \| (Y - \sum_{i=1}^{k-1} Y v_i v_i^T) v \|$$

- The matrix $P = [v_1, v_2, \ldots, v_k]$ is formed by the first $k$ principle components, which capture the dominant variance in the data

- Separate normal and anomalous network-wide traffic

$$Y = \hat{Y} + \tilde{Y}$$

$$\hat{Y} = PP^T Y, \text{ and } \tilde{Y} = (I - PP^T) Y$$

- $\tilde{Y}$: contains residual traffic.

- Volume anomaly will result in a large spike in $\tilde{Y}$
Outline

Introduction

Outlier Detection

Sequential Change Point Detection

High Dimensional Data

Summary and Outlook
Sequential Change Point Detection

To detect a change point in a time series \( \{X_1, X_2, \ldots, X_n\} \), it is assumed that the pre-change density is \( f_0 \), and if a change occurs at time \( \nu \), then the post-change density becomes \( f_1 \) beginning from moment \( \nu + 1 \).

The hypotheses are then formulated as:

- \( H_0: \quad \{X_1, X_2, \ldots, X_n\} \sim f_0 \)
- \( H_1: \quad \{X_1, X_2, \ldots, X_\nu\} \sim f_0, \) and \( \{X_{\nu+1}, X_{\nu+2}, \ldots, X_n\} \sim f_1 \)

The Change Point Detection Problem is to decide

1. which hypothesis is true?
2. if \( H_1 \) is true, \( \nu =? \)

The time instance \( \nu \), at which the state of the process changes is referred to as the *change point or time of change*. 
Sequential Change Point Detection

Algorithms

- Cumulative Sum Algorithm (CUSUM)
- Shiryaev-Roberts Procedure
- Sliding Window Algorithm

Performance Metrics

- False positive and false negative rates
  - If a change occurred but the detection procedure failed to detect it: false negative (misdetection)
  - If the detection time $N < \nu$: false positive (false alarm)
- Detection delay
  - If there is a true change and the time of change is $\nu$, the detection time is $N$, then detection delay $\tau = N - \nu$
CUSUM (Page, 1954)

Optimality
CUSUM is optimal in the sense of minimizing worst case detection delay.

Assumptions
- Observations $X_1, X_2, \ldots X_n$ are independent, iid pre-change and iid post-change
- Probability density functions: $f_0$ before change; $f_1$ after change
- Assume $f_0$ and $f_1$ are known
- The only thing unknown is $\nu$, the time of change

Parameter
- Detection threshold $h$
CUSUM

Parametric CUSUM: Based on Maximum Likelihood Principle

Detection statistics:

\[ W_n = \max(W_{n-1} + Z_n, 0) \text{ for } n \geq 1 \]

where \( W_0 = 0 \)
\[ Z_n = \log L_n \]

\( L_n \) is the likelihood ratio: \( L_n = \frac{f_1(X_n|X_{1}^{n-1})}{f_0(X_n|X_{1}^{n-1})} \), or \( L_n = \frac{f_1(X_n)}{f_0(X_n)} \) for i.i.d.

The procedure declares a change as soon as the detection statistics \( W_n \) exceeds a preset threshold \( h \):

\[ N = \min\{n \geq 1 : W_n \geq h\} \]
The Problem Setting

- Original Setting: "Quickest Detection of a Disorder in a Stationary Regime"
- The change is possibly taking place at a far horizon
- A randomized version for a general discrete time setting
- Applications: target detection and tracking, rapid detection of intrusions in communication networks, environmental monitoring
- Early detection of changes that may occur in a distinct future
Shiryaev-Roberts Procedure

The Algorithm

1. Shiryaev-Roberts statistic $R_n = \sum_{k=1}^{n} \frac{p(X_1,\ldots,X_n|\nu=k)}{p(X_1,\ldots,X_n|\nu=\infty)}$

2. From independence assumption: $R_n = \sum_{k=1}^{n} \prod_{i=k}^{n} \frac{f_1(X_i)}{f_0(X_i)}$

3. $R_n$ can be computed recursively: $R_n = (1 + R_{n-1}) \frac{f_1(X_n)}{f_0(X_n)}$, for $n \geq 1$; $R_0 = 0$

4. Stopping time: $R_{AB} = \min\{n \geq 1 : R_n \geq AB\}$

Parameter: $AB$ is chosen such that $E_\infty N_{AB} = B$

$B$ is a preset value before surveillance begins.
**Shiryaev-Roberts Procedure**

**Detection Delay**  Shiryaev-Roberts procedure is the best in terms of minimizing the expected detection delay (asymptotically).

**Theorem**

*Shiryaev-Roberts procedure minimizes*

\[
\sum_{k=1}^{\infty} E_k(N - k)^+
\]

over all stopping times \(N\) that satisfy \(E_{\infty}(N) \geq B\).

**CUSUM and S-R Procedure**

- Based on ratio of likelihoods
  - S-R procedure is a CUSUM-type of algorithm
- Difficult to apply when \(f_1\) and \(f_0\) are unknown
Sliding Window Algorithm

Preset Parameters

- Window size $m$ ($m \ll N$, the total number of data points)
- Significance level $\alpha$ (e.g., $\alpha = 0.05$)

Sliding Window Algorithm (Cheng et al., 2016)

1. Set window offset $d = 0$.
2. Compute the sum $S_1 = \sum_{i=d+1}^{d+m} X_i$, and $S_2 = \sum_{i=d+m+1}^{d+2m} X_i$.
3. If $|S_2 - S_1| \geq z\sigma \sqrt{2m}$, declare a change point $\hat{\nu} = d + m$
4. Else set $d = d + 1$, go to line 2.

Remarks:

- $z$ is the critical value that provides an area of $\alpha$ in the upper tail of the standard normal distribution.
- $\sigma^2$ is the variance, updated as the window moves.
Sliding Window Algorithm

Algorithm Properties

- Be able to detect a change in state without knowing the actual pre- and post-change densities
- Relate detection threshold to a tolerable false alarm rate — controlled trade-off
- Relate detection threshold to the dynamic characteristics of the data and not use a preset value
- Be able to detect abrupt changes as well as slow and subtle changes
- Avoid mistaking an isolated outlier as a change for a new state
Applications Using Sequential Change Point Detection

- DoS Attack Detection
  - SYN flood attack
- Attack Detection in Wireless Networks
  - Network layer
  - MAC layer
  - Physical layer
- Power Grid Anomaly Detection
DoS Attack Detection

A Common DoS Attack: SYN Flood Attack

- Attacker sends control packets to compromised nodes
- A large number of flooding sources send an excessive number of SYN requests to the victim
- The victim server returns SYN/ACK packet to the client waiting for ACK until timeout
- Flooding sources never return an ACK
- Exhaust the victim server’s backlog queue → all connection requests dropped

Challenges: preset threshold (X)

- Traffic patterns vary from site to site, from time to time
- Per-flow state information not known
- Normal traffic models hard to define
DoS Attack Detection

How to detect w/o prior knowledge of flow and traffic info?
Detection mechanism must be insensitive to site and traffic patterns.

- There is no normal traffic model or flow rate, but there is normal behavior
- Baseline: protocol behavior (TCP connection management)
  - Normal: FINs match with SYN requests from clients
  - Packet drop/retransmission cause small discrepancy
  - Under SYN flood attack: Large difference between the number of SYN and FINs received
SYN Flood Attack

- Attackers create a large number of "open" connections
- Change in network measurement: $|SYN - FIN|$ shows abrupt increase
**DoS Attack Detection**

**Detection Procedure**

- Monitor the number of SYNs and FINs
  - at egress router (near the flooding source)
  - at ingress router (near the victim server)

- Generate time series on (SYNs–FINs)

- Perform sequential change point detection on time series
  - Non-parametric version
DoS Attack Detection

Non-Parametric CUSUM for Change Point Detection (Wang et al., 2002)

- Tunable parameters: $a, N$
- Observations: $S$: number of SYNs; $F$: number of FINs

1. $D_n = S_n - F_n$
2. $R_n = \alpha(R_{n-1}) + (1 - \alpha)F_n$
3. $X_n = D_n / R_n$
4. choose constant $a > E(X_n)$
5. Test statistic: $y_n = (y_{n-1} + (X_n - a))^+, \; y_0 = 0$
6. Detection: first $n$ such that $y_n > N$

Remarks

- Algorithm very sensitive to $N$ and $a$.
- Difficulty: determining $N$ and $c$ before monitoring begins.
Wormhole Attack in Wireless Ad Hoc Networks

Routing: A category of routing protocols use shortest path routing.
- Nodes exchange local information and relay to others
- Nodes collectively decide a route towards a destination
- Select the ”best route” based on hop count (shortest path routing)

Wormhole Attack
- Adversary controls two end points and a tunnel between them
- Attract traffic to go through the controlled wormhole tunnel by making false route advertisement— a shorter path towards a destination
In-Band vs. Out-Band Wormhole Attack

In-Band Wormhole Attack

▶ Wormhole tunnel consists of other wireless nodes controlled by the adversary
▶ Re-routed packets go through these wireless nodes

Out-Band Wormhole Attack

▶ Wormhole tunnel is an external link
  - A wired link
  - A wireless link (e.g., a long-range directional link)

This talk: address in-band wormhole attack
Wormhole Attack Detection

Performance Degradation in an In-Band Wormhole Attack

- End-to-end delay increases
- Throughput decreases
- Packet Deliver Ratio drops (If the wormhole endpoints drops packets arbitrarily)
- and more ...

Proposed Method

- Model the end-to-end delay of a flow as a time series
- Perform Change Point Detection on the time series to detect the change
Stationary Network — Setup

- An in-band wormhole tunnel is established between node 1 and node 2 at 50 seconds.
- Wormhole tunnel: 1-19-25-23-2 advertised as one hop 1-2
- Two flows, without other traffic in the background

- Flow 18 ⇝ 28:
  - Before 50 seconds: use path 18-9-34-35-37-28
  - After 50 seconds: use path 18-1 ... 2-28
- Flow 17 ⇝ 38:
  - Before 50 seconds: use path 17-14-21-16-38
  - After 50 seconds: use path 17-1...2-38
Stationary Network—Result I

Sequential Change Point Detection
Stationary Network—Result II

- Three flows that changed routes: 9 ⇔ 2, 18 ⇔ 28, 17 ⇔ 38
- There are other flows in the background that stayed on the original routes

**Figure:** Packet size 256B, interval=0.01s, 0.025s

**Figure:** Packet size 256B, interval=0.02s, 0.01s
MAC-Layer Attack Detection in Wireless Networks

IEEE 802.11 MAC

- CSMA/CA
- RTS-CTS-DATA-ACK

(a) Computer Networking, Kurose & Ross

Sequential Change Point Detection
MAC Layer Misbehavior in IEEE 802.11 Networks

Sender Selfish Behavior
- Manipulation on carrier sense time
- Manipulation on back-off value during contention

Consequences
- Channel-capturing effect: other nodes have less chance to transmit

Receiver Selfish Behavior
- RTS dropping attack

Consequences
- Clear channel for itself
- Sender waste resource retransmit RTS
MAC Layer Misbehavior Detection

Other Flows Experience Performance Degradation
- End-to-end delay increases
- Throughput decreases
- Packet interval increases

Detection Method
- Monitor packets received
- Compute per-flow end-to-end delay (or throughput, packet interval) as a time series
- Use the sliding window change point detection method to detect the change on time series
Simulation Setup

- **Case 1: Shorter DIFS attack**
  - normal sender: DIFS > SIFS
  - attacker: switch to DIFS = SIFS starting at 50s

- **Case 2: Shorter DIFS and smaller back-off window $\gamma$**
  - normal sender: following binary exponential back-off, $\gamma \in [32, 1024]$
  - attacker: use fixed $\gamma = 2$

- **Case 3: RTS dropping attack**
  - normal receiver: respond CTS for every RTS request
  - attacker: RTS to CTS ratio 20:1
MAC Layer Misbehavior Detection Results

- Case 1: Five victim flows $19 	o 1$, $14 	o 1$, $12 	o 1$, $10 	o 1$, $6 	o 1$
- Case 2: Same as case 1
- Case 3: $20 \to 2$, $11 \to 2$, $8 \to 2$, $7 \to 2$, $5 \to 2$
- Node 2 is the attacker in all cases
Result I: Case 1

(d) Delay

(e) Throughput

(f) $\mu_D$

(g) $\mu_T$
Result II: Case 2

(h) Delay

(i) Throughput

(j) $\mu_D$

(k) $\mu_T$
Result III: Case 3

(l) Throughput

(m) Packet Interval

(n) $\mu_T$

(o) $\mu_I$
Jamming Attack Detection in Wireless Networks

Attacks

▶ All nodes exposed to open medium
▶ Jamming signals: using higher transmission power, do not have to follow MAC protocol
▶ Legitimate nodes suffer
  - TDMA: collision, increased packet error rate and drop rate
  - CSMA: collision, channel capturing

Detection Procedure

▶ Detect changes from network measurements (delay, throughput, error rate, packet delivery ratio, signal strength, IFS, etc)
▶ Distinguish
  - Jamming vs. weak signals from legitimate nodes
  - Jamming vs. network congestion among legitimate nodes
Jamming Attack Detection in Wireless Networks

Detection Methods

- Use summary information in a time interval, compare against a preset detection threshold
  - Not suitable for highly dynamic networks
- Use change point detection on time series
  - Test statistic: delay, throughput, received packets IFS

E.g.: Throughput when jamming signal duration varies

(p) 0.0005s
(q) 0.8s
(r) 1.5s
Anomaly Detection in Power Grids

Types of Anomalies

- **Line outage**
  - wild animals
  - weather
  - over-grown trees
  - coupled with aging infrastructure + lack of maintenance
- **Generator outage**
- **Transformer fault**
- **Human errors**
- **Cyber attacks**
Error Detection in Power Grids

Methods

- Detection Based on State Estimation (WLS)
  - Works for measurement errors (e.g., bad data detection)
  - Line outage: topology change often causes conforming errors

- Other Detection Methods
  - When the system matrix is known, e.g., (Wu et al., 2016)
  - When the system matrix is unknown:
    - Real-time change point detection + anomaly identification
Real-Time Anomaly Detection in Power Grids

What Feature to Use?

Sequential Change Point Detection
Outline

Introduction

Outlier Detection

Sequential Change Point Detection

High Dimensional Data

Summary and Outlook
Outlier Detection for High-Dimensional Data

- Notion of proximity not straightforward ($\times$)
- High-dimensional sparse data: sparsity makes every data point an outlier

Outlier detection algorithms for high-dimensional data
- Naïve brute force: exhaustive search (Slow!)
- Evolutionary algorithm (Aggarwal and Yu, 2001)
- Projection-based outlier detection (Huber, 1985)
  - E.g., principle component analysis

Outlook
- Detecting outliers among missing data
- Fast method for detecting outliers among a mixture of categorical and continuous variables
Change Point Detection in High-Dimensional Data

A Common Challenge: Scalability
Algorithms that work for univariate or low-dimensional time series may not work for high-dimensional time series.

Methods
- Sum CUSUM statistic from each series (Mei, 2010)
- Sum the local likelihood ratio statistic, then forming a CUSUM statistic (Tartakovsky and Veeravalli, 2008)

Assumptions
- Non-Structural problems
- Post-change distributions are prescribed
- All series are affected by the change
Change Point Detection in High-Dimensional Data

▶ Non-Structural Problem
- No spatial model relating the signal to observations at various locations
- Other work: Chen et al. (Aug. 2010); Petrov et al. (2003); Levy-Leduc and Roueff (2009)

▶ Structural Problem
- Has a spatial structure relating the signal to observations at various locations
- Other work: Rabinowitz (1994); Shafie et al. (2003); Siegmund and Yakir (2008)
Change Point Detection in High-Dimensional Data

Additional Challenge I: with missing data

- Detecting changes from high dimensional time series with missing data (Xie et al., 2013)
  - Use non-parametric submanifold model
  - Extract univariate detection test statistics from high-dimensional data

Additional Challenge II: change affects only a small subset of time series

- $M \ll N$, $M$ is unknown, the subset is unknown
- Unknown and non-homogeneous amplitudes at different series
- Xie and Siegmund (2012) developed a mechanism to suppress noise from unaffected sensors
  - First, compute a generalized likelihood ration (GLR) for each series, use it to suppress noise from non-affected sensors
  - Then, sum the GLRs to compare to a detection threshold
Outline

Introduction

Outlier Detection

Sequential Change Point Detection

High Dimensional Data

Summary and Outlook
Summary and Outlook

Anomaly Detection in Cyber Physical Systems

- Inherently high-dimensional
- Heterogeneous data streams
- Both structural problems and non-structural problems exist
  - Non-structural: some cyber attacks
  - Structural: some nature-induced faults in physical systems
- Real-time requirement
- False positives, false negatives, detection delay
- Causal analysis + anomaly analysis for meaningful results
Thank You!
References I


