Event Correlation & Anomaly Detection

Elizabeth Hou, Yasin Yilmaz, Tony Van, Taposh Banerjee, and Alfred Hero
University of Michigan

Nuclear Facility Monitoring

- Sensors are used to collect data of different forms:
  - electricity consumption
  - satellite imaging
  - radiation emissions
  - internal communication memos
  - seismic vibrations
  - shipment manifests
  - thermal output

Event Detection

- Information from sensors is filtered accordingly to obtain events of interests

System Model

- Normal power usage over a day
- Satellite images monitoring river flow patterns over time

Questions:
1. Can we learn the model for $P(\theta, \phi)$?
2. How many observations are necessary to identify $P(\theta|X)$?
3. What are the sufficient statistics for $\phi$ e.g., $P(S|\phi)$?
4. What are the statistical procedures for $\phi$ e.g., multi-view learning [1], hierarchical HMM [2]
5. Constraints: communication, computational complexity, missing data

References:

Event-Based Transmission of Test Statistics [4], [5]

- Non-independent sensors: Correlation Screening [3]
- Events composed of sequences of other events
- Communication and energy constraints on the sensors: Decentralized setup

References:

Quickest Change Detection [6], [7]

- Objective:
  - Find stopping time $\tau$ on $\{X_n\}$
  - Minimize delay {$\tau \rightarrow \gamma$}
  - Constraint on false alarm {$\tau < \gamma$}

References:

Human Aided Anomaly Detection

- Simulation: 5% is anomalous, 23% of anomalous points are considered important by a domain expert
- Want a model to automatically incorporate domain expert knowledge
- Learn utility function over space of anomalous points (constrained classification)
- Estimate modified minimum volume (MV) sets – high/low penalty for points inside/outside the MV set (constraint)
- Practical setting: majority of utility scores missing, domain expert only labels a few points

References: