Event Correlation & Anomaly Detection

Elizabeth Hou, Yasin Yılmaz, Tony Van, Taposh Banerjee, and Alfred Hero

University of Michigan

Nuclear Facility Monitoring

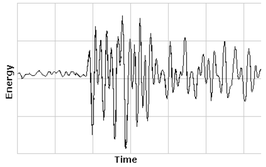
Sensors are used to collect data of different forms:

- electricity consumption
- satellite imaging
- radiation emissions
- internal communication memos
- seismic vibrations
- shipment manifests
- thermal output

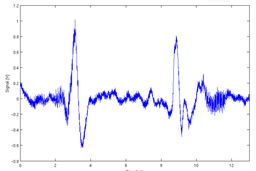
Watts Bar Nuclear Plant, Tennessee http://www.tva.gov/power/nuclear/wattsbar.htm

Event Detection

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



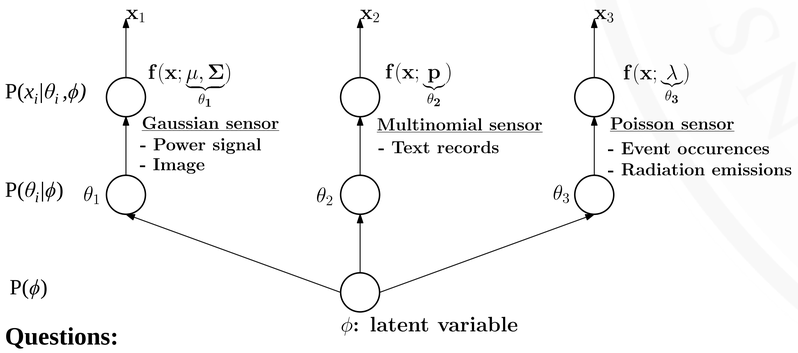
Normal power usage over a day



Unusual power surges are detected by filtering incoming signals against a baseline signal (e.g. matched filtering)

Satellite images monitoring river flow patterns over time

System Model



Questions:

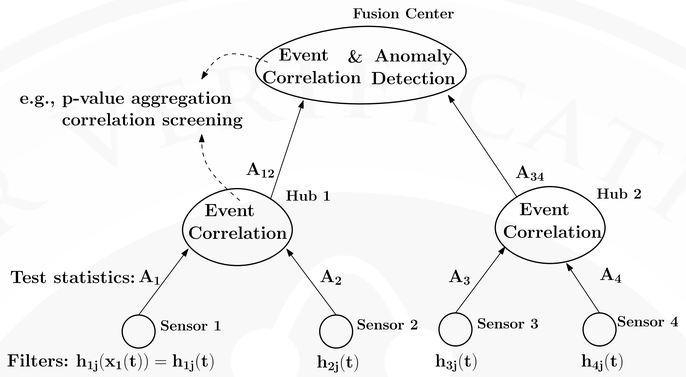
1. Can we learn the model for $P(\theta_1,\theta_2,\theta_3|\phi)$?
2. How many observations are necessary to identify $P(\phi|X)$
3. What are the sufficient statistics for ϕ? e.g., $P(S, V|\phi) = P(S|\phi)\ P(V|S)$
4. What are the statistical procedures for ϕ ? e.g., multi-view learning [1], hierarchical HMM [2]
5. Constraints: communication, computational complexity, missing data

References:

[1] C. Christoudias, R. Urtasun, T. Darrell, "Multi-View Learning in the Presence of View Disagreement", UAI 2008

[2] S. Fine, Y. Singer, N. Tishby, "The Hierarchical Hidden Markov Model: Analysis and Applications", Machine Learning 1998

Analysis Model



Due to multi-modality, a hierarchical approach is proposed:

1) Each sensor computes p-values of some events

$$A_i = \begin{bmatrix} a_{i,1}(1) & \cdots & a_{i,1}(T) \\ \vdots & \ddots & \vdots \\ a_{i,J}(1) & \cdots & a_{i,J}(T) \end{bmatrix}$$

$$a_{ij}(t) = P(h_{ij} \geq h_{ij}(t))$$

2) Event correlation at hubs, e.g., aggregate p-values

$$A_i A_k \rightarrow \text{Conjunctive}$$

$$1-(1-A_i)(1-A_k) \rightarrow \text{Disjunctive}$$

3) At fusion center, event correlation & anomaly or change point detection

Extensions:

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

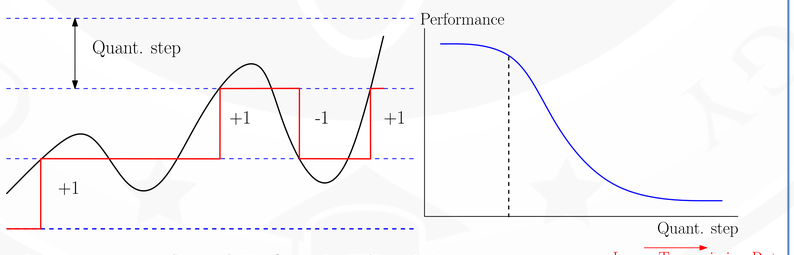
References:

[3] A. Hero, B. Rajaratnam, "Large- Scale Correlation Screening", AMSTAT 2011

Decentralized Data Collection: Low-Rate Communications

- Communication constraints
- Energy constraints on sensors
- Security concerns
- Big data

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



In contrast to traditional Uniform Sampling & Quantization

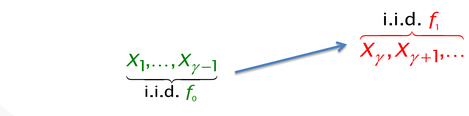
- Skip the uninformative samples
- Transmit only the change information

References:

[4] G. Fellouris and G. Moustakides, "Decentralized sequential hypothesis testing using asynchronous communication," IEEE Transactions on Information Theory, vol. 57, pp. 534-548, 2011

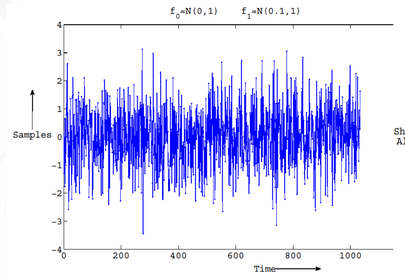
[5] Y. Yilmaz and X. Wang, "Sequential decentralized parameter estimation under randomly observed Fisher information" IEEE Transactions on Information Theory, vol. 60, pp. 1281-1300, Feb. 2014

Quickest Change Detection [6], [7]

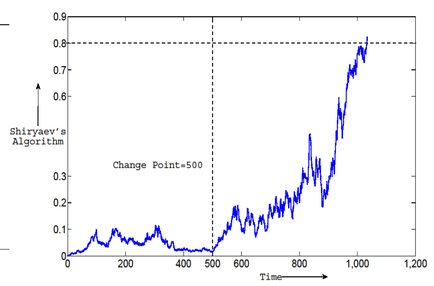


Objective:

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



Cannot detect change through visual inspection



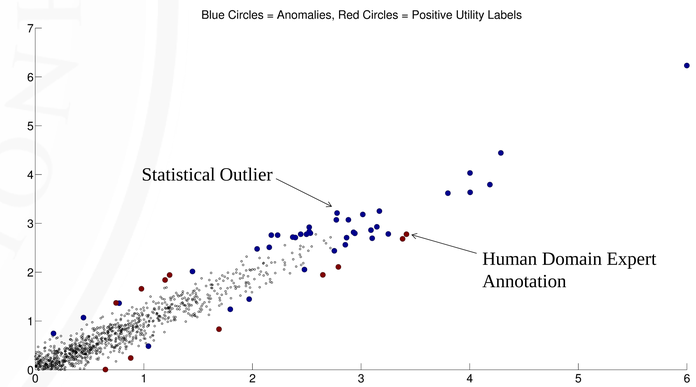
Shiryaev's Algorithm

References:

[6] M. Basseville and I. Nikiforov, "Detection of Abrupt Changes – Theory and Applications" 1998

[7] V. V. Veeravalli and T. Banerjee, "Quickest Change Detection," Elsevier: E-reference Signal Processing, 2013, http://arxiv.org/abs/1210.5552.

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:

[8] T. Jaakkola, M. Meila, T. Jebara, "Maximum entropy discrimination", 1999.

[9] K. Sricharan, A. Hero, "Efficient anomaly detection using bipartite k-nn graphs", NIPS 2011.

[10] T. Xie, N. Nasrabadi, A. Hero, "Learning to classify with possible sensor failures", ICASSP 2014

National Nuclear Security Administration