**Tutorial on event detection**

By Neill and Weng-Keen Wong, 2009

**0. INTRO**

**Applications**

* Outbreak of a disease
  + Detect spatial clusters of syndroms
* Crime hot-spot detection
* Pipe breaks
* Environment
* Container shipment

**Evaluation**

* ROC curve → accuracy
* AMOC curve : Activity Monitoring Operating Characteristic. Detection time / False positive rate → timeliness
  + Fawcett and Provost

**Event detection branches**

* Temporal
* Spatio-temporal

**1. TEMPORAL METHODS**

**1.1. Univariate temporal methods**

**Temporal univariate methods : framework**

1. Learn model to predict expected signal value
2. Measure difference between actual and expected value
3. Compute alarm value and compare to threshold

**Temporal univariate methods**

* Control Chart, 1931
* Moving Average
* Exponentially Weighted Moving Average - EWMA
* CUSUM
* Regression

**Other state-of-the-art univariate methods not presented in tutorial**

* Box-Jenkins models → ARMA, ARIMA
* Wavelets
* Change-point detection
* Kalman filters
* Hidden Markov Models

**Control Chart**

* Many false alerts
* Very basic : alarm activated when value above empirical mean + 3 \* empirical variance
* Or :

**Moving average**

* Takes seasonality into account
* Consider time window of W, and same methode as before
* Problem if recent history is anomalous

**EWMA Zi**



**Cumulative Sum Statistics**

* Detects shifts from the mean more quickly than control chart
* Cumulative sum of gap to mean
* Alarm reset whenever cumsum exceeds threshold

**Regression – linear**

* Can model trends : seasons, week, holiday...
  1. **Multivariate temporal methods**

**Methods**

* Multivariate Changepoint Detection

→ Detect a change has happened & does not identify the anomalous subgroup of data

* + Multivariate Statistical Quality Control
  + Other not presented
    - Cross-match test, Rosenbaum 2005
    - kdq-tree, Dasu et al. 2006
    - Density Test, Song et al. 2007
* Multivariate Event Detection

→ Detect a change has happened & identifies the anomalous subgroup of data

* + Emerging Patterns
  + STUCCO
  + WSARE 2.0
  + WSARE 3.0



**Multivariate Event Detection framework**

1. Learn model to predict expected signal value for the given subgroupe
2. Measure difference between actual and expected
3. Compute alarm value

**Idea**: find the rules with great count gap between train and test

**Emerging Patterns**



**STUCCO – Search and Testing for Understandable Consistent Contrasts**

* Define rules, « contrast set » // APD
* All rules are considered in a breadth-first tree search → *too high computation ?*
* Chi-Square test of independence of the variables Attribute (True or False) and Dataset (Train or Test)
* Compare X² statistic to chi-square distribution → compute p-value
* Problem : multiple hypothesis testing
* Bonferroni Correction for multiple hypothesis testing h1,..., hn → Randomization is better alternative



**WSARE 3.0**

* Uses Bayesian model, computed with Optimal Reinsertion, Moore & Wong 2003