

Event detection scope

Event

- Find all events or just a single one? → **ALL**
- What kind of events can be detected? → **ALL**
 - Regular / anomalous?
- Spatial & temporal extension predefined? → **ALL**
 - Recurrent event vs outbreak
- Specific or general event detection?

Assumption on the form of the event? → ***None, but a parametric method may be used in the general case?***

- Parametric vs non-parametric

Data

- What kind of dataset? → ***Spatial time series = Neill classic framework***
- Nature: counts vs categorical → ***Counts***
- Univariate vs Multivariate → ***Univariate***
- Scale
 - Space: 4000 locations
 - Time
 - Resolution: hourly/daily
 - Period: couple of years

Priority Neill techniques

Scan statistics / Fast subset scan with parametric modeling, 2012

- Kulldorff-like, faster
- Parametric modeling
 - Expectation-Based Poisson model
 - Expectation-Based Gaussian model
 - Parameters calculated by inference
- (Variants: different expectation forecasting time series techniques
 - *An empirical comparison of spatial scan statistics for outbreak detection, 2009*)
- Article: *Fast subset scan for spatial pattern detection, 2012*

Scan statistics / Non-parametric scan statistics, 2007

- Only assumption of the model: iid counts from stationary time series
- May benefit from fast subset scan 2012 progress
- Article: *A Nonparametric Scan Statistic for Multivariate Disease Surveillance, 2007*

Bayesian statistics / Bayesian parametric scan statistic – univariate, 2006

- Hierarchical Gamma-Poisson model – inference
- Better with prior knowledge of event, but can be used without it
- Looks for
 - a given event
 - a set of multiplicative increase event: {counts x 0.5, counts x 1.5, counts x 2,...}
- Article: *A Bayesian Scan Statistic for Spatial Cluster Detection, 2006*

Rather old, new papers focus on multivariate data

Extra - Scan statistics / Fast subset scan with connectivity constraint – GraphScan

- Extends fast subset scan space-time framework to graphs → road network
- Article: *Scalable detection of anomalous patterns with connectivity constraints, 2015*

Important questions

Should a parametric approach be used even if we have no information on the probability distribution of counts?

- What parametric model would best fit our taxi data? Poisson? Gaussian?
- Should instead a non-parametric approach be used?
- We can compare them

What are the non-parametric options?

What is the state-of-the-art in univariate event detection techniques?

- Papers compare univariate vs multivariate, but not much univariate techniques between each other

What is the code available?

- Almost all techniques are tested on real data in papers

What techniques did Neill use for urban data analysis?

- CityScan: monitors 311 data in Chicago, maybe other
- CrimeScan: monitors 911 data in Chicago

Prioritarian Neill techniques - backup

Univariate BSS → GO FOR IT, 2006

- Neill 2006
- Similar extension of Fast Subset sums? <http://www.cs.cmu.edu/~./neill/papers/StatMed2011.pdf>

Univariate FSS + (EBP / Gaussian) modeling, 2012

- k-NN/ r-circles
- model: poisson learnt on historical data → relevant model for taxi data?
- output: most at risk region → can it be adapted to get all significant clusters?

NPSS univariate before 2007 (with LTSS property)? 2009 + FGSS 2013

- based on p-values of the day(s) in the time window period: $(T_{\text{beat}} + 1) / (T + 1)$ where T is total number of days in window and T_{beat} number of days in window with higher counts // OR total past days??
- what randomization needed?
- hypothesis: independence and stationarity (each location has stationary time series)
- Then, for a region = locations x time, check whether a significant number of p-values are lower than expected
- statistical significance by randomization: generated $U[0,1]$ p-values
- univariate available

GraphScan (Maybe)

- LTSS on Graphs?
- Check

Ask for new features of old techniques / Check with Neill whether it is state of the art in univariate ED

- NPSS may have been combined with LTSS
- New features for Univariate BSS?

Prioritarian Neill techniques - backup 2

Neill DB, An empirical comparison of spatial scan statistics for outbreak detection, 2009 - quite old, before LTSS

<http://repository.cmu.edu/cgi/viewcontent.cgi?article=1373&context=heinzworke>

Comparison of ED techniques

- EBP
- EBG
- PBP/PBG (need population file → out)
- Robust Bernoulli-Poisson ??

Comparison of time series forecasting method

- x5 techniques

Possible Niell techniques backup 3 - Out-of-scope work

MBSS → Univariate BSS available !

- Relevant in univariate? → check article, maybe univariate tool used for comparison
- can display a heatmap of risk + most at risk region
- specify form of the events: multiplicative increase – relevant?
 - can monitor $\times 0.33$, $\times 0.5$, $\times 1.5$, $\times 2$, $\times 3$... events / Will not use FIT model which adapts to a single event
- test on 16×16 grid (but 3 variables)
- Powerful with prior: no prior for taxis but uniform prior works in test
- Gamma-Poisson expected probability distribution with inference *adapted to disease surveillance*

GFSS

- Similar to MBSS, useful for univariate?

FGSS