Monitoring the blogosphere for emerging, health related events, so Health Officials don‘t have to
Event-Based Biosurveillance

Monitor time series, textual data to provide early alerts to anomalies ... stimulate investigation of potential outbreaks

Shmueli 2010
What If ....?

Could we have detected the emergence of the 2009 Swine Flu Pandemic from Blog Social Media?
Questions to Consider?

• What approach can be used to create Simulated Outbreaks for blog text?
  – What outbreak patterns/signatures exist?
  – How can text be generated to simulate a given outbreak pattern?

• What tools would be useful in creating Simulated Textual and Numerical Outbreak Data?
Feature Selection and Counts for Outbreak Data

Event-Based (Text) Data
- Select (textual features)
  i.e.: Number documents containing mentions „flu“
- Get timestamps
- Create counts from features

Time Series Data
- Select features:
  i.e.: hospital visits, death rate
- Get timestamps
- Create counts from features
What is the Task?

Using existing data, tools and references:

• Part 1: Design an approach to creating Simulated Data from example data
• Part 2: Design an approach for adding Noise to the Simulated Data
• Part 3: List and summarize any additional tools and approaches that would be useful in creating Simulated Data

• Document your design:
  • Discuss the motivation for your work/results
  • Outline algorithms with PseudoCode
  • Provide several example input and outputs
  • Highlight strengths and shortcomings
What Will You Learn?

• Build Simulated Data from Event-Based Biosurveillance
• How to organize, design, implement, deliver small-scale project results
• That your contributions are valuable ....
Starting Points

Papers:
The Nature of Outbreaks and Their Determination (Shmueli, Section 3.2)
Characteristic shapes of outbreak news. (Collier, Figure 5)

Model-Specific Generation:
Data Simulation Using R
http://biostat.mc.vanderbilt.edu/wiki/pub/Main/AngelAn/myslides5.pdf

Random Generation:
“Generating Random Text with Bigrams“

“How to Generate Random Text in Word 2003”
http://www.ehow.com/how_2183058_generate-random-text-word.html
Background Reading

1. Statistical Challenges Facing Early Outbreak Detection in Biosurveillance, Galit Shmueli

2. What’s Unusual in Online Disease Outbreak News, Nigel Collier

3. Detecting Influenza Outbreaks by Analyzing Twitter Messages, Aron Culotta