Homeland Security: What Can Discrete Science Do?

Fred Roberts
Professor of Mathematics, Rutgers University
Director, DHS Center of Excellence for Discrete Dynamical Systems (DyDAn)

What is Discrete Science?

- Deals with data and information
- Massive, heterogeneous, dynamic
- Seeks patterns
- Seeks departure from patterns
- Is built on a foundation of the mathematical sciences:
  - Math, Computer Science, Statistics
- Develops powerful computer algorithms

Methods of discrete science have become important tools for homeland security, especially when combined with powerful, modern computer methods for analysis and simulation.

Are you Serious?? What Can Mathematics Do For Us?

After Pearl Harbor: Mathematical methods and mathematical scientists played a vitally important role in the US World War II effort.

Critical War-Effort Contributions Included:

•Code breaking.

Enigma machine

•Creation of the mathematical sciences-based field of Operations Research:

logistics

optimal scheduling

inventory

strategic planning

But: Homeland Security is Different.
Can Mathematics Really Help?

5 + 2 = ?

1, 2, 3, …

Example 1: Intelligence Analysis

Finding Patterns in Text

Objective:

Monitor huge amount of text (news stories, blogs, etc.) to automatically detect pattern changes and "significant" events.

Get early warning of terrorist plots.

The Daily

Sunday, August 30, 2006

Martians invade earth

Incredible as it may seem, it has been confirmed that a large martian invasion fleet has landed on earth tonight.

First vessels were sighted over Great Britain, Denmark and Norway already in the late evening from where, as further reports indicate, the fleet headed towards the North Pole and Santa Claus was taken hostage by the invaders.

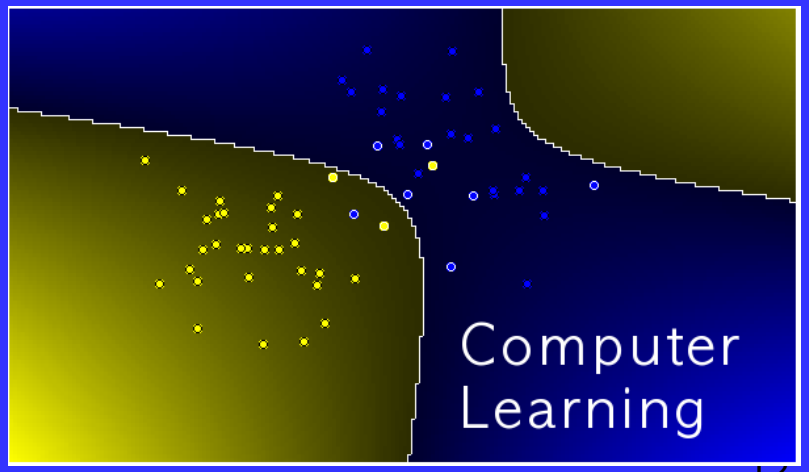
Afterwards they split apart in order to approach most major cities around the earth. The streets filled as thousands fled their homes, many only wearing their pajamas...

Technical Approaches:

- Given stream of text in any language.
- Decide whether new "events" are present in the flow of messages.
- ***New Event***: new topic or topic with unusual level of activity.
- Suppose events have been classified into classes or groups: group 1, group 2, …
- A new message comes in. Does it fit into group 1? Into group 2? Or does it (and related messages) define a new group of interest?

The Approach: Machine Learning

- Simplest approach: use word counts.
- Performance can improve over time.
- “Learn” how to classify better.
- Develop discrete science methods that will allow a machine to “learn” from past data.



Example 2: Disease Event Detection

Great concern about the deliberate introduction of diseases such as anthrax by bioterrorists has led to new challenges for discrete science.

anthrax

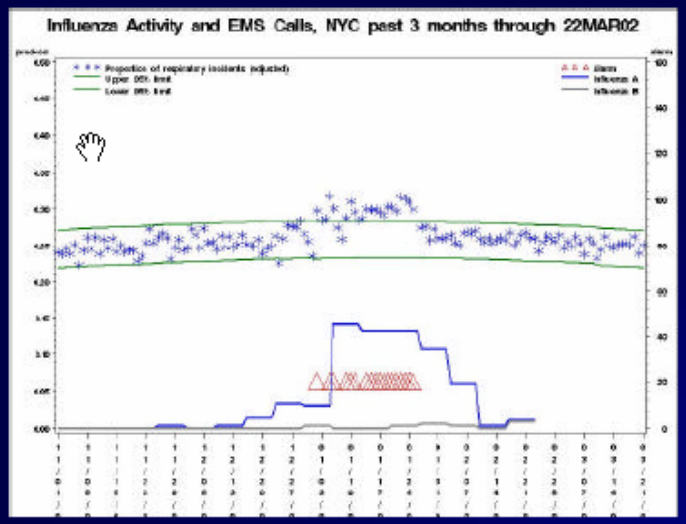
Disease Event Detection

- Use large data sets to detect "bioterrorist events" or "emerging diseases" (SARS, pandemic flu)
- Use machine learning methods
- **"syndromic surveillance"**

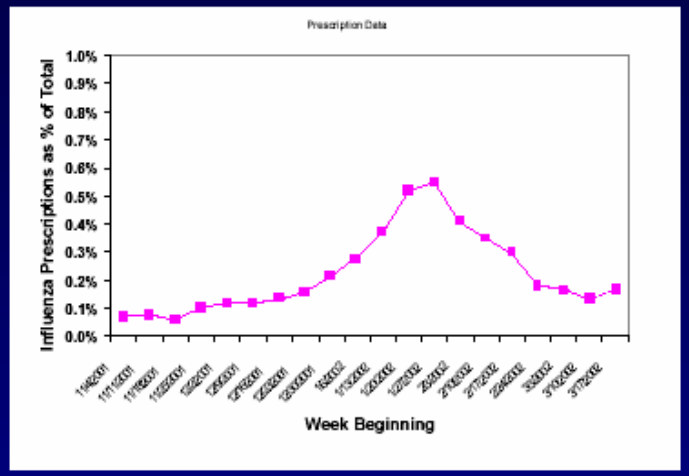
smallpox

Syndromic Surveillance: NYC Dept. of Health Data

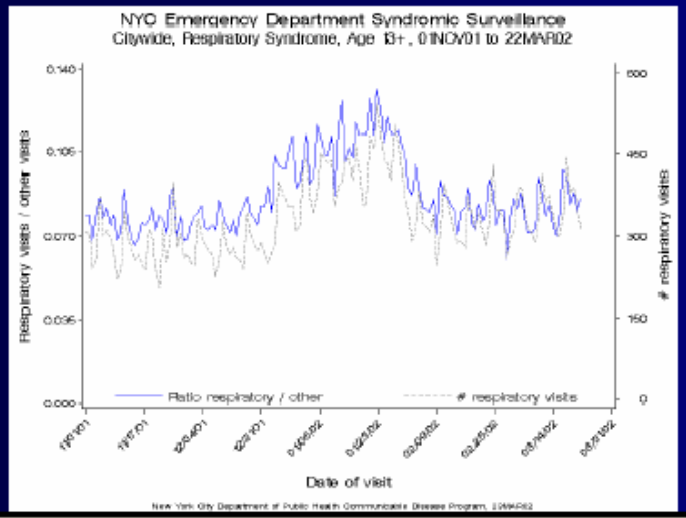
EMS calls



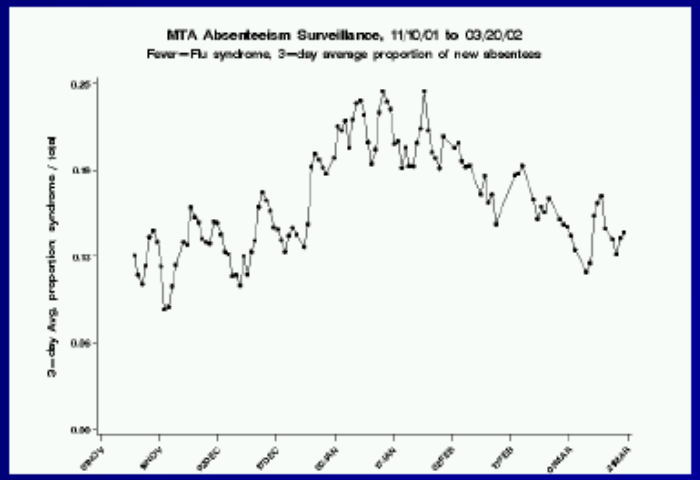
Pharmacy Antiviral Rx



ED respiratory visits



Subway worker- "flu"



Many New Discrete Science Methods and Approaches under Development

- Spatial-temporal “scan statistics”
- Statistical process control (SPC)
- Bayesian applications
- “Market-basket” association analysis
- Text mining
- Rule-based surveillance
- Change-point techniques

Example 3: Author Identification

Develop techniques for identifying authors in large collections of textual artifacts (e-mails, tapes, transcribed speech, etc.).

Questions Addressed:

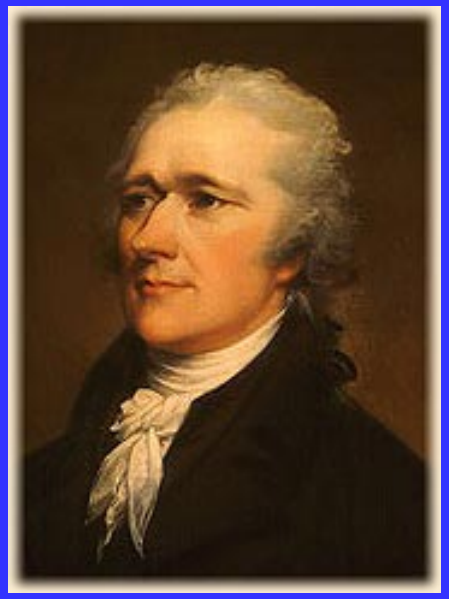
- Which of a set of authors wrote a document/message?
- Were two documents written by the same author?

Author Identification

Did Osama Bin Laden write this?

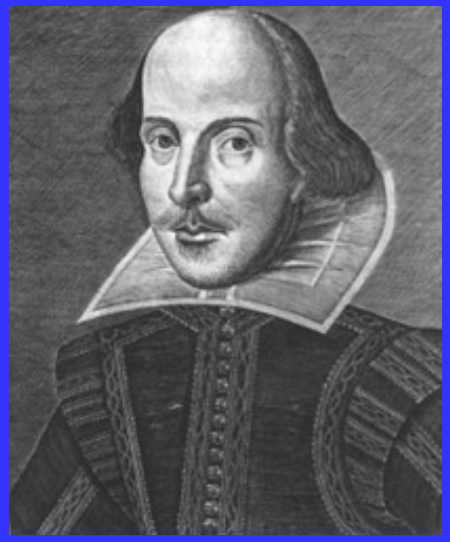
Author Identification

- Discrete science uses methods of machine learning related to the text analysis example.
- Building on classical work in Statistics:
  - Who wrote the Federalist papers, Hamilton or Madison? (85 papers, 12 disputed)
  - Who wrote Shakespeare's plays?



Alexander Hamilton

James Madison



Shakespeare

Author Identification

Challenges for discrete science: Author I.D. is more complicated than conventional text classification:

- Large number of possible authors
- Not much “training data”
- Authors write on multiple topics
- Authors write in different styles in different “genres”

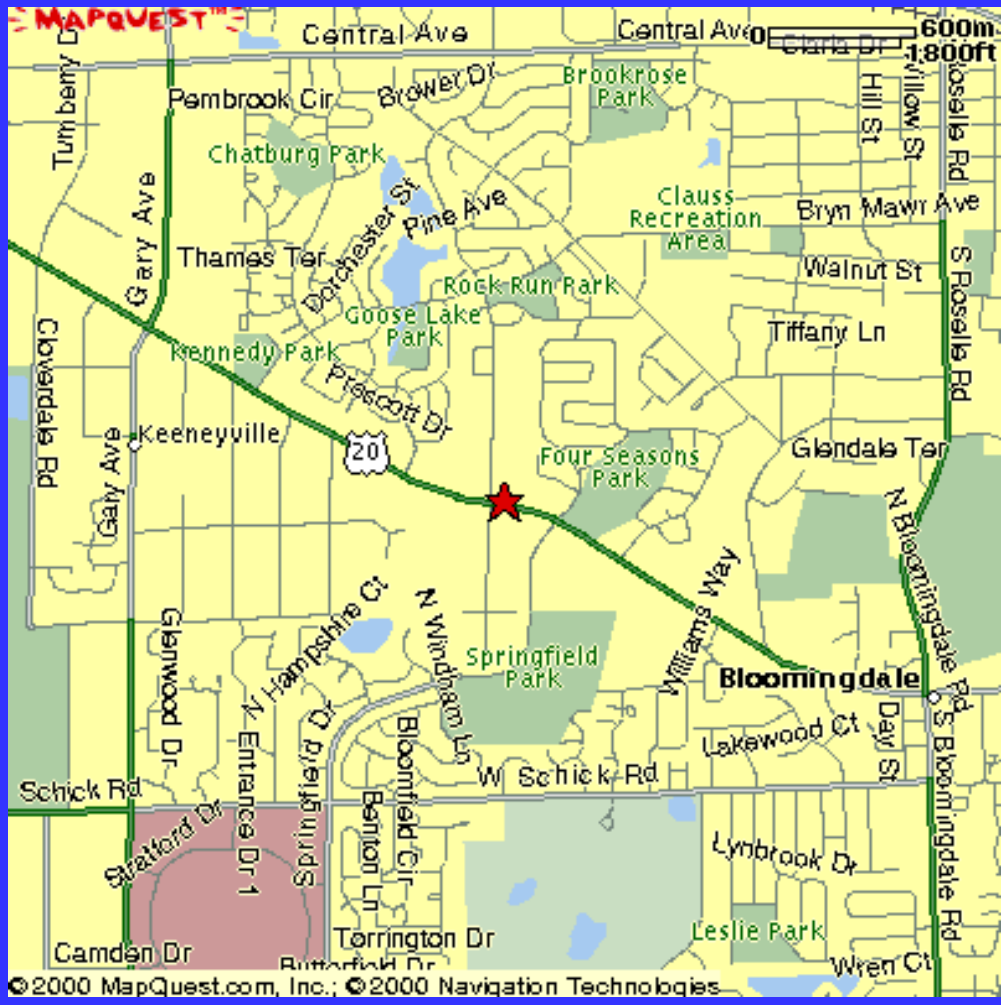
Example 4: Bioterrorism Sensor Location

- Early warning is critical in defense against terrorism
- This is a crucial factor underlying plans to place networks of sensors/detectors to warn of a bioterrorist attack

The BASIS System – Salt Lake City

Two Fundamental Problems

- ***Sensor Location Problem***
  - Choose an appropriate mix of sensors
  - decide where to locate them for best protection and early warning



Two Fundamental Problems

- ***Pattern Interpretation Problem:*** When sensors set off an alarm: Has an attack taken place?
  - What additional monitoring is needed?
  - What was its extent and location?
  - What is an appropriate response?

The Sensor Location Problem

- Challenge: develop **new algorithmic methods**.

- Greedy algorithms: e.g., steepest ascent
- "Location theory" methods from operations research: e.g., locating fire stations
- "Equipment placing" algorithms developed in telecom.

Approaches to the Pattern Interpretation Problem

- Using ***redundancy*** to minimize false alarms

- Using decision rules borrowed from missile defense.

Example 5: Port of Entry Inspection for WMDs

Port of Entry Inspection Algorithms

- ***Goal***: Find ways to efficiently intercept illicit nuclear materials and weapons destined for the U.S. via the maritime transportation system

Port of Entry Inspection Algorithms

•Aim: Find inspection schemes that minimize total "cost" including "cost" of false positives and false negatives and delays to commerce.

Mobile Vacis: truck-mounted gamma ray imaging system

Port of Entry Inspection Algorithms

- **Sequential decision making problem**
- Stream of containers arrives at a port
- Which inspections next based on previous results?
- ***Complications:***
  - If > 4 types of sensors, problem rapidly becomes computationally intractable

Example 6: Protection Against Invasive Species

- First line of defense: US Coast Guard.

hydrilla

Protection Against Invasive Species

- A ship arrives with a load of bananas from South America

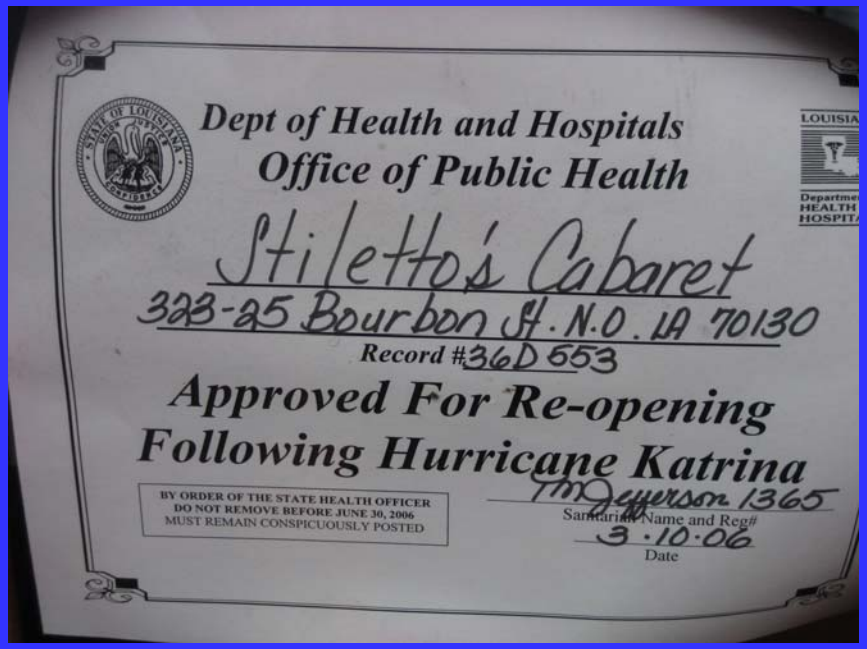
Protection Against Invasive Species

- A spider is discovered in the hold of the ship.
- What kind of spider is it?
- Is it safe to unload the cargo of bananas?

Protection Against Invasive Species

- Compare a database of images
- Use an efficient search feature.
- Is the spider dangerous?

Example 7: Response to Natural Disasters



Response to Natural Disasters

Discrete science methods include:

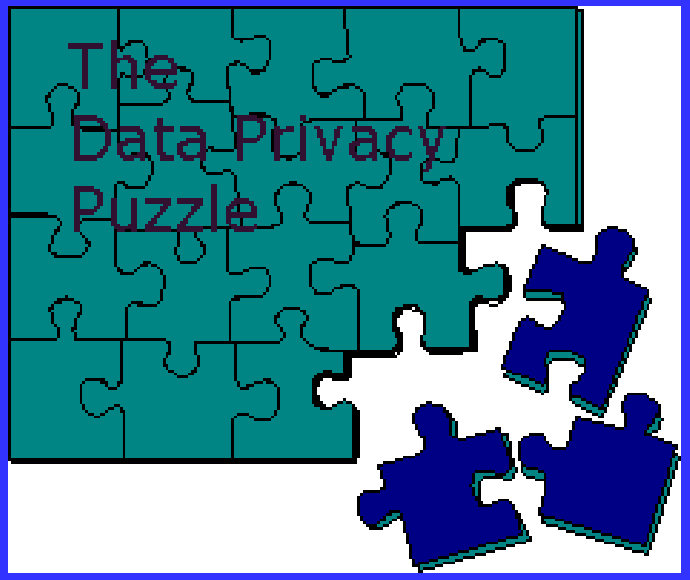
- Searching through a database of resources.
- Finding optimal strategies of distribution

Response to Natural Disasters

- Also: Situational awareness as an aid to emergency management

Example 8: Privacy-Preserving Data Analysis

Sample problem: How to share data between different agencies while protecting privacy?



Example: Privacy-Preserving Data Analysis

•Sample Challenge: How can we make a simultaneous query to two datasets without compromising information in those data sets? (E.g., is individual xx included in both sets?)

“Secure Multiparty Computation”

Discrete science provides an informatics tool for a wide range of homeland security problems