Anomaly Detection

National Defense and Homeland Security

SAMSI

Francisco Vera, postdoctoral fellow

What is an Anomaly?

- Definition from dictionary.com
  - Deviation or departure from the normal or common order, form, or rule.
  - One that is peculiar, irregular, abnormal, or difficult to classify.
- Statistical anomaly detection tries to spot out or identify anomalies based on data

Why are Anomalies Important?

- Anomalies may be due to randomness
- They could be a consequence of some action taken (or not taken)
- Some actions that can cause anomalies can come from malicious individuals, like terrorists
- The sooner an anomaly is detected, the quicker we can respond to it

Email example

- 1000 emails go out in one hour come from a computer that usually sends 2 emails a day.
- Possible reasons:
  - You are sending those emails to 1000 people you know
  - Virus has infected computer
  - Hacker is sending emails in your name

Stock Example

- The stock trade of a company increases to 1000 stocks per day, when it has never been above 20 stocks per day.
- Possible reasons:
  - Company launches new product, like when Microsoft launched Windows for the first time
  - Terrorist knew they were going to attack on September 11 and decided to sale all their American Airlines stock

Flu Medicine Example

- The amount of flu medicine sold in a week is three times as much what was sold the week before
- Possible reasons:
  - Flu season started
  - Bioterrorist attack with small pox, that has symptoms similar to the flu in its early stage
- Seasonality has to be taken into consideration
- Early detection of this anomaly can save millions of lives

Building Example

- A building full of people explodes
- Possible reasons:
  - Electrical mal function and poor design of building
  - A bomb implanted by terrorists goes off
  - An airplane hits building (9/11)
- Anomaly should be “detected” before it happens
  - Anomaly prediction

Malfunction Example

- An airplane engine stops running
- Possible reasons:
  - One component broke
  - A bird passed through
  - Terrorist attack
- Prevention of anomalies
  - Breaking components could be replaced before
  - Area of statistics dealing with this kind of problems: Reliability

Traditional Methods

- Intuition: 1000 seems anomalous if compared with 20
- Plotting: plot stock trades and see if they seem anomalous
- Expert opinion: get the president of the company and ask him if he thinks that the volume of stocks trade yesterday in his company is anomalous

Statistical Approach

- Today, computers have made it possible to have billions of records on huge databases
- It is impractical to see 1 million plots or ask an expert to see 2 million records
- Presence of randomness in data requires a statistical approach

Statistical AD Methods

- Series of numbers
  - Check all the numbers and see check if one is too large or too small
- Time problems
  - Check time of events and see if there is a period with lots of events
  - Events could be burglaries in Durham, improvised explosive devices in Iraq, etc

Statistical AD Methods

- Space problems
  - Similar to time problems
  - Locate hot-spots
  - Study event location features
    - Distance to closest US ambassy
    - Education of the population in the sorrounding area
    - Etc
    - Learn about the preference of attackers in selecting a location

Statistical AD Methods

- Social Networks
  - Look at communications between people
  - An anomalous communication pattern may be a signal of something bad about to happen
    - Company about to get indigment
    - Terrorist attack about to happen

Statistical AD Methods

- Image anomalies
  - Look for anomalies in pictures to improve them (remove red eye, remove wrinkles, etc)
  - Look at satellite images for anomalous activity
    - Terrorist attacks could be spotted out
    - Privacy could be violated

Some of My Work

Switch to Acrobat Reader