

Computers and Society

Privacy, Data Mining, and Social Networks

Notice: This set of slides is based on the notes by Professor Guattary of Bucknell and by the textbook author Michael Quinn

Data Mining

- Searching information in a collection of data, looking for patterns or relationships
- Can be used to profiles of individuals
- Allows companies to build more personal relationships with customers

A Few Stories

The Target Store Example

- From the book *Power of Habit: Why We Do What We Do in Life and Business* by *New York Times* reporter Charles Duhigg (February 2012, publisher: Random House Publishing Group)
 - The Target department store keeps track of what customers buy and when (consumer habits), its data mining group found out a teenage was pregnant before her parents did

...
http://www.nytimes.com/2012/02/19/magazine/shopping-habits.html?pagewanted=all&_r=0

The Twitter User Study

- Research done by a group of computer scientists at University of Rochester that was published in February 2012 show that with accuracy, the software can **infer** and **predict**
 - Your friends circle
 - Your exact location
 - as long as some of your friends are active on a social network, even you are not, or you hide your location

The Twitter Study (cont.)

- Researchers used users social network ties (e.g., users sent Twitter messages, or tweets, in the past) and geo-location of the devices (e.g., cell phone used for Twitter)
- Friends prediction (% of friends are given, accuracy): (0%, 78%), (10%, 82%), (25%, 88%), (50%, 95%)
- Location prediction (# of friends are given, accuracy): (1, 60%), (2, 77%), (9, 84%)

The Twitter Study Data

- Raw data statistics
 - The city of New York City & Los Angeles
 - Unique users 1,229,611
 - Unique geo-active users 11,380
 - Tweets total 26,118,084
 - GPS-tagged tweets 7,566,569
 - GPS-tagged tweets by geo-active users 4,016,286
 - Unique locations 89,077
 - Significant locations 25,830
 - "Follows" relationships 123,182 between geo-active users
 - "Friends" relationships 52,307 between geo-active users

The Twitter Study Demo

- Author's research website:
 - <http://cs.rochester.edu/u/sadilek/research/>
- YouTube demo of this project
 - http://www.youtube.com/watch?v=dMqs0voVu xc&feature=player_embedded#!

Credit Reports

- Example of how information about customers can itself become a commodity
- Credit bureaus
 - Keep track of an individual's assets, debts, and history of paying bills and repaying loans
 - Sell credit reports to banks, credit card companies, and other potential lenders
- System gives you more choices in where to borrow money
- Poor credit can hurt employment prospects

Microtargeting

- Political campaigns determine voters most likely to support particular candidates
 - Voter registration
 - Voting frequency
 - Consumer data
 - GIS data
- Target direct mailings, emails, text messages, home visits to most likely supporters

Facebook Beacon

- Fandango, eBay, and 42 other online businesses paid Facebook to do "word of mouth" advertising
- Facebook users surprised to learn information about their purchases was shared with friends
- Beacon was based on an opt-out policy
- Beacon strongly criticized by various groups
- Facebook switched to an opt-in policy regarding Beacon

Netflix Prize

- Netflix offered \$1 million prize to any group that could come up with a significantly better algorithm for predicting user ratings
- Released more than 100 million movie ratings from a half million customers
 - Stripped ratings of private information
- Researchers demonstrated that ratings not truly anonymous if a little more information from individuals was available
- U.S. Federal Trade Commission complaint and lawsuit
- Netflix canceled sequel to Netflix Prize

<http://www.wired.com/threatlevel/2009/12/netflix-privacy-lawsuit/>

Social Network Analysis

- Data mining now incorporating information collected from social networks
- Examples
 - Cell phone companies in India identify “influencers”
 - Police predict locations of big parties
 - Banks evaluate the riskiness of loans

Data Mining by the Government

IRS Audits

- IRS uses computer matching and data mining to look for possible income tax fraud
- Computer matching: matching tax form information with information provided by employers, banks, etc.
- Data mining: searching through forms to detect those that appear most likely to have errors resulting in underpayment of taxes

Telecommunications Records Database

- Created by National Security Agency after 9/11
- Contains phone call records of tens of millions of Americans
- NSA analyzing calling patterns to detect terrorist networks
- Phone records voluntarily provided by several major telecommunications companies
- *USA Today* revealed existence of database in May 2006
- Several dozen class-action lawsuits filed
- August 2006: Federal judge in Detroit ruled program illegal and unconstitutional
- July 2007: U.S. Court of Appeals overturned ruling, saying plaintiffs did not have standing to bring suit forward

Edward Snowden Incident

- In June 2013, Edward Snowden, a former NSA employee, released to various media (first to *The Guardian*) some of the details of various NSA surveillance programs
 - NSA collects a large number of phone communications records
 - NSA collects a large number of emails and other social media information (The PRISM program)

<http://www.theguardian.com/world/2013/jun/23/edward-snowden-nsa-files-timeline>

Family Education Rights and Privacy Act (FERPA)

- Rights given to
 - Students 18 years and older
 - Parents of younger students
- Rights include
 - Reviewing educational records
 - Requesting changes to erroneous records
 - Preventing release of records without permission

Video Privacy Protection Act

- Videotape service providers cannot disclose rental records without consumer's written consent
- Rental stores must destroy personal information related to rentals within a year of when it is no longer needed

Health Insurance Portability and Accountability Act

- Limits how doctors, hospitals, pharmacies, and insurance companies can use medical information
- Health care providers need signed authorization to release information
- Health care providers must provide patients with notice describing how they use medical information

Freedom of Information Act

- Federal law designed to ensure public has access to U.S. government records
- Signed by President Johnson (1966)
- Applies only to executive branch
- Nine exemptions
 - Classified documents
 - Trade secrets or financial information
 - Documents related to law enforcement investigations

Toll Booth Records

- E-ZPass: an automatic toll-collection system used on most toll roads, bridges, and tunnels between Illinois and Maine
- Drivers with E-ZPass tags pass through without stopping to pay attendant
- Records have been provided in response to court orders in criminal and civil cases

Advanced Imaging Technology Scanners

- Transportation Security Administration began installing AIT scanners in 2007
- AIT scanners reveal anatomical features
- Electronic Privacy Information Center sued government in 2010, saying systems violate 4th Amendment and various laws
- February 2011: TSA announced it was developing new software that would replace detailed image with generic outline of a person