Computers and Data Mining

By Sharon Tan
History

- Origins - late 80s
- Extract hidden information from data
- Originally directed towards tabular data
- Started with loyalty cards
Current Focus

- Effective mechanisms to mine larger data collections
- Application of data mining techniques to non-standard datasets
The Popular Mines

- Microsoft - identified women at risk of postpartum depression
- Facebook - analyzed posts and observed communication patterns
Problems for Society

- Privacy
  - Unknowing subject of research
- Strong extrapolations on weak data
  - Profiling
- Data Brokers
  - Selling information for profit
Proposed Solutions

- Boycott social media?
- Erase cyber existence?
- Use cash only?
- Enforce stronger privacy laws?
Proposed Solutions (Cont’d)

- Personal:
  - Recognize the risks
  - Be mindful of online activities
  - Read privacy agreements carefully
Proposed Solutions (Cont’d)

- Industry:
  - Develop privacy-preserving data mining algorithms
  - Enforce stronger ethics on data collection
Future Years

- Anonymity increases with volume of data collected
- Nothing is private anymore
References

References (Cont’d)
