Social and Privacy Issues of Data Mining

Patrick Vacek

21 April 2008

MATH 3210

Overview

- The Inference Problem

- Privacy

- Commercial and Government Uses

- Concerns

- The Future

The Inference Problem

- Deducing unauthorized information from legitimate query responses

- The influence of data mining

  - Easier to gather information

  - Easier to block gathering

Source: Thurasisingham 196-201

Privacy Issues

- World Wide Web

- Sensitive information

- Security

  - Encryption

  - Split databases

Sources: Thurasisingham 202-203; Han and Kamber 478

Commercial Uses

- Market basket analysis

- Billing information

- Amazon.com

Source: Streitfeld

Government Uses

- Terrorism prevention
  - Total Information Awareness
  - Commercial cooperation
  - Problems
    - Data quality and quantity
    - Interpretation
    - False positives
    - The Fourth Amendment
    - Mission creep

Sources: Pitts; Zobrist

Concerns for Data Mining

- Could become a derogatory term

- Consumer weariness and confidence

- "Big Brother"

Source: Yang and Wu

The Future

- Fair information practices
  - Openness
  - Limited use

- Opt-out

- Enhanced security

Source: Han and Kamber 477-478

Summary

- The Inference Problem

- Privacy

- Commercial and Government Uses

- Concerns

- The Future

References

- "Affinity analysis". 15 February 2008. *Wikipedia, The Free Encyclopedia*. Retrieved 12 April 2008, from http://en.wikipedia.org/w/index.php?title=Affinity_analysis &oldid=191707485.
- Han, Jiawei, and Micheline Kamber (2001). *Data mining: Concepts and Techniques*. San Francisco, CA: Morgan Kaufmann Publishers.
- Pitts, Joe W. (2007). Under Surveillance: The End of Illegal Domestic Spying? Don't Count on It. *The Washington Spectator: Online Edition*. Retrieved 12 April 2008, from http://www.washingtonspectator.com/articles/20070315surveillance_1.cfm.
- Streitfeld, David (1999). Amazon.com's data-mining technology stirs Internet privacy controversy. *Online Athens*. Retrieved 12 April 2008, from http://www.onlineathens.com/stories/082899/new_0828990006.shtml.
- Thuraisingham, Bhavani M (1999). *Data Mining: Technologies, Techniques, Tools, and Trends*. Boca Raton, FL: CRC Press.
- Yang, Qiang, and Xindong Wu (2006). 10 Challenging Problems in Data Mining Research [Electronic version]. *International Journal of Information Technology and Decision Making, 5* (4), 597-604.
- Zobrist, George W. (2003). Data Mining and Privacy Issues. *Today's Engineer Online*. Retrieved 12 April 2008, from http://www.todaysengineer.org/2003/Oct/data-mining.asp.