The Concept of Privacy in Data Mining

(2 hours)

Lila Ghemri

Department of Computer Science

Texas Southern University

ghemri_lx@tsu.edu

This work was supported by NSF grants 1241772

Any opinions, findings, conclusions, or recommendations expressed herein are those of the authors and do not reflect the views of the National Science Foundation

Data Mining

- Data Mining is the discipline of transforming Data into Information.
- Data mining processes large datasets in order to find patterns and commonalities in the dataset.
- Data mining techniques can be inadvertently or maliciously used to uncover data about the data holders

The Concept of privacy

Privacy is: The right to be left alone (Louis Brandeis)

The right to have some control over how your **information** is properly collected, stored, used or released.

freedom from excessive surveillance – the right to go about our daily lives without being watched or have all our actions caught on camera.

Types of Privacy

- **physical privacy** - such as bag searching, use of DNA
- **information privacy** – the way in which government agencies or organizations handle personal information such as age, address, physical or mental health records

Legislating Privacy

- **The U. S. Constitution contains no express right to privacy.**
- **The Bill of Rights, however protects specific aspects of privacy, such as:**
  - the privacy of beliefs (1st Amendment),
  - privacy of the home against demands that it be used to house soldiers (3rd Amendment),
  - privacy of the person and possessions as against unreasonable searches (4th Amendment),
  - privilege against self-incrimination, which provides protection for the privacy of personal information (the 5th Amendment)

Amendment I (Privacy of Beliefs)

Congress shall make no law respecting an establishment of religion, or prohibiting the free exercise thereof;
or abridging the freedom of speech, or of the press;
or the right of the people peaceably to assemble,
and to petition the Government for a redress of grievances.

Amendment III (Privacy of the Home)

No Soldier shall, in time of peace be quartered in any house, without the consent of the Owner, nor in time of war, but in a manner to be prescribed by law.

Amendment IV (Privacy of the Person and Possessions)

The right of the people to be secure in their persons, houses, papers, and effects, against unreasonable searches and seizures, shall not be violated, and no Warrants shall issue, but upon probable cause, supported by Oath or affirmation, and particularly describing the place to be searched, and the persons or things to be seized.

Amendment IX
(More General Protection for Privacy?)

The enumeration in the Constitution, of certain rights, shall not be construed to deny or disparage others retained by the people.

Liberty Clause of the Fourteenth Amendment

No State shall…deprive any person of life, **liberty**, or property, without due process of law.

http://www.ncsl.org/research/telecommunications-and-information-technology/privacy-protections-in-state-constitutions.aspx

Ninth Amendment

states that:

the "enumeration of certain rights" in the Bill of Rights "shall not be construed to deny or disparage other rights retained by the people."

This has been interpreted as justification for broadly reading the Bill of Rights to protect privacy in ways not specifically provided in the first amendments.

Invasion of Privacy

A person whose privacy has been invaded could sue the invader for damages. These torts exist as four separate branches:

1. Intrusion upon seclusion or solitude, or into private affairs;
2. Public disclosure of embarrassing private facts;
3. Publicity which places a person in a false light in the public eye; and
4. Appropriation of name or likeness.

Privacy in the workplace

- Many of the basic rights in the legislation are not protected in the workplace.
- Private businesses, particularly those whose employees are not unionized, are not limited by the Constitution and the Bill of Rights which address only state action.

Monitoring Communication

- Monitoring phone conversations:
  - Businesses routinely *monitor the telephone calls* of their employees.
  - Employers do not have to warn workers,
  - Employers can listen for up to five minutes to determine if the call is a work-related call or not.

Secret telephone monitoring invades the privacy of both the worker and the other person on the line.

Physical Intrusion

- Many employees are subjected to intrusive *physical searches* of their person, office or possessions in the workplace.
- Hidden cameras can be placed in any location, including locker rooms.
- Increasingly, some companies use imposter employees as spies to keep an eye on their workers.

Physical Intrusion

- Taking *written personality tests* to probe into the most intimate aspects of an employee life including hygiene habits, sexuality and family relationships.
- *Drug testing*
- *Lifestyle discrimination*,

Employees' Privacy

- Growing concern among private sector employers that their computer resources may be abused by employees either by accessing offensive material or jeopardizing the security of proprietary information—
- May provide an easy entry point into a company's electronic systems by computer trespassers.

Digital Privacy

- Most companies report that they :
- store their employees' electronic transactions:
  -e-mail messages,
- store information of Internet sites visited, and
- Monitor computer file activity.

If violations of company policies are found, result in a range of disciplinary actions.

Reasons for collecting this information

- Create duplicate or back-up files in case of system disruptions;
- Manage computer resources such as system capacity to handle routine e-mail and Internet traffic;
- Hold employees accountable for company policies.
- Check for employee visits to sites containing offensive or disruptive material and improper protection of proprietary information.

Companies Rights

Most company policies:

- affirm their rights to review employee use of company computer assets,
- describe appropriate employee uses of these assets, and
- Describe detailed penalties for misuse.

Digital Privacy Laws

- **The Computer Security Act Computer Security Law of 1987** was passed by the United States Congress, to:
- Improve the security and privacy of sensitive information in federal computer systems
- Establish a minimum acceptable security practices for such systems.
- Requires the creation of computer security plans, and the appropriate training of system users or owners where the systems house sensitive information

Privacy Act and Privacy Protection Act

Privacy Act of 1974:

- Establishes a Code of Fair Information Practice that governs the collection, maintenance, use, and dissemination of personally identifiable information about individuals that is maintained in systems of records by federal agencies

Privacy Protection Act

- Each agency that maintains a system of records shall—upon request by any individual ... permit him ... to review the record and have a copy made of all or any portion thereof in a form comprehensible to him .
- permit the individual to request amendment of a record pertaining to him.

The Electronic Communications Privacy Act

- **The Electronic Communications Privacy Act** (ECPA) was passed by Congress in 1986 to bring new communications technologies under the umbrella of the federal wiretap laws
- The law was designed to meet the challenge of sophisticated surveillance technologies, which create the opportunity for government surveillance beyond what is allowed by the Fourth Amendment.
- ECPA does not control government access to private communications strongly enough. An investigator only has to certify that information "relevant" to a criminal investigation will be collected.

HIPAA

- **Title I of HIPAA** protects health insurance coverage for workers and their families when they change or lose their jobs.
- **Title II of HIPAA**, known as the Administrative Simplification (AS) provisions, requires the establishment of national standards for electronic health care transactions and national identifiers for providers, health insurance plans, and employers.

Basic Online Privacy Principles

- *Purpose binding: personal data obtained for one purpose* **should not be used for another purpose without informed consent**.

- *Necessity of data collection and processing: the collection and processing of personal data* **shall only be allowed, if it is necessary for tasks falling within the responsibility of the data processing agency**.

Fair Information Principles

The Federal Trade Commission (FTC) identifies these principles as:

- Notice/Awareness (to the individual about information collected, maintained and used by the entity)
- Choice and Consent (on the part of the individual about that information, including whether it is collected in the first instance and how and under what circumstances it is disclosed to third parties)
- Access/Participation (whether the individual has access to that information and the ability to correct any mistakes)
- Integrity/Security (the administrative, technical and physical safeguards of the information, including notice if the information is breached)
- Enforcement/Redress (legal, policy, contractual or ethical)

Fair Information Practices

- Information should be regarded as held for a specific purpose and not to be used, without appropriate authorization, for other purposes.
- Access to information should be confined to those authorized to have it for the purpose for which it was supplied.
- The amount of information collected and held should be the minimum necessary for the achievement of the specified purpose.
- In computerized systems handling information for statistical purposes, adequate provision should be made in their design and programs for separating identities from the rest of the data.

Fair Information Practices

- There should be arrangements whereby the subject could be told about the information held concerning him.
- The level of security to be achieved by a system should be specified in advance by the user and should include precautions against the deliberate abuse or misuse of information.
- A monitoring system should be provided to facilitate the detection of any violation of the security system.
- In the design of information systems, periods should be specified beyond which the information should not be retained.
- Data held should be accurate. There should be machinery for the correction of inaccuracy and the updating of information.
- Care should be taken in coding value judgments.

- Privacy http://en.wikipedia.org/wiki/Privacy
- Privacilla http://www.privacilla.org/
- Fair Information Principles and Practice at http://www.it.cornell.edu/policies/infoprivacy/principles.cfm
- Workplace Privacy https://www.privacyrights.org/workplace-privacy-and-employee-monitoring

Exercise 1

- Perform an internet search on: Amy Boyer and Liam Youens. Summarize the case.
- What privacy laws were violated in your opinion?
- Should Docusearch have been blamed? Why or why not.
- What technological safeguards should be put in place to avoid a repeat of this situation?

Exercise 2

- After treating a patient injured in a rather unusual sporting accident, the hospital released to the local media, copies of the patient's skull x-ray as well as a description of the complainant's medical condition.
- The hospital asserted that the disclosures were made to avert a serious threat to health or safety.
- Was the hospital justified in their decision, were there any privacy violations and if so which ones.

Exercise 3

-Check the website “ Please Rob Me” at http://pleaserobme.com/

-Does the website violates any privacy laws?

-Research the term “oversharing” and write a paragraph about its meaning, how it happens and what can one do to prevent it.