Lecture 13
Revision & Privacy in DM

Dr.Vani Vasudevan

Privacy, Security and Social Impacts of Data Mining

- Many data mining applications do not touch personal data
  - E.g., meteorology, astronomy, geography, geology, biology, and other scientific and engineering data
- Many DM studies are on developing scalable algorithms to find general or statistically significant patterns, not touching individuals
- The real privacy concern: unconstrained access of individual records, especially privacy-sensitive information

Privacy, Security and Social Impacts of Data Mining

- Method 1: Removing sensitive IDs associated with the data
- Method 2: Data security-enhancing methods
  - Multi-level security model: permit to access to only authorized level
  - Encryption: e.g., *blind signatures*, *biometric encryption*, and *anonymous databases* (personal information is encrypted and stored at different locations)
- Method 3: Privacy-preserving data mining methods

Privacy-Preserving Data Mining

- Privacy-preserving (privacy-enhanced or privacy-sensitive) mining:
  - Obtaining valid mining results without disclosing the underlying sensitive data values
  - Often needs trade-off between information loss and privacy

- Privacy-preserving data mining methods:
  - Randomization (e.g., perturbation): Add noise to the data in order to mask some attribute values of records
  - distorting some classification models

Privacy-Preserving Data Mining

- K-anonymity and l-diversity: Alter individual records so that they cannot be uniquely identified
  - k-anonymity: Any given record maps onto at least k other records
  - l-diversity: enforcing intra-group diversity of sensitive values

- Distributed privacy preservation: Data partitioned and distributed either horizontally, vertically, or a combination of both

- Downgrading the effectiveness of data mining: The output of data mining may violate privacy
  - Modify data or mining results, e.g., hiding some association rules or slightly distorting some classification models

Reference

- Han and Kamber , Data Mining Concepts and Techniques , 3rd Edition