Chapter XIV

Improved Privacy: Preserving Techniques in Large Databases

K. Anbumani, Karunya University, India

R. Nedunchezian, Sri Ramakrishna Engineering College, India

Abstract

Data mining techniques have been widely used for extracting non-trivial information from massive amounts of data. They help in strategic decision-making as well as many more applications. However, data mining also has a few demerits apart from its usefulness. Sensitive information contained in the database may be brought out by the data mining tools. Different approaches are being utilized to hide the sensitive information. The proposed work in this article applies a novel method to access the generating transactions with minimum effort from the transactional database. It helps in reducing the time complexity of any hiding algorithm. The theoretical and empirical analysis of the algorithm shows that hiding of data using this proposed work performs association rule hiding quicker than other algorithms.
Introduction

Experts say that data mining in the wrong hands will end up in destruction. The main threat of data mining is to privacy and security of data residing in large data stores (Clifton & Marks, 1996; Atallah, Bertino, Elmagarmid, Ibrahim, & Verykios, 1999; Agrawal & Aggarwal, 2001; Muralidhar, Krishnamurty, Sarathy, Rathindira, Parsa, & Rahul, 2001; Oliveira & Za¨ıane, 2003a, 2003b; Verykios, Elmagarmid, Bertino, Saygin, & Dasseni, 2004; Ashrafi, Taniar, & Smith, 2005). Some of the information considered as private and secret can be brought out with advanced data mining tools; this is a real concern of people working in the field of database technology. Different research efforts are under way to deal with this problem of preserving security and privacy.

Sensitive information contained in a database can be extracted with the help of non-sensitive information. This is called the inference problem (Clifton et al., 1996; Marks, 1996; Verykios et al., 2004). Different concepts have been proposed to handle the inference problem. The process of modifying the transactional database to hide some sensitive information is called sanitization. By sanitizing the original transactional database, the sensitive information can be hidden. In the sanitization process, selective transactions are retrieved and modified before handing over the database to a third party.

Modification of transaction involves removing an item from a transaction or adding an element to the transaction. In some cases, transactions are either added to or removed from the database as suggested in Clifton et al. (1996). The modified database is called sanitized database or released database.

Several approaches have been proposed to hide sensitive data with good accuracy. The efficiency of a privacy-preserving algorithm is measured based on (1) the time taken to hide the data, (2) the number of new rules introduced as a result of the hiding process, and (3) the number of legitimate rules lost or which cannot be extracted from the released database.

The task of locating a transaction for sanitization from a massive amount of data is not a trivial process and it is certainly a time consuming one. In many research efforts, the highly time consuming process of retrieving the transactional database is not taken into account efficiently. This chapter proposes two methods to hide sensitive association rule in a faster manner. The first method uses the advantages of frequent pattern growth tree (FPT) to identify and retrieve the generating transactions directly from the transactional database without exhaustive search. An array is used to keep track of the identifiers of the required transactions for sanitization. The second method proposes another approach to hide rules using a partitioning approach.

This chapter is organized as follows: the subsequent section discusses the existing related works. The proposed approach is discussed in third section. The performance
Related Content

On Dependability Issues in Ambient Intelligence Systems
[www.igi-global.com/article/dependability-issues-ambient-intelligence-systems/58337?camid=4v1a](www.igi-global.com/article/dependability-issues-ambient-intelligence-systems/58337?camid=4v1a)

The Bees Algorithm and Its Applications
[www.igi-global.com/chapter/the-bees-algorithm-and-its-applications/123078?camid=4v1a](www.igi-global.com/chapter/the-bees-algorithm-and-its-applications/123078?camid=4v1a)

Display Content Adaptation Using a Force Sensitive Office Chair
Andreas Riener (2011). *International Journal of Ambient Computing and Intelligence* (pp. 8-17).
[www.igi-global.com/article/display-content-adaptation-using-force/58336?camid=4v1a](www.igi-global.com/article/display-content-adaptation-using-force/58336?camid=4v1a)
Agent-Oriented Methodologies: An Introduction

www.igi-global.com/chapter/agent-oriented-methodologies/24275?camid=4v1a