Chapter 33
Survey on Privacy Preserving Association Rule Data Mining

Geeta S. Navale
Smt. Kashibai Navale College of Engineering, India

Suresh N. Mali
Sinhgad Institute of Technology and Science, India

ABSTRACT

The progress in the development of data mining techniques achieved in the recent years is gigantic. The collative data mining techniques makes the privacy preserving an important issue. The ultimate aim of the privacy preserving data mining is to extract relevant information from large amount of data base while protecting the sensitive information. The togetherness in the information retrieval with privacy and data quality is crucial. A detailed survey of the present methodologies for the association rule data mining and a review of the state of art method for privacy preserving association rule mining is presented in this paper. An analysis is provided based on the association rule mining algorithm techniques, objective measures, performance metrics and results achieved. The metrics and the short comings of the various existing technologies are also analysed. Finally, the authors present various research issues which can be useful for the researchers to accomplish further research on the privacy preserving association rule data mining.

INTRODUCTION

With advancement in cloud computing based on the internet and data centres, the data availability in the outsourcing is aggressively expanding and is expected to skyrocket in near future. The fact is evident that the amount of the data in the world in doubling every two months (Geng & Hamilton, 2006). From the bulk data available because of the rapid growth in the information and e-commerce applications, useful and user expected information cannot be easily discovered. The user expectation is to make use of the sophisticated information from the bulk mass of data. Data mining is the finest solution for the retrieval of the sophisticated information. Data mining is a technique that benefits to extract the important data from a large data base. The sorting out of irrelevant information by the data mining is finished by the

DOI: 10.4018/978-1-5225-7113-1.ch033

use of different retrieving techniques or by the use of certain sophisticated algorithms. The data mining process is explained as follows: 1) Understanding the Job: The data needs to be in line with the job objectives, Data Mining objectives, risk analysis involving the data privacy etc. 2) Understanding the data: Describe, explore and verify data. 3) Preparation of data: Data is filtered, consolidated, cleaned, and formatted. 4) Process Modelling: Modelling techniques based on identified in data mining objectives and business domain in step 1, parameter setting and test designing is done. 5) Process Evaluation: Evaluation of results to establish accuracy, and review the process. 6) Deployment: Monitoring and maintenance, Review reports etc. The tasks in the data mining process are categorized into:

- Classification
- Clustering
- Association rule mining
- Sequential pattern mining
- Regression

In classification task, the data in the corpus is classified into predefined classes. In clustering task, the pattern of the partition is set into disjoint and homogeneous groups. Association rule mining task aids in information retrieval, by identifying the frequent patterns in the data and in the form of the dependencies among concepts-attributes association. Data mining can be achieved using many approaches such as statistical, machine learning, database-oriented, neural networks, rough sets, and visualization. The useful applications of data mining are discussed with respect to Classification, Forecasting, Association and Clustering.

Purpose of the data mining is to discover the unknown information. The discovery of the implicit information also reveals the private or secure information in the data base such as credit card numbers, personal identification numbers, telephone numbers and other confidential data. In business sector, the information retrieval using data mining technique reveals sensitive knowledge about the corporate to the competitors. Verykios, Elmagarmid, Bertino, Saygin, and Dasseni (2004), put forward data mining potentially reveals the privacy as an open source. There is a need to protect the private data during the data mining process. This problem is called the privacy preserving data mining (PPDM). Privacy preserving in the data mining is important task to overcome the problem associated with privacy revealing. The main aim of the privacy preserving in the data mining is to hide or secure the confidential information during the discovery of the information from the database. Privacy securement with the permission for the data analysis is the ultimate aim of PPDM. This can be done by hiding the sensitive association rules. According to Wong, Cheung, Hung, Kao, and Mamoulis (2007), association rule mining is the important task which can help in overcoming the privacy problem associated with data mining. In large databases, associate rule mining is used for discovering the frequently co-occurring data items and interesting association relationship between the data item. Association rule mining is a technique in data mining that identifies the regularities found in large volume of data. Such a technique may identify and reveal hidden information that is private for an individual or organization.

Verykios, Elmagarmid, Bertino, Saygin, and Dasseni (2004), divulge the Association rule hiding as one technique to PPDM. The sanitization of the original data in the corpus is achieved by association rule hiding technique. It hides the sensitive rules in the data base containing sensitive information during the data analysis.