Abstract

Data mining is the extraction of the hidden information from large databases. It is a powerful technology with new great potential to analyze important information in the data warehouse.
Preserving privacy against data mining algorithms is a new research area. It investigates the side-effects of data mining methods that derive from the privacy diffusion of persons and organizations. Privacy preserving data mining is the emerging field that protects sensitive data. Classification is one of the popular techniques of data mining. Classification is a data mining technique used to predict group membership for data instances. Classification involves finding rules that partition the data into disjoint groups. Many classification rule algorithms are used to generate the classification rules such as OneR, Ridor, and Conjuctive Rule. In this paper, we focus on the problem of privacy preservation in classification rules. The rule based classification algorithms namely C4.5, Ripper and Part algorithms are used for generating rules. The privacy is preserved by hiding the sensitive rules and the new dataset is reconstructed from the non-sensitive rules. In this paper the experimental results shows the effectiveness of each algorithm.

Reference

An Efficient Algorithm for Classification Rule Hiding

- Juggapong Natwichai, Xingzhi Sun, Xue Li, “A Heuristic Data Reduction Approach for Associative Classification Rule Hiding”, IBM Research Laboratory, Beijing, China.
- Juggapong Natwichai, Xingzhi Sun, Xue Li, “Associative classification rules hiding for privacy preservation”, IBM Research Laboratory, Beijing, China.
- Mary DeRosa. “Data Mining and Data Analysis for Counterterrorism”.

Index Terms

Computer Science

Data Mining

Key words

Rules

C4.5

Ripper

Part