

数据库、信息处理

隐私保护关联规则挖掘算法的研究

王锐, 刘杰

哈尔滨工程大学 计算机科学与技术学院, 哈尔滨 150001

收稿日期 2008-5-14 修回日期 2008-9-1 网络版发布日期 2009-9-15 接受日期

摘要 针对MASK算法的不足, 将随机响应技术与关联规则挖掘算法相结合, 提出一个多参数随机扰动算法—MRD算法。当以不同的随机参数对数据集进行处理时, 可以实现对原始数据的干扰或隐藏, 解决了单一使用数据干扰策略和数据隐藏策略的缺陷, 有效地提高了算法的隐私保护度。在此基础上, 给出了在伪装后的数据集上生成频繁项集的挖掘算法。最后, 通过具体实验验证, 证明了当随机参数选择合适时, MRD算法的隐私性和准确性均优于原算法。

关键词 [数据挖掘](#) [关联规则](#) [频繁项集](#) [隐私保护](#) [随机响应](#)

分类号 [TP301](#)

Research of privacy preserving association rules mining algorithm

WANG Rui, LIU Jie

College of Computer Science and Technology, Harbin Engineering University, Harbin 150001, China

Abstract

In view of the insufficiency of MASK algorithm, randomized response technology and association rule mining algorithm are integrated and a multi-parameters randomized disturb algorithm is proposed, which is called MRD algorithm. When the data sets are processed with different random parameters, the original data can be disturbed and hidden, and the defects of the simplex using of data diturb and data hiding strategy are solved, and the privacy-preserving degree of the algorithm is improved effectively. On this basis, the algorithm of generating frequent items from transformed data sets is proposed. Finally, through specific certification of examples, it can be proved that when the random parameters are chosen suitably, the privacy and accuracy of MRD algorithm are both better than the original algorithm.

Key words [data mining](#) [association rule](#) [frequent itemset](#) [privacy preservation](#) [randomized response](#)

DOI: 10.3778/j.issn.1002-8331.2009.26.037

通讯作者 王锐 liujie@hrbeu.edu.cn

扩展功能

本文信息

- ▶ [Supporting info](#)
- ▶ [PDF\(905KB\)](#)
- ▶ [\[HTML全文\]\(0KB\)](#)
- ▶ [参考文献](#)

服务与反馈

- ▶ [把本文推荐给朋友](#)
- ▶ [加入我的书架](#)
- ▶ [加入引用管理器](#)
- ▶ [复制索引](#)
- ▶ [Email Alert](#)
- ▶ [文章反馈](#)
- ▶ [浏览反馈信息](#)

相关信息

- ▶ [本刊中 包含“数据挖掘”的相关文章](#)
- ▶ [本文作者相关文章](#)

- [王锐](#)
- [刘杰](#)