

数据库、信号与信息处理

Apriori-sort算法研究

汪维清, 罗先文, 胡继宽

西南大学 信息管理系, 重庆 402460

收稿日期 2008-7-31 修回日期 2008-10-20 网络版发布日期 2008-12-18 接受日期

摘要 Apriori算法虽然在候选集的产生时利用了剪支技术, 但每次扫描数据库时都必须扫描整个数据库, 因此扫描的数据量大, 速度较慢。Apriori-sort算法是在Apriori算法基础上的改进, 基本思想是把事务数据库变为以度表示的事务度数据库, 并对事务度数据库进行排序。Apriori-sort算法查找频繁项集时, 只扫描数据库 Dd 中满足 $d(Ck) \leq d(Ti)$ 的事务。对扫描数据库进行了有效剪支, 因此Apriori-sort算法的计算效率高。并用仿真数据对Apriori-sort算法和Apriori算法进行了仿真对比实验, 实验结果证明了新算法的高效性。

关键词 [排序](#) [Aprior](#) [关联规则](#) [购物篮分析](#) [数据库知识发现](#)

分类号

Apriori-sort algorithm

WANG Wei-qing, LUO Xian-wen, HU Ji-kuan

Department of Information Management, Southwest University, Chongqing 402460, China

Abstract

Although Apriori algorithm uses cut-technology when it generate item sets of candidates, it has to scan the entire database while scanning the transaction database each time. The scanning speed is very slow for its large amount of data. The Apriori-sort algorithm is improved from the Apriori algorithm. Its basic idea is to transform a transaction database into a transaction degree database, and sort the transaction-degree database. When the Apriori-sort algorithm searches the frequent item sets, it only scans the transactions in the database Dd that appeares $d(Ck) \leq d(Ti)$. It scans the database with an effective cut. Therefore the Apriori-sort algorithm's computing time is very fast. The results of simulation experiments for the Apriori-sort algorithm and the Apriori algorithm show the new algorithm's efficiency.

Key words [sort](#) [Apriori](#) [association rule](#) [market basket analysis](#) [Knowledge Discovery in Database \(KDD\)](#)

DOI: 10.3778/j.issn.1002-8331.2008.36.044

通讯作者 汪维清 wwqlhy@163.com

扩展功能

本文信息

▶ [Supporting info](#)

▶ [PDF\(562KB\)](#)

▶ [\[HTML全文\]\(0KB\)](#)

▶ [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“排序”的 相关文章](#)

▶ 本文作者相关文章

· [汪维清](#)

· [罗先文](#)

· [胡继宽](#)