

博士论坛

一种基于模式树的频繁项集快速挖掘算法

战立强 刘大昕 张健沛

哈尔滨工程大学计算机科学与技术学院 哈尔滨工程大学计算机科学与技术学院

收稿日期 2007-1-18 修回日期 网络版发布日期 2007-4-4 接受日期

摘要 模式树是目前频繁项集挖掘最常用的数据结构,使用模式树可以有效地将数据库压缩于内存,并在内存中完成对频繁项集的挖掘。为了进一步提高频繁项集挖掘算法的可扩展性,本文对模式树进行了细致的研究,在此基础上提出了一种挖掘频繁项集的新算法,FP-DFS算法。该算法通过对模式树的各种操作简化了对频繁项集的搜索过程。实验表明,该算法对于频繁项集挖掘具有比较高的效率。

关键词 [关联规则](#) [频繁项集挖掘](#) [可扩展性](#) [模式树](#)

分类号

A Fast algorithm for frequent item-set mining based on Pattern tree

Da-xin LUI

Abstract

Pattern tree was the most frequently used data structure in frequent item-set mining. By using Pattern tree, database could be effectively compressed into main memory, and the subsequence mining task could be completed in main memory. To make further improvement on the scalability of the algorithm, we made a further study on the Pattern tree, and proposed a new algorithm called FP-DFS based on the study. FP-DFS simplifies the mining processes through applying various operations on Pattern tree. The experiments show that FP-DFS had good efficiency in frequent item-set mining.

Key words [association rule](#) [frequent item-set mining](#) [scalability](#) [Pattern tree](#)

DOI:

通讯作者 战立强 topdream@people.com.cn

扩展功能

本文信息

▶ [Supporting info](#)

▶ [PDF\(563KB\)](#)

▶ [\[HTML全文\]\(0KB\)](#)

▶ [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“关联规则”的相关文章](#)

▶ [本文作者相关文章](#)

· [战立强 刘大昕 张健沛](#)