

本期目录 | 下期目录 | 过刊浏览 | 高级检索

[打印本页] [关闭]

软件技术与数据库

基于动态项集计数的加权频繁项集算法

秦丽君^{1,2}, 罗雄飞¹

(1. 中国科学院软件研究所, 北京 100190; 2. 中国科学院研究生院, 北京100190)

摘要: 基于Apriori的加权频繁项集挖掘算法存在扫描数据集次数多的问题。为此, 提出一种基于动态项集计数的加权频繁项集算法。该算法采用权值键树的数据结构和动态项集计数的方法, 满足向下闭合特性, 并且动态生成候选频繁项集, 从而减少扫描数据集的次数。实验结果证明, 该算法生成的加权频繁项集具有较高的效率和性能。

关键词: 数据挖掘 加权频繁项集挖掘 动态项集计数 加权支持度 权值键树 向下闭合特性 最大权值

Weighted Frequent Itemset Algorithm Based on Dynamic Itemset Counting

QIN Li-jun^{1,2}, LUO Xiong-fei¹

(1. Institute of Software, Chinese Academy of Sciences, Beijing 100190, China; 2. Graduate University of Chinese Academy of Sciences, Beijing 100190, China)

Abstract: The existing weighted frequent itemset mining algorithms which are based on Apriori require multiple dataset scans. This paper proposes a weighted frequent itemset algorithm weighted frequent itemset mining based on dynamic itemset counting which uses the structure of weighted trie tree and the method of dynamic itemset counting. This algorithm satisfies the downward closure property and dynamically generates candidate frequent itemsets, thereby reduces the number of scanning datasets and improves the performance. Experimental results show that the proposed algorithm not only generates the weighted frequent itemsets, but also has high efficiency and time performance.

Keywords: data mining weighted frequent itemset mining dynamic itemset counting weighted support degree weighted trie tree downward closure property maximum weight

收稿日期 2011-06-29 修回日期 网络版发布日期 2012-02-05

DOI: 10.3969/j.issn.1000-3428.2012.03.011

基金项目:

国家“863”计划基金资助项目(2007AA040702)

通讯作者:

作者简介: 秦丽君(1983—), 女, 硕士研究生, 主研方向: 数据挖掘, 数据可视化; 罗雄飞, 博士

通讯作者E-mail: weiweiqueeena@gmail.com

参考文献:

- [6] Brin S.[J].Motwani R, Ullman J, et al. Dynamic Itemset Counting and Implication Rules for Market Basket Data[C]//Proc. of 1997 ACM SIGMOD Int’1 Conf. on Management of Data. [S. 1.]: ACM Press.1997,.- 
- [8] SPSS. Data Mining(Clementine)[DB/OL]. (2008-10-30). http:// www.spss.com/data

扩展功能

本文信息

- ▶ Supporting info
- ▶ [PDF\(252KB\)](#)
- ▶ [\[HTML\] 下载](#)
- ▶ [参考文献\[PDF\]](#)
- ▶ [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](#)
- ▶ [引用本文](#)
- ▶ [Email Alert](#)
- ▶ [文章反馈](#)
- ▶ [浏览反馈信息](#)

本文关键词相关文章

- ▶ [数据挖掘](#)
- ▶ [加权频繁项集挖掘](#)
- ▶ [动态项集计数](#)
- ▶ [加权支持度](#)
- ▶ [权值键树](#)
- ▶ [向下闭合特性](#)
- ▶ [最大权值](#)

本文作者相关文章

- ▶ [秦丽君](#)
- ▶ [罗雄飞](#)

PubMed

- ▶ [Article by Qin, L. J.](#)
- ▶ [Article by Luo, X. F.](#)

本刊中的类似文章

1. 朱小栋, 樊重俊, 杨坚争. 面向机场场区管理的数据挖掘系统[J]. 计算机工程, 2012,38(3): 224-227
2. 李光, 王亚东, 苏小红. 基于概率论的隐私保持分类挖掘[J]. 计算机工程, 2012,38(3): 12-13,18
3. 刘佳新, 严书亭, 任家东. 缩减投影数据库规模的增量式序列模式算法[J]. 计算机工程, 2012,38(3): 28-30
4. 潘冬生, 章昭辉, 代秀娟, 杨娟. 基于层次梯度分析的协同数据挖掘算法[J]. 计算机工程, 2012,38(2): 72-74
5. 刘芳. 基于图和双向搜索的频繁项集挖掘算法[J]. 计算机工程, 2012,38(01): 59-61
6. 姜丽莉, 孟凡荣, 周勇. 多值属性关联规则挖掘的 Q-Apriori算法?[J]. 计算机工程, 2011,37(9): 81-83
7. 刘文远, 张亮, 孙德杰, 陈子军. 改进的SOD孤立点检测算法[J]. 计算机工程, 2011,37(9): 93-94,97
8. 花红娟, 张健, 陈少华. 基于频繁模式树的约束最大频繁项集挖掘算法[J]. 计算机工程, 2011,37(9): 78-80
9. 神鹏飞, 王希武, 耿志广, 王创伟, 李国良. 一种无阈值的频繁模式生成算法[J]. 计算机工程, 2011,37(8): 31-33
10. 马博, 杨雅婷, 周喜, 胡斌华. 改进的关联分类算法在信息化评估中的应用[J]. 计算机工程, 2011,37(7): 7-9,17

文章评论

反馈人	<input type="text"/>	邮箱地址	<input type="text"/>
反馈标题	<input type="text"/>	验证码	<input type="text" value="7741"/>
<input type="button" value="5"/>			