

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

缩减投影数据库规模的增量式序列模式算法

刘佳新^a, 严书亭^b, 任家东^a

(燕山大学 a. 信息科学与工程学院; b. 科学技术研究院, 河北 秦皇岛 066004)

摘要: 在增量式序列模式挖掘算法中, 数据库更新只有插入和扩展2种操作, 未考虑序列删除的情况。为此, 提出一种基于频繁序列树的增量式序列模式更新算法(IUFST)。在数据库和支持度发生变化时, IUFST算法分不同情况对频繁序列树进行更新操作, 缩减投影数据库的规模, 提高算法效率。实验结果表明, 该算法在时间性能上优于PrefixSpan算法和IncSpan算法。

关键词: 数据挖掘 增量式挖掘 序列模式 投影数据库 频繁序列树 深度优先

Incremental Sequential Pattern Algorithm of Reducing Projected Database Size

LIU Jia-xin^a, YAN Shu-ting^b, REN Jia-dong^a

(a. College of Information Science and Engineering; b. Science and Technology Administration Office, Yanshan University, Qinhuangdao 066004, China)

Abstract: This paper proposes an incremental sequential patterns updating algorithm based on frequent sequence tree, called IUFST, in order to solve the problem that when the database is updated, the existed incremental mining algorithms of sequential patterns only mention two kinds of database updates, insert and append rather than the delete operation. When the database is updated and the support is changed, IUFST is divided into four kinds of situations to update the frequent sequence tree. It reduces the size of the projected database and improves the efficiency. Experimental results show that IUFST outperforms PrefixSpan and IncSpan in time cost.

Keywords: data mining incremental mining sequential pattern projected database frequent sequence tree depth-first

收稿日期 2011-06-03 修回日期 网络版发布日期 2012-02-05

DOI: 10.3969/j.issn.1000-3428.2012.03.010

基金项目:

河北省教育厅科学研究计划基金资助项目(2008498); 河北省自然科学基金资助项目(F2010001298); 秦皇岛市科学技术研究与发展计划基金资助项目(201001 A018);

通讯作者:

作者简介: 刘佳新(1978—), 女, 博士研究生, 主研方向: 数据挖掘; 严书亭, 硕士; 任家东, 教授、博士生导师

通讯作者E-mail: ljx@ysu.edu.cn

参考文献:

[1] Zou Xiaohong, Zhao Li, Guo Jingfeng, et al. An Advanced Algorithm of Frequent Subgraph Mining

扩展功能

本文信息

- ▶ Supporting info
- ▶ PDF(212KB)
- ▶ [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
- ▶ 增量式挖掘
- ▶ 序列模式
- ▶ 投影数据库
- ▶ 频繁序列树
- ▶ 深度优先

本文作者相关文章

- ▶ 刘佳新
- ▶ 严书亭
- ▶ 任家东

PubMed

- ▶ Article by Liu, J. X.
- ▶ Article by Yan, S. T.
- ▶ Article by Lin, J. D.

Based on ADI[J]. ICIC Express Letters, 2009, 3(3): 639-644.

[2] Agrawal R, Srikant R. Mining Sequential Patterns[C]//Proc. of the 11th International Conference on Data Engineering. [S. 1.]: IEEE Press, 1995.

[3] Cheng Hong, Yan Xifeng, Han Jiawei. IncSpan: Incremental Mining of Sequential Patterns in Large Database[C]//Proc. of the 10th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. Seattle, USA: [s. n.], 2004.

[4] Nguyen S N, Sun Xingzhi, Orłowska M E. Improvements of IncSpan: Incremental Mining of Sequential Patterns in Large Database[C]//Proc. of the 9th Pacific-Asia Conference on Knowledge Discovery and Data Mining. Hanoi, Vietnam: [s. n.], 2005.

[5] Ren Jiadong, Sun Yafei, Guo Sheng. Incremental Sequential Pattern Mining Based on Constraints[J]. Journal of Computational Information Systems, 2008, 4(2): 571-576.

[6] Ezeife C I, Liu Yi. Fast Incremental Mining of Web Sequential Patterns with PLWAP Tree[J]. Data Mining and Knowledge Discovery, 2009, 19(3): 376-416.

[7] Liu Jiaxin, Yan Shuting, Ren Jiadong. The Design of Frequent Sequence Tree in Incremental Sequential Patterns Mining[C]//Proc. of International Conference on Software Engineering and Service Sciences. Beijing, China: [s. n.], 2011.

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. 秦丽君, 罗雄飞. 基于动态项集计数的加权频繁项集算法[J]. 计算机工程, 2012, 38(3): 31-33
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. 姚全珠, 李如琼, 王美君. 项约束先过滤的最大频繁项集挖掘算法[J]. 计算机工程, 2012, 38(04): 73-75
6. 张玉冲, 蔡皖东, 丁要军. 一种适用于应用层协议的特征提取算法[J]. 计算机工程, 2012, 38(04): 266-268
7. 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
8. 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
9. 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
10. 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

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="6569"/>
<input type="text"/>			