

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

支持实时增量更新的闭子树聚类算法

黄 伟, 郭 鑫, 周清平

(吉首大学信息管理与工程学院, 湖南 张家界 427000)

摘要: 现有的树聚类算法在树数据库实时更新后无法及时更新已有的聚类结果。为此, 建立一种支持实时增量更新的闭子树聚类模型, 以解决闭子树的增量聚类问题并提高聚类效率。针对树的半结构化特性, 将结点语义和结点-边的结构特性结合在一起, 提出一种准确率更高的树相似性度量方法, 在此基础上, 利用CTUM算法、TC算法和UTC算法, 分别解决闭子树增量更新、聚类和增量聚类等问题。实验结果表明, 该算法具有较高的运行效率和聚类准确率。

关键词: 聚类算法 数据挖掘 闭子树 增量更新

Closed Subtree Clustering Algorithm Supporting Real-time Incremental Update

HUANG Wei, GUO Xin, ZHOU Qing-ping

(School of Information Management and Engineering, Jishou University, Zhangjiajie 427000, China)

Abstract: In real application environment, when tree database carries out live updating, the present tree-cluster algorithm can not update existing cluster result. In consideration of the semi-structured characteristics and the lower accuracy rate of similarity measurement of tree, this paper puts forward a similarity measuring method of combining knot semantics and structure feature of one side of knot. On that basis, it brings forth Closed Tree Update Mining (CTUM) algorithm, Tree Cluster(TC) algorithm and Update Tree Cluster(UTC) algorithm, which can separately solve problems of increment updating, clustering and increment clustering of close subtree. Experimental result proves that the novel algorithms are efficient and practicable to own higher executing efficiency and better cluster accuracy rate.

Keywords: clustering algorithm data mining closed subtree incremental update

收稿日期 2011-05-19 修回日期 网络版发布日期 2011-12-20

DOI: 10.3969/j.issn.1000-3428.2011.24.009

基金项目:


吉首大学校级科研基金资助项目(11JD051); 吉首大学教学改革研究基金资助项目(10JD043)


通讯作者:

作者简介: 黄 伟(1981—), 男, 讲师、硕士, 主研方向: 聚类算法, 数据挖掘; 郭 鑫, 助教、硕士; 周清平, 教授、博士

通讯作者E-mail: jianghai079@126.com

参考文献:

[1] Zaki M J. Efficiently Mining Frequent Trees in a Forest: Algorithms and Applications[J]. IEEE Transactions on Knowledge and Data Engineering. 2005, 17(8): 1021-1035 

[3] 赵传申, 孙志挥, 张 净. 基于投影分支的快速频繁子树挖掘算法[J]. 计算机研究与发展. 2006, 43(3): 456-462 

扩展功能

本文信息

- ▶ Supporting info
- ▶ PDF(273KB)
- ▶ [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

- ▶ 聚类算法
- ▶ 数据挖掘
- ▶ 闭子树
- ▶ 增量更新

本文作者相关文章

- ▶ 黄伟
- ▶ 郭鑫
- ▶ 周清平

PubMed

- ▶ Article by Huang, W.
- ▶ Article by Guo, X.
- ▶ Article by Zhou, Q. B.



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. 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
5. 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
6. 刘永文, 李天瑞, 陈红梅, 高子喆, 谷小广. 覆盖广义粗糙集中近似集增量更新方法研究[J]. 计算机工程, 2012,38(2): 156-158
7. 向河林, 张明西, 李珀瀚, 何震瀛, 汪卫. 一种基于聚类的语义检索算法[J]. 计算机工程, 2012,38(2): 36-38
8. 王亮, 王士同. 基于成对约束的动态加权半监督模糊核聚类[J]. 计算机工程, 2012,38(01): 148-150
9. 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
10. 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

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="4069"/>
	<input type="text"/>		