

基于 MapReduce 的并行加权 FIUT 算法

张航, 张欣, 张平康, 李琪

(贵州大学 大数据与信息工程学院, 贵州 贵阳 550025)

摘要: 针对目前大数据环境下, 传统频繁项集挖掘算法效率低下的问题, 在 MapReduce 框架的基础上, 结合加权模型提出了一种并行加权频繁项集挖掘算法 PWFIUT(Parallel Weighted Frequent Itemset Ultrametric Tree). 该算法将候选项映射到哈希表中以划分支持度, 同时避免构建条件模式和实现压缩存储. 最后, 对 PWFIUT 算法在 Hadoop 平台进行了测试与分析, 实验结果表明所提出的算法具有较好的运行效率和扩展性.

关键词: 大数据; 关联规则; MapReduce; 加权模型; FIUT

Parallel Weighted FIUT Algorithm Based on MapReduce

ZHANG Hang, ZHANG Xin, ZHANG Ping-kang, Li Qi

(College of Big Data and Information Engineering, Guizhou University, Guiyang 550025, China)

Abstract: Aiming at the inefficiency of traditional frequent itemsets mining algorithm in view of the big data environment, a solution to this problem parallel weighted mining of frequent itemsets using PWFIUT(Parallel Weighted Frequent Itemset Ultrametric Tree) algorithm is implemented on MapReduce framework. Support is counted by mapping the items from the candidate list into the buckets which is divided according to support known as Hash table structure, also to avoid building conditional patterns and to achieve compressed storage. Finally, the Algorithm is verified and analyzed on Hadoop platform. According to the compared experiment results, it shows that the proposed algorithm has high efficiency and good scalability.

Key words: big data; association rule; MapReduce; weighted model; FIUT

作者简介:

张航男, (1991-), 硕士研究生. 研究方向为云计算与大数据.

张欣 (通讯作者) 男, (1976-), 博士, 副教授. 研究方向为下一代无线通信及应用等. E-mail: 2225400115@qq.com.

张平康男, (1993-), 硕士研究生. 研究方向为图像处理.

李琪男, (1990-), 硕士研究生. 研究方向为并行化计算.