URN etd-0407109-183702-104
Author Hui-Ling Chang
Author's Email Address No Public.
Statistics This thesis had been viewed 48 times. Download 12 times.
Department
Year 2007
Semester 2
Degree Master
Type of Document Master's Thesis
Language zh-TW.Big5 Chinese
Title Mining Consecutive Sequential Patterns Based on Projection Tree
Date of Defense 0000-00-00
Page Count 147
Keyword
data mining
consecutive sequential patterns
project tree
compressed tree

Abstract
Recently the technologies of data mining are used in many applied sciences and become indispensable. As Internet has gained popularity among society, a great number of data are accumulated during the users browse websites. The owners of websites can use those data to discover the interests and improve the structures of their website.
In this thesis, we propose four Consecutive Sequence Patterns algorithm of TPCT(Traversal Patterns Using Compressed Tree)、PTP(Project Tree Pruning)、JCT(Joined Compressed Tree) and CPFP-Growth(Continuous Patterns Using FP-Growth). TPCT and JCT which uses compressed tree structures to mine the consecutive sequential patterns. The compressed tree technology can save more memory than other tree technologies. The TPCT algorithm needs to spend N-1 time of decomposition when building TPCP-tree, and the performance is becoming worse when the depth of the tree is becoming larger. Therefore we propose an algorithm JCT(Joined Compressed Tree) to improve the shortcomings of TPCT algorithm.ICT algorithm still uses Compressed Tree structures. It uses the compression tree structure which reduces the size of tree and uses filtering mechanism to reduce the infrequent patterns, and therefore it can obtain the consecutive sequential patterns efficiently. CPFP-Growth and PTP algorithm maily use project tree technologies to obtain the frequent itemsets. The Concept of CPFP-Growth algorithm is similar to FP-Growth which is used in association rules mining. In this thesis, CPFP-Growth algorithm uses the methods of FP-Growth to build trees and project trees for mining traversal patterns. The differences of CPFP-Growth algorithm and FP-Growth are CPFP-Growth does not need to sort the items of transaction records and it uses depth-first earch to discover the Conditional pattern base. CPFP-Growth algorithm is good in performance, but it needs to spend more time to build and mine project trees when support is low or the dataset is sparse. In this thesis, we propose an algorithm PTP (Projected Tree Pruning) to improve the shortcomings of CPFP-Growth algorithm. CPFP-Growth algorithm produces a projected tree for every item in the head table, but not every subtree of projected tree is frequent. In order to avoid building any projected tree with redundant branches which are infrequent. PTP algorithm uses filtering mechanism to reduce the size of projected trees, and therefore it saves the time of building projected trees and increases the performance of mining.
Comprehensive experiments have been conducted to assess the performance of the proposed algorithm. The experimental results show that the algorithms which we proposed outperform previous algorithms in the experiments.
Advisory Committee Jen-Peng Huang - advisor

Files etd-0407109-183702-104-001.pdf

indicate access worldwide
Date of Submission 2008-07-05