URN etd-0407109-183719-1561
Author Chang-En Ke
Author's Email Address No Public.
Statistics This thesis had been viewed 43 times. Download 47 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 Tree Decomposing
Date of Defense 0000-00-00
Page Count 157

Keyword
data mining
consecutive traversal paths
data stream
filtration mechanism

Abstract
Due to the science and technology make a great progress, transactions, documents and data are transformed into electronic types, the large number of data has been accumulated. Therefore, data mining technology becomes more important than before in recent years. In data mining territory, sequence pattern and data stream mining plays a quite important position. In this paper we research consecutive traversal paths were proposed to improve the efficiency of data mining.
In this paper we propose new algorithms CTP (An Efficient Algorithm For Mining Consecutive Traversal Paths), CTP-M(An Efficient Algorithm For Mining Consecutive Traversal Paths-Modified Version) and DSCP(An Efficient Algorithms For Mining Consecutive Traversal Paths From Data Stream). CTP algorithm scans database once and saves them in CTP-tree structure which can save space. Second, it uses a hash based technique, Hash MAP, which is similar to Hash Table to increase the access efficiency. Third, CTP algorithm uses an infrequent consecutive sequences filtering mechanism to avoid generating a great deal of infrequent or duplicate sub-sequences of transaction records. The CTP algorithm avoids wasting a lot of I/O time and increases the efficiency and the practicability in application. It wastes a lot of time to construct CTP-tree of all transactions, so we propose a modified version of CTP algorithm - CTP-M (An Efficient Algorithm For Mining Consecutive Traversal Paths-Modified Version). It uses the reduced transation records which are reduced by the frequent 1-itemset L1 to build CTP-tree. Therefore CTP-M algorithm reduces the size of CTP-tree and saves the time of tree construction.
DSCP algorithm uses the concept of CTP, DSCP-tree structure and sliding window model for mining consecutive traversal paths from data stream.
Keywords: data mining, consecutive traversal paths, data stream,filtration mechanism.
Advisory Committee Jen-Peng Huang - advisor

Files etd-0407109-183719-1561-001.pdf

indicate access worldwide
Date of Submission 2008-07-04