URN etd-0407109-183709-681
Author Ho-Yi chang
Author's Email Address No Public.
Statistics This thesis had been viewed 49 times. Download 15 times.
Department
Year 2007
Semester 2
Degree Master
Type of Document Master's Thesis
Language zh-TW.Big5 Chinese
Title Efficient Approaches for Mining Consecutive Sequential Patterns
Date of Defense 0000-00-00
Page Count 114
Keyword
data mining
web mining
traversal sequential patterns
the filtration the filtration mechanism
Consecutive sequential patterns

Abstract
Consecutive sequential patterns exist in many applications. Air routes, ship routes, or websites browsing routes are kinds of consecutive sequential patterns. However, few studies have been done on consecutive sequential patterns. The TFA algorithm is the one of few algorithms which attempts to mine consecutive sequential patterns.
The purpose of this thesis is to improve the defects of the TFA algorithm. The TFA algorithm is one of the most efficient algorithms, but it still has two problems. First, the performance of TFA algorithm is getting worse when transaction length is getting longer. Second, it generates the consecutive sub-sequences by decompose duplicate transactions. Therefore performance of TFA algorithm can be improved.
First, we propose a new algorithm – CFP（Mining Consecutive Sequence Using Filtered and Pruning the mechanism）that is improved from TFA algorithm. Using Filtered and Pruning mechanism in earlier stage CFP algorithm accumulates the same shortened transactions of databases. CFP algorithm is very suitable to mine the sequences of databases whose record length is very long, and it will not generate the consecutive sub-sequences by decompose duplicate transactions. It only scans database four times and avoids generating any unnecessary candidate consecutive sub-sequence in mining process. The size of the databases in the real world is always greater than the size of the memory. In order to solve this problem, the CFP algorithm divides a large database into many sub-databases and mines rules from those sub-databases. The CFP algorithm only scans database four times and will not be affected by the length of frequent sub-sequences. The CFP algorithm avoids wasting a lot of I/O time and increases the efficiency and the practicability in application.
The CFP algorithm is one of the most efficient algorithms, but it still has one problem which is that CFP needs to inspect the same sub-sequence repeatedly in order to eliminate infrequent consecutive sub-sequences. Therefore performance of CFP algorithm can be improved. So we propose a new algorithm – CNP（Mining Consecutive Sequence Using New Filtered and Produce the mechanism）which uses a new filtering and producing mechanism to reduce frequency of inspection when eliminating infrequent consecutive sub-sequences.
Comprehensive experiments have been conducted to assess the performance of the proposed algorithm. The experimental results show that algorithms which we proposed outperform previous algorithms in the experiments.

Advisory Committee Jen-Peng Huang - advisor

Files etd-0407109-183709-681-001.pdf

indicate access worldwide
Date of Submission 2008-07-07