Abstract

Sequential Pattern Mining (SPM) is one of the key areas in Web Usage Mining (WUM) with broad applications such as analyzing customer behavior from weblog files. The current algorithms in this area can be classified into two broad areas, namely, apriori-based and pattern-growth based. Apriori-based algorithms for mining sequential patterns need to scan the database many times as they focus on candidate generation and test approach. A lot of research has been done so far, but even the best apriori-based algorithm for SPM in terms of number of database scans is SPADE that scans the database three times for discovering sequential patterns. Pattern growth-based algorithms avoid the candidate generation step and the best pattern growth algorithm known so far is Prefix Span that needs to scan the database at least twice. In this paper, a novel algorithm for SPM is proposed called FP-Split SPADE that reduced the database scan to only one by creating an FP-Split tree and applying SPADE algorithm on the tree instead of sequence database that greatly improved the efficiency of mining sequential patterns.
References

7. Han, J., Pei, J., Yin, Y. and Mao R. 2004. Mining Frequent Patterns without candidate generation: A Frequent pattern tree approach, Data Mining and Knowledge Discovery, Springer, 53-87.

Index Terms

Computer Science  Algorithms

Keywords
Sequential Pattern Mining, Web Mining, SPADE, Apriori, FP-Split tree