Abstract

Data mining refers to extracting or mining knowledge from large amounts of data. Among the various data mining tasks sequential pattern mining is one of the most important tasks. It has broad applications in several domains such as the analysis of customer purchase patterns, web access patterns, seismologic data, and weather observations. Sequential pattern mining consists of mining subsequences that appear frequently in a set of sequences. Sequential pattern mining was first introduced by Rakesh Agarwal and Ramakrishnan Srikant in 1995. Many novel approaches for sequential pattern mining were proposed like Apriori, AprioriALL, GSP, SPADE, SPAM and PrefixSpan. In this paper, the performance of state-of-the-art sequential pattern mining algorithms PrefixSpan and SPAM is evaluated. From the comprehensive experiments what have been done several phenomena were observed which are different from the traditional standpoint will be explained in this paper.
Performance Evaluation on State of the Art Sequential Pattern Mining Algorithms

- JAY AYRES, JOHANNES GEHRKE, TOMI YIU, JASON FLANNICK. Sequential pattern mining using a bitmap representation. In Proceedings of the 8th ACM SIGKDD, International Conference on Knowledge Discovery and Data Mining.
- ZHENGLU YANG, 2008. Fast Algorithms for Sequential Pattern Mining, pp 19, Section 2.2.4, Para 5, lines 1 to 4, Section 3.1.2, Figure 3.1 (a & b), pp 24. Figure 3.2 (a & b), pp 25.
- YANG TANG, FEIFEI LI, HONGYAN LI. Mining Scalable Pattern Based on Temporal Logic over Data Streams. 2012, 9th International conference on Fuzzy Systems and knowledge discovery (FSKD) 2012.

Index Terms

Computer Science
Algorithms

Keywords

Data Mining  Sequential Pattern Mining  PrefixSpan  SPAM