Abstract

During the last ten years, many algorithms have been proposed to mine sequential patterns. Some of them are based on the *Apriori* algorithm, developed to iteratively mine frequent itemsets, for example the *GSP* algorithm. Results obtained from experiments using these category of algorithms have shown that the candidate support count phase spends a huge part of the execution time. In this work, aiming at reducing the computational cost of multiple database scans and the computational effort to count the support of the candidate sequences, typical of iterative algorithms for the problem of mining sequential patterns, we propose the progressive reduction of the database during the execution of the algorithm. Therefore, fewer transactions are read at each iteration and the computational cost of counting the support of each candidate is reduced. Results obtained from evaluating different combinations of databases and minimum supports have shown that the database pruning techniques, adopted by the proposed algorithm *GSP2P*, significantly reduce the total execution time of the *GSP2* algorithm (implementation of *GSP*), which does not use pruning mechanisms. In this same work, aiming at validating the use of the proposed database pruning techniques and extending their applications, the techniques were applied to the problem of constraint-based sequential patterns mining. Results obtained from evaluating different combinations of databases and constraint selectivity values have shown that the database pruning techniques, adopted by the proposed algorithm *GSP2P-F*, significantly reduce the total execution time of the *GSP2-F* algorithm, which does not use pruning mechanisms.