Abstract

The need for real-time data mining has long been recognized in various application domains. Some of the applications like customer shopping sequences, medical treatments, natural disasters, telephone calling patterns, Weblog click streams, DNA sequences and gene structures require sequential pattern mining techniques. These techniques find the complete set
Recursive Prefix Suffix Pattern Detection Approach for Mining Sequential Patterns

of frequent subsequences for the given set of sequences and support threshold. Traditional pattern growth based approaches for sequential pattern mining derive length (n+1) pattern based on projected databases of length n-patterns recursively. As result lot of recursions occur which may lead to certain complexities. Thus, in order to reduce the number of iterations, an efficient bidirectional sequential pattern mining approach namely Recursive Prefix Suffix Pattern detection, RPSP algorithm is proposed. The RPSP algorithm first finds all Frequent Itemsets (FI's) according to the given minimum support and transforms the database such that each transaction is replaced by all the FI's it contains and then finds the patterns. The pattern further is detected based on ith projected databases, and constructs suffix and prefix databases based on the apriori property. RPSP will increase the number of frequent patterns by reducing the minimum support and vice versa. Recursion is terminated when the detected FI set of prefix or suffix projected database of parent database is null. All the patterns that correspond to a particular ith projected database of transformed database are formed into a set, which is disjoint from all other sets. The union of all the disjoint subsets is the resultant set of frequent patterns. The proposed algorithm was tested on the hypothetical data and results obtained were found satisfactory. Thus, RPSP algorithm can be applicable to many real world sequential data sets.

Reference

- R.Agarwal and R.Srikanth, “Mining Sequential Patterns” ICDE’95, Pg 3-14,1995.
Index Terms

Computer Science

Pattern Recognition

Key words

Sequential patterns

Frequent itemsets

Prefix

and Suffix databases

Projected database

Transformed database