Mining Frequent Itemsets without candidate Generation (FP Growth)

Appriori algorithm generates candidate itemsets & significantly reduces the size of candidate sets. leading to good performance. However it can suffer from two nontrivial costs.

1. It may still need to generate a huge number of candidate sets.
2. It may need to scan the whole database repeatedly & check large set of candidates by pattern matching.

FP-Growth, Frequent Pattern Growth adopts divide and conquer strategy to overcome this drawback First it compresses the database representing frequent items into a frequent pattern tree or FP tree, which retains the itemset association information. It then divides the compressed database into a set of conditional databases, each associated with one frequent item or pattern fragment. for each pattern fragment only its associated data set need to be examined. Therefore, this approach may substantially reduce the size of the data sets to be searched.