

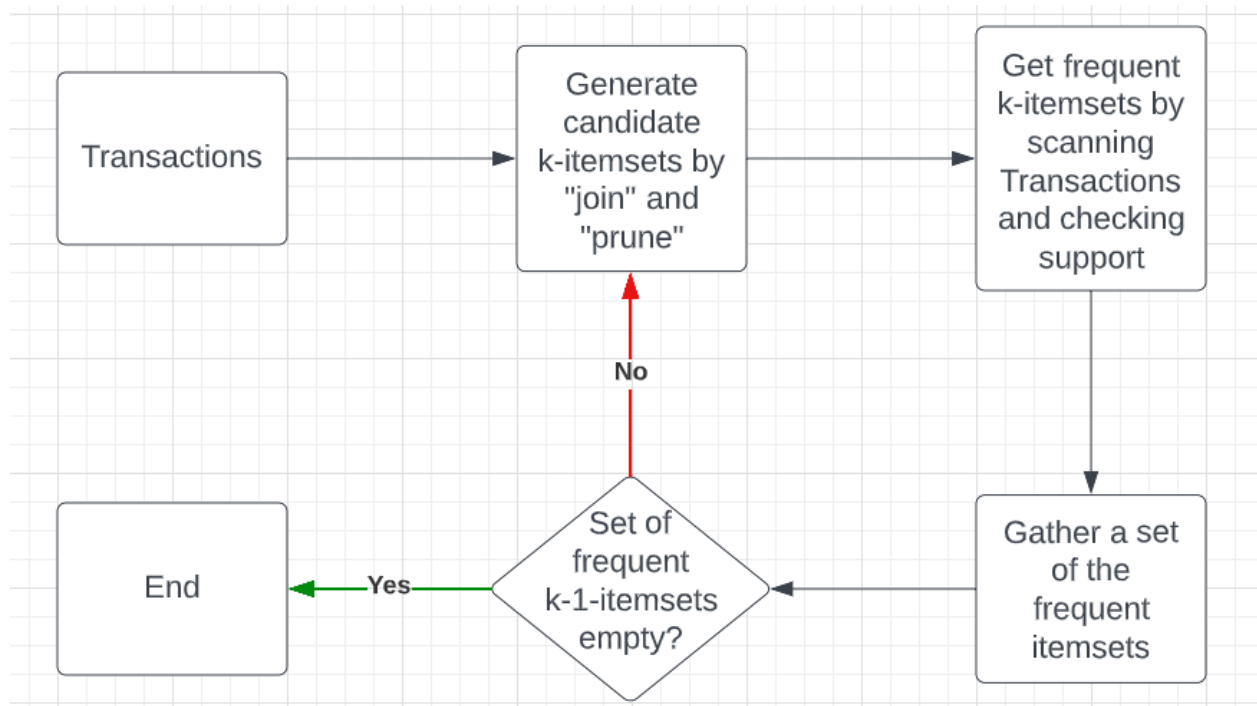
Apriori Algorithm

Yihan Yin (yy3114)

Background

Apriori Algorithm is designed for frequent item set mining and association rule learning over relational databases that contain transactions. Each transaction is seen as a set of items, that is, itemset. Given a threshold, the Apriori Algorithm is to identify frequent itemsets that are subsets of the transactions above the threshold.

Flow Chart



1. Scan the transactions to get the level 1 itemsets whose support is larger than the threshold.
2. Use "join" to generate a set of candidate level k itemsets and "prune" the unfrequented level k-itemsets by the idea that any subset of a frequent itemset must be a frequent itemset.
3. Scan the transactions again to filter those whose support is lower than the threshold.
4. Gather those frequent itemsets in one set.
5. If the set of frequent level k-1 itemsets is not empty, then go to 2 and generate the candidate level k itemsets, else the algorithm ends.

Goal

Since Apriori Algorithm applies a “bottom up” approach where frequent itemsets are extended one itemset at a time, it has a great potential in parallel improvement. The project aims to give implementations of both single thread and parallel Haskell programs and introduce comparisons between Java single thread version, Haskell single thread version and parallel Haskell version to possibly give a better illustration of functional programming and parallel processing.

References

https://en.wikipedia.org/wiki/Apriori_algorithm

<https://dwgeek.com/mining-frequent-itemsets-apriori-algorithm.html/>