Report for Assignment 2: Implementing Association Rule Mining (Apriori) Algorithm

Nawshad Farruque
Department of Computing Science,
University of Alberta

For this assignment I implemented Apriori algorithm mentioned at [1]. Additionally, In transaction look up and support count phase this algorithm can avoid looking up the transactions which do not contribute to frequent itemset generation and thus achieves scalability for large transaction size as inspired by *AprioriTid* mentioned also at [1].

Results:

Support scores were 0.1, 0.01, and 0.001 for 100, 1000, and 10000 transactions and confidence score 0.8 for all.**

Itemsets and Rules Counts:

Number of Transactions	Number of Frequent Itemsets	Number of Strong Association Rules
10000	40377	161856
1000	508	0
100	10	0

Execution Time:

Apriori (for sets of 100 transactions):

Transaction Sets	Time (in milliseconds)
100 (set 1)	0
100 (set 2)	0
100 (set 3)	0
Average	**0**

Apriori (for sets of 1000 transactions):

Transaction Sets	Time (in milliseconds)
1000 (set 1)	2576
1000 (set 2)	2561
1000 (set 3)	2474
Average	**2537**

Apriori (for sets of 10000 transactions):

Transaction Sets	Time
10000	**9 minutes**

*See Readme file for compilation instruction.
**All the fractional values are rounded to nearest integers.

References:

[1] Agrawal, Rakesh, and Ramakrishnan Srikant. "Fast algorithms for mining association rules." *Proc. 20th int. conf. very large data bases, VLDB*. Vol. 1215. 1994.