CS422-01: Homework #5

Vijay K. Gurbani

Sujad Patel

October 30, 2022

Part I
Exercises

1.1 Tan, Chapter 3 (Association Analysis)

Question 15

(a) Which data set(s) will produce the most number of frequent itemsets?

Data set (e) will produce the most number of frequent itemsets, since it contains the most frequently occurring items(dense columns).

(b) Which data set(s) will produce the fewest number of frequent itemsets?

Data set (d), because it contains the least frequently occurring items and the items are sparsely distributed in the columns and don't fulfill the minimum support.

(c) Which data set(s) will produce the longest frequent itemset?

Since Data set (e) has multiple overlapping frequent items, it will produce the longest frequent itemset.

(d) Which data set(s) will produce frequent itemsets with highest maximum support?

In Data set (b), around 100 has the most number of transactions, which is the highest support among the datasets.

(e) Which data set(s) will produce frequent itemsets containing items with wide-varying supportlevels (i.e., items with mixed support, ranging from less than 20% to more than 70%)?

Data set (e) has the most columns with varying length.

1.2 Zaki, Chapter 8(Frequent Pattern Mining)

Question 1

(a) Usin g minsup = 3/8, show how the Apriori algorithm enumerates all frequentpatterns from this dataset.

1-itemset Frequency

Item	Count
A	5
B	4
C	5
D	6
E	1
F	4
G	5

Candidate E is the only one that is lower than min sup limit, therefore it is ignored for rest of analysis.

2-itemset Frequency that is greater than or equal to min sup

Itemset	Count
A,B	3
A,C	3
A,D	4
C,D	4
C,G	3
D,F	4
D,G	3

Itemset (A,C,D) is the most frequent pattern in the dataset.

Question 4

Show all rules that one can generate from the set ABE

1-item Frequency for A, B, and E

Item	Count/Support
A	4
B	5
E	4

2-itemset Frequency

Itemset	Support
A,B	3
A,E	2
B,E	4

3-itemset Frequency

Itemset	Support
A,B,E	2

$$Confidence(A \longrightarrow B) = \frac{Support(AB)}{Support(B)}$$

Rules for set *ABE*.

Rules	Confidence
A->B,E	50%
B->A,E	100%
E->A,B	66%
A,B->E	50%
A,E->B	40%
B,E->A	50%