Question 1: Tan Chapter 5

Question 15: Answer the following questions using the data sets shown in Figure 5.6. Note that each data set contains 1000 items and 10,000 transactions. Dark cells indicate the presence of items and white cells indicate the absence of items. We will apply the Apriori algorithm to extract frequent itemsets with minsup = 10% (i.e., itemsets must be contained in at least 1000 transactions)?

1. Which data set(s) will produce the most number of frequent itemsets?
    a. Answer: Data set (e) because it has to generate the longest frequent itemset along with its subsets
2. Which data set(s) will produce the fewest number of frequent itemsets?
    a. Answer: Data set (d) which does not produce any frequent itemsets at 10% support threshold.
3. Which data set(s) will produce the longest frequent itemset?
    a. Answer: Data set (e).
4. Which data set(s) will produce frequent itemsets with highest maximum support?
    a. Answer: Data set (b).
5. Which data set(s) will produce frequent itemsets containing items with wide-varying support levels (i.e., items with mixed support, ranging from less than 20% to more than 70%).
    a. Answer: Data set (e).

Question 2: Zaki Chapter 8

Question 1(a): Given the database in Table 8.2.

Table 8.2. Transaction database for Q1

tid	itemset
t_1	$ABCD$
t_2	$ACDF$
t_3	$ACDEG$
t_4	$ABDF$
t_5	BCG
t_6	DFG
t_7	ABG
t_8	$CDFG$

(a) Using minsup = 3/8, show how the Apriori algorithm enumerates all frequent patterns from this dataset

Solution: The min-sup count is 3. Since the minsup is the minimum support count / total transactions. Do the following to find how the apriori algorithm works to enumerate all frequent patterns from this dataset. Since L3 has set {ACD} with support count 3, {ACD} is the frequent item set.

- Create 1-itemset,C1 with support count.
- Scan the database and determine the frequent 1-items set ,L1.
- Prune the 1-itemsets having support count less than minimum support count, from L1.
- Join L1 with itself to get 2-items set, C2
- Then scan the database again to determine the frequent 2-items set, L2.
- Prune the 2-itemsets having support count less than minimum support count, from L2.
- Join L2 with itself to get 3-items set, C3
- Then scan the database to determine the frequent 3-itemsets, L3
- Prune the 3-itemsets having support count less than minimum support count, from L3.
- Further 4-itemsets are not possible, since L3 has only one set.

Question 4: Given the database in Table 8.4. Show all rules that one can generate from the set ABE

Table 8.4. Dataset for Q4

tid	itemset
t_1	ACD
t_2	BCE
t_3	$ABCE$
t_4	BDE
t_5	$ABCE$
t_6	$ABCD$

Solution:

L_1

Frequent set	Support-count
{A}	4
{B}	5
{E}	4

L_2

Frequent Set	SC
{A,B}	3
{A,E}	2
{B,E}	4

L_3

FS	SC
{A,B,E}	2

{A,B,E} Rules

{A} ⇒ {B,E} ; confidence = 2/4 = 50%.

{B} ⇒ {A,E} ; Confidence = 2/5 = 40%.

{E} ⇒ {A,B} ; Confidence = 2/4 = 50%.

{A,B} ⇒ {E} ; Confidence = 2/3 = 66%.

{A,E} ⇒ {B} ; Confidence = 2/2 = 100%.

{B,E} ⇒ {A} ; Confidence = 2/4 = 50%.