Assignment - 5

1702001070044

Q-1 State the apriori property. Generate candidate item sets, frequent itemsets and strong association rules using apriori algorithm on the following data set with minimum support = 60% and minimum confidence = 80%.

Tid	Items Bought
T_{100}	{M, O, N, K, E, Y}
T_{200}	{D, O, N, K, E, Y}
T_{300}	{M, A, K, E}
T_{400}	{M, U, K, C, Y}
T_{500}	{C, O, O, K, I, E}

Min support = 60%

$= \frac{60 \times 5}{100}$

$= 3$

C_1 ↓

Itemset	Supp. Count
A	1
C	2
D	1
E	4
I	1
K	5
M	3
N	2
O	3
U	1
Y	3

→ Comparing candidat S.C. with min. sup. count

L_1 ↓

Itemset	Supp. count
E	4
K	5
M	3
O	3
Y	3

vision

170200107044

C_2 ↓

Itemset	Supp. Count
E, K	4
E, M	2
E, O	3
E, Y	2
K, M	3
K, O	3
K, Y	3
M, O	1
M, Y	2
O, Y	2

L_2 ↓

Itemset	Supp. Count
E, K	4
E, O	3
K, M	3
K, O	3
K, Y	3

C_3 ↓

Itemset	Supp. Count
EKO	3

L_3 ↓

Itemset	Supp. count
EKO	3

C_4 ↓

—

L_4 ↓

—

=> From EKO, $P(A \cup B) / P(A) = F_i$

$E \rightarrow K \wedge O = 3/4 = 75\%$ $\quad E \wedge K \rightarrow O = 3/4 = 75\%$

$K \rightarrow E \wedge O = 3/5 = 60\%$ $\quad E \wedge O \rightarrow K = 3/3 = 100\%$

$O \rightarrow E \wedge K = 3/3 = 100\%$ $\quad K \wedge O \rightarrow E = 3/3 = 100\%$

☆ Strong association rule : $O \rightarrow E \wedge K$

$E \wedge O \rightarrow K$

$K \wedge O \rightarrow E$

vision

1702001070 44

Q-2 Apply apriori algorithm to generate frequent itemsets (min. sup. ≥ 33.3%) for the following transaction database.

T_{id}	Item
T_1	{A, B, D, K}
T_2	{A, B, C, D, E}
T_3	{A, B, C, E}
T_4	{B, D}
T_5	{A, C}
T_6	{A, D}

$$\text{min. sup.} = \frac{33.3 \times 6}{100} = 1.9998$$

C_1 ↓

Item	S.C
A	4
B	5
C	3
D	4
E	2
K	1

→

L_1 ↓

Item	S.C
A	4
B	5
C	3
D	4
E	2

→

C_2 ↓

Item	S.C
AB	3
AC	3
AD	2
AE	2
BC	2
BD	4
BE	2
CD	1
CE	2
DE	1

→

L_2 ↓

Item	S.C
AB	3
AC	3
AD	2
AE	2
BC	2
BD	4
BE	2
CE	2

170200107044

C_3 ↓

Item	S.C.
ABC	2
ABD	2
ABE	2
ACE	2
BCE	2

→

L_3 ↓

Item	S.C.
ABC	2
ABD	2
ABE	2
ACE	2
BCE	2

C_4 ↓

Item	S.C.
ABCE	2

L_4 ↓

Item	S.C.
ABCE	2

C_5 ↓

—

L_5 ↓

—

=> Here, ABCE is biggest frequent set.

170200107044

Q-3 Generate frequent itemsets and generate association rules based on it using apriori algorithm. minimum support is 50% and minimum confidence is 70%.

T_{ID}	Items
100	1, 3, 4
200	2, 3, 5
300	1, 2, 3, 5
400	2, 5

min. sup. = $\frac{50 \times 4}{100} = 2$

C_1 →

Item	S.C
1	2
2	3
3	3
4	1
5	3

→ L_1 →

Item	S.C.
1	2
2	3
3	3
5	3

C_2 →

Item	S.C.
1 2	1
1 3	2
1 5	1
2 3	2
2 5	3
3 5	2

→ L_2 →

Item	S.C.
1 3	2
2 3	2
2 5	3
3 5	2

C_3 →

Item	S.C.
2 3 5	2

→ L_3 →

Item	S.C.
2 3 5	2

C_4 → — → L_4 → —

⇒ from largest itemset 2 3 5 conf = P(A∪B)/P(A)

$2 \rightarrow 3 \wedge 5 = 2/3 = 66.67\%$ $\quad 2 \wedge 3 \rightarrow 5 = 2/2 = 100\%$

$3 \rightarrow 2 \wedge 5 = 2/3 = 66.67\%$ $\quad 2 \wedge 5 \rightarrow 3 = 2/3 = 66.67\%$

$5 \rightarrow 2 \wedge 3 = 2/3 = 66.67\%$ $\quad 3 \wedge 5 \rightarrow 2 = 2/2 = 100\%$

Strong assoc. rule : $2 \wedge 3 \rightarrow 5 = 100\%$

: $3 \wedge 5 \rightarrow 2 = 100\%$

VISION

1702001070444

Q-4 Solve apriori algorithm on the following dataset with min. sup. value and min. conf. value set as 50% and 75% respectively to generate large itemset and association rules.

T_{ID}	Items purchased
T_{101}	Cheese, Milk, Cookies
T_{102}	Butter, Milk, Bread
T_{103}	Cheese, Butter, Milk, Bread
T_{104}	Butter, Bread

Min. sup. $= \frac{50 \times 4}{100} = 2$

$C_1 \downarrow$

Itemset	Sup. Count
A	1
C	2
D	1
E	4
I	1
K	5
M	3
N	2
O	3
U	1
Y	3

$L_1 \downarrow$

Itemset	Sup. Count
E	4
K	5
M	3
O	3
Y	3

$C_1 \downarrow$

Item	S.C.
Bread	3
Butter	3
Cheese	2
Cookies	1
Milk	3

$L_1 \downarrow$

Item	S.C.
Bread	3
Butter	3
Cheese	2
Milk	3

$C_2 \downarrow$

{Bread, Butter}	3
{Bread, Cheese}	1
{Bread, Milk}	2
{Butter, Cheese}	1
{Butter, Milk}	2
{Cheese, Milk}	2

170200107044

L_2 ↓

Item	S.C.
{Bread, Butter}	3
{Bread, Milk}	2
{Butter, Milk}	2
{Cheese, Milk}	2

C_2 ↓

Item	S.C.
{Bread, Butter}	3
{Bread	

L_3 ↓

Item	S.C.
{Bread, Butter}	3
{Bread, Milk}	2
{Butter, Milk}	2
{Cheese, Milk}	2

C_3 ↓

Item	S.C.
{Brea	

L_3 ↓

Item	S.C.
{Bread, Butter, Milk}	2

C_4 ↓ —

L_4 ↓ —

⇒ Largest itemset {Bread, Butter, Milk} for this assc. rule.

Bread → Butter ^ Milk = 2/3 = 66.67 %
Butter → Bread ^ Milk = 2/3 = 66.67 %
Milk → Butter ^ Bread = 2/3 = 66.67 %
Bread ^ Butter → Milk = 2/3 = 66.67 %
Butter ^ Milk → Bread = 2/2 = 100 %
Bread ^ Milk → Butter = 2/2 = 100 %

170200107044

Q-5 What is "Market Basket Analysis"? Explain the two measures of rule interestingness: support and confidence with suitable example.

→ *Market Basket Analysis:

- It is an example of association rule mining.
- This process analyzes customer buying habits by finding associations between items that customers place in their "shopping carts".
- The discovery of such associations help retailers develop marketing strategies by gaining insight into which items are frequentally bought together.
- Such information can lead to increased sales by helping retailers do selective marketing and plan their self space.
- For example, placing milk and bread within close, they may encourage the sale of these items together within single visits to the store.
- It helps to plan marketing or advertising strategies or in the design of a new catalog.
- It is also used to help you design different store layouts.
- Market basket analysis can also be helpful to retailers plan which items to put on sale at reduced prices.

VISION

17o2oo1o7o44

* Support :

=> Percentage of transactions in D, that contain both x and y.

=> Support $(x \Rightarrow y) = p(x \cup y) / \text{total tuples}$

$= (\#\text{ tuple having } x \& y) / (\#\text{ total tuples})$

* Confidence :

=> Percentage of transactions in D, containing x that also contain y.

=> Confidence $(X \Rightarrow Y) = P(X \cup Y) / P(X)$

$= (\#\text{ tuples contain both } x \& y) / \#\text{ tuples having } x$

Ex:

T_{ID}	Itemset
1	Bread, Milk
2	Bread, Diapers, Beer, Eggs
3	Milk, Diapers, Beer, Coke
4	Bread, Milk, Diapers, Beer
5	Bread, Milk, Diapers, Coke

{Milk, Diapers} => Beer (Suppose this is association rule)

$$\text{Support } (S) = \frac{\sigma\{\text{Milk, Diapers, Beer}\}}{\sigma|T| \text{ (total tuple)}}$$

$$= \frac{2}{5} = 0.40 = 40\%$$

vision

170200107044

Confidence (C) = $\frac{\text{C(Milk, Diapers, Beers)}}{\text{C(Milk, Diapers)}}$

$= \frac{2}{3}$

$= 0.67 = 67\%$