

Apriori Algorithm

Q1) Consider the foll transaction database with minimum support 50%. and minimum confidence 66%. Find the frequent patterns and the strong association rules.

Tid	Items
10	A, C, D
20	B, C, E
30	A, B, C, E
40	B, E

$$\Rightarrow \text{minimum support} = \frac{50}{100} \times 4$$

$$= 0.5 \times 4$$

$$= 2$$

Step 1: counting each itemset we get

Itemset	Count
A	2
B	3
C	3
D	1
E	3

Step 2: D itemset does not meet min support = 2, so after discarding it, we get

Itemset	Count
A	2
B	3
C	3
E	3

Step 3: Join step

Itemset	Count
A, B	1
A, G	2
A, E	1
B, C	2
B, E	3
C, E	2

Step 4: $\{A, B\}$ & $\{A, E\}$ doesn't meet min support, so after discarding, we get

Itemset	Count
A, C	2
B, C	2
B, E	3
C, E	2

Step 5: Join step

Itemset	Count
A, B, C	1
A, C, G	1
A, B, G	1
B, C, E	2

Step 6: $\{A, B, C\}$, $\{A, C, E\}$, $\{A, B, E\}$ doesn't meet min support, so after discarding, we get

Itemset	Count
B, C, E	2

Thus $\{B, C, E\}$ is frequent

If $x \rightarrow y$ then $\text{sup}(x \rightarrow y)$

$\text{sup}_{\text{frequent}}$

(line)

Step 7: Generate association rules

i) $\{B, C\} \rightarrow \{C\}$

confidence = $\frac{\text{support } \{B, C\}}{\text{support } \{B, C\}} \times 100$

$$= \frac{2}{2} \times 100$$

$$= 100\%$$

Strong Association rule (confidence $> 66\%$)

ii) $\{C, E\} \rightarrow \{B\}$

confidence = $\frac{\text{support } \{B, C, E\}}{\text{support } \{C, E\}} \times 100$

$$= \frac{2}{2} \times 100 = 100\%$$

Strong association rule (confidence $> 66\%$)

iii) $\{B, E\} \rightarrow \{C\}$

confidence = $\frac{2}{3} \times 100 = 66.67\%$

Strong association rule (confidence $> 66\%$)

iv) $\{B\} \rightarrow \{C, E\}$

confidence = $\frac{\text{support } \{B, C, E\}}{\text{support } B} \times 100$

$$= \frac{2}{3} \times 100 = 66.67\%$$

Strong association rule

vi)

$$\{E\} \rightarrow \{B, C\}$$

confidence = $\frac{2}{3} \times 100$

$$= 66.67\%$$

Strong association rule

v) $\{C\} \rightarrow \{B, E\}$

confidence = $\frac{2}{3} \times 100 = 66.67\%$

Strong association rule

(Q2) A database has five transactions. Let min support = 60%. and min confidence = 80%. Find all frequent item sets by using apriori algorithm. T_id is the transaction ID

T_id	Items brought
T1000	M, O, N, K, E, Y
T1001	D, O, N, K, G, Y
T1002	M, A, K, E
T1003	M, U, C, K, Y
T1004	C, O, U, K, E

$$\Rightarrow \text{minimum support} = \frac{60}{100} \times 5 \\ = 3$$

Step 1:

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

Step 2:

Itemset	Count
M	3
O	3

K	5
E	4
Y	3

Step 3 :

Itemset	Count
{M, OY}	1
{M, KY}	3
{M, EY}	2
{M, Y}	2
{O, KY}	3
{O, EY}	3
{O, Y}	2
{K, EY}	4
{K, Y}	3
{E, Y}	2

Step 4 :

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

Step 5 :

Itemset	Count
M, O, K	1
M, E, Y	1
M, K, Y	2
M, K, E	2
M, O, Y	1
M, O, E	1

O, K, E	3
O, K, Y	2
O, E, Y	2
K, E, Y	2

Step 6:

Itemset	Count
O, K, E	3

Step 7:

$$\textcircled{1} \{O, K\} \rightarrow \{E\}$$

Confidence = $\frac{\text{support } \{O, K, E\}_N}{\text{support } \{O, K\}} \times 100 = \frac{3}{3} \times 100 = 100\%$.

Strong association rule.

$$\textcircled{2} \{O, E\} \rightarrow \{K\}$$

Confidence = $\frac{3}{3} \times 100 = 100\%$.

Strong association rule.

$$\textcircled{3} \{K, E\} \rightarrow \{O\}$$

Confidence = $\frac{3}{4} \times 100 = 75\%$.

not a Strong association rule. (less than 80%).

$$\textcircled{4} \{O\} \rightarrow \{K, E\}$$

Confidence = $\frac{\text{support of } \{O, K, E\}_N \times 100}{\text{support of } \{O\}}$

$$\frac{2}{3} \times 100 = 100\%$$

Strong association rule

⑤ $\{k\} \rightarrow \{0, E\}$

$$\text{Confidence} = \frac{3}{5} = 60\%$$

not a strong association rule. (less than 80%)

⑥ $\{E\} \rightarrow \{0, k\}$

$$\text{Confidence} = \frac{3}{4} \times 100 = 75\%$$

not a strong association rule. (less than 80%)

$$d \times 88.88 = \text{trigger minimum}$$

$$d \approx 0.001 =$$

: 1 opt + 2

time)

P

S

S

E

NE

format

asb to H

mid

quintet

e(a)

quad

: 2 opt + 2

time)

P

S

S

format

asb to H

mid

Q3) Apply apriori algorithm on the following dataset to find strong association rules. minimum support threshold ($s = 33.33\%$) and minimum confident threshold ($c = 60\%$).

Transaction ID	Items
T1	Hot Dogs, Buns, Ketchup
T2	Hot Dogs, Buns
T3	Hot dogs, Coke, Chips
T4	Coke, Chips
T5	Chips, Ketchup
T6	Hot dogs, Coke, Chips

$$\Rightarrow \text{minimum support} = \frac{33.33}{100} \times 6 \\ = 1.999 \approx 2$$

Step 1 :

Itemset	Count
Hot dogs	4
Buns	2
Ketchup	2
Coke	3
Chips	4

Step 2 :

Itemset	Count
Hot dogs	4
Buns	2
Ketchup	2
Coke	3
Chips	4

Hotdogs, buns, coke, chips

Page No.

Date

Step 3 :

Itemset ← Count

Hot dogs, Buns	2
Hot dogs, ketchup	1
Hot dogs, Coke	2
Hot dogs, chips	2
Buns, Ketchup	1
Buns, Coke	0
Buns, chips	0
ketchup, Coke	0
ketchup, chips	1
Coke, chips	3

Step 4 :

Itemset ← Product count

Hot dogs, Buns	2
Hot dogs, Coke	2
Hot dogs, chips	2
Coke, chips	3

Step 5 :

Itemset ← Count

Hot dogs, buns, coke	0
Hot dogs, coke, chips	2
Hot dogs, buns, chips	0
buns, coke, chips	0

Step 6 :

Itemset ← Count

Hot dogs, coke, chips	2
-----------------------	---

Step 7 :

(1) $\{ \text{hotdogs, Coke} \} \rightarrow \{ \text{chips} \}$

$$\text{confidence} = \frac{\text{support}(\text{hotdogs, buns, chips})}{\text{support}(\text{hotdogs, buns})} \times 100$$

$$= \frac{2}{2} \times 100 = 100\%$$

Strong association rule

(2) $\{ \text{hotdogs, chips} \} \rightarrow \{ \text{Coke} \}$

$$\text{confidence} = \frac{2}{2} \times 100 = 100\%$$

Strong association rule

(3) $\{ \text{Coke, chips} \} \rightarrow \{ \text{hotdogs} \}$

$$\text{confidence} = \frac{2}{3} \times 100 = 66.67\%$$

Strong association rule

(4) $\{ \text{hotdogs} \} \rightarrow \{ \text{Coke, chips} \}$

$$\text{confidence} = \frac{2}{4} \times 100 = 50\%$$

not a strong association rule (less than 60%)

(5) $\{ \text{Coke} \} \rightarrow \{ \text{hotdogs, chips} \}$

$$\text{confidence} = \frac{2}{3} \times 100 = 66.67\%$$

Strong association rule

(6) $\{ \text{chips} \} \rightarrow \{ \text{hotdogs, Coke} \}$

$$\text{confidence} = \frac{2}{4} \times 100 = 50\%$$

not a strong association rule (less than 50%)