

Frequent item set.

Problem solved with Algorithms

TID	item	Frequency
1	E, K, M, N, O, Y	A 1
2	D, E, K, O, Y	C 2
3	A, E, K, M	P 1
4	C, K, M, U, N, Y	E 4
5	C, E, I, K, O, O	I 1
	K	K 5
	M	M 3
	N	N 2
	O	O 4
	U	U 1
	Y	Y 3

Now take support frequency set = 3

For now

→ ~~remove~~ make set of equal to 3 or greater than 3

$$L = \{E=4, K=5, M=3, O=4, Y=3\}$$

↓
Frequent pattern result (FPS)

Now make it decending order

$$L = \{K=5, E=4, O=4, M=3, Y=3\}$$

Map this set on table 1

The value which are not present in set, Simply drop them and write if in order of sets order

- 1
- 2
- 3
- 4
- 5

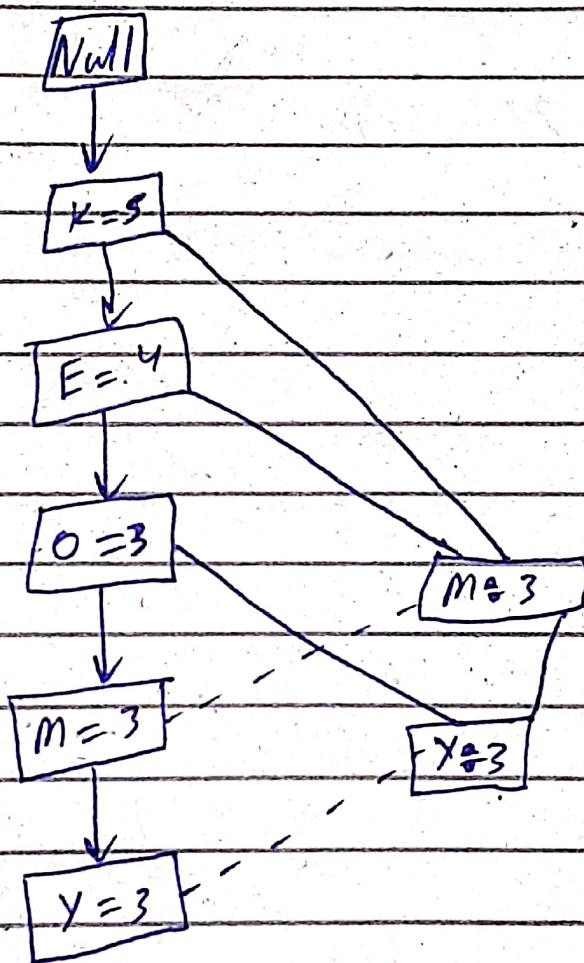
K, E, O, M, ~~S~~, X

K, E, O, Y

K, E, M

K, M, Y

K, E, O



T₁ E I L M N O X

T₂ D E K N O Y

T₃ A E K M

T₄ C I L M U Y

T₅ S C E T K O O

Frequency table min support value = 3

A 1

C 2

D 1

E 4

I 1

K 5

M 3

N 2

O 4

U 1

Y 3

Update table



convert to Data Structure (tree)

T₁ | K, E, O, M, Y

T₂ | K, E, O, X

T₃ | K, E, M

T₄ | K, M, X

T₅ | K, E, O

[Null]

K: S

E: 4

O: 3

M: 1

Y: 1

using L set:

Next
Second page

m: 1

m: 1

y: 1

y: 1

→ all possible paths to reach this list -

Y	$\{K, E, O, M : 1\}$, $\{K, E, O : 1\}$, $\{K, M : 1\}$	^{Common set} K:3	$\langle K, Y : 3 \rangle$
M	$\{K : 1\}$, $\{K, E : 1\}$, $\{K, E, O : 1\}$	K:3	$\langle K, M : 3 \rangle$
O	$\{K, E : 3\}$	$K, E : 3$	$\langle K, O : 3 \rangle$ $\langle E, O : 3 \rangle$
E	$\{K : 4\}$	K:4	$\langle K, E : 4 \rangle$
K	∅		

Frequency Table

$T_1 : M, E, B, BU$

$M: 9$

$T_2 : M, BN, E, LC$

$B: 10$

$BU: 10$

$T_3 : B, BU, LC$

$E: 3$

$LC: 3$

$T_4 : M, B, BV$

$C: 5$

$T_5 : B, BV, C$

$$\text{support} = \frac{F(x, y)}{N} = 30\%$$

$T_6 : M, B, BV, C$

$$\text{confidence} = \frac{F(x, y)}{F(x)} = 60\%$$

$T_7 : M, C$



	2-item set	Frequency	
$T_8 : M, B, BV$	M, B	?	$7/12 = ? > 30\%$
	M, BV	?	$7/12 = ? < 30\% \times$
$T_9 : B, BV, E, C$	M, C	3	$3/12 = ? \times$
	B, BV	9	$9/12 = ? \checkmark$
$T_{10} : M, BV, B$	B, C	4	$4/12 = ? \checkmark$
	B, C	3	$3/12 = ? \times$

$T_{11} : MB, BV$

$T_{12} : M, B, C, V$

$$L = \{MB: 7, MBV: 3, BBV: 9, B, C: 4\}$$

	3-item set	Frequency	
	M, B, BV	6	
	M, B, C	2	\times
	B, BV, C	3	\times

$$L = \{M, B, BV\}$$

$$L = \{M, B, BU\}$$



Subsets

$$\{\{M\}, \{B\}, \{BU\}, \{M, B\}, \{M, BU\}, \{B, BU\}\} = \boxed{T}$$

Association Rule:

$$\textcircled{1} \quad S \rightarrow (I-S)$$

$$S(B, BU) = 9/12 = 75 \quad \text{valid}$$

These are above 60%.

$$C(B, BU) = 9/10 = 90 \quad \text{valid}$$

$$C(BU, B) = 9/10 = 90\% \quad \text{valid} \quad \text{Do it}$$

$$S(M, BU) = 7/12 = 58\% \quad \text{valid}$$

$$C(M, BU) = 7/9 = 77\% \quad \text{valid}$$

$$C(BU, M) = 7/10 = 70\% \quad \text{valid}$$

$$S(M, B) = 7/12 = 58 \quad \text{valid}$$

$$C(M, B) = 7/9 = 77\% \quad \text{valid}$$

$$C(B, M) = 7/10 = 70\% \quad \text{valid}$$

$$\textcircled{1} \quad S \rightarrow (I-S)$$

$$\{Milk\} \rightarrow \{Milk, Bread, Butter\} - \{Milk\}$$

$$\{Milk\} \rightarrow \{Bread, Butter\}$$

$$\textcircled{2} \quad \text{if } \frac{\text{support}(I)}{\text{support}(S)} > \min \text{ support} \quad \tau = \min \text{ confidence}$$

$$\text{Support} = \frac{F(x, y, z)}{N}$$

$$\frac{6/12}{9/12} \div \frac{6}{9} = 0.66$$

(2) $B \rightarrow \{M, B, Bu\} - \{B\}$

$$B \rightarrow \{M, Bu\} = 58\%$$

(2) $\frac{\{M, B, Bu\}}{S(B)} = \frac{6/12}{10/12} = 0.6$

(3) $Bu \rightarrow \{M, B, Bu\} - \{Bu\}$

$$Bu \rightarrow \{M, B\} = \cancel{58\%}$$

~~$\frac{6/12}{10/12} = 0.6$~~

(4) $\{M, B\} \rightarrow \{M, B, Bu\} - \{M, B\}$

$$\{M, B\} \rightarrow \{Bu\}$$

$$\frac{6/12}{7/12} = 85\%$$

(4) $\{M, BV\}$

$$\frac{6/12}{7/12} = 85\%$$

(5) $\{B, BV\}$

$$\frac{6/12}{9/12} = 66\%$$

Example

(1)

T, F, A, C, D, G, T, M, P

T₂ A, C, F, I, M, O

T₃ B, E, H, J, D

T₄ B, C, K, S, P,

Frequency Table

T₅ A, F, C, E, I, P, M, N

L = { A 3 } \rightarrow L-set

B 3

C 4

Support 9

F 4

I 4

M 3

P 3 3

Pick

2-item Set Table

C 4 2-item CF : 3 3-item
F 4 \rightarrow CI : 3 \rightarrow C, FI : 3
I 4 FI : 3

$S \rightarrow \{\{C\}, \{F\}, \{I\}, \{CF\}, \{CI\}, \{FI\}\}$

① $S \rightarrow (I - S)$

$\hookrightarrow \{CFI\} - \{C\}$

$\hookrightarrow \{FI\}$

② $\frac{S(CFI)}{S(C)} = \frac{3/S}{4/S} = 3/4$

Big data

T₁ 1, 2, 5

Frequency table

T₂ 2, 4

Min support = 2

$$\text{confidence} = \frac{2}{9} \text{ or } 2/9 = 22\%$$

T₃ 2, 3

2 : 7

3 : 6

T₄ 1, 2, 4

4 : 2

5 : 2

T₅ 1, 3

2-item set | Frequency

T₆ 2, 3

1, 2

4

1, 3

4

T₇ 1, 3

1, 4

1 x

1, 5

2

T₈ 1, 2, 3, 5

2, 3

4

2, 4

2

T₉ 1, 2, 3

2, 5

2

3, 4

0 x

3, 5

1 x

4, 5

0 x

3-item set | Frequency

1, 2	4	1, 2, 3	2
------	---	---------	---

1, 3	4	1, 2, 3	2
------	---	---------	---

1, 5	2 →	1, 2, 4	1 x
------	-----	---------	-----

2, 3	4	2, 3, 4	0 x
------	---	---------	-----

2, 4	2	1, 3, 5	1 x
------	---	---------	-----

2, 5	2	1, 3, 4	0 x
------	---	---------	-----

2, 3, 5		1, 3, 4	1 x
---------	--	---------	-----

3, 4, 5		0 x	→
---------	--	-----	---

4-item set

↑ benefit

1, 2, 3 → set.

Freq

1, 2, 3, 5 → 1 X

1, 2, 5



continue with 3 item set:

{1, 2, 3} / 5 → {1, 2, 5}

Make subset

→ S³

{1}, {2}, {3}, {1, 2}, {1, 3}, {2, 3}, ~~{1, 2, 3}~~

① S → {I - S}

→ remain constant.

② Confidence = $\frac{\text{Support}(I)}{\text{Support}(S)}$

i³ → i

→ {1} → {(1, 2, 3) - (1)}

{1} → {2, 3}

Support = 6/9 = 66%

Confidence = $\frac{2/9}{6/9} = 33\%$

→ {2} → {(1, 2, 3) - (2)}

{2} → {1, 3}

Support = 7/9 = 77%

Confidence = $\frac{2/9}{7/9} = 28\%$

$$\rightarrow \{3\} \rightarrow \{(1, 2, 3) - (3)\}$$

$$\{3\} \rightarrow \{1, 2\}$$

$$\text{Support} = 6/9 = 66\%$$

$$\text{confidence} = \frac{2/9}{6/9} = 33\%$$

$$\rightarrow \{1, 2\} \rightarrow \{(1, 2, 3) - (1, 2)\}$$

$$\{1, 2\} \rightarrow \{3\}$$

$$\text{Support} = 4/9 = 44\%$$

$$\text{confidence} = \frac{2/9}{4/9} = 50\%$$

$$\rightarrow \{2, 3\} = \{1, 2, 3\} - \{(2, 3)\} =$$

$$\{2, 3\} \rightarrow \{1\}$$

$$\text{Support} = 4/9 = 44\%$$

$$\text{confidence} = \frac{2/9}{4/9} = 50\%$$

$$\{1, 2, 5\}$$

$$\{1\} \rightarrow \{1, 2, 5\} - \{1\} =$$

$$\{1\} \rightarrow \{2, 5\}$$

$$\text{support} = \frac{6}{9} = 66\%$$

$$\text{confidence} = \frac{2/9}{6/9} = 33\%$$

$$\{2\} \rightarrow \{1, 2, 5\} - \{2\}$$
$$\{2\} \rightarrow \{1, 5\}$$

$$\text{support} = \frac{7}{9} = 77\%$$

$$\text{confidence} = \frac{2/9}{7/9} = 28\%$$

$$\{5\} \rightarrow \{1, 2, 5\} - \{5\}$$
$$\{5\} \rightarrow \{1, 2\}$$

$$\text{support} = \frac{2}{9} = 22\%$$

$$\text{confidence} = \frac{2/9}{2/9} = 100\%$$

$$\{1, 2\} \rightarrow \{1, 2, 5\} - \{1, 2\} =$$
$$\{1, 2\} \Rightarrow \{5\}$$

$$\text{Support } \frac{4}{9} = 44\%$$

$$\text{confidence} = \frac{2/9}{4/9} = 50\%$$