



8 Data Mining

- Exercise 2: Multi MinSup

- $M = \{3, 2, 5, 4, 6, 1\}$

- Read transactions:

Item	Count	SUP %	MIS %
1	8	80	70
2	2	20	17
3	2	20	15
4	2	20	30
5	3	30	30
6	6	60	35

- $L = \{3, 2, 5, 4, 6, 1\}$

- $F_1 = \{3, 2, 5, 6, 1\}$

Transactions	Item	MIS %
1, 4, 6	1	70
1	2	17
1, 5, 6	3	15
1, 6	4	30
4, 6	5	30
1, 2, 3, 5	6	35
1, 2, 3, 5		
6		
1		
1, 6		

F	Item	SUP %	MIS %
F1	3	20	15
	2	20	17
	5	30	30
	6	60	35
	1	80	70



8 Data Mining

– $L = \{3, 2, 5, 4, 6, 1\}$

– Candidate gen., $K=2$

- $\{3, 2\}$: $\text{sup}(2) = 20\%$
 $20\% > \text{MIS}(3) = 15$ and
 $|\text{sup}(3) - \text{sup}(2)| = |20 - 20| = 0 < \varphi = 20\%$
so $\{3, 2\}$ is a good candidate
- $\{3, 5\}$: is a good candidate
- $\{3, 4\}$: is a good candidate
- $\{3, 6\}$: is NOT a good candidate ($> \varphi$)
- $\{3, 1\}$: is NOT a good candidate ($> \varphi$)

Item	Count	SUP %	MIS %	Transactions
1	8	80	70	1, 4, 6
2	2	20	17	1
3	2	20	15	1, 5, 6
4	2	20	30	1, 6
5	3	30	30	4, 6
6	6	60	35	1, 2, 3, 5
				1, 2, 3, 5
				6
				1
				1, 6

$\varphi = 20\%$



8 Data Mining

– $L = \{3, 2, 5, 4, 6, 1\}$

- $\{2, 5\}$: is a good candidate
- $\{2, 4\}$: is a good candidate
- $\{2, 6\}$: is NOT a good candidate ($> \varphi$)
- $\{2, 1\}$: is NOT a good candidate ($> \varphi$)

$$\varphi = 20\%$$

Item	Count	SUP %	MIS %
1	8	80	70
2	2	20	17
3	2	20	15
4	2	20	30
5	3	30	30
6	6	60	35



8 Data Mining

$\varphi = 20\%$

Transactions
1, 4, 6
1
1, 5, 6
1, 6
4, 6
1, 2, 3, 5
1, 2, 3, 5
6
1
1, 6

- $L = \{3, 2, 5, 4, 6, 1\}$
 - $\{5, 4\}$: $\text{sup}(4) = 20\% < \text{MIS}(5) = 30\%$
so $\{5, 4\}$ is NOT a good candidate
 - $\{5, 6\}$: is NOT a good candidate
 - $\{5, 1\}$: is NOT a good candidate ($> \varphi$)
 - 4 can't be used as seed since $\text{sup}(4) < \text{MIS}(4)$
 - $\{6, 1\}$: is a good candidate
- $C2 = \{\{3, 2\}, \{3, 5\}, \{3, 4\}, \{2, 5\}, \{2, 4\}, \{6, 1\}\}$

Item	Count	SUP %	MIS %
1	8	80	70
2	2	20	17
3	2	20	15
4	2	20	30
5	3	30	30
6	6	60	35



8 Data Mining

- $C2 = \{\{3, 2\}, \{3, 5\}, \{3, 4\}, \{2, 5\}, \{2, 4\}, \{6, 1\}\}$
- Read Transactions to calculate F2
 - $F2 = \{\{3, 2\}, \{3, 5\}, \{2, 5\}, \{6, 1\}\}$

F	Item	SUP %	MIS %
F1	3	20	15
	2	20	17
	5	30	30
	6	60	35
	1	80	70
F2	{3, 2}	20	15
	{3, 5}	20	15
	{2, 5}	20	17
	{6, 1}	40	35

Transactions
1, 4, 6
1
1, 5, 6
1, 6
4, 6
1, 2, 3, 5
1, 2, 3, 5
6
1
1, 6



8 Data Mining

– $F2 = \{\{3, 2\}, \{3, 5\}, \{2, 5\}, \{6, 1\}\}; k = 3$

– Join:

- $\{3, 2, 5\}$: $MIS(2) < MIS(5)$ and $|\text{sup}(2) - \text{sup}(5)| = 10 < \varphi$, so it can be joined
- Nothing else can be joined

– Prune

- $\{3, 2\}$ and $\{3, 5\} \in F2$
- Since $\{2, 5\} \in F2$ the head problem is avoided otherwise we should have recorded also $\text{sup}(\{2, 5\})$

– $C3 = \{3, 2, 5\}$

Transactions
1, 4, 6
1
1, 5, 6
1, 6
4, 6
1, 2, 3, 5
1, 2, 3, 5
6
1
1, 6

Item	Count	SUP %	MIS %
1	8	80	70
2	2	20	17
3	2	20	15
4	2	20	30
5	3	30	30
6	6	60	35



8 Data Mining

minconf = 60%

- Scan transactions, $F3 = \{3, 2, 5\}$
 - $\text{Sup}(\{3, 2, 5\}) = 20\% > \text{MIS}(3) = 15$
- Step 2: rule generation from $F3 = \{3, 2, 5\}$
 - Non-empty subsets: $\{3, 2\}, \{3, 5\}, \{2, 5\}, \{3\}, \{2\}, \{5\}$
 - Possible rules derived from F_3 :
 - $\{3, 2\} \rightarrow \{5\}$, [sup = 20%, conf = 100%]
 - $\{3, 5\} \rightarrow \{2\}$, [sup = 20%, conf = 100%]
 - $\{2, 5\} \rightarrow \{3\}$, [sup = 20%, conf = 100%]
 - $\{3\} \rightarrow \{2, 5\}$, [sup = 20%, conf = 100%]
 - $\{2\} \rightarrow \{3, 5\}$, [sup = 20%, conf = 100%]
 - $\{5\} \rightarrow \{3, 2\}$, [sup = 20%, conf = 67%]
 - All are valid since minconf = 60%

F	Item	SUP %	MIS %
F1	3	20	15
	2	20	17
	5	30	30
	6	60	35
	1	80	70
F2	{3, 2}	20	15
	{3, 5}	20	15
	{2, 5}	20	17
	{6, 1}	40	35
F3	{3, 2, 5}	20	15



8 Data Mining

- Possible rules derived from F_2 :
 - $\{3\} \rightarrow \{2\}$, [sup = 20%, conf = 100%]
 - $\{2\} \rightarrow \{3\}$, [sup = 20%, conf = 100%]
 - $\{3\} \rightarrow \{5\}$, [sup = 20%, conf = 100%]
 - $\{5\} \rightarrow \{3\}$, [sup = 20%, conf = 67%]
 - $\{2\} \rightarrow \{5\}$, [sup = 20%, conf = 100%]
 - $\{5\} \rightarrow \{2\}$, [sup = 20%, conf = 67%]
 - $\{6\} \rightarrow \{1\}$, [sup = 40%, conf = 67%]
 - $\{1\} \rightarrow \{6\}$, [sup = 40%, conf = 50%]
- Except $\{1\} \rightarrow \{6\}$, all are valid

minconf = 60%

F	Item	SUP %	MIS %
F1	3	20	15
	2	20	17
	5	30	30
	6	60	35
	1	80	70
F2	{3, 2}	20	15
	{3, 5}	20	15
	{2, 5}	20	17
	{6, 1}	40	35
F3	{3, 2, 5}	20	15