

Problem set 7

Association pattern mining

Exercise 1

Consider the transaction database in the table below:

id	Transaction
1	A, B, C, D
2	B, C, E, F
3	A, D, E, F
4	A, E, F
5	B, D, F

Determine the support of itemsets A,E,F and D,F.

Solution $Sup(A) = \frac{3}{5}$, $Sup(B) = \frac{2}{5}$, $Sup(C) = \frac{2}{5}$, $Sup(D) = \frac{3}{5}$, $Sup(E) = \frac{3}{5}$, $Sup(F) = \frac{4}{5}$

Exercise 2

For the database in Exercise 1

1. compute all frequent itemsets at frequency thresholds 0.4, 0.6, and 0.8;
2. determine all the maximal frequent itemsets at frequency thresholds 0.4, 0.6, and 0.8.

Solution	Frequent itemset	0.4 (2)	{A},{B},{C},{D}, {E}, {F},{AD},{AE},{AF}, {BC},{BD},{BF},{EF},{DF},{AEF}
		0.6(3)	{A}, {B}, {D}, {E}, {F}, {EF}
		0.8 (4)	{F}
Maximal Frequent Itemset		0.4 (2)	{AD},{BD},{DF},{AEF}
		0.6(3)	{A}, {B}, {D}, {EF}
		0.8 (4)	{F}

Exercise 3

For the database in Exercise 1, determine the confidence and support of the rules $A \Rightarrow F$, and $A,E \Rightarrow F$

Solution $conf(A \Rightarrow F) = \frac{Sup(A \cup F)}{Sup(A)} = \frac{2}{3}$
 $conf(A,E \Rightarrow F) = \frac{Sup(A \cup E \cup F)}{Sup(A \cup E)} = \frac{2}{5}$

Exercise 4

For the database in Exercise 1 and frequency threshold 0.4, show the candidate itemsets and the frequent itemsets in each level-wise pass of the Apriori algorithm.

Solution $F_1 = \{\{A\}, \{B\}, \{C\}, \{D\}, \{E\}, \{F\}\}$
 $C_2 = \{\{A,B\}, \{A,C\}, \{A,D\}, \{A,E\}, \{A,F\}, \{B,C\}, \{B,D\}, \{B,E\}, \{B,F\}, \{C,D\}, \{C,E\}, \{C,F\}, \{D,E\}, \{D,F\}, \{E,F\}\}$
 $F_2 = \{\{A,D\}, \{A,E\}, \{A,F\}, \{B,C\}, \{B,D\}, \{B,F\}, \{D,F\}, \{E,F\}\}$
 $C_3 = \{\{A,D,E\}, \{A,D,F\}, \{A,E,F\}, \{B,C,D\}, \{B,C,F\}, \{B,D,F\}\}$
 $F_3 = \{\{A,E,F\}\}$ $C_4 = \phi$, $F_4 = \phi$