Introduction to Data Mining

Homework 1A

Homework group 4 :

Arnarinur Einarsson
Troy Macasland
Nick Geertjens
Guðmundur Pálsson
Stefan Jonsson

Apriori

	RS	Rv	SPL	MMT	TB	YS	AR	LB
Z_1	1	1	0	1	1	0	1	1
Z_2	0	1	0	1	0	1	1	1
Z_3	0	1	0	0	0	0	0	0
Z_4	0	1	0	0	0	0	1	0
Z_5	1	0	1	1	1	0	0	0
Z_6	1	1	0	1	1	0	0	1
Z_7	1	1	1	1	1	1	0	1
Z_8	1	0	1	1	1	0	0	1
Z_9	1	0	0	1	0	1	0	1
Z_{10}	0	0	0	0	1	0	0	0

The table translates to the following groups

{RS, RV, MMT, TB, AR, LB}

{RV, MMT, YS, AR, LB}

{RV}

{RV, AR}

{RV, SDL, MMT, TB}

Apriori

a) $s_{min} = \frac{2}{5}$ $\quad$ $s_{min} = \frac{3}{5}$

Support Count

$s_{min} = \frac{2}{5}$

Album	Freq	Support
RS	6	6/10
R	6	6/10
~~SgP~~	~~3~~	~~3/10~~
MMT	7	7/10
TB	6	6/10
~~YS~~	~~3~~	~~3/10~~
~~AR~~	~~3~~	~~3/10~~
LB	6	6/10

→

Album	Freq	Support
RS	6	6/10
R	6	6/10
MMT	7	7/10
TB	6	6/10
LB	6	6/10

Candidate set

Candidate set

Album	Freq	Support
~~RS, R~~	~~3~~	~~3/10~~
RS, MMT	6	6/10
RS, TB	5	5/10
RS, LB	5	5/10
R, MMT	4	4/10
~~R, TB~~	~~3~~	~~3/10~~
R, LB	4	4/10
MMT, TB	5	5/10
MMT, LB	6	6/10
TB, LB	4	4/10

→

Set of frequent itemsets

Album	Freq	Support
RS, MMT	6	6/10
RS, TB	5	5/10
RS, LB	5	5/10
R, MMT	4	4/10
R, LB	4	4/10
MMT, TB	5	5/10
MMT, LB	6	6/10
TB, LB	4	4/10

Candidate set

Albums	Freq	Support
RS, MMT, TB	5	5/10
RS, MMT, LB	5	5/10
RS, TB, LB	4	4/10
~~R, MMT, LB~~	~~3~~	~~3/10~~
MMT, TB, LB	4	4/10

→

Set of frequent itemsets

Albums	Freq	Support
RS, MMT, TB	5	5/10
RS, MMT, LB	5	5/10
RS, TB, LB	4	4/10
MMT, TB, LB	4	4/10

Album	Freq
RS, MMT, TB, LB	4

Rubber Soul, Magical Mystery Tour, The Beatles, Let it be

Support Count

$s_{min} = \frac{3}{5}$

Album	Freq	Support
RS	6	6/10
R	6	6/10
SgP	3	3/10
MMT	7	7/10
TB	6	6/10
YS	3	3/10
AR	3	3/10
LB	6	6/10

→

Album	Freq	Support
RS	6	6/10
R	6	6/10
MMT	7	7/10
TB	6	6/10
LB	6	6/10

Candidate set

Album	Freq	Support
~~RS, R~~	~~3~~	~~3/10~~
RS, MMT	6	6/10
~~RS, TB~~	~~5~~	~~5/10~~
~~RS, LB~~	~~5~~	~~5/10~~
~~R, MMT~~	~~4~~	~~4/10~~
~~R, TB~~	~~3~~	~~3/10~~
~~R, LB~~	~~4~~	~~4/10~~
~~MMT, TB~~	~~5~~	~~5/10~~
MMT, LB	6	6/10
~~TB, LB~~	~~4~~	~~4/10~~

→

Set of frequent itemsets

Album	Freq	Support
RS, MMT	6	6/10
MMT, LB	6	6/10

Candidate set

Albums	Freq	Support
~~RS, MMT, LB~~	~~5~~	~~5/10~~

→ Empty!

Set of frequent itemsets

B) We have for $s_{min} = \frac{2}{5}$:

Album	Freq
RS, MMT, TB, LB	4

Rubber Soul, Magical Mystery Tour, The Beatles, Let it be

We check $conf_1$ and if it's below 0.8 then we don't need to check $conf_2 \dots conf_{k-1}$ because $conf_1 \geq conf_2 \dots \geq conf_{k-1}$

We get

$l_1 = \{RS, MMT, TB, LB\}$ $c_1 = \{RS, MMT, TB\} \rightarrow \{LB\}$ $conf_1 = \frac{4/10}{5/10} = \frac{0.4}{0.5} = \underline{0.8}$

$l_1 = \{RS, MMT, TB, LB\}$ $c_2 = \{RS, MMT\} \rightarrow \{LB, TB\}$ $conf_2 = \frac{4/10}{6/10} = \frac{0.4}{0.6} = 0.67$

$l_1 = \{RS, MMT, TB, LB\}$ $c_3 = \{RS\} \rightarrow \{LB, TB, MMT\}$ $conf_2 = \frac{4/10}{6/10} = \frac{0.4}{0.6} = 0.67$

$l_1 = \{RS, MMT, LB, TB\}$ $c_1 = \{RS, MMT, LB\} \rightarrow \{TB\}$ $conf_1 = \frac{4/10}{5/10} = \frac{0.4}{0.5} = \underline{0.8}$

already checked

$l_1 = \{RS, LB, MMT, TB\}$ $c_1 = \{RS, LB, MMT\} \rightarrow \{TB\}$ $conf_1 = \frac{4/10}{5/10} = \frac{0.4}{0.5} = \underline{0.8}$

$l_1 = \{RS, LB, MMT, TB\}$ $c_2 = \{RS, LB\} \rightarrow \{MMT, TB\}$ $conf_2 = \frac{4/10}{5/10} = \frac{0.4}{0.5} = \underline{0.8}$

already checked

$l_1 = \{RS, LB, TB, MMT\}$ $c_1 = \{RS, LB, TB\} \rightarrow \{MMT\}$ $conf_1 = \frac{4/10}{4/10} = \frac{0.4}{0.4} = \underline{1}$

already checked

$l_1 = \{RS, TB, MMT, LB\}$ $c_1 = \{RS, TB, MMT\} \rightarrow \{LB\}$ $conf_1 = \frac{4/10}{5/10} = \frac{0.4}{0.5} = \underline{0.8}$

$l_1 = \{RS, TB, MMT, LB\}$ $c_2 = \{RS, TB\} \rightarrow \{MMT, LB\}$ $conf_2 = \frac{4/10}{5/10} = \frac{0.4}{0.5} = \underline{0.8}$

already checked

$l_1 = \{RS, TB, LB, MMT\}$ $c_1 = \{RS, TB, LB\} \rightarrow \{MMT\}$ $conf_1 = \frac{4/10}{4/10} = \frac{0.4}{0.4} = \underline{1}$

already checked

$l_1 = \{MMT, LB, RS, TB\}$ $c_1 = \{MMT, LB, RS\} \rightarrow \{TB\}$ $conf_1 = \frac{4/10}{5/10} = \frac{0.4}{0.5} = \underline{0.8}$

$l_1 = \{MMT, LB, RS, TB\}$ $c_2 = \{MMT, LB\} \rightarrow \{RS, TB\}$ $conf_2 = \frac{4/10}{6/10} = \frac{0.4}{0.6} = 0.67$

$l_1 = \{MMT, LB, RS, TB\}$ $c_3 = \{MMT\} \rightarrow \{LB, RS, TB\}$ $conf_2 = \frac{4/10}{7/10} = \frac{0.4}{0.7} = 0.57$

$l_1 = \{MMT, LB, TB, RS\}$ $c_1 = \{MMT, LB, TB\} \rightarrow \{RS\}$ $conf_1 = \frac{4/10}{4/10} = \frac{0.4}{0.4} = \underline{1}$

:

already checked

:

$l_1 = \{MMT, TB, LB, RS\}$ $c_1 = \{MMT, TB, LB\} \rightarrow \{RS\}$ $conf_1 = \frac{4/10}{4/10} = \frac{0.4}{0.4} = \underline{1}$

$l_1 = \{MMT, TB, LB, RS\}$ $c_2 = \{MMT, TB\} \rightarrow \{LB, RS\}$ $conf_2 = \frac{4/10}{5/10} = \frac{0.4}{0.5} = \underline{0.8}$

already checked

$l_1 = \{MMT, TB, RS, LB\}$ $c_1 = \{MMT, TB, RS\} \rightarrow \{LB\}$ $conf_1 = \frac{4/10}{5/10} = \frac{0.4}{0.5} = \underline{0.8}$

:

already checked

:

$l_1 = \{MMT, RS, TB, LB\}$ $c_1 = \{MMT, RS, TB\} \rightarrow \{LB\}$ $conf_1 = \frac{4/10}{5/10} = \frac{0.4}{0.5} = \underline{0.8}$

$l_1 = \{MMT, RS, TB, LB\}$ $c_2 = \{MMT, RS\} \rightarrow \{TB, LB\}$ $conf_2 = \frac{4/10}{6/10} = \frac{0.4}{0.6} = 0.67$

already checked

$l_1 = \{MMT, RS, LB, TB\}$ $c_1 = \{MMT, RS, LB\} \rightarrow \{TB\}$ $conf_1 = \frac{4/10}{5/10} = \frac{0.4}{0.5} = \underline{0.8}$

$\frac{4/10}{4/10}$

:

already checked

:

$l_1 = \{TB, MMT, LB, RS\}$ $c_1 = \{TB, MMT, LB\} \rightarrow \{RS\}$ $conf_1 = \frac{4/10}{4/10} = \frac{0.4}{0.4} = 1$

$l_1 = \{TB, MMT, LB, RS\}$ $c_2 = \{TB, MMT\} \rightarrow \{LB, RS\}$ $conf_2 = \frac{4/10}{5/10} = \frac{0.4}{0.4} = 0.8$

$l_1 = \{TB, MMT, LB, RS\}$ $c_3 = \{TB\} \rightarrow \{MMT, LB, RS\}$ $conf_2 = \frac{4/10}{6/10} = \frac{0.4}{0.6} = 0.67$

$l_1 = \{TB, MMT, RS, LB\}$ $c_1 = \{TB, MMT, RS\} \rightarrow \{LB\}$ $conf_1 = \frac{4/10}{5/10} = \frac{0.4}{0.5} = 0.8$

:

already checked

:

$l_1 = \{TB, RS, MMT, LB\}$ $c_1 = \{TB, RS, MMT\} \rightarrow \{LB\}$ $conf_1 = \frac{4/10}{5/10} = \frac{0.4}{0.5} = 0.8$

$l_1 = \{TB, RS, MMT, LB\}$ $c_2 = \{TB, RS\} \rightarrow \{MMT, LB\}$ $conf_2 = \frac{4/10}{5/10} = \frac{0.4}{0.5} = 0.8$

:

already checked

:

$l_1 = \{TB, RS, LB, MMT\}$ $c_1 = \{TB, RS, LB\} \rightarrow \{MMT\}$ $conf_1 = \frac{4/10}{4/10} = \frac{0.4}{0.4} = 1$

:

already checked

:

$l_1 = \{TB, LB, RS, MMT\}$ $c_1 = \{TB, LB, RS\} \rightarrow \{MMT\}$ $conf_1 = \frac{4/10}{4/10} = \frac{0.4}{0.4} = 1$

$l_1 = \{TB, LB, RS, MMT\}$ $c_2 = \{TB, LB\} \rightarrow \{RS, MMT\}$ $conf_2 = \frac{4/10}{4/10} = \frac{0.4}{0.4} = 1$

:

already checked

:

$l_1 = \{TB, LB, MMT, RS\}$ $c_1 = \{TB, LB, MMT\} \rightarrow \{RS\}$ $conf_1 = \frac{4/10}{4/10} = \frac{0.4}{0.4} = 1$

:

already checked

:

$l_1 = \{LB, TB, MMT, RS\}$ $c_1 = \{LB, TB, MMT\} \rightarrow \{RS\}$ $conf_1 = \frac{4/10}{4/10} = \frac{0.4}{0.4} = 1$

$l_1 = \{LB, TB, MMT, RS\}$ $c_2 = \{LB, TB\} \rightarrow \{MMT, RS\}$ $conf_2 = \frac{4/10}{4/10} = \frac{0.4}{0.4} = 1$

$l_1 = \{LB, TB, MMT, RS\}$ $c_3 = \{LB\} \rightarrow \{TB, MMT, RS\}$ $conf_2 = \frac{4/10}{6/10} = \frac{0.4}{0.6} = 0.67$

$l_1 = \{LB, TB, RS, MMT\}$ $c_1 = \{LB, TB, RS\} \rightarrow \{MMT\}$ $conf_1 = \frac{4/10}{4/10} = \frac{0.4}{0.4} = 1$

:

already checked

:

$l_1 = \{LB, RS, TB, MMT\}$ $c_1 = \{LB, RS, TB\} \rightarrow \{MMT\}$ $conf_1 = \frac{4/10}{4/10} = \frac{0.4}{0.4} = 1$

$l_1 = \{LB, RS, TB, MMT\}$ $c_2 = \{LB, RS\} \rightarrow \{TB, MMT\}$ $conf_2 = \frac{4/10}{5/10} = \frac{0.4}{0.5} = 0.8$

:

already checked

:

$l_1 = \{LB, RS, MMT, TB\}$ $c_1 = \{LB, RS, MMT\} \rightarrow \{TB\}$ $conf_1 = \frac{4/10}{5/10} = \frac{0.4}{0.5} = 0.8$

:

already checked

:

$l_1 = \{LB, MMT, TB, RS\}$ $c_1 = \{LB, MMT, TB\} \rightarrow \{RS\}$ $conf_1 = \frac{4/10}{4/10} = \frac{0.4}{0.4} = 1$

$l_1 = \{LB, MMT, TB, RS\}$ $c_2 = \{LB, MMT\} \rightarrow \{TB, RS\}$ $conf_2 = \frac{4/10}{6/10} = \frac{0.4}{0.6} = 0.67$

:

already checked

:

$l_1 = \{LB, MMT, RS, TB\}$ $c_1 = \{LB, MMT, RS\} \rightarrow \{TB\}$ $conf_1 = \frac{4/10}{5/10} = \frac{0.4}{0.5} = 0.8$

:

already checked

:

In the above calculations we have calculated the association rules with minimal confidence 0.8, these rules are:

$c_1 = \{RS, MMT, TB\} \rightarrow \{LB\}$ $conf_1 = \frac{4/10}{5/10} = \frac{0.4}{0.5} = \underline{0.8}$

$c_1 = \{RS, MMT, LB\} \rightarrow \{TB\}$ $conf_1 = \frac{4/10}{5/10} = \frac{0.4}{0.5} = \underline{0.8}$

$c_2 = \{RS, LB\} \rightarrow \{MMT, TB\}$ $conf_2 = \frac{4/10}{5/10} = \frac{0.4}{0.5} = \underline{0.8}$

$c_1 = \{RS, LB, TB\} \rightarrow \{MMT\}$ $conf_1 = \frac{4/10}{4/10} = \frac{0.4}{0.4} = \underline{1}$

$c_2 = \{RS, TB\} \rightarrow \{MMT, LB\}$ $conf_2 = \frac{4/10}{5/10} = \frac{0.4}{0.5} = \underline{0.8}$

$c_1 = \{MMT, LB, TB\} \rightarrow \{RS\}$ $conf_1 = \frac{4/10}{4/10} = \frac{0.4}{0.4} = \underline{1}$

$c_2 = \{MMT, TB\} \rightarrow \{LB, RS\}$ $conf_2 = \frac{4/10}{5/10} = \frac{0.4}{0.5} = \underline{0.8}$

$c_2 = \{TB, LB\} \rightarrow \{MMT, RS\}$ $conf_2 = \frac{4/10}{4/10} = \frac{0.4}{0.4} = 1$

FP - Growth

Data:

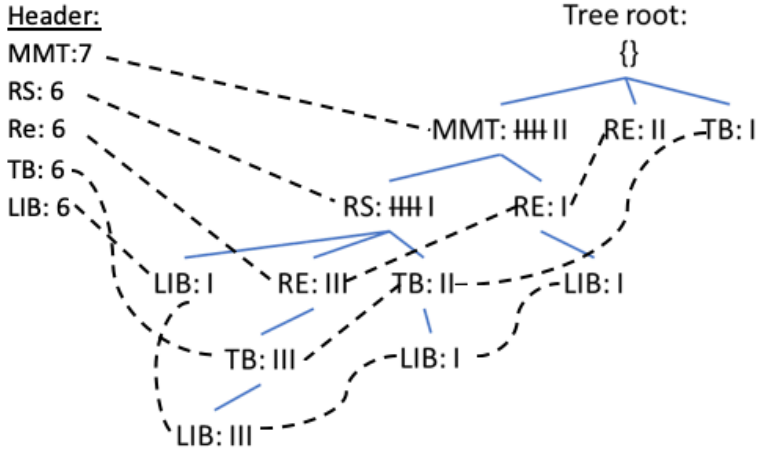
Album	Z1	Z2	Z3	Z4	Z5	Z6	Z7	Z8	Z9	Z10
RS	■				■	■	■	■	■	
Re	■	■	■	■		■	■			
SPL					■		■	■		
MMT	■	■			■	■	■	■	■	
TB	■				■	■	■	■		■
YS		■					■		■	
AR	■	■		■						
LIB	■	■				■	■	■	■	

γ.) Sorted by frequency:

Album	Z1	Z2	Z3	Z4	Z5	Z6	Z7	Z8	Z9	Z10	Total:
MMT	■	■			■	■	■	■	■		7
RS	■				■	■	■	■	■		6
Re	■	■	■	■		■	■				6
TB	■				■	■	■	■		■	6
LIB	■	■				■	■	■	■		6
SPL					■		■	■			3
YS		■					■		■		3
AR	■	■		■							3

Z1: MMT-RS-Re-TB-LIB
Z2: MMT-Re-LIB
Z3: Re
Z4: Re
Z5: MMT-RS-TB
Z6: MMT-RS-Re-TB-LIB
Z7: MMT-Rs-Re-TB-LIB
Z8: MMT-RS-TB-LIB
Z9: MMT-RS-LIB
Z10: TB

δ.)

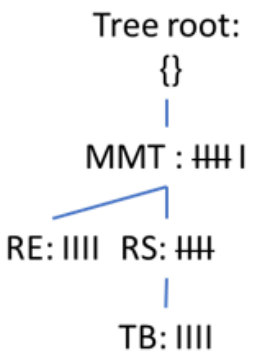


ε.)

Item:	Conditional pattern base:
MMT:	{}
RS:	M:6
Re:	MMT-RS: 3, MMT:1
TB:	MMT-RS-RE: 3, MMT-RS: 2
LIB:	MMT-RS:1, MMT-RS-RE-TB:3, MMT-RS-TB:1, MMT-RE:1

ζ.) Conditional FP-Trees:

MMT:	{}
RS:	{M:6}|RS
Re:	{M:4}|Re
TB:	{(MMT:5, RS:5)}|TB
LIB:	



η.) Determining frequent itemsets:

Of course the single items are frequent: MMT, RS, Re, TB,LIB

From the RS conditional tree (single branch, all combinations frequent):

MMT-RS

From the Re conditional tree (also single branch, all combinations frequent):

MMT-Re

From the TB conditional tree (also a single branch, all combinations frequent):

MMT-TB, RS-TB, MMT-RS-TB

The LIB conditional tree is not a single branch, will need to be recursively mined:

Item:	Conditional pattern base:	Conditional FP-tree:
TB-LIB	MMT-RS: 4	{(MMT:4, RS:4)}|TB-LIB
Re-LIB	MMT:4	{MMT:4}|Re-LIB

From the TB-LIB conditional tree (which is now a single branch):

MMT-TB-LIB, RS-TB-LIB, MMT-RS-LIB, MMT-LIB, RS-LIB, TB-LIB, MMT-RS-TB-LIB

From the RE-LIB conditional tree (also a single branch):

MMT-Re-LIB, ~~MMT-LIB~~, Re-LIB