TDT4300

Assignment 2

Corrected task 1 b) and 4

Andreas Amundsen

Magnus Lauritzen Holtet

Task 1

a)

Minimum support count: 0.5
Transactions: 10

Candidate 1-itemsets

Item	**Count**
A	6
B	8
C	10
G	8
H	7

Candidate 2-itemsets

Items	**Count**
A, B	5
A, C	6
A, G	6
A, H	5
B, C	8
B, G	7
B, H	5
C, G	8
C, H	7
G, H	5

Candidate 3-itemset

Items	**Count**
A, B, C	5
A, B, G	5
A, C, G	6
A, C, H	5
A, G, H	5
B, C, G	7
B, C, H	5

C, G, H	5

Candidate 4-itemset

Items	**Count**
A, B, C, G	5
A, C, G, H	5

b)

Confidence threshold: 0.8

Valid rules for {ABCG}

Rule	Conf	
A -> BCG	5/5 = 1	σ (ABCG) / σ (A)
AG -> BC	5/6 = 0,83	σ (ABCG) / σ (AG)
AC -> BG	5/6 = 0,83	σ (ABCG) / σ (AC)
AB -> CG	5/5 = 1	σ (ABCG) / σ (AB)
BCG -> A	5/7 = 0,71	σ (ABCG) / σ (BCG)
ABC -> G	5/5 = 1	σ (ABCG) / σ (ABC)
ABG -> C	5/5 = 1	σ (ABCG) / σ (ABG)
ACG -> B	5/6 = 0,83	σ (ABCG) / σ (ACG)

Valid rules for {ACGH}

Rule	Conf	
CG => AH	5/8 = 0,63	σ (ACGH) / σ (CG)
CH => AG	5/7 = 0,71	σ (ACGH) / σ (CH)
GH => AC	5/5 = 1	σ (ACGH) / σ (GH)
A => CGH	5/6 = 0,83	σ (ACGH) / σ (A)
ACG => H	5/6 = 0,83	σ (ACGH) / σ (ACG)
ACH => G	5/5 = 1	σ (ACGH) / σ (ACH)
AGH => C	5/5 = 1	σ (ACGH) / σ (AGH)
CGH => A	5/5 = 1	σ (ACGH) / σ (CGH)
AC => GH	5/6 = 0,83	σ (ACGH) / σ (AC)
AG => CH	5/6 = 0,83	σ (ACGH) / σ (AG)
AH => CG	5/5 = 1	σ (ACGH) / σ (AH)

Task 2

Original table

TID	Items
110	A,C,F,G,H
111	B,C,D,E,G
112	B,C,E,F,H
113	A,B,C,G
114	C,D,E,H
115	A,B,C,G,H
116	A,B,C,D,G,H
117	B,C,E,G
118	A,B,C,F,G,H
119	A,B,C,D,E,G,H

Step 1: Calculate support count and remove elements whose support count < 0.5

Item	Support count
A	6
B	8
C	10
D	4
E	5
F	3
G	8
H	7

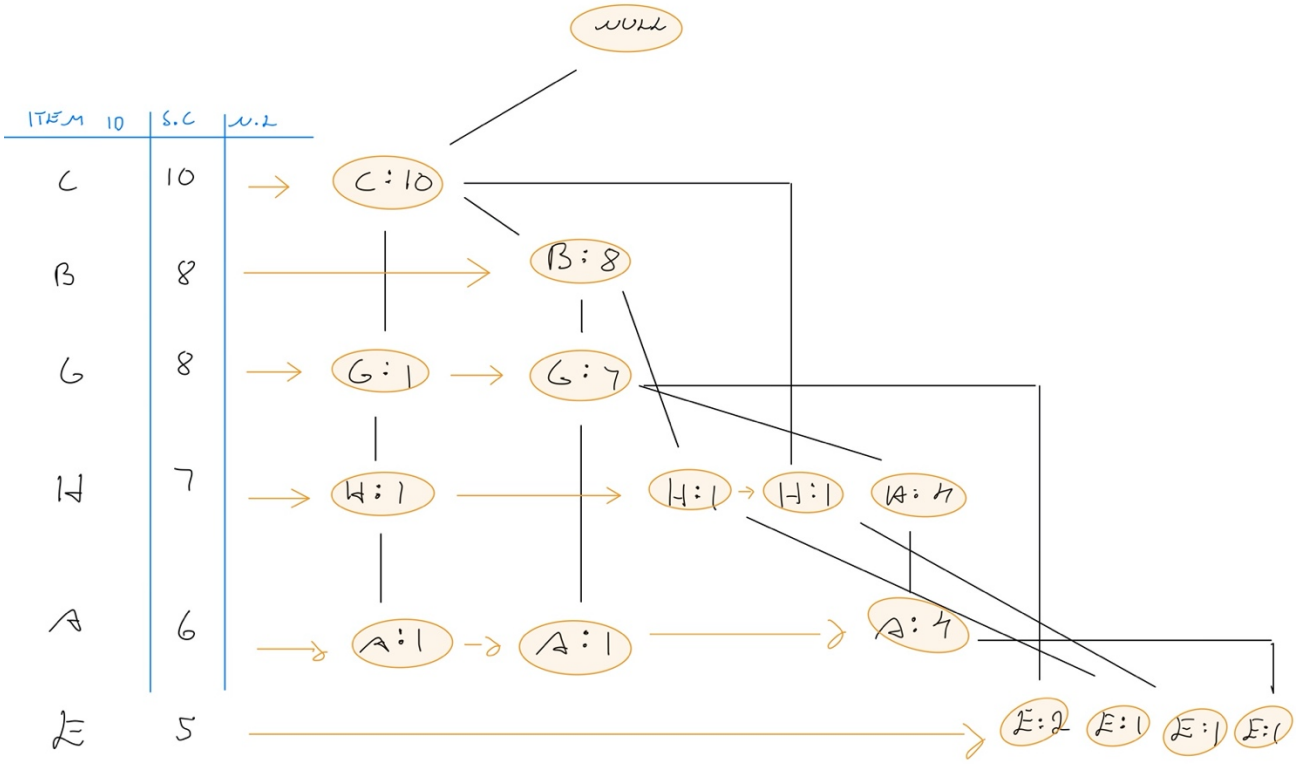
Step 2: Generate frequent patterns in descending order

Item	Support count
C	10
B	8
G	8
H	7
A	6
E	5

Step 3: Generate ordered item set based on table generated from step 2

TID	Items	Ordered item set
110	A,C,F,G,H	C,G,H,A
111	B,C,D,E,G	C,B,G, E
112	B,C,E,F,H	C,B,H, E
113	A,B,C,G	C,B,G,A
114	C,D,E,H	C,H, E
115	A,B,C,G,H	C,B,G,H,A
116	A,B,C,D,G,H	C,B,G,H,A
117	B,C,E,G	C,B,G, E
118	A,B,C,F,G,H	C,B,G,H,A
119	A,B,C,D,E,G,H	C,B,G,H,A, E

Step 4: Generate FP tree



Step 4: Generate conditional pattern base

Items	**Conditional pattern base**
E	{C,B,G: 2}, {C,B,H: 1}, {C,B,H: 1}, {C,B,G,H,A: 1}
A	{C,G,H: 1}, {C,B,G: 1}, {C,B,G,H: 4}
H	{C: 1}, {C,G: 1}, {C,B: 1}, {C,B,G: 4}
G	{C: 1}, {C,B: 7}
B	{C: 8}
C	-

Step 5: Generate conditional FP tree

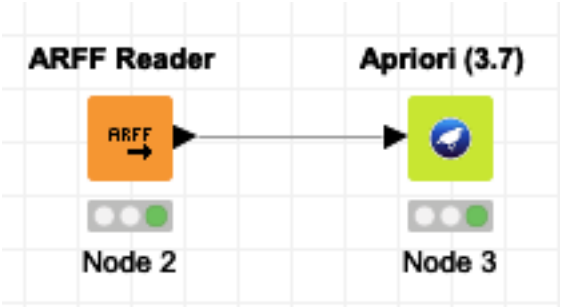
Items	**Conditional pattern base**	**Conditional FP tree**
E	{C,B,G: 2}, {C,B,H: 1}, {C,B,H: 1}, {C,B,G,H,A: 1}	{C: 5}
A	{C,G,H: 1}, {C,B,G: 1}, {C,B,G,H: 4}	{C,G: 6}
H	{C: 1}, {C,G: 1}, {C,B: 1}, {C,B,G: 4}	{C: 7}
G	{C: 1}, {C,B: 7}	{C: 8}
B	{C: 8}	{C: 8)
C	-	

Step 6: Generate frequent patterns

Items	**Conditional pattern base**	**Conditional FP tree**	**Frequent patterns**
E	{C,B,G: 2}, {C,B,H: 1}, {C,B,H: 1}, {C,B,G,H,A: 1}	{C:5}	{C,E: 5}
A	{C,G,H: 1}, {C,B,G: 1}, {C,B,G,H: 4}	{C,G: 6}	{C,A: 6} {G,A: 6] {C,G,A: 6}
H	{C: 1}, {C,G: 1}, {C,B: 1}, {C,B,G: 4}	{C: 7}	{C,H: 7}
G	{C: 1}, {C,B: 7}	{C: 8}	{C,G: 8}
B	{C: 8}	{C: 8)	{C,B: 8}
C	-		

Task 3

Apriori



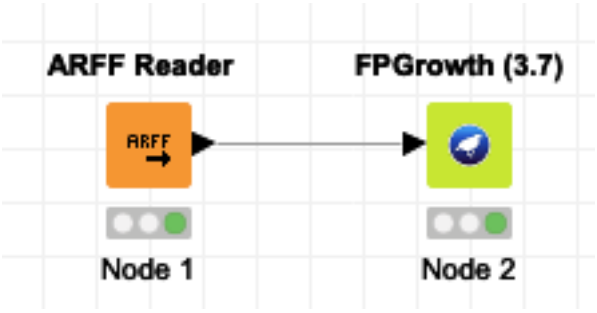
Weka Node View - 0:3 - Apriori (3.7)

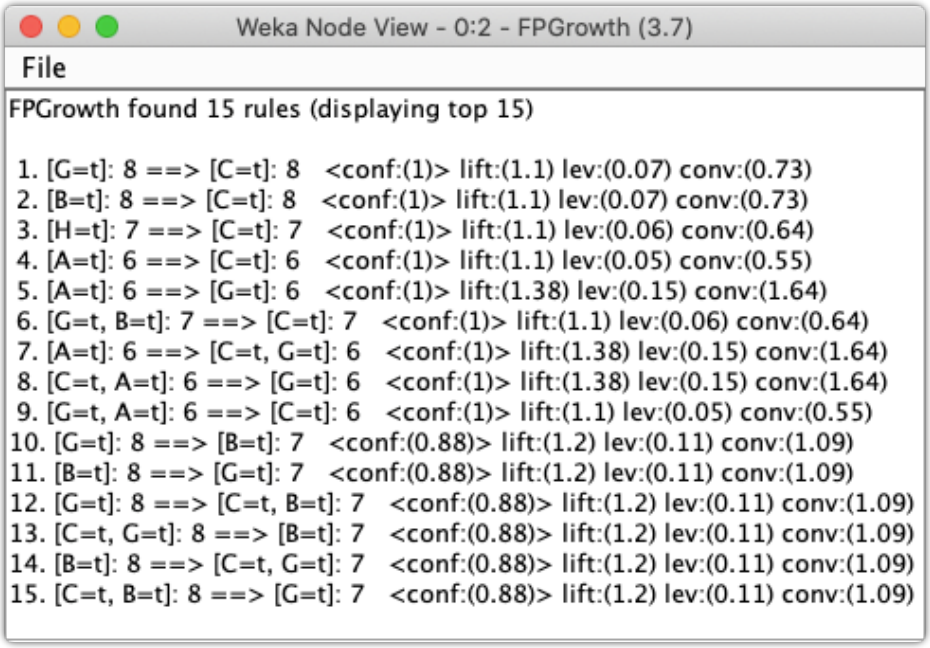
File

```
Apriori
=======

Minimum support: 0.65 (7 instances)
Minimum metric <confidence>: 0.8
Number of cycles performed: 7

Generated sets of large itemsets:

Size of set of large itemsets L(1): 4

Size of set of large itemsets L(2): 4

Size of set of large itemsets L(3): 1

Best rules found:

 1. B=t 8 ==> C=t 8    <conf:(1)> lift:(1.1) lev:(0.07) [0] conv:(0.73)
 2. G=t 8 ==> C=t 8    <conf:(1)> lift:(1.1) lev:(0.07) [0] conv:(0.73)
 3. H=t 7 ==> C=t 7    <conf:(1)> lift:(1.1) lev:(0.06) [0] conv:(0.64)
 4. B=t G=t 7 ==> C=t 7    <conf:(1)> lift:(1.1) lev:(0.06) [0] conv:(0.64)
 5. G=t 8 ==> B=t 7    <conf:(0.88)> lift:(1.2) lev:(0.11) [1] conv:(1.09)
 6. B=t 8 ==> G=t 7    <conf:(0.88)> lift:(1.2) lev:(0.11) [1] conv:(1.09)
 7. C=t G=t 8 ==> B=t 7    <conf:(0.88)> lift:(1.2) lev:(0.11) [1] conv:(1.09)
 8. B=t C=t 8 ==> G=t 7    <conf:(0.88)> lift:(1.2) lev:(0.11) [1] conv:(1.09)
 9. G=t 8 ==> B=t C=t 7    <conf:(0.88)> lift:(1.2) lev:(0.11) [1] conv:(1.09)
10. B=t 8 ==> C=t G=t 7    <conf:(0.88)> lift:(1.2) lev:(0.11) [1] conv:(1.09)
```

Exported KNIME workflow is located in ZIP folder

FP Growth





Exported KNIME workflow is located in ZIP folder

Task 4

The procedure goes like this:

- Find itemsets with size 4
- Find itemsets with size 3
- Find itemsets with size 2
- Find itemsets with size 1
- Combine the results

If a given itemset is not located in the table, the support for the maximum itemset is used.

For size 4

{a, c, d, e}

For size 3

Find every subset of itemset {a, c, d, e}

{ {a,d,e} , {c,d,e} , {a,b,e} , {b,d,e}, {a,c,d}, {a,c,e} }

Itemset	Max of itemset	σ
a, c, d		6
a, c, e	a, c, d, e	5
a, d, e	a, c, d, e	5
c, d, e	a, c, d, e	5
a, b, e		7
b, d, e		4

For size 2

{ {a, c}, {a, d}, {a, e}, {c, d}, {c, e}, {d, e}, {b, d}, {b, e} }

Itemset	Max of itemset	σ
a, c	{a, c, e} and {a, c, d}	6
a, d		11
a, e	{a, c, e}, {a, b, e} and {a, d, e}	7
c, d	{c, d, e} and {a, c, d}	6
c, e	{c, d, e} and {a, c, e}	5
d, e		6
b, d		7
b, e		8

For size 1

{ {a}, {b}, {d}, {e} }

Itemset	Max of itemset	σ
a	{a, d}, {a, e} and {a, c}	11
c	{a, c}, {c, e} and {c, d}	6
d		13
e	{b, e}, {c, e}, {a, e} and {d, e}	8

Total:

Itemset	σ
a	11
c	6
d	13
e	8
c, e	5
c, d	6
a, e	7
a, c	6
a, d	11
d, e	6
b, d	7
b, e	8
a, c, d	6
a, c, e	5
a, d, e	5
c, d, e	5
a, b, e	7
b, d, e	4
a, c, d, e	5