Exercise 4

Data Mining Algoritms 1 – WS 2015/16

Davide Pedranz (362504) | Jakub Amanowicz (362594) | Hongmei Liang (292520)

Exercise 4.1

Let D = given dataset
Knowing that minimal support is equal to 10% we can evaluate

$$\lceil minsup * |D| \rceil = \lceil 0.01 * 13 \rceil = 2$$

which states that an itemset I needs to occur at least 2 times in tn D to be treated as “frequent”.

Discarding non frequent items → Ordering and pruning D

D
{b,f,g,m}
{d,g,h,l}
{b,c,d}
{d,g,h,i}
{d,g,h,j}
{a,b,d,g,k}
{a,d,h}
{e,f,g}
{b,d}
{a,f,g}
{a,b,d,g}
{b,f,g}
{f,g}

Item	*Frequency*
a	4
b	6
c	1
d	8
e	1
f	5
g	10
h	4
i	1
j	1
k	1
l	1
m	1

Item	*Frequency*
g	10
d	8
b	6
f	5
a	4
h	4

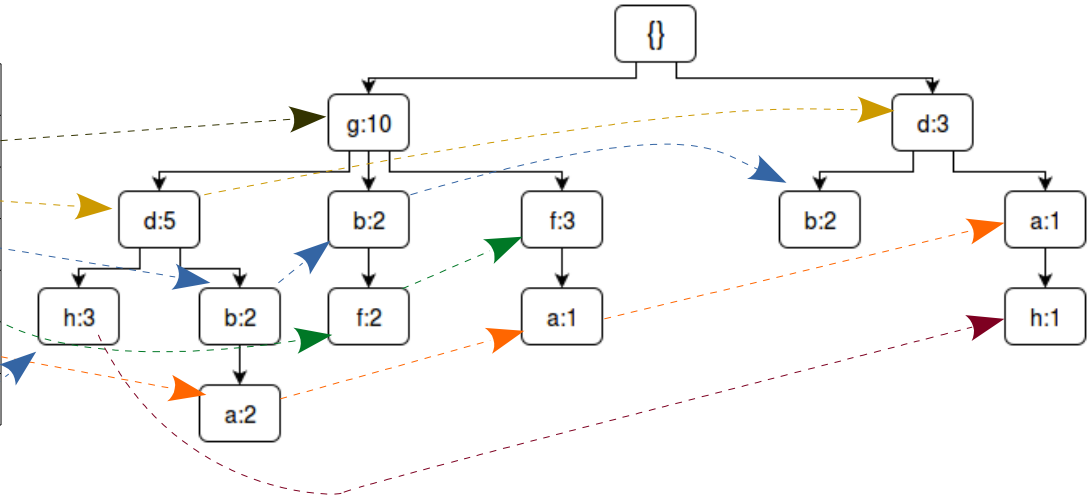
↑
Header table

D'
{g,b,f}
{g,d,h}
{d,b}
{g,d,h}
{g,d,h}
{g,d,b,a}
{d,a,h}
{g,f}
{d,b}
{g,f,a}
{g,d,b,a}
{g,b,f}
{g,f}

Having the dataset sorted and header table composed, we are able to use them to build the fp-tree.

Header table

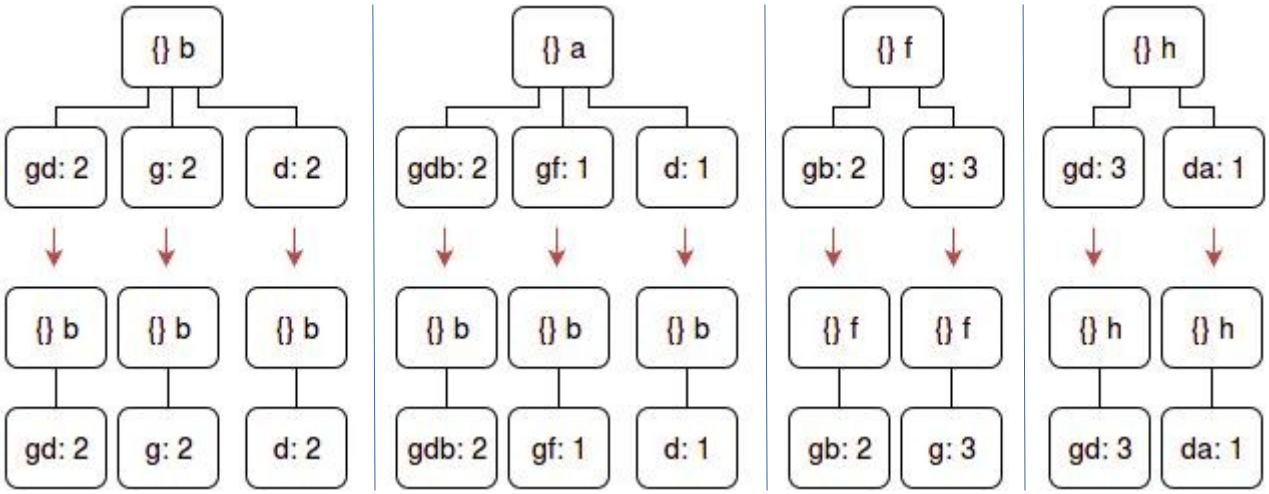
Item	*Fr.*	*head*
g	10	●
d	8	●
b	6	●
f	5	●
a	4	●
h	4	●

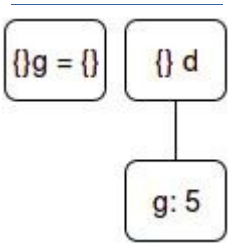


After we've constructed the fp-tree we can compose a conditional pattern base table.
We will need it to construct mutliple fp-trees
Which will be recursively searched so we can
Obtain frequent patterns from dataset D.

Item	Conditional pattern base
g	{}
d	g: 5,{}
b	gd: 2, g: 2, d: 2
f	gb: 2, g: 3
a	gdb: 2, gf: 1, d: 1
h	gd: 3, da: 1

Below there are fp-trees built basing on the table on the right.
Each complex tree was desintegrated into smaller, single path, ones.





We have to remember about minimal support and verify which of the obtained itemsets are frequent. Therefore, we discard the ones with count less than 2.
What we are left with are following itemsets.

{b,g,d}, {b,g}, {b,d}, {a,g,b,d,},{f,g,b},{f,g},{h,g,d},{g},{d,g}

Of course these are not the only frequent itemsets in the dataset – we have to remember about their subsets which, by definition, are frequent as well.