

Assignment

Data Mining

Topic

*Weka Implementation of Algorithms
(Apriori, Apriori TID and FP-Growth)*

Submitted by

Junaid Iqbal (70070342)

Q.1. Consider an example with the following set of transactions. We wish to find associations with at least 30% support and 60% confidence. Find the frequent item sets and then the association rules. Implement using Weka with different algorithms of association rule mining (Apriori, Apriori TID and FP-Growth) and differentiate among them.

Transaction ID	Items bought
1	B M T Y
2	B M
3	A T S P
4	A B C D
5	A B
6	T Y E M
7	A B M
8	B C D T P
9	D T S
10	A B M

Solution:

Frequency Item Sets

Items	Frequency/ Support count	Support
A	5	50%
B	7	70%
C	2	20%
D	3	30%
E	1	10%
M	5	50%
P	2	20%
S	2	20%
T	5	50%
Y	2	20%

1 Apriori Algorithm

Data for Apriori [1]

TID	A	B	C	D	E	M	P	S	T	Y
1		YES				YES			YES	YES
2		YES				YES				
3	YES						YES	YES	YES	
4	YES	YES	YES	YES						
5	YES	YES								
6					YES	YES			YES	YES
7	YES	YES				YES				
8		YES	YES	YES			YES		YES	
9				YES				YES	YES	
10	YES	YES				YES				

Results:

Minimum support: 0.3 (4 instances)

Minimum metric <confidence>: 0.6

Number of cycles performed: 14

Generated sets of large itemsets:

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

Large Itemsets L(1):

A=YES 5

B=YES 7

M=YES 5

T=YES 5

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

Large Itemsets L(2):

A=YES B=YES 4

B=YES M=YES 4

Best rules found:

1. A=YES 5 ==> B=YES 4 <conf:(0.8)> lift:(1.37) lev:(0.09) [1] conv:(1.04)

2. M=YES 5 ==> B=YES 4 <conf:(0.8)> lift:(1.37) lev:(0.09) [1] conv:(1.04)

2 AprioriTID Algorithm

The difference between Apriori and AprioriTID is only that it doesn't use the database for counting the support of candidate itemsets after the first pass. It is implementable by using SPMF Open Source Data Mining Library [2].

3 FP Growth Algorithm

Data for FP Growth Algorithm [3]

```
@relation supermarket
@attribute 'A' { t }
@attribute 'B' { t }
@attribute 'C' { t }
@attribute 'D' { t }
@attribute 'E' { t }
@attribute 'M' { t }
@attribute 'P' { t }
@attribute 'S' { t }
@attribute 'T' { t }
@attribute 'Y' { t }
@data
?,t,?,?,t,?,?,t,t|
?,t,?,?,t,?,?,?,?
t,?,?,?,t,t,t,?
t,t,t,t,?,?,?,?
t,t,?,?,?,?,?,?
?,?,?,t,t,?,?,t,t
t,t,?,?,t,?,?,?
?,t,t,t,?,t,?,t,?
?,?,?,t,?,t,t,?
t,t,?,?,t,?,?,?
```

Results

FPGrowth found 2 rules (displaying top 2)

1. [M=t]: 5 ==> [B=t]: 4 <conf:(0.8)> lift:(1.14) lev:(0.05) conv:(0.75)
2. [A=t]: 5 ==> [B=t]: 4 <conf:(0.8)> lift:(1.14) lev:(0.05) conv:(0.75)

References:

- [1] <https://github.com/junaideqbal/data-mining/blob/main/assignment1/Apriori-Data.csv>
- [2] <http://www.philippe-fournier-viger.com/spmf/>
- [3] <https://github.com/junaideqbal/data-mining/blob/main/assignment1/FP-Growth-Data.arff>