Week-11
Association Rule Mining
Solution
Answer the following Multiple-Choice Questions

1. Association Rule Mining involves,
   a) Same as frequent itemset mining
   b) Finding Strong association rules using frequent itemsets
   c) Both a and b
   d) None of the above

2. What do we mean by Support(X)?
   a) Total number of transactions containing in X
   b) Total number of transactions not containing in X
   c) Number of Transactions containing X/ Total number of transactions
   d) Number of Transactions not containing X/ Total number of transactions

3. Which of the following is direct application of frequent itemset mining?
   a) Social Network Analysis
   b) Market Basket Analysis
   c) Outlier Detection
   d) None of the Above

4. Following is a transaction data given:

TID	Items
10	Ball, Nuts, Pen
20	Ball, Coffee, Pen, Nuts
30	Ball, Pen, Eggs
40	Ball, Nuts, Egg, Milk
50	Nuts, Coffee, Pen, Eggs, Milk

For the given, support threshold 0f 40%, how many 3-items/item sets can be obtained?
   a) 1
   b) 0
   c) 2
   d) 3

5. What will be the effect of reducing the Confidence Threshold on the same dataset?
   a) Number of Association rules will get added to the currently available rule
   b) Number of Association rules will get reduced from the currently available rule
   c) Association Rules become invalid
   d) None of the above

6. Choose the following analysis method that you think will be more appropriate for the scenario, first buys iPhone and then buys Air pod.
   a) Classification
   b) Association Rule
   c) Prediction
   d) None of the above

7. Choose which data mining task is the most suitable for the following scenario:
   To identify items that are bought concomitantly by a reasonable fraction of customers that they can be shelved.
   a) Association Rule
   b) Classification
   c) Clustering
   d) Noe of the above

8. What does Apriori Algorithm do?
   a) It mines all frequent patterns through pruning rules with lesser support & confidence
   b) It mines all frequent patterns through pruning rules with higher support & confidence
   c) Both of the above
   d) None of the above

9. When do you consider an association to be appropriate for its interestingness to be measured?
   a) Satisfies Support threshold
   b) Satisfies Confidence threshold
   c) Both a and b
   d) None of the above

10. **Our use of association analysis will yield the same frequent item sets and strong association rules whether a specific item occurs once or three times in an individual transaction**
   a) True
   b) False

Answer the following Question

Measure the Associations for the following Transactional Data given below using Support, Confidence and Lift:

Transaction ID	Items
1	Jam, Perfume, Snacks, Soap
2	Beer, Bread, Jam, Snacks
3	Potato, Onion
4	Bread, Chocolate, Coffee, Jam
5	Beer, Coffee, Perfume, Snacks, Soap
6	Bread, Chocolate, Jam, Soap
7	Onion, Perfume
8	Bread, Chocolate, Snacks
9	Chocolate, Perfume, Soap
10	Beer, Bread, Jam, Snacks

Kimia Aksir

The threshold for Support is estimated as: 30%

The threshold for Confidence is estimated as: 70%

Answer:

Calculating Support:

Step:1 Finding the support of each distinct item (or, for 1-item set):

The threshold for Support will be 30%.

Items	Support (%)
(Soap}	40
{Bread}	50
{Jam}	50
{Beer}	30
{Onion}	20
{Potato}	10
{Snacks)	50
{Coffee}	20
{Perfume}	40
{Chocolate}	40

So, the candidate solution of this step will look like:

Items	Support (%)
(Soap}	40
{Bread}	50
{Jam}	50
{Beer}	30
{Snacks}	50
{Perfume}	40
{Chocolate}	40

Step:2 Finding the support of associated items (or, for 2-items set):

The items already filtered out in the previous step will no longer be considered here and only the ones sustained having support threshold 30% or, more will be used in this step.

Items	Support (%)
{Soap, Bread}	10
{Soap, Jam}	10

Kimia Aksir

{Soap, Beer}	10
{Soap, Snacks}	20
{Soap, Perfume}	30
{Soap, Chocolate}	20
{Bread, Jam}	40
{Bread. Beer}	20
{Bread, Snacks}	30
{Bread, Perfume)	0
{Bread, Chocolate}	30
{Jam, Beer}	20
{Jam, Snacks}	30
{Jam, Perfume}	10
{Jam, Chocolate}	20
{Beer, Snacks}	30
{Beer, Perfume}	10
{Beer, Chocolate}	0
{Snacks, Perfume}	20
{Snacks, Chocolate}	10
{Perfume, Chocolate}	10

So, the candidate solution of this step will look like:

Items	Support (%)
{Soap, Perfume}	30
{Bread, Jam}	40
{Bread, Snacks}	30
{Bread, Chocolate}	30
{Jam, Snacks}	30
{Beer, Snacks}	30

Step:3 Finding the support of associated items (or, for 3-items set):

Again, the items already filtered out in the previous step will no longer be considered here and only the ones sustained having support threshold 30% or, more will be used in this step.

Items	Support (%)
{Bread, Jam, Snacks}	20
{Bread, Jam, Chocolate}	20
{Bread, Snacks, Chocolate}	10
{Soap, Perfume, Bread}	…
…..	…
…..	…
…..	…

Kimia Aksir

In this step, we found the support for "3-items set" are always less than the support threshold, so we will carry-forward to our next step with the 2-items sets having support greater than the support threshold.

Calculating Confidence and Lift:

Rules	Support (%)	Confidence (%)	Lift
Soap > Perfume	30	75	1.87
Perfume > Soap	30	75	1.87
Bread Jam	40	80	1.6
Jam › Bread	40	80	1.6
Bread > Snacks	30	60	1.2
Snacks › Bread	30	60	1.2
Bread › Chocolate	30	60	1.5
Chocolate › Bread	30	75	1.5
Jam › Snacks	30	60	1.2
Snacks > Jam	30	60	1.2
Beer > Snacks	30	100	2
Snacks › Beer	30	60	2

So, the finalized rules or, set of rules for the dataset will be:

Rules	Confidence (%)
Soap ⇒ Perfume	75
Perfume ⇒ Soap	75
Bread ⇒ Jam	80
Jam ⇒ Bread	80
Chocolate ⇒ Bread	75
Beer ⇒ Snacks	100

Also, from the lift values of the selected rules we can determine the interestingness of each rule.

Kimia Aksir