Applications of Market Basket Analysis

1. Credit card transactions done by a customer may be analysed. → Analysis of credit & debit purchase.
2. Phone calling patterns.
3. Fraudulent Medical insurance claims can be identified.
4. Analysis of service/products → customers who have taken executive credit card are more likely to take personal loan of $5000
5. Special combo offers to the customers on the products sold together.
6. Placement of items on the shelf
7. Inventory management.

An itemset X is frequent if X's

A set of items is referred to as an itemset.

An itemset that contains K items is a K-itemset.

The set {computer, antivirus_sw} is a 2-itemset. The occurrence frequency of an itemset is the number of transactions that contain the itemset.

This is known as support count.

An itemset K is a frequent itemset if K's support is no less than a minimum support threshold.

Frequent itemsets are used to generate association rules.

$$\text{confidence}(A \rightarrow B) \Rightarrow \frac{\text{support\_count}(A \cup B)}{\text{support\_count}(A)}$$

$(A \cup B)$ → indicates that it contains every item in A and B

In general, association rule mining can be viewed as a two-step process

1. Find all frequent itemsets :
2. Generate strong association rules from the frequent itemset.

Association Rule :

Given a set of items $I = \{I_1, I_2, \dots I_m\}$ and a database of transactions $D = \{t_1, t_2, \dots t_n\}$ where $t_i = \{I_{i_1}, I_{i_2}, I_{ik}\}$ and $I_{ij} \in I$, an association rule is an implication of the form $X \rightarrow Y$ where $X, Y \subset I$ are sets of items called itemsets and $X \cap Y = \phi$

The support(S) for an association rule $X \rightarrow Y$ is the percentage of transactions in the database that contain $X \cup Y$

confidence or strength (α) for an association rule $X \rightarrow Y$ is the ratio of the number of transactions that contain $X \cup Y$ to the number of transactions that contain X.

Subject: Data Warehousing & Mining Sem: V

Topic:

As discussed earlier, in frequent mining usually interesting association & correlation between item sets & transactional database are found.

Frequent Mining shows which items appear together in a transaction or relation.

Transaction	Items
1	{A, C, D}
2	{B, C, D}
3	{A, B, C, D}
4	{B, D}
5	{A, B, C, D}

Let's say minimum support count is 3

{A} = 3 ⟶ Frequent.

~~{A, B}~~ = {A, C} = 3 ⟶ Frequent

Support count of {A} & {A, C} are same hence we call {A} ⟶ not

a closed itemset.

similarly $\{B\} = 4$, $\{B, D\} = 4$ hence

$\{B\}$ & $\{B, D\}$ are not closed.

If you consider $\{A, B, C, D\} = 2$

not frequent as does not satify

min threshold criterion.

$\{A, C, D\} = 3 \longrightarrow$ Maximum / frequent / closed.

$\{A, B, C, D\} = 2$

~~Maximal Itemset : An itemset is closed if none of its immediate supersets have same support count as itemset~~

Closed Itemset : An itemset is closed if none of its immediate supersets have same support count same as itemset.

Subject: Data Warehousing & Mining Sem: V

Topic:

so as per the definition we need to identify the large itemsets. & validate them against minsup & minConf to classify them as strong association rules.

Finding large itemsets is easy but could be expensive. A naive approach would be to count all itemsets that appear in any transactions.

Given a set of items of size m, there are 2^m subsets. As we are not interested in empty sets, the potential number of large itemsets is $2^m - 1$

Finding Large / frequent Itemset

1. Find all the possible association rule.
2. Calculate the support & confidence of each rule generated in the above step.
3. The rules that fail the minSup & minConf are prunned from the above list.
4. As this is a time consuming approach, we need a better approach.