Association Analysis

"association" in a rule
quantify the strength of this association.
mining of these rules from a list of thousands of items
using Apriori Algorithm.

Market basket Analysis

↓

Association Rules.

helping stores cross-sell in the process.

Association – Do not tie back a users different transactions overtime to identify relationships.
– List of items with unique transaction IDs from all users are studied as one group.

Collaborative filtering – Ties back all transactions corresponding to a user ID to identify similarity between user preferences.

Association Rules
- Antecedent {items}
- Consequent . {items}

itemset – Items of both Ant. and Cons.

Various metrics to understand strength b/w Ant. and Cons.

1) Support
2) Confidence
3) Lift.

{Bread, Egg} ⟶ {milk}
Antecedant — Consequent

Itemset = {Bread, Egg, milk}

Scanned with CamScanner

$$\text{Support}(\{x\} \rightarrow \{y\}) = \frac{\text{no}(X \text{ and } Y)}{\text{Total Transactions}}$$

$$\text{Support}(\{y\}) = \frac{\text{no}\cdot(Y)}{\text{Total}}$$

$$\text{Confidence}(\{x\} \rightarrow \{y\}) = \frac{\text{no}\cdot(X \text{ and } Y)}{\text{no}\cdot(X)}$$

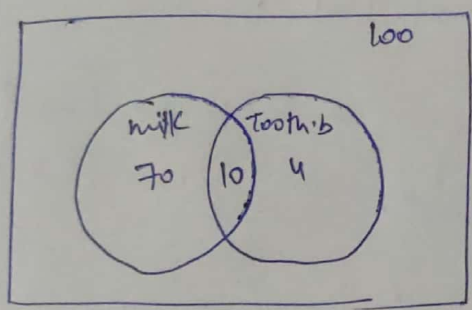
Similar to $P(y|x)$

$\{\text{Yogurt}\} \rightarrow \{\text{Milk}\}$

$\{\text{Butter}\} \rightarrow \{\text{Bread}\}$

$\{\text{Toothbrush}\} \rightarrow \{\text{Milk}\} \longrightarrow ???$

Very frequent Consequent $\Rightarrow$ high confidence



$$\text{Confidence}(\{\text{Toothbrush}\} \rightarrow \{\text{milk}\}) = \frac{10}{10+4} = 0.7$$

$$\text{Lift} = \frac{\text{Confidence}(\{x\} \rightarrow \{y\})}{\text{Fraction of } \{y\}} = \frac{0.7}{80/100} = 0.87 < 1$$

$$\text{Lift} = \frac{\text{Confidence}(\{x\} \rightarrow \{y\})}{\text{Support}(\{y\})}$$

Next step $\rightarrow$ Association Rule mining for entire list of items.

10 items $\longrightarrow$ will lead to 57000 rules.

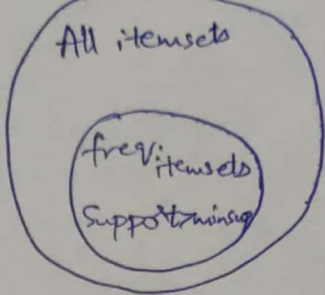
Hence $\boxed{\text{Apriori Algorithm}}$

Scanned with CamScanner

Rule generation process

1) Generate an item set.
2) Generate a rule from each itemset.

Step 1



Brute force technique → form all possible itemsets

Apriori principle
- All subsets of a frequent itemset must also be frequent.

Anti-monotone property of support.

Support({a}, {b}, {c}) < Support({a}, {b}) < Support({a}).

Helpful in pruning of infrequent itemsets.

Step 2 Generating all possible rules from the frequent itemsets.

Do a binary partition of each itemset.

Confidence of rules generated from the same itemset also follows an anti-monotone property (w.r.t consequent)

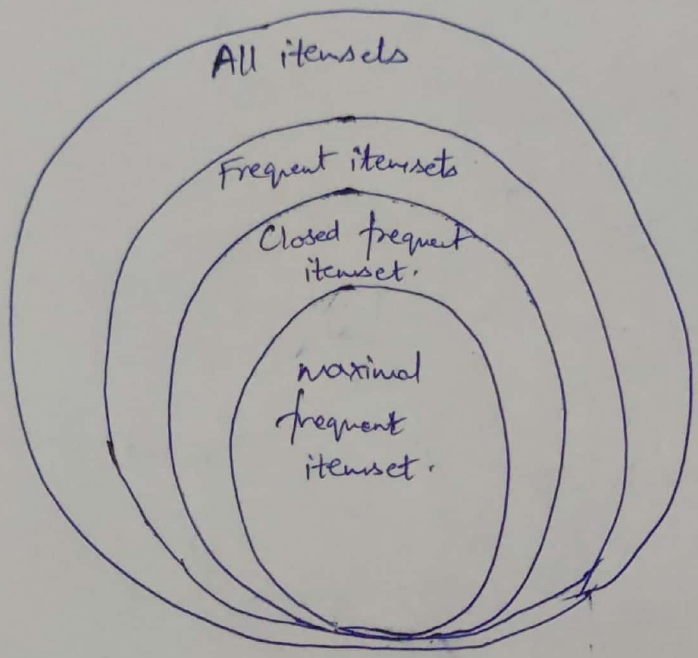
$conf(A, B, C \to D) \geq conf(B, C \to A, D) \geq conf(C \to A, B, D)$

$$conf(X \to Y) = \frac{Support(X, Y)}{Support(X)}$$

Scanned with CamScanner

Thus after step 1 and 2, we get subset of Rules.

In these subset of Rules, search for the ones which give highest lift.



max. frequent itemset

$\{x\}$ minsup ✓ $\{x, y\}$ minsup ✗ → frequent itemset with none of the immediate supersets are frequent.

Closed frequent itemset

$$Support(\{x\}) \neq Support(\{x, y\})$$

Means X is not accompanied by occurrence of y.

apyori, mlxtend.

Scanned with CamScanner

Eclat Algorithm

↳ Equivalence Class Clustering and bottom-up Lattice traversal.

Apriori algorithm works in a horizontal sense
(Breadth first Search of a graph)

Eclat algorithm works in a vertical sense
(Depth-first Search of a graph)

The basic idea is to use transaction ID sets (tidsets) intersections to compute the support value of a candidate and avoiding the generation of subsets which do not exist in the prefix tree.

FP-Growth tree → No library for this available.

Scanned with CamScanner