

Apriori 101

Apriori

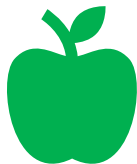
Introduction

- Algorithms for analysing itemsets for association rules
- Identifies frequent individual items and aggregates them to larger itemsets
- Assumptions
 - All subsets of frequent item sets are also frequent.
 - For an infrequent item set, all supersets are infrequent.

Apriori

Example: Itemset

- Itemset:
 - Banana (B)
 - Tomato (T)
 - Orange (O)
 - Apple (A)



Apriori

Example: Transactions

Transaction Number	Items Bought
1	B
2	T, A
3	O, A
4	T, A
5	O, T, A

Apriori

Example: Support

- Minimum Support = 2
- Candidate List

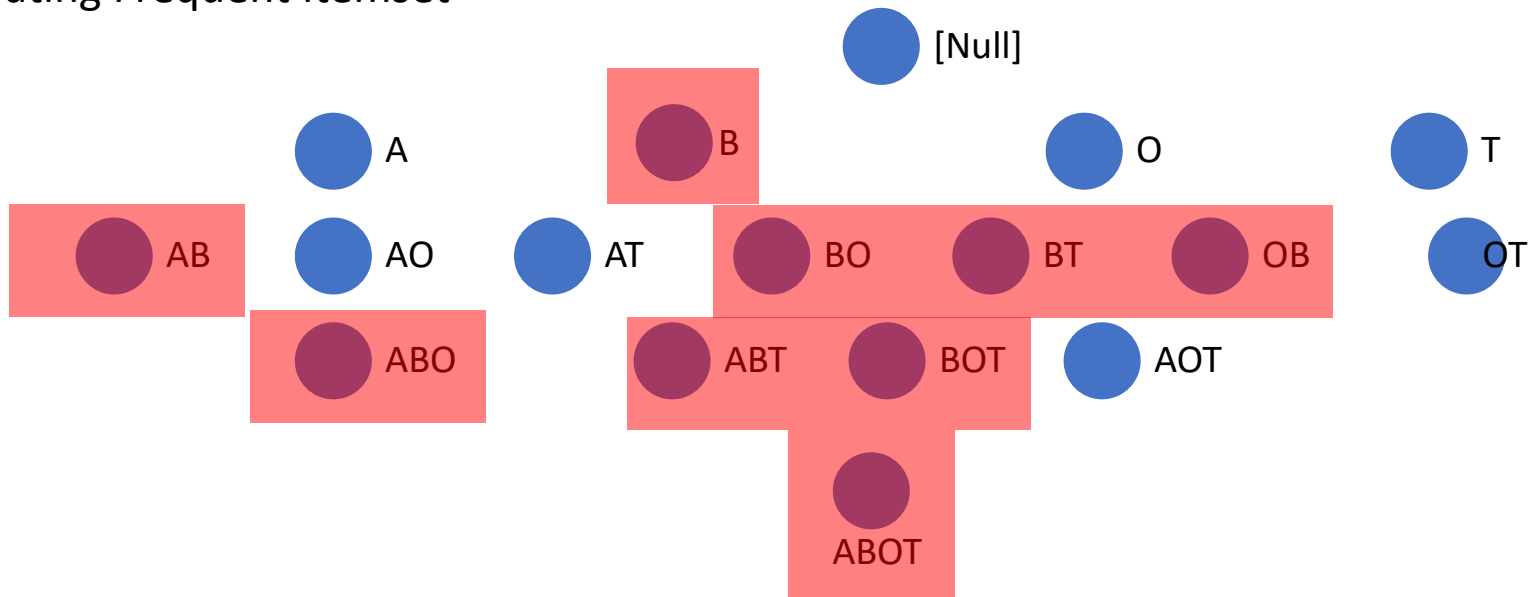
Itemset	Items Bought
A	4
B	1
O	2
T	3

- All infrequent candidates removed (B)

Apriori

Example: Support

- Generating Frequent Itemset



- B is infrequent \rightarrow All supersets are infrequent

Apriori

Example: Support

- 2-Items Itemset created

2-Items Itemset	Items Bought
A, O	2
A, T	3
O, T	1

- Removed because Min Support not fulfilled

Apriori

Example: Support

- 3-Items Itemset created

3-Items Itemset	Items Bought
A, O, T	1

- Removed because Min Support not fulfilled

Apriori

Example: Confidence

- Generate Rules

Itemset	Items Bought
A	4
O	2
T	3
A, O	2
A, T	3

- Min. Confidence = 75 %

- Remove Rules with Confidence < Minimum Confidence

Rule	Confidence
$\{A\} \rightarrow \{O\}$	$2 / 4 = 50 \%$
$\{O\} \rightarrow \{A\}$	$2 / 2 = 100 \%$
$\{A\} \rightarrow \{T\}$	$3 / 4 = 75 \%$
$\{T\} \rightarrow \{A\}$	$3 / 3 = 100 \%$

$$\text{Confidence}(X \rightarrow Y) = \frac{\text{Support}(X, Y)}{\text{Support}(X)}$$

Apriori

Example: Lift

Itemset	Items Bought	P
A	4	80%
O	2	40%
T	3	60%
A, O	2	40%
A, T	3	60%

- Total Nr. Transactions: 5

$$Lift(X \rightarrow Y) = \frac{Support(X, Y)}{Support(X) * Support(Y)}$$

Rule	P(LHS, RHS) (Denominator)	Lift
{O} → {A}	=0.4*0.8=0.32	0.4/0.32=1.25
{A} → {T}	=0.8*0.6=0.48	0.6/0.48=1.25
{T} → {A}	=0.6*0.8=0.48	0.6/0.48=1.25

P(T) P(A)

Actual probability
of A and T occurring
together

Apriori

Advantages / Disadvantages



- Simple to understand
- Simple to implement
- Finds all rules
- Intuitive results



- High computational effort