

**Sudha Amarnath**



**SAN JOSÉ STATE**  
UNIVERSITY

**Graduate and Extended Studies**

**FA19: CMPE-297 Sec 01 - Special Topics**

**Prof. Chandrasekar Vuppalapati**

## Learning Objective: Apriori Algorithm

Given the user-specified minsup = 30% and minconf = 80%, generate all association rules (**sup** is the support, and **conf** is the confidence)

Transaction	Items
T1	Noodles, Pickles, Milk
T2	Noodles, Cheese
T3	Cheese, Shoes
T4	Noodles, Pickles, Cheese
T5	Noodles, Pickles, Clothes, Cheese, Milk
T6	Pickles, Clothes, Milk
T7	Pickles, Milk, Clothes

**The association rules problem is as follows:**

Let  $I = \{i_1, i_2, \dots, i_n\}$  be a set of literals call items. Let  $D$  be a set of all transactions where each transaction  $T = \{t_1, t_2, \dots, t_n\}$  is a set of items such that  $T \subseteq I$ . Let  $X, Y$  be a set of items such that  $X, Y \subseteq I$ . An association rule is an implication in the form  $X \Rightarrow Y$ , where  $X \subset I, Y \subset I, X \cap Y = \emptyset$ .  $X$  and  $Y$  are called antecedent and consequent of the rule respectively.

**Step 1:** Scans all of the transactions in order to count the number of occurrences of each item.

Items	Count
Noodles	4
Pickles	5
Milk	4
Cheese	4
<del>Shoes</del>	<del>1</del>
Clothes	3

**Step 2:** Given the minimum support count required is 3, that is, minsup = 3. (Here, we are referring to absolute support because we are using a given support count 30%). Eliminate all items that are less than minsup.

Items	Count
Noodles	4
Pickles	5
Milk	4
Cheese	4
Clothes	3

**Step 3:** Compute 2 itemset.

Items	Count
Noodles, Pickles	3
<del>Noodles, Milk</del>	<del>2</del>
Noodles, Cheese	3
<del>Noodles, Clothes</del>	<del>1</del>
Pickles, Milk	4
<del>Pickles, Cheese</del>	<del>2</del>
Pickles, Clothes	3
<del>Milk, Cheese</del>	<del>1</del>
Milk, Clothes	3
<del>Cheese, Clothes</del>	<del>1</del>

**Step 4:** Eliminate items that are less than minimum support ( $<3$ ).

Items	Count
Noodles, Pickles	3
Noodles, Cheese	3
Pickles, Milk	4
Pickles, Clothes	3
Milk, Clothes	3

**Step 5:** Construct triplets:

Items	Count
<del>Noodles, Pickles, Cheese</del>	<del>2</del>
<del>Noodles, Pickles, Milk</del>	<del>2</del>
<del>Noodles, Pickles, Clothes</del>	<del>1</del>
Pickles, Milk, Clothes	3

**Step 6:** Eliminate items that are less than minimum support ( $<3$ ).

Items	Count
Pickles, Milk, Clothes	3

Frequent bought items: **Pickles, Milk, Clothes.**

Apriori uses a Level-wise search, where k-itemsets (An itemset that contains k items is a k-itemset) are used to explore (k+1)-itemsets.

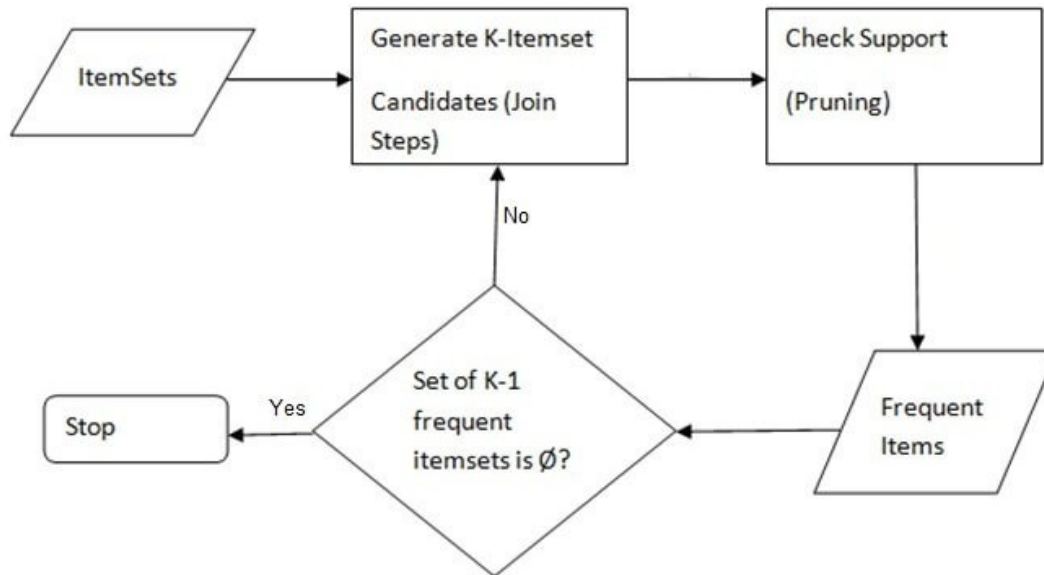
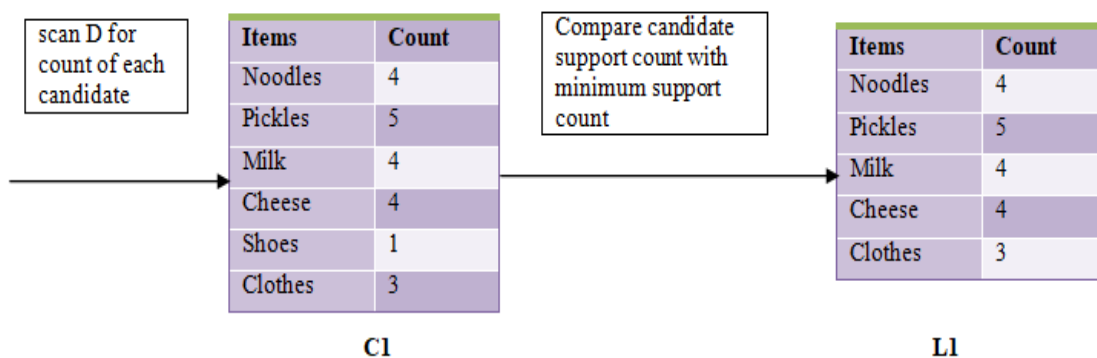


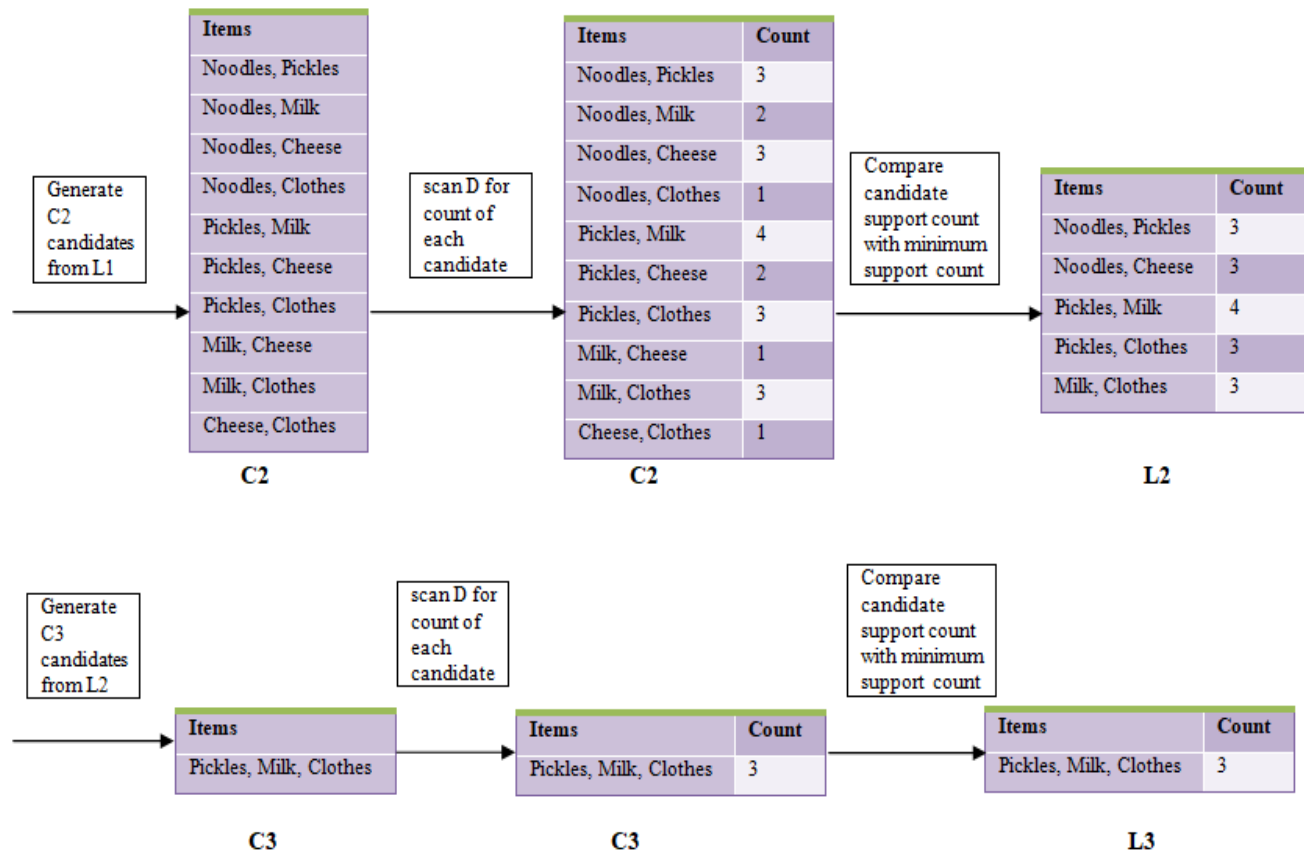
Fig 1: Apriori algorithm: Level-wise search

**Apriori property:** All nonempty subsets of a frequent itemset must also be frequent.

The two step process is followed:

- **The join step:** To find  $L_k$ , a set of candidate K-items is generated by joining  $L_{k-1}$  with itself. This is set of candidates and is denoted by  $C_k$ .
- **The prune step:**  $C_k$  is a superset of  $L_{k-1}$ , that is, its members may or may not be frequent but all of the frequent K- items are included in  $C_k$ . To reduce the size of  $C_k$ , Apriori property is used.





The final step is to generate strong association rules from frequent itemsets (where strong association rules satisfy both minimum support and minimum confidence). Association rules can be generated as follows:

- Support =  $P(A \cap B)$  = Number of transactions containing both A and B / Total number of transactions. For frequently bought items {**Pickles, Milk, Clothes**},

Support =  $3/7 = 42.84\%$  ( $> 30\%$ ) for all the associations containing these combinations.

- For each frequent itemset  $l$ , generate all nonempty subsets of  $l$ .
- For every nonempty subsets  $s$  of  $l$ , output the rule " $s \Rightarrow (l - s)$ " if  $\text{support\_count}(l) / \text{support\_count}(s) \geq \text{minconf}$  (where  $\text{minconf}$  is minimum confidence threshold).

Confidence ( $A \Rightarrow B$ ) =  $p(A|B) = \text{support\_count}(A \cap B) / \text{support\_count}(A)$

In this case, for frequently bought items  $l = \{\text{Pickles, Milk, Clothes}\}$ , we generate subsets  $s$ :

$\{\text{Pickle}\}$

$\{\text{Milk}\}$

$\{\text{Clothes}\}$

$\{\text{Pickle, Milk}\}$

$\{\text{Pickle, Clothes}\}$

$\{\text{Milk, Clothes}\}$

We calculate the Confidence for each of these subsets:

$\{\text{Pickle}\} \Rightarrow \{\text{Milk, Clothes}\}$ : Confidence is  $3/5 \times 100\% = 60\%$

$\{\text{Milk}\} \Rightarrow \{\text{Pickle, Clothes}\}$ : Confidence is  $3/4 \times 100\% = 75\%$

$\{\text{Clothes}\} \Rightarrow \{\text{Pickle, Milk}\}$ : Confidence is  $3/3 \times 100\% = 100\%$

$\{\text{Pickle, Milk}\} \Rightarrow \{\text{Clothes}\}$ : Confidence is  $3/4 \times 100\% = 75\%$

$\{\text{Pickle, Clothes}\} \Rightarrow \{\text{Milk}\}$ : Confidence is  $3/3 \times 100\% = 100\%$

$\{\text{Milk, Clothes}\} \Rightarrow \{\text{Pickle}\}$ : Confidence is  $3/3 \times 100\% = 100\%$

According to the above calculation, given the user-specified minsup= 30% and minconf= 80%, the following association rule for below subsets is valid as its support is 42.84% (> 30%) and its confidence is 100% (> 80%).

**$\{\text{Clothes}\} \Rightarrow \{\text{Pickle, Milk}\}$** : ‘Clothes’ are bought in 3 transactions and in all those transactions ‘Pickle’ and ‘Milk’ are also bought i.e. Support =  $3/7 = 42.84\%$ , Confidence =  $3/3 \times 100 = 100\%$ .

**$\{\text{Pickle, Clothes}\} \Rightarrow \{\text{Milk}\}$** : ‘Pickles’ and ‘Clothes’ are bought in 3 transactions and in all those transactions ‘Milk’ is also bought i.e. Support =  $3/7 = 42.84\%$ , Confidence =  $3/3 \times 100 = 100\%$ .

**$\{\text{Milk, Clothes}\} \Rightarrow \{\text{Pickle}\}$** : ‘Milk’ and ‘Clothes’ are bought in 3 transactions and in all those transactions ‘Pickle’ is also bought i.e. Support =  $3/7 = 42.84\%$ , Confidence =  $3/3 \times 100 = 100\%$ .