April 2018

Apriori Algorithm

- Apriori is a frequent pattern mining algorithm used in association rule learning.
- It helps uncover which items frequently co-occur in a transactional dataset.

- Used in
  - Market Basket Analysis
  - Recommender Systems
  - Inventory Mgmt
  - Web Usage Mining
  - Medical Diagnosis

- **Frequent pattern** mining is a fundamental data mining technique used to find frequently co-occurring items ⓞⓡ patterns in a dataset.

Ex: Market baskets, web logs etc

Pattern: people who buys bread often by Butter

Association Rule Learning is a rule based ML technique
- that discovers interesting relationships (association) b/w variables in large dataset.

Apriori Algorithm Steps

① Generate 1-itemsets that meet minimum support

② Generate 2-itemsets from frequent 1-itemsets
   & filter again

③ Continue to generate k-itemsets until no more frequent itemsets are found

④ Generate association rules from these itemsets.

Data Mining is the process of discovering patterns, correlations, anomalies & useful info from datasets using stats & ML

May
Week M T W T F S S   Week M T W T F S S
13/18 30   18 1 2 3 4 5 6
14 2 3 4 5 6 7 8   19 7 8 9 10 11 12 13
15 9 10 11 12 13 14 15 20 14 15 16 17 18 19 20
16 16 17 18 19 20 21 22 21 21 22 23 24 25 26 27
17 23 24 25 26 27 28 29 22 28 29 30 31

Wednesday 11
(101-264) Wk 15

Thursday 12
(102-263) Wk 15

feature	Apriori
Type	FP Mining
Data Required	Transactional
output	Frequent Itemsets, rules
pros	Simple, interpretable
cons	Slow on large datasets.

Key Metrics.

Support — How often the rule occurs in the dataset

confidence — How much more likely Y is to occur with X

problem

Generate Rules using Apriori Algorithm. Consider the values on Support = 50% & Confidence = 70%.

ID	Items
1	Bread, Butter, Jam, Milk
2	Bread, Butter, Milk
3	Bread, Juice, Curd
4	Bread, Milk, Juice
5	Butter, Milk, Juice

Step 1 : 1-Itemset Support Count (L1)

Items	Count	Support (50%) (no of occurrence/total trans)
Bread	4	4/5 = 80%
Butter	3	3/5 = 60%
X Jam	1	1/5 = 20%
milk	4	4/5 = 80
Juice	3	3/5 = 60
X Curd	1	1/5 = 20

Step2: 2-itemsets Support Count (L2)

	Item	Count	Support
✗	{Bread, Butter}	2	2/5 = 40
✓	{Bread, milk}	3	3/5 = 60
✗	{Bread, Juice}	2	2/5 = 20
✓	{Butter, milk}	3	3/5 = 60
✗	{Butter, Juice}	1	1/5 = 20
✗	{Milk, Juice}	2	2/5 = 40

13 Friday

(103-262) Wk 15

Step3: {Bread, milk} ⎫ 3-itemsets
  {Butter, milk} ⎭ ✗ {Bread, milk, Butter}

Support = 2/5 = 40%

(Checked from original ds)

Stop calculating the Support.

---

14 Saturday

Sinhala & Tamil New Year (Aluth Avurudda) / Vaisakhi / Vishu (Sri Lanka / India)

(104-261) Wk 15

Last step: Calculating the Confidence of 2-itemsets whose support is greater than 60%.

Cof (Bread, milk) = 75

Cof (milk, Bread) = 60/80 = 75%   Cof (Butter, milk) = 60/80 = 60

Cof (milk, Butter) = 60/80 = 75

$$Cof(x \rightarrow y) = \frac{S(x \cup y)}{S(x)}$$

① $Cof(Bread, milk) = \frac{S(B \cup M)}{S(B)} = \frac{60}{80} = 75\%$

Confidence Requirement is 75%.

Cof (Bread → milk) $= \frac{60}{80} = 75\%$.

① There is 75% of chances that if we buy Bread we will buy milk also.

② Cof (milk → Bread) $= 60\!/\!80 = 75\%$.

There is 75% of chances that if we buy milk we will buy Bread also.

③ Cof (Butter → milk) $= \frac{80}{80} = 100\%$.

There is 100% of chances that if We buy Butter we will buy milk.

④ . Cof (milk → Butter)

there is $\frac{100\%}{}$ of chances that if we buy milk we will buy Butter