Association Rule mining

Association Rules: Relations between Items

↳ mining mean: sets of methods to find this relations

↳ Also called: Market Basket Analysis

~~Some definitions~~

Items → $I = \{i_1, i_2, \dots, i_m\}$

Transactions → $t \subseteq I$
↳ Baskets ↳ subset

Imagine: a shop

$t1: \{bread, cheese, milk\}$
$t2: \{apple, eggs, salt, yogurt\}$
⋮
$tn: \{biscuit, eggs, milk\}$

Subsets of Items: X

$X \subseteq t$

↕

transaction

Association rule:

X & Y are subset of all items → X & Y don't share items

$X \rightarrow Y$, where $X, Y \subseteq I$, $X \cap Y = \emptyset$

↳ Subset of items X is related to subset of items Y

K-itemset ⇝ itemset that has K items

{milk, bread, cereal} is a 3-Itemset

Assume: Bread → Milk [sup = 5%, Conf = 100%]

support: chance of buying X & Y together

$sup = Pr(X \cap Y)$

Confidence: chance that if x is in t then y will be too

$Conf = Pr(Y|X)$

Rule: $X \Rightarrow Y$

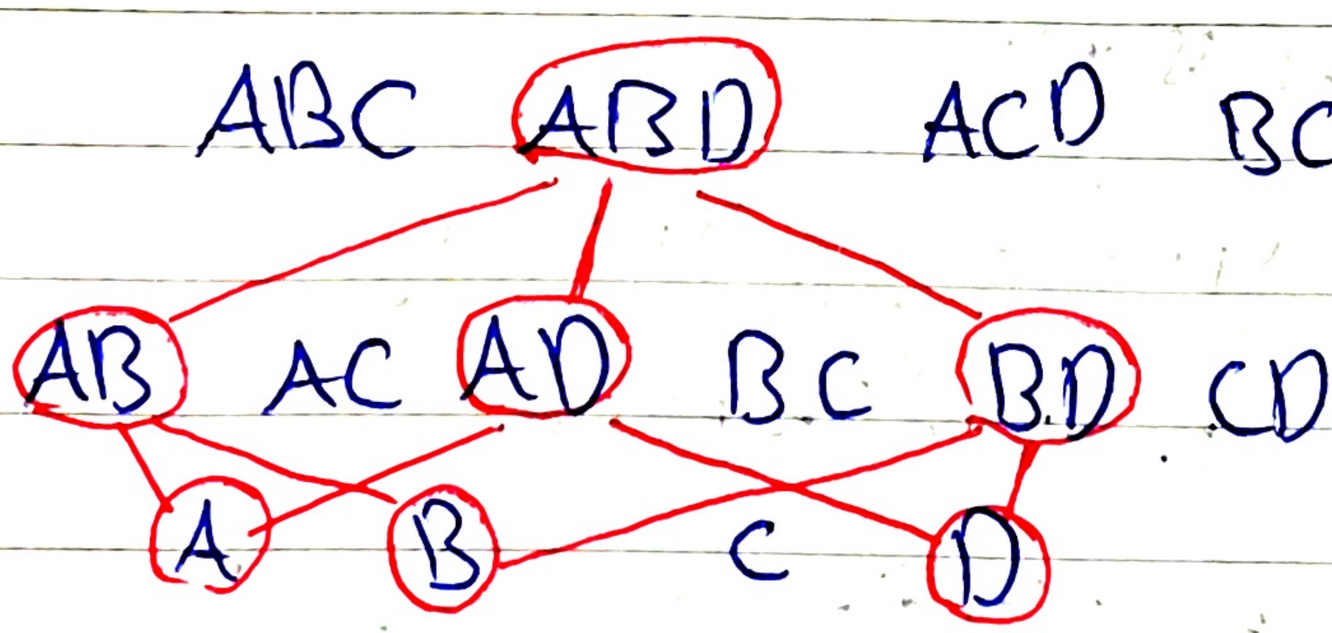
$$\text{support} = \frac{\text{frquncy}(X,Y)}{\text{sizeofall transaction}}$$

$$\text{confidence} = \frac{\text{frq}(X,Y)}{\text{frq}(X)}$$

frequent itemset: set of items that their support is bigger minsup

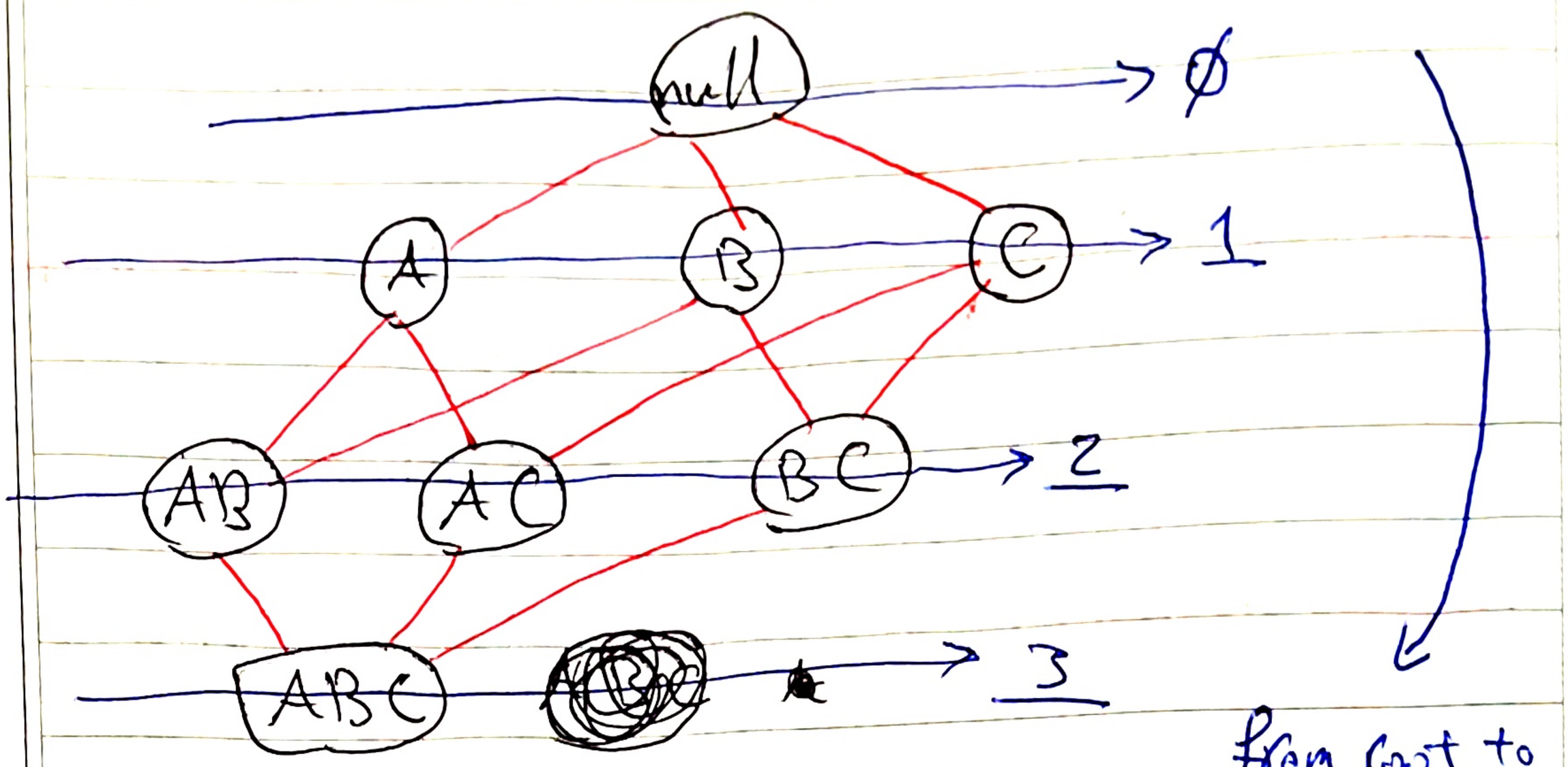
Apriori Algorithm

each subset of a frequent itemset is a frequent set itself.



↳ each of these red circles must be frequent themself

so ABD can be frequent itemset.



if A is infrequent then AB, AC, ABC will be cancelled and ignored.

Algorithm Apriori(t)

$C_1 \leftarrow$ init-pass(t);

$F_1 \leftarrow \{f \mid f \in C_1, f.count/n \geq minsup\}$;

for ($k=2$; $F_{k-1} \neq \emptyset$; $k++$) do

$C_k \leftarrow$ candidate-gen(F_{k-1});

for each transaction $t \in T$ do

for each candidate $c \in C_k$ do

if c is contained in t then

c.Count++;

end

end

end

$F_k \leftarrow \{c \in C_k \mid c.count/n \geq minsup\}$

end

return $F \leftarrow \bigcup_k F_k$;

Candidate-gen function

Function candidate-gen(F_{k-1})

$C_k \leftarrow \emptyset$;

forall $f_1, f_2 \in F_{k-1}$

with $f_1 = \{i_1, \dots, i_{k-2}, i_{k-1}\}$

and $f_2 = \{i_1, \dots, i_{k-2}, i'_{k-1}\}$

and $i_{k-1} < i'_{k-1}$ do

$c \leftarrow \{i_1, \dots, i_{k-1}, i'_{k-1}\}$; } Join f_1, f_2

$C_k \leftarrow C_k \cup \{c\}$;

for each $(k-1)$-subset s of c do

if ($s \notin f_{k-1}$) then

delete c from C_k; } delete non related combinations

end

end

end

return C_k;

eg. $F_3 = \{\{1,2,3\}, \{1,2,4\}, \{1,3,4\}, \{1,3,5\}, \{2,3,4\}\}$

After Join: $C_4 = \{\{1,2,3,4\}, \{1,3,4,5\}\}$

After pruning: $C_4 = \{\{1,2,3,4\}\}$ → 1,4,5 not in F_3

تاريخ :

موضوع :

generating rules from frequent itemsets

* for each frequent itemset X,
  for each proper nonempty subset A of X
  Let $B \leftarrow X - A$
  $A \rightarrow B$ is associative rule if:
    * $Conf(A \rightarrow B) \geq minConf$
    * $Sup(A \rightarrow B) = Sup(A \cup B) = Sup(X)$
    * $Conf(A \rightarrow B) = \frac{Sup(A \cup B)}{Sup(A)}$
  end
end
end

eg. DB

minsup = 2

~~A,~~ A, C, D	
B, C, E	
A, B, C, E	
B, E	

1st → C_1

{A}, 2
{B}, 3
{C}, 3
~~{D}, 1~~
{E}, 3

→ L_1

{A}, 2
{B}, 3
{C}, 3
{E}, 3

2nd → C_2

~~{A, B}, 1~~
{A, C}, 2
~~{A, E}, 1~~
{B, C}, 2
{B, E}, 3
{C, E}, 2

→ L_2

{B, C, E}, 2

eg. generating rule

$\{A, B, C\} \rightarrow$ (A, B)

$\{A\} \rightarrow \{B, C\}$

$\{B\} \rightarrow \{A, C\}$

$\{C\} \rightarrow \{A, B\}$

if $conf(A \rightarrow B) \geq minConf$ ⇝ it's rule

$sup(A \rightarrow B)$ is $sup(X)$

$conf(A \rightarrow B)$ is $\frac{sup(X)}{sup(A)}$

X

eg. $\{2,3,4\}$ is freq, $sup = 50\%$

(A, B)

$\{2\} \rightarrow \{3,4\} \rightarrow conf = \frac{50\%}{75\%} = 67\%$

$\{3\} \rightarrow \{2,4\} \rightarrow conf = 67\%$

$\{4\} \rightarrow \{2,3\} \rightarrow conf = 67\%$

$\{2,3\} \rightarrow \{4\} \rightarrow conf = 100\%$

$\{2,4\} \rightarrow \{3\} \rightarrow 100\%$

$\{3,4\} \rightarrow \{2\} \rightarrow 67\%$

* $2 \rightarrow sup = 75\%$
  $3 \rightarrow sup = 75\%$
  $4 \rightarrow sup = 75\%$

$\{2,3\} \rightarrow 50\%$
$\{2,4\} \rightarrow 50\%$
$\{3,4\} \rightarrow 75\%$