

FP - GROWTH Algorithm

ou voir
études

- Algorithme de génération des règles d'association.
- Cet algo a le même principe presque que l'algo a priori sauf cet algo a le rôle de éliminer (recours) les inconvénients de l'algo a priori

↳ L'algo (a priori) a 2 inconvénients

↳ Les inconvénients

1) Le nombre d'accès au base de donnée est très lent

2) Le nb des itemset candidats: base de donnée \Rightarrow nb itemset candidats

↳ L'algo FP growth

↳ générer les itemset fréquents sans utilisation des itemset candidats

↳ nb d'accès au base de donnée = 2

\Rightarrow Exemple ^{transaction}

Nb Tickets	Produits
T1	A, B, E
T2	B, D
T3	B, C
T4	A, B, D
T5	A, C
T6	B, C
T7	A, C
T8	A, B, C, E
T9	A, B, C, F

Min Support Count = 2

Support = $\frac{2}{9}$

Support = $\frac{nbTA}{nbT}$

\Rightarrow Appliquer FP - Growth

{ < > , { > } , { < > } }

Branches Branches

Phase 1 : Construction de FP: Theo

(Min Support Count = 2)

Le 1^{er} Scan de la base de données pour extraire les itemsets fréquents de taille 1.

	A	B	C	D	E	F
T1	1	1	0	0	1	
T2	0	1	0	1	0	
T3	0	1	1	0	0	
T4	1	1	0	1	0	
T5	1	0	1	0	0	
T6	0	1	1	0	0	
T7	1	0	1	0	0	
T8	1	1	1	0		
T9	1	1	1	0		

Matrice de données formelle

Entrée

Représentation

Binair

Tableaux des Supports : extraire les itemsets fréquents de taille 1

Produit	Support Count
A	6 ≥ 2 : fréquent
B	7 ≥ 2 : fréquent
C	6 ≥ 2 : fréquent
D	2 ≥ 2 : fréquent
E	2 ≥ 2 : fréquent

$$T_1 = A B E$$

$$T_2 = B D$$

$$T_3 = \dots$$

$$T_9 = A B C F$$

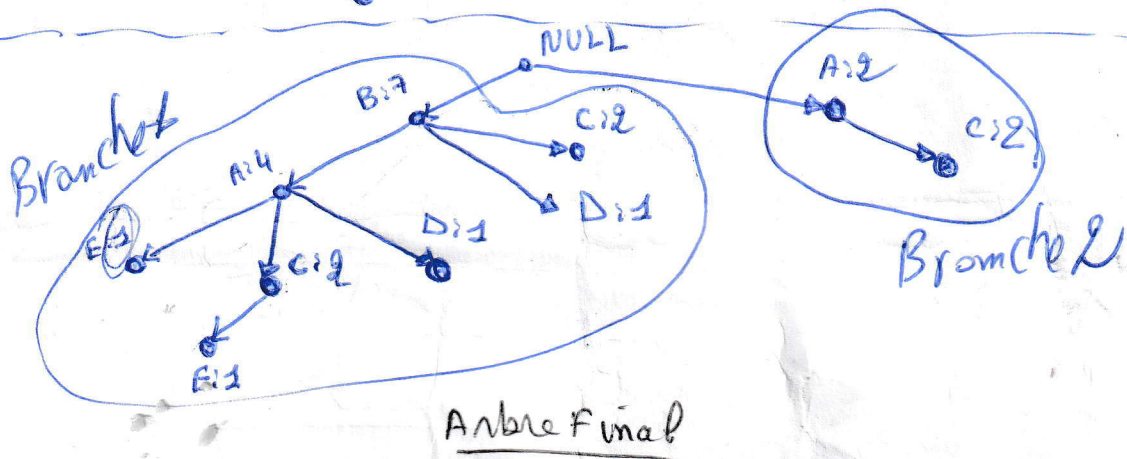
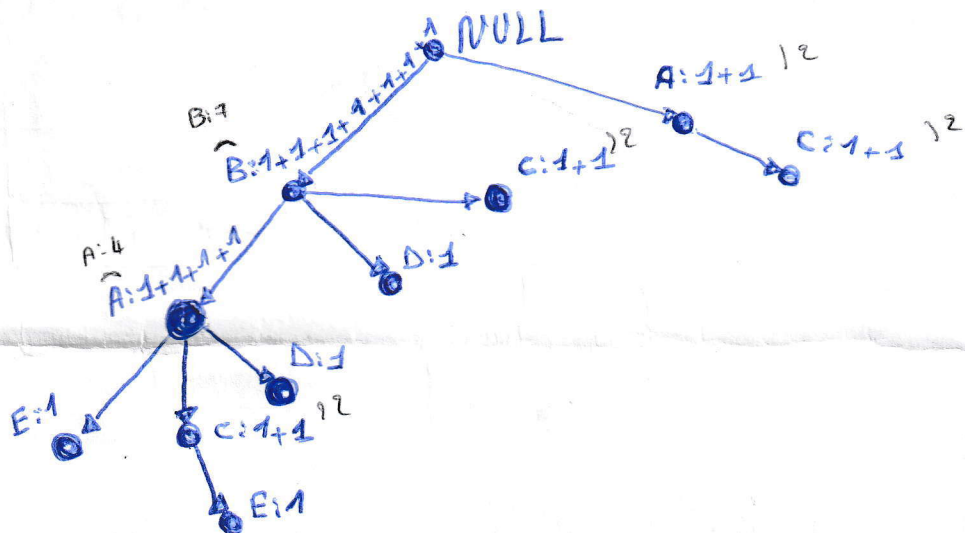
Mise à jour du Tableau des Supports de façon

Produit	Support
B	7 \leftarrow <u>Peut</u>
A	6
C	6
D	2
E	2

Ordonner les Transactions selon le Support des item sets ③

	Produits	Produits selon Poids plus fort
T1	A, B, E	B, A, E
T2	B, D	B, D
T3	B, C	B, C
T4	A, B, D	B, A, D
T5	A, C	A, C
T6	B, C	B, C
T7	A, C	A, C
T8	A, B, C, E	B, A, C, E
T9	A, B, C	B, A, C

⇒ Construction FP-Tree



Arbre Final

Phase 2: Exploration de FP-Tree

on a le Support de $E = 2$: a le Support Minimal parmi B, A, C, D .

\Rightarrow debut par E (plus bas Support)

\Rightarrow Pour l'item E: on a deux chemins:

$B \rightarrow A \rightarrow E$

$B \rightarrow A \rightarrow C \rightarrow E$

Après
FP-Tree
Déterminer les Patterns
Frequent

Nb: On ne prend pas B car B a un support = 7 : plus gros Support \Rightarrow associée automatiquement à une autre chose.

Toutes les combinaisons possibles

Item	Conditional Pattern Base	Conditional FP-Tree	Frequent Pattern
E	$\{B, A: 1\} \{B, A, C: 1\}$	$\{B: 2, A: 2, C: 1\}$	$\{E: 2\}, \{A, E: 2\}, \{B, A, E: 2\}$
D	$\{B: 1\} \{B, A: 1\}$	$\{B: 2\}$	$\{B, D: 2\}$
C	$\{B: 2\} \{A: 2\} \{B, A: 2\}$	$\{B: 2\} \{A: 2\} \{B, A: 2\}$	$\{B, C: 2\}, \{A, C: 2\}, \{B, A, C: 2\}$
A	$\{B: 4\}$	$\{B: 4\}$	$\{B, A: 4\}$

L'ensemble des itemsets fréquents:

$A, B, C, D, E, AB, AC, AE, BC, BD, BE, ABC, ABE$

Tableau trié

$4 + 2 + 4 + 6 + 2$

12
6
18

B
A
BA
BAC

B: 4
A: 4

BAE

B \rightarrow AC
AC \rightarrow B
A \rightarrow BC
BC \rightarrow A
C \rightarrow BA
BA \rightarrow C