

Série de TD N°4

Règles d'association

Exercice 1 :

On considère le tableau d'achats suivants réalisés par un client dans un supermarché.

	Produit A	Produit B	Produit C	Produit D	Produit E	Produit F	Produit G	Produit H
Achat 1	X	X	X	X				
Achat 2	X	X	X		X	X		
Achat 3	X	X	X			X	X	
Achat 4		X		X				X
Achat 5					X		X	X

- 1- Pour un support minimal 0.6, trouver tous les itemsets fréquents.
- 2- Pour une confiance minimale 0.7, trouver toutes les règles d'association de la forme $item1 \rightarrow \{item2, item3\}$
- 3- Combien de règles d'association certaines (confiance=1.0) peut-on trouver?

Exercice 2 :

Soit la table suivante des transactions d'un magasin. En utilisant l'algorithme Apriori , déterminer l'ensemble des itemsets fréquents et extraire les différentes règles possibles. (Support Minimum=10%).	Transaction ID	Items
	001	A, C, D
	002	B, C, E
	003	A, B, C, E
	004	B, E

Exercice 3 :

Soit la base de données suivante ayant cinq transactions. Soient $min_sup=60\%$ (support minimum) et $min_conf=80\%$ (confiance minimale).

(a) Trouver tous les itemsets fréquents ayant un support s supérieur ou égal à min_sup , en utilisant l'algorithme Apriori.

TID	Items_achetés
T1	K, E, Y, M, O, N
T2	K, E, Y, D, O, N
T3	M, A, K, E
T4	M, U, C, K, Y
T5	C, O, O, K, I, E

(b) Lister toutes les règles pertinentes (avec support s et confiance qui sont de la forme:

$$\forall x \in \text{transaction}, \text{achète}(x, \text{item1}) \wedge \text{achète}(x, \text{item2}) \Rightarrow \text{achète}(x, \text{item3}) [s, c]$$

Où X est la variable qui représente les clients, et $itemi$ désigne la variable qui représente les différents items (e.g., "A", "B", etc.).

Série de TD N°4 (Corrigé)

Règles d'association

Exercice 1 :

1- On commence par trouver tous les 1-itemsets (itemsets de taille 1) et leurs supports.

{A} – 3/5, support 0.6 {B} – 4/5, support 0.8 {C} – 3/5, support 0.6 {D} – 2/5, support 0.4
 {E} – 2/5, support 0.4 {F} – 2/5, support 0.4 {G} – 2/5, support 0.4 {H} – 2/5, support 0.4

On retient les itemsets ayant un support 0.6

{A} – 3/5, support 0.6 {B} – 4/5, support 0.8 {C} – 3/5, support 0.6

Sur la base des 1-itemsets fréquents, on génère les 2-itemsets.

{A, B} – 3/5, support 0.6 {B,C} – 3/5, support 0.6 {A, C} – 3/5, support 0.6

Tous ces itemsets ont un support 0.6, on les utilise pour générer les 3-itemsets. Il existe un seul itemset.

{A, B, C} – 3/5, support 0.6

Ceci termine les processus de génération des itemsets fréquents.

{A} – 3/5, support 0.6 {B} – 4/5, support 0.8 {C} – 3/5, support 0.6

{A, B} – 3/5, support 0.6 {B,C} – 3/5, support 0.6 {A, C} – 3/5, support 0.6

{A, B, C} – 3/5, support 0.6

2- Puisque la règle demandée comporte 3 items, on génère les règles d'association possibles à partir de l'itemset qui contient 3 items, c-à-d l'itemset {A, B, C} et on calcule pour chaque règle générée sa confiance.

$A \rightarrow \{B, C\}$ - confiance $3/3 = 1.0$ $C \rightarrow \{A, B\}$ - confiance $3/3 = 1.0$ $B \rightarrow \{A, C\}$ - confiance $3/4 = 0.75$

Toutes ces règles satisfont la confiance minimale de 0.7.

3- Les règles d'association certaines sont:

$A \rightarrow \{B, C\}$ - confiance $3/3 = 1.0$ et $C \rightarrow \{A, B\}$ - confiance $3/3 = 1.0$.

Exercice 2 :

1-Itemset :

Item	Fréquence	Support
A	2	2/5=0.5 (50%)
B	3	3/5=0.6 (60%)
C	3	0.6 (60%)
D	1	0.25 (25%)
E	3	0.6 (60%)

On garde tous les 1-itemset car ils ont un support > 10%.

2-Itemset :

	A	B	C	D	E
A	X	1	2	1	1
B	X	X	2	0	2
C	X	X	X	1	2
D	X	X	X	X	0

On garde les 2-itemset suivants : AB, AC,AD, AE, BC, BE, CD et CE (8 2-itemset).

3-Itemset

	AB	AC	AD	AE	BC	BE	CD	CE
AB	X	1	0	1	Déjà vu	Déjà vu	X	X
AC	X	X	1	1	Déjà vu	X	Déjà vu	Déjà vu
AD	X	X	X	0	X	X	Déjà vu	X
AE	X	X	X	X	X	Déjà vu	X	Déjà vu
BC	X	X	X	X	X	2	0	Déjà vu
BE	X	X	X	X	X	X	X	Déjà vu
CD	X	X	X	X	X	X	X	0

On garde les 5 3-itemset suivants : ABC, ABE, ACD, ACE, BCE.

4- Itemset :

	ABC	ABE	ACD	ACE	BCE
ABC	X	1	0	Déjà vu	Déjà vu
ABE	X	X	X	Déjà vu	Déjà vu
ACD	X	X	X	0	X
ACE	X	X	X	X	Déjà vu

Un seul 4-itemset, ABCE.

Génération des règles : (n'oubliez pas de calculer le support et la confiance pour chaque règle)

1- Règles générées à partir des 2-itemset : il y a $8 * 2 = 16$ règles au total.

Si A Alors B	Si B Alors A (sup ??, conf ??)
Si A Alors C	Si C Alors A
Si A Alors D	Si D Alors A
Si A Alors E	Si E Alors A
Si B Alors C	Si C Alors B
Si B Alors E	Si E Alors B
Si C Alors D	Si D Alors C
Si C Alors E	Si E Alors C

1- Règles générées à partir des 3-itemset : il y a $3 * 5 * 2 = 30$ règles au total.

Si A Alors BC	Si BC Alors A
Si A Alors BE	Si BE Alors A
Si A Alors CD	Si CD Alors A

Si	A	Alors	CE	Si	CE	Alors	A
Si	B	Alors	CE	Si	CE	Alors	B
Si	B	Alors	AC	Si	AC	Alors	B
Si	B	Alors	AE	Si	AE	Alors	B
Si	C	Alors	AD	Si	AD	Alors	C
Si	C	Alors	AE	Si	AE	Alors	C
Si	C	Alors	BE	Si	BE	Alors	C
Si	C	Alors	AB	Si	AB	Alors	C
Si	E	Alors	AB	Si	AB	Alors	E
Si	D	Alors	AC	Si	AC	Alors	D
Si	E	Alors	AC	Si	AC	Alors	E
Si	E	Alors	BC	Si	BC	Alors	E

2- Règles générées à partir du 4-itemset : il y a 14 règles au total.

Si	A	Alors	BCE	Si	BCE	Alors	A
Si	B	Alors	ACE	Si	ACE	Alors	B
Si	C	Alors	ABE	Si	ABE	Alors	C
Si	E	Alors	ABC	Si	ABC	Alors	E
Si	AB	Alors	CE	Si	CE	Alors	AB
Si	AC	Alors	BE	Si	BE	Alors	AC
Si	AE	Alors	BC	Si	BC	Alors	AE