

# Correction - Examen final

Fouille de Données

le 18 décembre 2014

1 point d'office

1. (3 points) Dans toutes les transactions, il suffit d'introduire des attributs (items) indiquant un aspect temporel ; dans une même transaction on peut en avoir 0, 1 ou plusieurs de ces nouveaux items. *Exemples :*

$\{Vendredi, Lait, Glace\}, \{Vendredi, Soir, Vin, Glace\}$

2. (6 points) Pour ce point il n'y a pas de solution unique. Ci-joint une solution basée sur l'adaptation de l'algorithme Apriori.

Rappel – la confiance d'une règle  $X \rightarrow \alpha$  se calcule avec la formule :

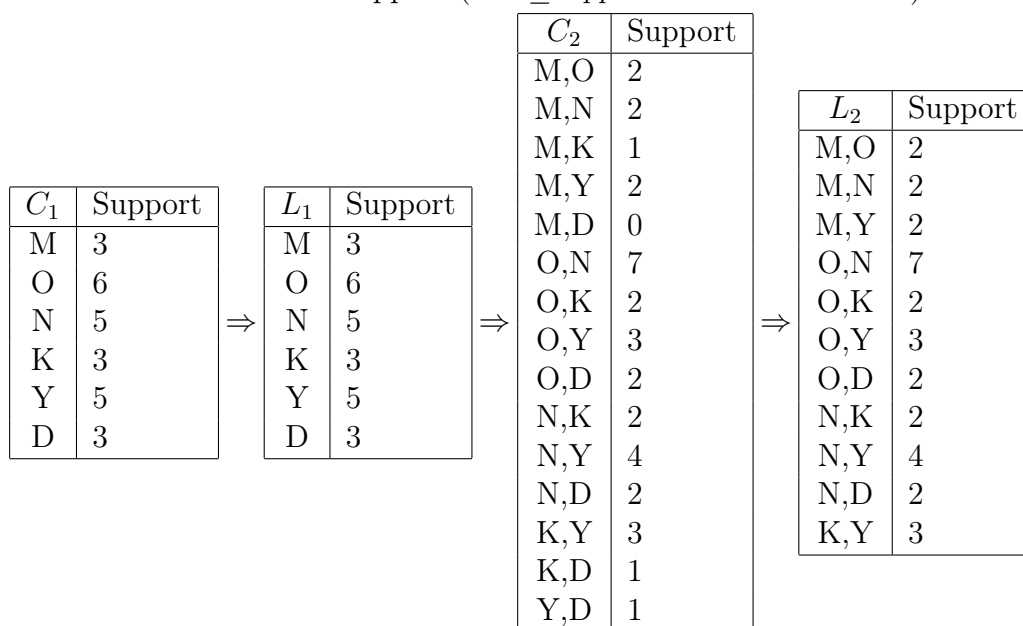
$$Confiance(X \rightarrow \alpha) = \frac{support(X \cup \alpha)}{support(\alpha)}$$

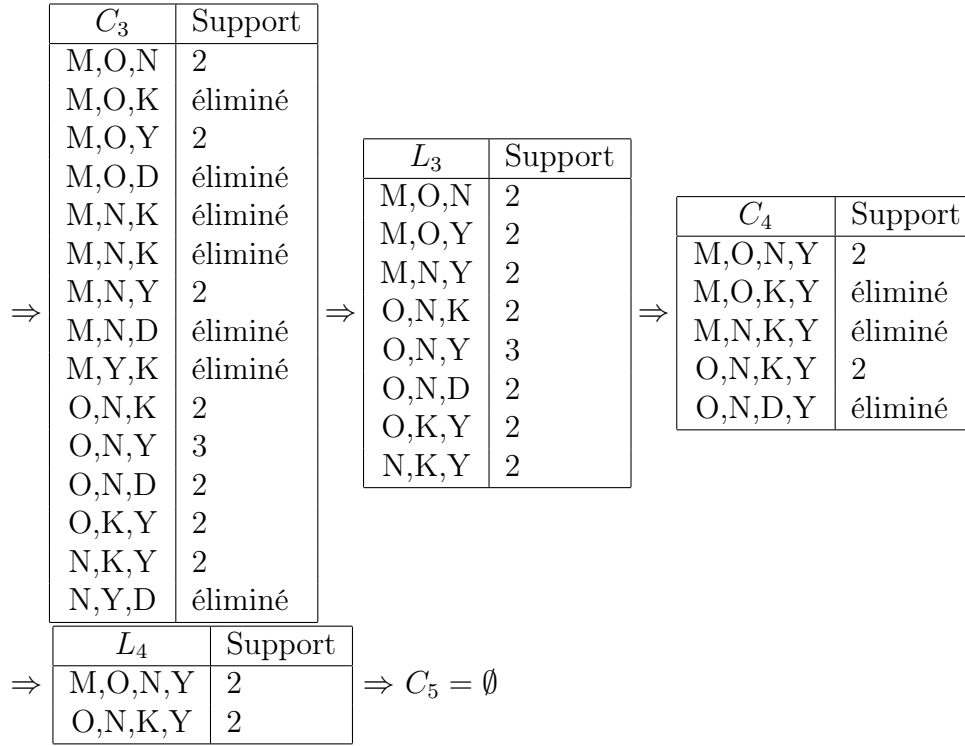
Pour en extraire des règles de cette forme il faut donc explorer les itemsets de forme  $X \cup \alpha$ . Pour calculer les itemsets fréquents d'au moins deux éléments, il faut pouvoir calculer tous les itemsets fréquents d'un élément. L'algorithme proposé est donc le suivant :

- pas 1 : on calcule tous les 1-itemsets fréquents ( $C_1$  d'abord, puis  $L_1$ )
  - pas 2 : pour un indice  $k$ ,  $k \geq 2$ , parmi les candidats  $C_k$  on garde uniquement ceux qui contiennent au moins un élément de  $Sink$
  - on répète le pas 2 jusqu'à plus de candidats  $C_k = \emptyset$  ou  $L_k = \emptyset$
3. (10 points) On remarque que  $support\_card(E) = 8$ , à savoir  $support(E) = 100\%$ , donc il suffit de faire les calculs en éliminant l'item  $E$ . Pour le nouvel ensemble de transactions :

Id	Transaction
T100'	M, O, N, K, Y
T200'	D, O, N, K, Y
T300'	K, Y
T400'	M, O, N, Y
T500'	D, O, N
T600'	M, O, D
T700'	O, N
T800'	Y, D

L'application de l'algorithme A priori sur ce nouvel ensemble de transaction et avec le même support ( $min\_support = 2$  en cardinalité) :





- (a) Les itemsets fréquents ( $min\_support = 20\%$ ) sont : M, O, N, K, Y, D, E, MO, MN, MY, ON, OK, OY, OD, NK, NY, ND, KY, EM, EO, EN, EK, EY, ED, MON, MOY, MNY, ONK, OND, MOY, MNY, EMO, EMN, EMY, EON, EOK, EOY, EOD, ENK, ENY, END, EKY, MONY, ONKY, EMONY, EONKY.
- (b) Les règles d'association qui contiennent E dans la partie droite et deux items dans la partie gauche sont issus des itemsets : MO, MN, MY, ON, OK, OY, OD, NK, NY, ND, KY. Toutes ces règles ont la confiance à 1.

Pour Y dans la partie droite, les règles qui nous intéressent doivent contenir trois ou quatre itemsets ; on examine donc uniquement les règles obtenues à partir de MOY, MNY, ONY, OKY, NKY :

Règle d'association	Confiance	LIFT	
M,O $\rightarrow$ Y	1	1,6	éliminé
M,N $\rightarrow$ Y	1	1,6	
O,N $\rightarrow$ Y	3/7	–	
O,K $\rightarrow$ Y	1	1,6	
N,K $\rightarrow$ Y	1	1,6	

On peut garder donc :

- $M, O \rightarrow Y$
- $M, N \rightarrow Y$
- $O, K \rightarrow Y$
- $N, K \rightarrow Y$

Comme  $support(E) = 100\%$  on déduit aussi comme règles avec  $E, Y$  ans la partie droite :

- $M, O \rightarrow E, Y$
- $M, N \rightarrow E, Y$
- $O, K \rightarrow E, Y$
- $N, K \rightarrow E, Y$