

Notations

You

8 février 2016

1 Frequent Item Set

1.1 Définitions

- Soit $B = \{i_1, \dots, i_m\}$ un ensemble de m items
- Un sous-ensemble $I \subseteq B$ est appelé *item set*
- Soit $T = (t_1, \dots, t_n)$ un ensemble de transactions, avec $t_k \subseteq B$

Definition 1. Une transaction $t \in T$ **couvre** un item set I ssi $I \subseteq t$

Definition 2. L'ensemble $K_T(I) = \{k \in \{1..n\} | I \subseteq t\}$ est appelé *couverture* de I par rapport à T

Definition 3. $s_T(I) = |K_T(I)|$ est le *support absolu* de I dans T

Definition 4. $\sigma_T(I) = \frac{1}{n}|K_T(I)|$ est le *support relatif* de I dans T

1.2 Frequent Item Set

Definition 5. Etant donnée :

- une valeur $s_{min} \in [0; n]$
- ou une valeur $\sigma_{min} \in [0; 1]$

l'ensemble des *Frequent Item Sets* est :

- L'ensemble $F_T(s_{min}) = \{I \subseteq B | s_T(I) \geq s_{min}\}$
- ou $\Sigma_T(\sigma_{min}) = \{I \subseteq B | \sigma_T(I) \geq \sigma_{min}\}$

1.3 Algorithme de Recherche

1.4 Idée

La recherche par force brute impossible.

Definition 6. Principe de monotonie : $\forall J \subseteq I, J \subseteq t \Rightarrow I \subseteq t$ et donc $K_T(J) \subseteq K_T(I)$

Definition 7. Corrolaire : $\forall J \subseteq I, s_T(J) \leq s_T(I)$

Si un item set est étendu, son support ne peut pas augmenter.

Definition 8. Propriété APriori : $\forall J \subseteq I, s_T(I) < s_{min} \Rightarrow s_T(J) < s_{min}$

N.B : Tout sous-ensemble d'un frequent item set est aussi un frequent item set.

1.5 Treillis

2 Règles d'association

Nous allons considérer des règles du type $X \rightarrow Y$ avec $X, Y \subseteq B$ et $X \cap Y = \emptyset$

Definition 9. Support d'une règle : $sup_T(X \rightarrow Y) = \sigma_T(X \cup Y)$ (Règle correcte)

Definition 10. Support d'une règle : $sup_T(X \rightarrow Y) = \sigma_T(X)$ (Règle acceptable)

Definition 11. Confiance d'une règle : $c_T(X \rightarrow Y) = \frac{\sigma_T(X \cup Y)}{\sigma_T(X)} = \frac{s_T(X \cup Y)}{s_T(X)}$

La confiance correspond à la probabilité : $P(Y|X)$.

Definition 12. Etant donnée :

- une valeur de support $sup_{min} \in [0; 1]$
- et une valeur de confiance $c_{min} \in [0; 1]$

l'ensemble des règles d'association est :

$$\mathcal{R} = \{R : X \rightarrow Y \mid sup_T(R) \geq sup_{min} \text{ et } c_T(R) \geq c_{min}\}$$

2.1 Découverte des règles d'association

1. Trouver les frequent item sets
2. Construire les règles d'association "par dessus" ces items set

Definition 13. Propriété de monotonie : Soit $Z = X \cap Y = X' \cap Y'$ avec $X' \subseteq X$ alors :

- $sup_T(X \Rightarrow Y) = sup_T(X' \Rightarrow Y')$
- $c_T(X \Rightarrow Y) \geq c_T(X' \Rightarrow Y')$

Déplacer un item de l'antécédent (X) vers la conséquence Y ne pas pas augmenter la confiance dans la règle.

Definition 14. Corollaire : Si une règle n'a pas la confiance minimale, aucune règle définie sur le même ensemble, mais avec une conséquence plus grande n'a besoin d'être considérée.