

Recherche des motifs fréquents

Exercice 1 Soit la table des transactions suivante :

TID	Liste des items	TID	Liste des items
T100	I1, I2, I5	T600	I2, I3
T200	I2, I4	T700	I1, I3
T300	I2, I3	T800	I1, I2, I3, I5
T400	I1, I2, I4	T900	I1, I2, I3
T500	I1, I3		

1. Représenter cette table par une table formelle.
2. Trouver, en utilisant l'algorithme Apriori, les motifs fréquents sachant que $\sigma_s = 2/9$.
3. En déduire les motifs fréquents fermés et maximaux.
4. Calculer à partir du motifs fréquent le plus long, les règles d'associations les plus fortes sachant que le seuil de confiance est de 70%.

Exercice 2 Supposons qu'il existe 6 articles numérotés de 1 à 6 (a_1, \dots, a_6), et 12 paniers numérotés de 1 à 12 (p_1, \dots, p_{12}). L'article a_i est dans le panier p_j si et seulement si i divise j , par conséquent, a_1 est dans tous les paniers et a_2 est dans les paniers de numéro pair, et ainsi de suite. Avec un support minimum de 25% et une confiance minimale de 90%, donner :

1. La base de données formelle.
2. L'ensemble des motifs fréquents.
3. Les motifs fréquents fermés et les motifs fréquents maximaux.
4. Les règles solides.

Exercice 3 Soit la table suivante :

Patient	Age	Résidus	Tension	Infection
01	Jeune	Positif	Élevée	oui
02	Jeune	Négatif	Élevée	non
03	Vieux	Positif	Élevée	non
04	Vieux	Positif	Normale	non
05	Jeune	Positif	Élevée	oui
06	Jeune	Négatif	Normale	non

1. Transformer la table en une base formelle
2. En considérant un support minimum de 50 %, trouver :
 - les motifs fréquents,

Client	M	A	R	E	I
01	moyen	moyen	village	oui	oui
02	élevé	moyen	bourg	non	non
03	faible	âgé	bourg	non	non
04	faible	moyen	bourg	oui	oui
05	moyen	jeune	ville	oui	oui
06	élevé	âgé	ville	oui	non
07	moyen	âgé	ville	oui	non
08	faible	moyen	village	non	non

- Les motifs fréquents fermés,
- Les motifs fréquents maximaux,
- Les motifs rares minimaux.

Exercice 4 Une banque dispose des informations suivantes sur un ensemble de clients : L'attribut client indique le numéro du client ; l'attribut M indique la moyenne des crédits sur le compte du client ; l'attribut A donne la tranche d'âge ; l'attribut R décrit la localité du client ; l'attribut E possède la valeur oui si le client possède un niveau d'études supérieur au bac ; l'attribut I (la classe) indique si le client effectue ses opérations de gestion de compte via Internet.

1. Donner la base de données formelle correspondant à cette base
2. Calculer les motifs fréquents correspondant à un minimum de support $\sigma_s = 0.4$
3. En déduire les motifs fréquents fermés et maximaux
4. Calculer les règles solides correspondant à un seuil de confiance = 0.9 à partir du motif fréquent le plus long

Responsable de la matière :
Dr A.Djeffal
