

L3 informatique, L3 mathématiques

Examen

Unité M.MIM5E2 : Aide à la décision et intelligence artificielle
2 h — Tous documents autorisés

Chaque candidat doit, au début de l'épreuve, porter son nom dans le coin de la copie qu'il cachera par collage après avoir été pointé. Il devra en outre porter son numéro de place sur chacune des copies, intercalaires, ou pièces annexées.

1 Fouille de données (6 points)

On considère le jeu de données de la table suivante.

	Items				
1	A	B			E
2	A	B			
3			C		
4			C		
5	A	B	C	D	
6		B	C	D	E
7			C		
8				D	E

Les lignes sont des transactions et chaque transaction représente, par exemple, des achats d'un client dans un supermarché. A , B , C , D et E désignent les items, par exemple les produits achetés par les clients.

Question 1 Rappeler la définition d'une règle d'association ainsi que les définitions des mesures de fréquence et de confiance d'une règle.

Question 2 Donner deux exemples de règles d'association avec leurs valeurs de fréquence et de confiance dans le jeu de données ci-dessus.

Question 3 En TP, nous avons codé l'algorithme de recherche de règles d'association *Apriori*. Une fois les paramètres donnés par l'utilisateur, cet algorithme donne-t-il les meilleures règles ou toutes les règles vérifiant les valeurs de ces paramètres ? Justifier.

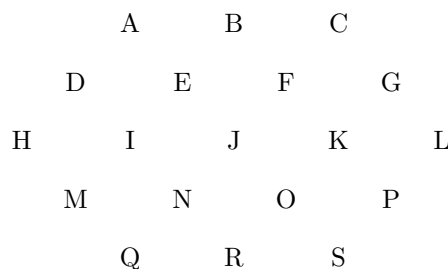
Question 4 On pose $\text{minfr} = 2$ (valeur absolue). Pour le jeu de données ci-dessus, donner la liste des motifs fréquents et leurs fréquences suivant le seuil minfr .

Question 5 Quel est le principe clé des algorithmes extrayant les motifs fréquents ?

Question 6 On pose $\text{minfr} = 3$ (valeur absolue) et $\text{minconf} = 80\%$. Pour le jeu de données ci-dessus, donner une règle d'association valide pour ces seuils. Combien ce jeu de données comporte-t-il de règles d'association valides avec ces seuils ?

2 Modélisation sous forme d'un CSP (7 points)

On veut placer les nombres de 1 à 19 selon le motif hexagonal suivant :



afin que la somme sur chaque ligne, chaque diagonale principale et chaque diagonale secondaire soit égale à 38.

Question 7 *Modéliser ce problème sous forme d'un CSP.*

Indication : Il n'est pas demandé de lister toutes les contraintes. On pourra se limiter à deux exemples pour les lignes, idem pour les diagonales principales et idem pour les diagonales secondaires.

Ce problème possède 12 solutions, reportées à la table 1 (ci-dessous). Mais ces 12 solutions représentent en fait un même hexagone. Pour éliminer les solutions « similaires », on procède en deux étapes.

Étape 1 On considère les 6 coins de la figure, c'est-à-dire les sommets A, C, L, S, Q et H, et l'on ajoute au CSP de la question 7 les cinq contraintes suivantes :

$$A < C, \quad A < L, \quad A < S, \quad A < Q \quad \text{et} \quad A < H.$$

Ce nouveau CSP ne possède plus que 2 solutions (ce sont les 2 premières lignes de la table 1).

Question 8 *Comment caractériser les solutions qui ont disparu ?*

Étape 2 On s'intéresse maintenant aux deux solutions restantes.

Question 9 *Quelle propriété relie ces deux solutions ?*

Question 10 *En déduire la contrainte qu'il faudrait ajouter pour avoir une unique solution au problème.*

A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S
3	17	18	19	7	1	11	16	2	5	6	9	12	4	8	14	10	13	15
3	19	16	17	7	2	12	18	1	5	4	10	11	6	8	13	9	14	15
9	11	18	14	6	1	17	15	8	5	7	3	13	4	2	19	10	12	16
9	14	15	11	6	8	13	18	1	5	4	10	17	7	2	12	3	19	16
10	12	16	13	4	2	19	15	8	5	7	3	14	6	1	17	9	11	18
10	13	15	12	4	8	14	16	2	5	6	9	19	7	1	11	3	17	18
15	13	10	14	8	4	12	9	6	5	2	16	11	1	7	19	18	17	3
15	14	9	13	8	6	11	10	4	5	1	18	12	2	7	17	16	19	3
16	12	10	19	2	4	13	3	7	5	8	15	17	1	6	14	18	11	9
16	19	3	12	2	7	17	10	4	5	1	18	13	8	6	11	15	14	9
18	11	9	17	1	6	14	3	7	5	8	15	19	2	4	13	16	12	10
18	17	3	11	1	7	19	9	6	5	2	16	14	8	4	12	15	13	10

TABLE 1 – Les 12 solutions du CSP de la question 7

3 Raisonnement (7 points)

On rappelle qu'on dit qu'une base de connaissances KB implique un littéral ℓ de la forme x ou $\neg x$ (« non x ») si tous les modèles de KB sont aussi des modèles de ℓ . On rappelle également que les notations $\neg, \vee, \wedge, \top, \perp$ désignent le « non » logique, le « ou » logique (inclusif), le « et » logique, le vrai et le faux, respectivement.

Par exemple, la base de connaissances $KB = (x \vee \neg y) \wedge (\neg x \vee \neg y)$ implique le littéral $\ell = \neg y$ puisque, parmi les 4 affectations des variables x, y , seules $(x \leftarrow \top, y \leftarrow \perp)$ et $(x \leftarrow \perp, y \leftarrow \perp)$ sont des modèles de KB , et les deux sont des modèles de ℓ . Une autre façon d'arriver à cette conclusion consiste à constater que toute affectation des variables doit mettre x à vrai ou x à faux (par définition d'une affectation), que dans le premier cas, il faut mettre y à faux pour obtenir un modèle de $(\neg x \vee \neg y)$, et que dans le deuxième cas, il faut à nouveau mettre y à faux, pour obtenir un modèle de $(x \vee \neg y)$; on déduit donc que tous les modèles de KB sont des modèles de ℓ , sans avoir besoin de les énumérer.

Question 11 Pour chacun des exemples suivants sur les variables x, y, z , déterminer si la base de connaissances KB implique ou non le littéral ℓ ; justifier.

1. $KB = (x \vee y) \wedge (\neg x \vee z) \wedge (\neg y \vee z)$ et $\ell = z$,
2. $KB = (x \vee y) \wedge (\neg x \vee z) \wedge (y \vee z)$ et $\ell = z$,
3. $KB = (x \vee y) \wedge (\neg x \vee z) \wedge (\neg y \vee z)$ et $\ell = \neg z$,
4. $KB = (x \vee y) \wedge (\neg y \vee z) \wedge (\neg x)$ et $\ell = z$,
5. $KB = (x) \wedge (\neg y) \wedge (\neg x \vee y) \wedge (x \vee z)$ et $\ell = z$.

On considère l'algorithme suivant pour décider l'implication. L'algorithme prend en entrée une base de connaissances KB et un littéral ℓ .

```

res ← ⊥;
pour chaque affectation a des variables faire
    si a est un modèle de KB et un modèle de ℓ alors
        res ← ⊤;
    fin
fin
return res;
```

Question 12 On suppose que la boucle « pour » énumère correctement toutes les affectations aux variables, et que la condition du « si » est correctement testée. Montrer que malgré tout, cet algorithme n'est pas correct, c'est-à-dire qu'il ne retourne pas nécessairement la bonne réponse à la question « KB implique-t-elle ℓ ? », c'est-à-dire \top si KB implique ℓ et \perp si KB n'implique pas ℓ . Le montrer en donnant son déroulement sur l'un des exemples de la question 11.

Question 13 Proposer un algorithme correct pour décider l'implication d'un littéral par une base de connaissances, basé sur l'algorithme donné ci-dessus.