
Examen

Question de cours (4 pts: 1 + 1,5 + 1,5)

1. Présentez les idées principales (sous forme de pseudocode) de l'algorithme qui vous permet de trouver le motif le plus fréquent dans une série temporelle.
2. Quels sont les principaux avantages et inconvénients de l'algorithme GSP de fouille de motifs séquentiels ?
3. Quelle est la principale différence entre les deux algorithmes GSP et SPADE ?

Exercice 1 (6 pts: 1 + 1 + 2 + 1 + 1)

Le tableau suivant donne le chiffre d'affaires (CA) y d'une entreprise (en milliers d'euros) selon le mois de l'année x :

	Année 1												Année 2			
x_i	J	F	M	A	M	J	J	A	S	O	N	D	J	F	M	A
y_i	106	107	119	121	117	121	130	143	146	146	145	155	161	173	189	194

1. Représenter graphiquement cette série (en commençant par 1 dans l'axe x).
2. Que peut-on dire de la tendance ?
3. Calculer la distance entre les 4 premiers mois des deux années en utilisant la méthode Dynamic Time Warping (DTW). Que déduisez vous ?
4. Proposer un ajustement linéaire par la méthode des moyennes mobiles à 4 périodes.
5. Proposer une prévision pour les mois de Mai et Juin de la 2ème année.

Exercice 2 (10 pts: 2 + 2 + 2 + 4)

Soit la base de données suivante représentant des événements générés par divers capteurs triés par rapport au temps.

Capteur	pas de temps	évènement
S1	1	A, B
	2	C
	3	D, E
	4	C
S2	1	A, B
	2	C, D
	3	E
S3	1	B
	2	A
	3	B
	4	D, E
S4	1	C
	2	D, E
	3	C
	4	E
S5	1	B
	2	A
	3	B, C
	4	A, D

1. Transformer la base de données en une base de données séquentielle.
2. Calculer pour le capteur S3 la matrice des taux de transitions.
3. Calculer et comparer les entropies longitudinales des séquences de données des capteurs S1 et S2.
4. En utilisant l'algorithme AprioriAll et un support minimum de 45 %, trouver tous les motifs fréquents séquentiels.

Bonne chance!

Dr D.AKROUR

Examen

Question de cours (4 pts: 1 + 1,5 + 1,5)

1. Présentez les idées principales (sous forme de pseudocode) de l'algorithme qui vous permet de trouver le motif le plus fréquent dans une série temporelle.

Réponse: (1pt)

```
Algorithme Meilleur_Motifs_C1(T, n, R)
NbreMatchMeilleurMotif ← 0;
IndiceMeilleurMotif ← NULL;
pour i = 1 jusqu'à Long(T) - n + 1 faire
    NbreMotifs ← 0;
    ListeIndices ← Nil;
    pour j = i jusqu'à Long(T) - n + 1 faire
        si Matching_Non_Trivial(T[i:i+n-1], T[j:j+n-1], R) alors
            NbreMotifs ← NbreMotifs + 1;
            ListeIndices.AjouterIndice(j);
        fin si
    fin pour
    si NbreMotifs > NbreMeilleurMotifs alors
        NbreMatchMeilleurMotif ← NbreMotifs;
        IndiceMeilleurMotif ← i;
        ListeMotifsMatching ← ListeIndices;
    fin si
fin pour
Fin.
```

2. Quels sont les principaux avantages et inconvénients de l'algorithme GSP de fouille de motifs séquentiels ?

Réponse: il réduit l'espace de recherche par rapport à AprioriAll (avantage) et scanne la base de données plusieurs fois et génère un vaste ensemble de séquences candidates (inconvénients) **(1,5pt)**

3. Quelle est la principale différence entre les deux algorithmes GSP et SPADE ?

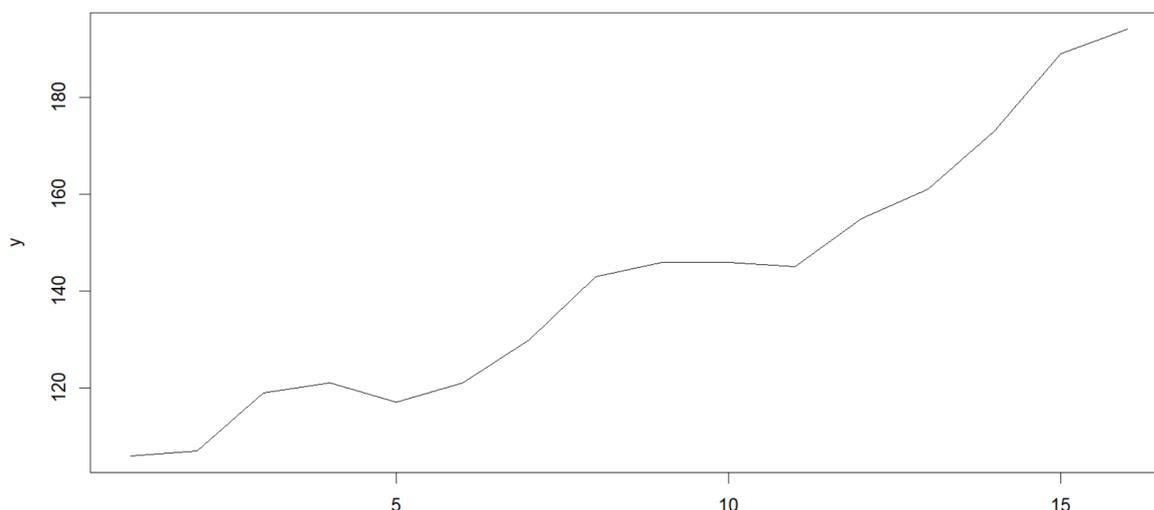
Réponse: GSP est un algorithme de recherche en largeur et SPADE est un algorithme de recherche en profondeur. **(1,5pt)**

Exercice 1 (8 pts: 1 + 1 + 2 + 1 + 1)

Le tableau suivant donne le chiffre d'affaires (CA) y d'une entreprise (en milliers d'euros) selon le mois de l'année x :

	Année 1												Année 2			
x_i	J	F	M	A	M	J	J	A	S	O	N	D	J	F	M	A
y_i	106	107	119	121	117	121	130	143	146	146	145	155	161	173	189	194

1. Représenter graphiquement cette série (en commençant par 1 dans l'axe x). **(1pt)**



2. Que peut-on dire de la tendance ?

Réponse: croissante **(1pt)**

3. Calculer la distance entre les 4 premiers mois des deux années en utilisant la méthode Dynamic Time Warping (DTW). Que déduisez vous ?

réponse:(2pt)

194	293	290	266	264
189	205	203	191	231
173	122	121	163	203
161	55	109	151	191

106 107 119 121

la distance est $1/264 = 0.003787$

4. Proposer un ajustement linéaire par la méthode des moyennes mobiles à 4 périodes.

réponse: (1pt)

	Année 1												Année 2			
x_i	J	F	M	A	M	J	J	A	S	O	N	D	J	F	M	A
y_i	106	107	119	121	117	121	130	143	146	146	145	155	161	173	189	194
<i>Moy</i>					113.25	116	119.5	122.25	127.75	135	141.25	145	148	151.75	158.5	169.5

5. Proposer une prévision pour les mois de Mai et Juin de la 2ème année.

réponse:

pour le mois de Mai: **179.25 (0.5pt)**

pour le mois de juin: **183,8125 (0.5pt)**

Exercice 2 (10 pts: 2 + 2 + 2 + 4)

Soit la base de données suivante représentant des événements générés par divers capteurs triés par rapport au temps.

Capteur	pas de temps	évènement
S1	1	A, B
	2	C
	3	D, E
	4	C
S2	1	A, B
	2	C, D
	3	E
S3	1	B
	2	A
	3	B
	4	D, E
S4	1	C
	2	D, E
	3	C
	4	E
S5	1	B
	2	A
	3	B, C
	4	A, D

1. Transformer la base de données en une base de données séquentielle.

Réponse: (2pt)

Capteur	évènement
S1	(A B) C (D E) C
S2	(A B) (C D) E
S3	B A B (D E)
S4	C (D E) C E
S5	B A (B C) (A D)

2. Calculer pour le capteur S3 la matrice des taux de transitions.

Réponse: (2pt)

S3 < B A B (D E)>

	A	B	D	E
A	0	1	0	0
B	1/3	0	1/3	1/3
D	0	0	0	0
E	0	0	0	0

3. Calculer et comparer les entropies longitudinales des séquences de données des capteurs S1 et S2.

Réponse:

S1 <(A B) C (D E) C > Entropie: $-(1/6 \log_2 1/6) * 4 + 2/6 \log_2 2/6 = 2.25$ (1pt)

S2 <(A B) (C D) E > Entropie: $-(1/5 \log_2 1/6) * 5 = 2.32$ (1pt)

4. En utilisant l'algorithme AprioriAll et un support minimum de 45 %, trouver tous les motifs fréquents séquentiels.

Réponse:

Étape L-itemset (trouver les motifs de longueur 1)

motif	A	B	C	D	E	(A B)	(D E)	(C D)	(B C)	(AD)
supp	4	4	4	5	4	2	3	1	1	1

F1= { A, B, C, D, E, (DE) } (1pt)

Étape de transformation

Mapper chaque motif de longueur 1 à un entier (0,5pt)

motif	A	B	C	D	E	(D E)
mappé à	1	2	3	4	5	6

Puis mapper chaque transaction en les séquences fréquentes qu'elle contient **(0,5pt)**

C	séquence d'évènement originale	séquence d'évènement transformé	Après mappage
S1	(A B) C (D E) C	{A, B} {C} {D, E, (D E)} {C}	{1, 2} {3} {4, 5, 6} {3}
S2	(A B) (C D) E	{A, B} {C, D} {E}	{1, 2} {3, 4} {5}
S3	B A B (D E)	{B} {A} {B} {D, E, (D E)}	{2} {1} {2} {4, 5, 6}
S4	C (D E) C E	{C} {D, E, (D E)} {C} {E}	{3} {4, 5, 6} {3} {5}
S5	B A (B C) (A D)	{B} {A} {B, C} {A, D}	{2} {1} {2, 3} {1, 4}

Trouver les motifs de longueur 2

F1	jointure	élagage	Support	F1	jointure	élagage	Support
1	11	11	1	2	21	21	2
	12	12	2		22	22	2
	13	13	3		23	23	3
	14	14	4		24	24	4
	15	15	3		25	25	3
	16	16	2		26	26	2
F1	jointure	élagage	Support	F1	jointure	élagage	Support
3	31	31	1	4	41	41	0
	32	32	0		42	42	0
	33	33	2		43	43	2
	34	34	3		44	44	0
	35	35	3		45	45	2
	36	36	2		46	46	0
F1	jointure	élagage	Support	F1	jointure	élagage	Support
5	51	51	0	6	61	61	0
	52	52	0		62	62	0
	53	53	2		63	63	2
	54	54	0		64	64	0

	55	55	1		65	65	1
	56	56	0		66	66	0

$F2 = \{13, 14, 15, 23, 24, 25, 34, 35\}$ (1pt)

Trouver les motifs de longueur 2

F2	jointure	élagage	Support	F2	jointure	élagage	Support
13	134	134	2	24			
	135	135	2	25			
14				34			
15				35			
23	234	234	2				
	235	235	2				

$F3 =$ ensemble vide (1pt)

Les motifs fréquents sont: $\{ A, B, C, D, E, (DE), AC, AD, AE, BC, BD, CD, CE \}$

Bonne chance!

Dr D.AKROUR