
Examen

Questions de cours (5 pts : 2 + 1.5 + 1.5)

1. Expliquer brièvement l'**utilité** de la méthode des moyennes mobiles.
2. Citer deux approches pour la détection des anomalies dans les séries temporelles.
3. Expliquer l'avantage principal de la mesure de similarité DTW par rapport aux autres méthodes vues en cours.

Exercice 1 (Séries temporelles 7 pts : 2 + 1.5 + 1.5 + 1 + 1)

Soit les cinq séries temporelles suivantes avec leurs classes :

t	1	2	3	4	5	Classe
S_1	2	4	3	2	2	A
S_2	0	4	3	1	2	B
S_3	4	4	2	3	1	B
S_4	1	3	0	4	3	A
S	3	1	2	3	1	?

1. Trouver la classe de la série S en utilisant la méthode 3-KPPV et la distance Euclidienne.
2. Calculer la distance DTW entre la série S et la série S_1 .
3. Trouver le chemin de coût minimal utilisée pour l'agrégation par la méthode de Rémi Gaudin vue en cours à partir de la distance DTW entre S et S_1 .
4. Dessiner sur un plot les arcs obtenus par ce chemin.
5. Donner les groupes d'arcs conclus à partir de ce chemin.

Exercice 2 (8 pts : 1.5 + 1 + 2.5 + 3)

Nous souhaitons dans cet exercice analyser l'assiduité des étudiants de la classe M2SIOD dans le module ADS durant les mois de novembre et décembre (7 semaines). Nous utilisons pour cela les listes de présences de six étudiants résumées dans la table suivante (P : présent, A :Absent) :

Étudiant	S1	S2	S3	S4	S5	S6	S7
E1	A	P	P	A	P	P	P
E2	A	A	P	P	A	P	P
E3	A	P	P	P	P	P	P
E4	A	P	A	A	A	P	P
E5	A	A	A	A	A	P	P
E6	P	A	A	A	A	P	P

1. Dessiner les graphiques suivants :
 - D-plot
 - Série des entropies
 - Séquence des états modaux
2. Dessiner les taux de transition des deux premiers étudiants.
3. Supposant que les étudiants E1, E2 et E3 ont réussi le module tandis que les étudiants E4 et E5 l'ont raté. En utilisant la méthode 3-KPPV avec la distance 2-spectrum, dites si l'étudiant E6 réussira le module ADS ou non.
4. En utilisant l'algorithme AprioriAll et un support minimum de 50 %, trouver les motifs fréquents séquentiels.

Bonne Chance

Dr A.Djeffal

Corrigé type

Questions de cours (5 pts : 2 + 1.5 + 1.5)

1. Expliquer brièvement l'**utilité** de la méthode des moyennes mobiles.
 - La méthode des moyennes mobile est utilisée pour la prédiction et pour la réduction du bruit.
2. Citer deux approches pour la détection des anomalies dans les séries temporelles.
 - Par clustering : en supposant les petits clusters de sous séries comme étant des anomalies.
 - Par motifs rares : en recherchant les motifs rares et les considérer comme des anomalies.
 - En construisant un modèle monoclasse qui reconnaît les sous séries normales et rejette les autres.
3. Expliquer l'avantage principal de la mesure de similarité DTW par rapport aux autres méthodes vues en cours.
 - L'avantage principal de la distance DTW est la capacité de gérer les décalages temporels qui peuvent éventuellement exister entre deux séries.

Exercice 1 (Séries temporelles 7 pts : 2 + 1.5 + 1.5 + 1 + 1)

1. Trouver la classe de la série S en utilisant la méthode 3-KPPV et la distance Euclidienne.
 - $D(S, S1) = \sqrt{(3-2)^2 + (1-4)^2 + (2-3)^2 + (3-2)^2 + (1-2)^2} = 3.60$
 - $D(S, S2) = \sqrt{(3-0)^2 + (1-4)^2 + (2-3)^2 + (3-1)^2 + (1-2)^2} = 4.89$
 - $D(S, S3) = \sqrt{(3-4)^2 + (1-4)^2 + (2-2)^2 + (3-3)^2 + (1-1)^2} = 3.16$
 - $D(S, S4) = \sqrt{(3-1)^2 + (1-3)^2 + (2-0)^2 + (3-4)^2 + (1-3)^2} = 4.12$ **1 pt**
 - $D(S, S1) < D(S, S3) < D(S, S4) < D(S, S2)$
 - Les 3 plus proches voisins : S3 classe B, S1 classe A, S4 classe A
 - \Rightarrow **Classe se S est A** **1 pt**
2. Calculer la distance DTW entre la série S et la série S₁.

	<i>S</i>					
<i>S1</i>		3	1	2	3	1
2		1	2	2	3	4
4		2	4	4	3	6
3		2	4	5	3	5
1		3	3	3	4	4
2		3	4	3	4	5

1 pt

$$DTW(S, S1) = 5$$

0.5 pt

3. Trouver le chemin de coût minimal utilisée pour l'agrégation par la méthode de Rémi Gaudin vue en cours à partir de la distance DTW entre S et S_1 .

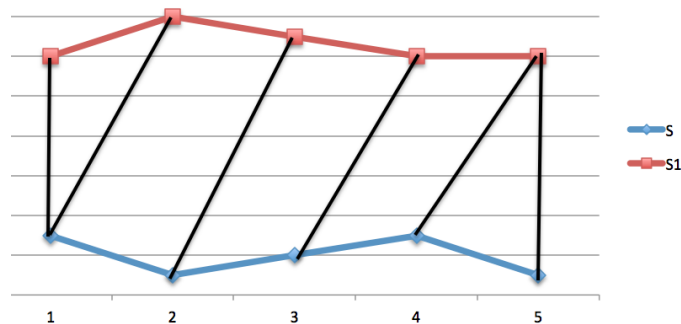
		1	2	3	4	5
	$S_1 \backslash S$	3	1	2	3	1
1	2	1	2	2	3	4
2	4	2	4	4	3	6
3	3	2	4	5	3	5
4	1	3	3	3	4	4
5	2	3	4	3	4	5

1 pt

Chemin = $\{(1, 1), (1, 2), (2, 3), (3, 4), (4, 5), (5, 5)\}$

0.5 pt

4. Dessiner sur un plot les arcs obtenus par ce chemin.



1 pt

5. Donner les groupes d'arcs conclus à partir de ce chemin.

Groupes d'arcs

- $A^1 = \{(1, 1), (1, 2)\}$
- $A^2 = \{(2, 3)\}$
- $A^3 = \{(3, 4)\}$
- $A^5 = \{(4, 5), (5, 5)\}$

1 pt

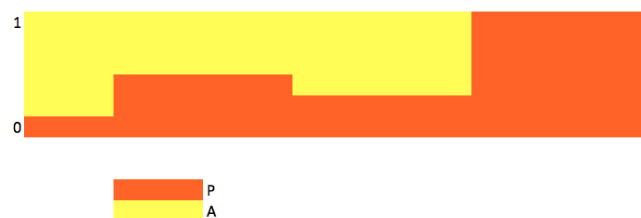
Exercice 2 (8 pts : 1.5 + 1 + 2.5 + 3)

Étudiant	S1	S2	S3	S4	S5	S6	S7
E1	A	P	P	A	P	P	P
E2	A	A	P	P	A	P	P
E3	A	P	P	P	P	P	P
E4	A	P	A	A	A	P	P
E5	A	A	A	A	A	P	P
E6	P	A	A	A	A	P	P

1. Dessiner les graphiques suivants :

– D-plot

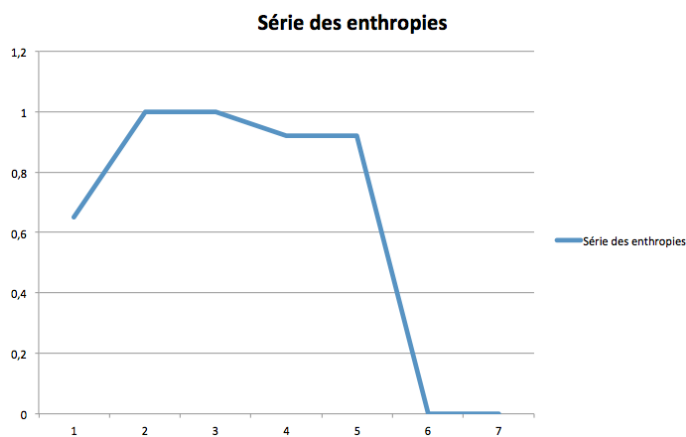
Etat \ t	S1	S2	S3	S4	S5	S6	S7
A	5/6	3/6	3/6	4/6	4/6	0/6	0/6
P	1/6	3/6	3/6	2/6	2/6	6/6	6/6



0.5 pt

– Série des entropies

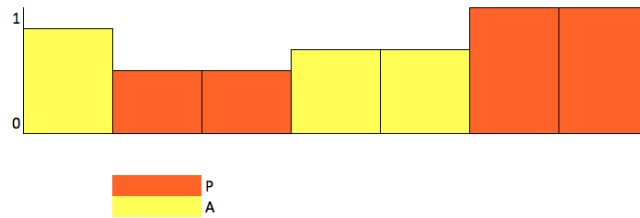
	S1	S2	S3	S4	S5	S6	S7
A	5	3	3	4	4	0	0
P	1	3	3	2	2	6	6
Entropie	0.65	1	1	0.92	0.92	1	1



0.5 pt

– Séquence des états modaux

	S1	S2	S3	S4	S5	S6	S7
Etat	A	P	P	A	A	P	P
Fréquence	5/6	3/6	3/6	4/6	4/6	6/6	6/6



0.5 pt

2. Dessiner les taux de transition des deux premiers étudiants.

– E1

	A	P
A	0/2	2/2
P	1/4	3/4

0.5 pt

– E2

	A	P
A	1/3	2/3
P	1/3	2/3

0.5 pt

3. Supposant que les étudiants E1, E2 et E3 ont réussi le module tandis que les étudiants E4 et E5 l'ont raté. En utilisant la méthode 3-KPPV avec la distance 2-spectrum, dites si l'étudiant E6 réussira le module ADS ou non.

– Vecteurs 2-spectrums

	AA	AP	PA	PP
E1	0	2	1	3
E2	1	2	1	2
E3	0	1	0	5
E4	2	2	1	1
E5	4	1	0	1
E6	3	1	1	1

0.75 pt

– Distances 2-spectrum

- $Distance(E6, E1) = 3 \times 0 + 1 \times 2 + 1 \times 1 + 1 \times 3 = 6$
- $Distance(E6, E2) = 3 \times 1 + 1 \times 2 + 1 \times 1 + 1 \times 2 = 8$
- $Distance(E6, E3) = 3 \times 0 + 1 \times 1 + 1 \times 0 + 1 \times 5 = 6$
- $Distance(E6, E4) = 3 \times 2 + 1 \times 2 + 1 \times 1 + 1 \times 1 = 10$
- $Distance(E6, E5) = 3 \times 4 + 1 \times 1 + 1 \times 0 + 1 \times 1 = 14$

1.25 pt

– Les 3 plus proche voisins à E6 sont E5 (Raté), E4 (Raté) et E2 (Réussi)

– La classe de E6 et donc **Raté**

0.5 pt

4. En utilisant l'algorithme AprioriAll et un support minimum de 50 %, trouver les motifs fréquents séquentiels.

Étudiant	S1	S2	S3	S4	S5	S6	S7
E1	A	P	P	A	P	P	P
E2	A	A	P	P	A	P	P
E3	A	P	P	P	P	P	P
E4	A	P	A	A	A	P	P
E5	A	A	A	A	A	P	P
E6	P	A	A	A	A	P	P

- La base est déjà séquentielle
- Motifs de longueur 1

Motif	fréquence
A	6/6
P	6/6

$$F_1 = \{A, P\}$$

0.5 pt

- Le mapping n'est pas nécessaire car on a des motifs de longueur 1 simple
- AprioriAll : Séquences candidates de longueur 2

Motif	Supp
AA	5/6
AP	6/6
PA	4/6
PP	6/6

$$F_2 = \{AA, AP, PA, PP\}$$

0.5 pt

- Candidats de longueur 3

2-Itemset fréquent	Supp	Candidat 3-itemset après jointure	Candidat 3-itemset après élagage	Supp
AA	4	AAA	AAA	4/6
		AAP	AAP	4/6
AP	6	APA	APA	3/6
		APP	APP	6/6
PA	4	PAA	PAA	2/6
		PAP	PAP	4/6
PP	6	PPA	PPA	2/6
		PPP	PPP	5/6

$$F_3 = \{AAA, AAP, APA, APP, PAP, PPP\}$$

0.5 pt

- Candidats de longueur 4

3-Itemset fréquent	Supp	Candidat 4-itemset après jointure	Candidat 4-itemset après élagage	Supp
AAA	4	AAAA	AAAA	3/6
		AAAP	AAAP	4/6
AAP	4	AAPA	AAPA	1/6
		AAPP	AAPP	5/6
APA	3	APAA		
		APAP	APAP	3/6
APP	6	APPP	APPP	4/6
		APPA		
PAP	4	PAPP	PAPP	4/6
PPP	6	PPPP	PPPP	3/6

$$F_4 = \{AAAA, AAAP, AAPP, APAP, APPP, PAPP, PPPP\}$$

0.5 pt

– Candidats de longueur 5

4-Itemset fréquent	Supp	Candidat 5-itemset après jointure	Candidat 5-itemset après élagage	Supp
AAAA	3	AAAAA	AAAAA	1/6
		AAAAP	AAAAP	3/6
AAAP	4	AAAPP	AAAPP	4/6
		AAAPA		
AAPP	4	AAPPP	AAPPP	2/6
APAP	3	APAPP	APAPP	3/6
APPP	6	APPPP	APPPP	3/6
PAPP	4	PAPPP	PAPPP	1/6
PPPP	3	PPPPP	PPPPP	2/6

$$F_5 = \{AAAAP, AAAPP, APAPP, APPPP\}$$

0.5 pt

– Candidats de longueur 6

5-Itemset fréquent	Supp	Candidat 6-itemset après jointure	Candidat 6-itemset après élagage	Supp
AAAAP	3	AAAAPP	AAAAPP	3/6
AAAPP	4	AAAPPP		
APAPP	3	APAPPP		
APPPP	3	APPPPP		

$$F_6 = \{AAAAPP\}$$

0.5 pt

– Candidats de longueur 7

6-Itemset fréquent	Supp	Candidat 7-itemset après jointure	Candidat 7-itemset après élagage	Supp
AAAAPP	3	AAAAPPP		

$$F_7 = \{\phi\}$$

– $F = \{A, P, AA, AP, PA, PP, AAA, AAP, APA, APP, PAP, PPP, AAAA, AAAP, AAPP, APAP, APPP, PAPP, PPPP, AAAAP, AAAPP, APAPP, APPPP, AAAAPP\}$