
Examen

Barème: 2 + 6 + 5 + 6 + 1

Nous souhaitons analyser les mots arabes coraniques suivants :

- | | |
|-------------|-----------|
| 1. كفيٓناك | 5. عاكفين |
| 2. سيكفيكهم | 6. يكيكيم |
| 3. كفوا | 7. كفي |
| 4. كفيلا | 8. كفيه |

Questions :

1. Dessiner le d-plot de ces mots.
2. Calculer les distances sacs de caractères, LLCP et LLCS entre les deux mots n° 2 et 6.
3. Calculer l'entropies et la turbulence du mot n° 6 .
4. En utilisant l'algorithme AprioriAll avec un support minimum de 30 %, trouver les motifs fréquents séquentiels en détaillant les différentes étapes.
5. En déduire les séquences fréquentes maximales.

Bonne Chance

Pr A.Djeffal

Corrigé type

1. Dessiner le d-plot de ces mots.

On prends par exemple l'ordre suivant $\text{ى ع ل و ا ن م ه ف ك ي س}$

8	ع	ا	ى						
7	ك	ف	و	ل					
6	ك	ف	ف	ا					
5	ك	ف	ك	ن	ا				
4	ك	ف	ك	ه	ا	ن			
3	ك	ف	ي	ف	ك	م			
2	ي	ك	ي	ف	ي	ك			
1	س	ي	ي	ي	ي	ك	ه	م	
	1	2	3	4	5	6	7	8	

2pts

2. Calculer les distances sacs de caractères, LLCP et LLCS entre les deux mots n° 2 et 6.

E_2 : سيكفيكهم

E_6 : يكفيكم

– Sac de caractères

– Représentation en sac de caractères

	س	ي	ك	ف	ه	م
E_2	1	2	2	1	1	1
E_6	0	2	2	1	0	1

$$\text{Distance} = \frac{1 \times 0 + 2 \times 2 + 2 \times 2 + 1 \times 1 + 1 \times 0 + 1 \times 1}{\sqrt{(1+4+4+1+1+1) \times (4+4+1+1)}} = \frac{10}{10.95} = 0.91 \quad \text{(2 pts)} \quad \text{2pts}$$

– LCP = ""

$$\text{LLCP}(\text{سيكفيكهم}, \text{يكفيكم}) = 0 \quad \text{2pts}$$

– LCS = مكيفكي

$$\text{LLCS}(\text{سيكفيكهم}, \text{يكفيكم}) = 6 \quad \text{2pts}$$

3. Calculer l'entropies et la turbulence du mot n° 6 .

(a) Entropie (يكفيكم) = $-2/6 \log_2(2/6) - 1/6 \log_2(1/6) - 1/6 \log_2(2/6) - 1/6 \log_2(1/6) = 1.04 + 0.82 = 1.86$

(b) Turbulence :

– Nombre de sous séquences ($\Phi(x)$) = 4 (Long 1) + 11 (Long 2) + 18 (long 3) + 15 (long 4) + 6 (long 5) + 1 (long 6) = 55

– Vecteur des durées passées dans les états = [1,1,1,1,1,1]

– Moyenne $\bar{X} = 1$

– Variance $s_t^2(x) = 0$

– $\bar{t} = 6/4 = 1.5$

– $s_{t,max}^2(x) = (6 - 1)(1 - 1.5)^2 = 1.25$

– Turbulence $T(x) = \log_2(\phi(x) \frac{s_{t,max}^2(x)+1}{s_t^2(x)+1}) = \log_2(55 \times \frac{1.25+1}{0+1}) = 6.95$

2pts

4. En utilisant l'algorithme AprioriAll avec un support minimum de 30 %, trouver les motifs fréquents séquentiels en détaillant les différentes étapes.

item En utilisant l'algorithme AprioriAll avec un support minimum de 40 %, trouver les motifs fréquents séquentiels en détaillant les différentes étapes.

- Support min = 30% \Rightarrow Fréquence min = $8 \times 0.3 = 2.4 \approx 3$ (0.5 pt)

- Motifs de longueur 1 = {س : 1, ي : 6, ك : 8, ف : 8, ه : 2, م : 2, ن : 2, ا : 4, و : 1, ل : 1, ع : 1, ى : 1}

- $F_1 = \{س : 1, ي : 6, ك : 8, ف : 8, ا : 4\}$ 2pts

- Pas besoin de faire le mapage puisque les motifs de longueur 1 sont simples

- Candidats de longueur 2

Motifs L2	Freq	Motifs L2	Freq
يي	2	في	6
يك	3	فك	3
يف	2	فف	0
يا	2	فا	3
كي	6	اي	1
كك	2	اك	2
كف	8	اف	1
كا	3	اا	0

- $F_2 = \{يك : 3, كي : 6, كف : 8, كا : 3, في : 6, فك : 3, فا : 3\}$ 2pts

- Candidats de longueur 3

Motif L2	Jointure	Elagage	Fréq	Motif L2	Jointure	Elagage	Fréq
يك	يكك	x		في	فيي	x	2
كي	ككي	x			فيك	x	
	كيف	x			فيا	x	
	كيا	x		فك	فكي	فكي	0
كف	كفي	كفي	6		فكك		
	كفف	x			فكا	فكا	0
	كفا	كفا	3	فا	فاي	x	
كا	كاي	x			فاك	x	
	كاف	x			فاا	x	
	كاا	x					

- $F_3 = \{كفي : 6, كفا : 3\}$ 1pt

- Candidats de longueur 4

Motif L3	Jointure	Elagage	Fréquence
كفي	كفيي	x	
	كفيا	x	
كفا	كفاي	x	
	كفاا	x	

$$- F_4 = \{\Phi\}$$

$$- F = \{ي:6, ك:8, ف:8, ا:4, ي:3, ك:6, ف:8, ك:3, في:6, ف:3, فا:3, كفي:6, كفا:3\}$$

5. En déduire les séquences fréquentes maximales.

$$Fmax = \{ي:3, ف:3, كفي:6, كفا:3\}$$

0.5pt