

RATTRAPAGE DE WEB MINING

Code UE : SIN 553

EPREUVE THEORIQUE

Par : Pr Kaladzavi / M. Touza Isaac

Durée : 3h

Exercice 1 : Web Content Mining

7pts

Vous disposez sur un site web d'un ensemble de données représentant les symptômes de patients atteints de différentes maladies. Chaque patient est décrit par les symptômes qu'il présente. Utilisez l'algorithme Apriori pour trouver les ensembles de symptômes fréquents et générer des règles d'association basées sur l'ensemble de données donné.

Patient 1 : {Fièvre, Maux de tête, Fatigue}

Patient 2 : {Maux de tête, Toux, Frissons, Fatigue}

Patient 3 : {Fièvre, Toux, Douleurs musculaires}

Patient 4 : {Fièvre, Maux de tête, Toux, Frissons}

Patient 5 : {Toux, Frissons, Fatigue}

1. Donner une représentation binaire de ces données. **2pts**
2. Utilisez l'algorithme Apriori pour trouver tous les ensembles fréquents avec un support minimum de 2 patients. **2pts**
3. En utilisant les ensembles fréquents trouvés à l'étape précédente, générez toutes les règles d'association possibles avec une confiance minimale de 70%. **2pts**
4. Interprétez les règles d'association intéressantes trouvées en termes de symptômes de maladies. **1pt**

Exercice 3 : Le Web Structure Mining

7pts

Considérons un site web contenant 5 pages, représentées par les identifiants de page suivants : P1, P2, P3, P4 et P5. Les liens entre ces pages sont les suivants :

- P1 contient des liens vers P2 et P3.
- P2 contient un lien vers P4.
- P3 contient un lien vers P4.
- P4 contient des liens vers P2 et P5.
- P5 contient un lien vers P3.

1. Dessinez un graphe où chaque page est représentée par un nœud, et les liens entre les pages sont représentés par des arêtes. Utilisez les informations fournies pour connecter les nœuds correspondant aux pages concernées. **2pts**
2. Quel nom donne-t-on à cette représentation ? **1pt**
3. Construisez la matrice d'adjacence A en utilisant la représentation du graphe. La matrice d'adjacence est une matrice carrée où chaque élément $A[i, j]$ indique s'il existe un lien de la page i à la page j (1 pour un lien existant, 0 sinon). **2pts**

4. Utilisez une méthode de clustering, pour regrouper les pages en fonction de leurs liens. Choisissez le nombre de clusters à 2. **2pts**

Exercice 3 : Web Usage Mining

6pts

Considérons l'utilisateur **u** qui a généré la session serveur suivante : $v = \{u, 16 : 09 : 10, < (A, 16 : 09 : 10), (B, 16 : 09 : 43), (C, 16 : 12 : 02), (A, 18 : 32 : 02), (C, 18 : 33 : 05), (E, 18 : 47 : 12), (C, 18 : 48 : 20), (H, 19 : 15 : 49), (C, 19 : 51 : 32) >\}$

1. Définir les notions suivantes :

3pts

- Utilisateur
- Session
- Visite

2. En considérant le seuil temporel standard $\Delta t = 30$ minutes, déterminer toutes les visites de l'utilisateur **u**. **3pts**