

Sécurité Appliquée

Protection de la vie privée-PVP

Jean-François COUCHOT
couchot@arobase.femto-st.fr

16 novembre 2020

1 *l*-diversité

La table 1 reprend un jeu de données minime issu de ¹

QID			SA	
Gender	ZIP Code	Age	Disease	Salary
Male	400071	35	bronchitis	10k
Male	400182	37	pneumonia	11k
Male	400095	39	stomach cancer	12k
Female	440672	54	gastritis	12k
Female	440123	58	Flu	15k
Male	440893	54	bronchitis	16k
Male	400022	41	gastric ulcer	16k
Male	400135	46	gastritis	17k
Female	400182	44	stomach cancer	18k

TABLE 1 – Dataset original

Gender	Zip Code	Age	Disease	Salary
*	400***	[30, 40[bronchitis	1k0
*	400***	[30,40[pneumonia	11k
*	400***	[30, 40[stomach cancer	12k
*	400***	[40, 50[stomach cancer	18k
*	400***	[40, 50[gastric ulcer	16k
*	400***	[40, 50[gastritis	17k
*	440***	[50, 60[gastritis	12k
*	440***	[50, 60[Flu	15k
*	440***	[50, 60[bronchitis	16k

TABLE 2 – Une version 3-diverse

- Exercice 1.1.**
1. *Quelle stratégie de généralisation proposez-vous ?*
 2. *On considère la proposition donnée dans le tableau 2.*
 - (a) *Est-elle 3-anonyme ?*
 - (b) *Pourquoi est-elle 3-diverse ?*
 - (c) *Quelle est sa valeur de Loss.*
 - (d) *Trouver le coefficient c de $(c, 3)$ -diversité récursive.*

1. Elabd, E., Abdulkader, H., & Mubark, A. (2015). L-diversity-based semantic anonymization for data publishing. *IJ Information Technology and Computer Science (IJITCS)*, 10, 1-7.

2 t -proximité

Exercice 2.1. On reprend le même jeu de données que précédemment.

1. Calculer l'indice de proximité de la proposition donnée.
2. Existerait-il une autre proposition (différente de la généralisation extrême de tous les QID) qui aurait un indice t de proximité plus faible ? Dans ce cas, quel serait sa valeur de Loss ?

3 Distances entre bases de données, bases voisines

Dans cette section, on nomme D_1 la base de données donnée dans le tableau 1.

Exercice 3.1 (Insertion d'une ligne). Soit D_2 la base obtenue en ajoutant à D_1 la ligne suivante.

Gender	Zip Code	Age	Disease	Salary
Female	400022	38	gastric ulcer	13k

1. Exprimer les bases D_1 et D_2 comme deux vecteurs de 6 éléments.
2. Calculer $\|D_1\|_1$, $\|D_2\|_1$ et $\|D_1 - D_2\|_1$. Lorsque la distance entre deux bases vaut l'unité, on dit que ces deux bases sont voisines.
3. Dans ces deux bases, on voudrait compter le nombre de personnes qui ont des soucis gastriques, c.-à-d. qui ont l'une des maladies suivantes : stomach cancer, gastris ou gastric ulcer.
 - (a) Exprimer cette requête S comme un vecteur de 6 bits, en reprenant le même ordre que dans la première question.
 - (b) A l'aide d'un produit scalaire entre 2 vecteurs, répondre à cette question pour D_1 et D_2 .
4. Calculer $\Delta_{Q_S} = |Q_S(D_1) - Q_S(D_2)|$.
5. La quantité Δ_{Q_S} dépend-elle des valeurs de la ligne ajoutée.

Exercice 3.2 (Modification d'une ligne). Soit D_3 la base obtenue en modifiant dans D_1 la ligne

Gender	Zip Code	Age	Disease	Salary
Female	440672	54	gastris	12k

par

Gender	Zip Code	Age	Disease	Salary
Female	440672	54	gastric ulcer	12k

1. Calculer $\|D_3\|_1$ et $\|D_1 - D_3\|_1$. Comparer cette dernière valeur à celle obtenue dans l'exercice précédent. D_3 et D_1 sont-elles voisines ?
2. Calculer $|Q_S(D_1) - Q_S(D_3)|$.