

Travaux Dirigés n° 1 : CAH-Association

Objectifs : pratiques de quelques algorithmes de base en fouille de données

1 CAH

On considère un ensemble $E = \{w_1, \dots, w_7\}$ de 7 individus, ainsi qu'une application d définie sur l'ensemble des couples de E par le tableau 1 ci-dessous (par exemple, $d(w_2, w_4) = 3.5$).

d	w_1	w_2	w_3	w_4	w_5	w_6	w_7
w_1	0	2	4.5	5.5	7.5	9.5	4
w_2	2	0	2.5	3.5	5.5	7.5	4
w_3	4.5	2.5	0	3	5	7	6.5
w_4	5.5	3.5	3	0	2	4	7.5
w_5	7.5	5.5	5	2	0	4	9.5
w_6	9.5	7.5	7	4	4	0	5.5
w_7	4	4	6.5	7.5	9.5	5.5	0

TABLE 1 – Données pour l'exercice 1.

1. L'application d est-elle une dissimilarité ? une distance ultramétrique ?
(Vous justifierez dans chaque cas votre réponse).
2. Construire et représenter graphiquement les deux hiérarchies indicées suivantes obtenues chacune par une Classification Ascendante Hiérarchique (C.A.H.) avec :
 α l'indice d'agrégation du lien maximum,
 β l'indice d'agrégation du lien minimum.
 Pour chacune des méthodes α et β , vous indicerez la hiérarchie à l'aide de l'indice f défini par :

$$f(A \cup B) = \delta(A, B)$$

où A et B désignent deux classes réunies lors d'une itération quelconque de l'algorithme de la CAH.

3. Donner les deux partitions en 3 classes déduites respectivement des deux hiérarchies indicées obtenues selon les deux méthodes α et β . Vous expliquerez (brièvement) la méthode utilisée pour déterminer ces deux partitions.

2 Association

Inférer les règles d'association de la base de données de la tabel 2 avec un support=2 et une confiance=70%.

TID	Items
1	a,b,c
2	b,c,d,e
3	c,d
4	a,b,d
5	a,b,c

TABLE 2 – Données pour l'exercice 2.