ALGORITMOS EN METODOS DE CONGLOMERADOS SUBDOMINANTES

Por Miguel Sánchez García. Analista del Centro de Cálculo de la Universidad Complutense.

1.- Introducción.

Desde que el hombre está sobre la tierra; ha intentado relacionar los multiples fenómenos que desde ella se observan para tratar; ya sea de predecir cuando determinados fenómenos iban a tener lugar, ya de influir sobre otros con el objeto de lograr que su evolución estuviera más en consonancia con los deseos de la humanidad. Las relaciones ó dependencias entre distintos fenómenos se han construido mediante funciones ó relaciones matemáticas; para lo cual se han utilizado distintas técnicas. Una de estas, de gran desarrollo actual, está constituida por el análisis de conglomerados ó clustering; que trata de agrupar elementos teniendo en cuenta ciertas medidas de semejanza entre ellos. En el presente artículo nos ocuparemos de los conceptos fundamentales de estas técnicas; así como de la construcción de diversos algoritmos de métodos de conglomerados subdominantes; tales como el ultramétrico y los k-ultramétricos fuertes., u-diametricos y obtención de conglomerados.

2.- Definiciones Fundamentales.

En el artículo supondremos que vamos a construir grupos ó conglomerados sobre un conjunto $P=\{P_1,P_2...,P_n\}$ de n elementos; para ello disponemos del conocimiento de alguna medida de desemejanza ó distinción entre los elementos de P; y esto nos lleva a los siguientes conceptos:

Def 1. Un coeficiente de desemejanza sobre P es una función

$$d: PxP \dashrightarrow R^+$$

que cumple las dos condiciones siguientes.

DC1 $\quad d(P_i,P_i)=0$ para todo $P_i \varepsilon P$

DC2 $\quad d(P_i,P_j)= d(P_j,P_i)$ para todo $P_i,P_j \varepsilon P$

Llamaremos $\mathcal{C}(P)=\{ d/d$ es un coeficiente de desemejanza sobre $P\}$

Para todo $h \varepsilon R^+$ definimos el siguiente conjunto

$$r_d(h)=\{ (P_i,P_j) \ / \ d(P_i,P_j) \leq h; \ P_i,P_j \varepsilon P\}$$

y es evidente que $r_d(h)$ es una relación binaria, reflexiva, y simétrica sobre P, mediante la siguiente definición:

$$P_i \ r_d(h) \ P_j \Longleftrightarrow (P_i,P_j) \ \varepsilon \ r_d(h)$$

Llamaremos $\Sigma(P)=\{r/r$ es una relación binaria reflexiva y simétrica sobre $P\}$

El concepto definido a continuación nos sirve para relacionar $\mathcal{C}(P)$ con otro nuevo vonjunto; y poder obtener de esta forma una visión más amplia del concepto de coeficiente de desemejanza

Def. Un conglomerado estratificado numericamente; (CEN) es una aplicación

$$r: [0,\infty) \longrightarrow \Sigma(P)$$

cumpliendo las tres propiedades siguientes:

I Para todo $h,h' \in [0,\infty)$ con $h \geq h'$ se cumple que $r(h) \supset r(h')$

II Existe un h_0 de tal forma que si $h \geq h_0$, entonces $r(h)=P\times P$

III Para todo $h>0$, existe un $\delta(h)>0$ tal que $r(h)=r(h+\delta(h))$

Llamaremos $CEN(P)=\{r/r$ es un CEN sobre $P\}$; y construiremos dos aplicaciones inyectivas entre los conjuntos $\mathcal{C}(P)$ y $CEN(P)$ que probará que ambos son biyectivos, viniendo dadas estas por:

$$T: \mathcal{C}(P) \longrightarrow CEN(P)$$

$$T: d \longrightarrow T(d)(h)=r(h)=\{(a,b)/d(a,b)\leq h\}$$

y evidentemente T es inyectiva; y

$$U: CEN(P) \longrightarrow \mathcal{C}(P)$$

$$U: r \longrightarrow U(r)(a,b)=\inf\{h/(a,b)\in r(h)\}$$

siendo también evidente que U es inyectiva.

A veces nos interesa reducir el conjunto $\mathcal{C}(P)$ de coeficientes de desemejanza a otros más pequeños; lo cual lleva consigo una reducción del conjunto $CEN(P)$ mediante la biyección existente entre ambos. Si llamamos

$$U(P)=\{r/r \text{ es una relación de equivalencia sobre } P\}$$

podemos considerar unicamente los C.E.N. cuya imegen está contenida en $U(P)$; recibiendo entonces cada elemento el nombre de dendograma.

Por tanto un dendograma es un C.E.N. cuya imagen está contenida en $U(P)$; y si llamamos $\mathcal{D}^0(P)=\{r/\ r$ es un dendograma$\}$ el subconjunto de $\mathcal{C}(P)$ biyectivo con $\mathcal{D}^0(P)$ que llamaremos $\mathcal{D}(P)$, es el de las ultramétricas sobre P; ya que para todo $h\geq 0$

$$r(h)=\{(a,b)\ /\ a,b\in P \text{ y } d(a,b)\leq h\}$$

es una relación de equivalencia sobre P; para lo cual se necesita que d cumpla la sigüiente propiedad:

Para todo $a,b,c \in P$ $\quad d(a,b) \leq \max \{d(a,c),\ d(c,b)\}$

<u>Def.</u> Un coeficiente de desemejanza es una ultramétrica si cumple la propiedad anterior.

Para la comprensión del método de conglomerados subdominantes necesitamos los **siguientes conceptos:**

<u>Def.1</u> Diremos que un coeficiente de desemejanza d_1 domina a otro d_2 , si para todo $a,b \in P$ se cumple:

$$d_1(a,b) \geq d_2(a,b)$$

También diremos que d_2 está dominado por d_1

<u>Def.2</u> Un subconjunto $Y \subset \mathcal{C}(P)$ se dice que está acotado, si existe un coeficiente de desemejanza d_1, de tal forma que para todo $d \in Y$, d_1 domina a d.

Si Y está acotado **definiremos** $(\sup Y) \in \mathcal{C}(P)$ como

$$(\sup Y)(a,b) = \sup_{d \in Y} \{ d(a,b) \}$$

<u>Def.</u> Un subconjunto $Z \subset \mathcal{C}(P)$ se dice que es cerrado, si para todo $Y \subset Z$ acotado se tiene:

$$\sup Y \in Z$$

<u>Def.</u> Un método de conglomerados subdominante consiste en elegir un subconjunto Z de $\mathcal{C}(P)$ cerrado, y en construir partiendo de un $d \in \mathcal{C}(P)$ un $d_1 \in Z$ de tal forma que:

$d_1 = \sup(Y) = \sup \{d' / d' \text{ es dominado por } d_1 \text{ y } d' \in Z\}$

Hasta aquí hemos visto unicamente los conceptos y propiedades relacionados con los coeficientes de desemejanza; vamos a ver a continuación los conceptos relacionados con la construcción efectiva de conglomerados.

<u>Def.</u> Un conjunto de conglomerados sobre P, es una familia $\mathcal{A}$ de subconjuntos de P de tal forma que $P = \bigcup_{A \in \mathcal{A}} A$; y que en $\mathcal{A}$ no existan dos elementos A_1 y A_2 con $A_1 \neq A_2$ y $A_1 \subset A_2$.

El conjunto de conglomerados se llama jerarquico, si la familia es una partición de P; y no jerarquico en otro caso.

Sea $r \in \Sigma(P)$. Un subconjunto M de P, es enlazado maximal para r si cumple las dos condiciones siguientes:

I. Para todo $a,b \in M$, $(a,b) \in r$

II. Para todo $a \notin M$, existe un $b \in M$ de tal forma que $(a,b) \notin r$

Sea d un coeficiente de desemejanza para P.

Def. Un conjunto de conglomerados para el par (P,d) a nivel h, donde $h \in R^+$, es la familia de conjuntos enlazados maximales de la relación $r(h)=\{(a,b) / a,b \in P \text{ y } d(a,b) \leq h\}$
Es evidente que el conjunto de conglomerados definido anteriormente lo es para P; como también es evidente que si d es una ultramétrica, entonces la familia anterior es una partición de P; formando en este caso esta familia un conjunto de conglomerados jerarquicos.

En general los conglomerados obtenidos a partir de un coeficiente de desemejanza cualquiera d, no tienen buenas propiedades; de ahí que se construyan los conglomerados en algunas ocasiones a partir de algún elemento de un cierto conjunto A, que dé lugar a conglomerados con buenas propiedades; quedando así justificado el interés de los métodos subdominantes; que construyen un coeficiente de desemejanza d_1 perteneciente a un cierto conjunto Z, a partir de un d cualquiera perteneciente a $\mathcal{C}(P)$.

Aquí estudiaremos los métodos ultramétrico y K-ultrametrico fuerte, u-diamétricos y la obtención de conglomerados.

3 Método subdominante Ultramétrico.

Estudiaremos en este apartado algunas propiedades de la ultramétrica d; así como un algoritmo que construye la máxima ultramétrica a partir de un coeficiente de desemejanza cualquiera.

PROPOSICION. El númerc máximo de valores diferentes que puede tomar una ultramétrica definida sobre un conjunto P de n elementos es $n-1$.

DEMOSTRACION . por inducción.

Suponemos que $P=\{P_1,P_2,P_3\}$ y puede darse dos casos a) $d(P_1,P_2)=d(P_2,P_3)=d(P_1,P_3)$ en cuyo caso es cierta la proposición; ó por el contrario dos de las desemejanzas son distintas; que por comodidad supondremos que son $d(P_1,P_2) \neq d(P_2,P_3)$; en este caso

$$d(P_2,P_3)=\max\left[d(P_1,P_2),\ d(P_1,P_3)\right]$$

y el número de valores diferentes de d son 2

Supongamos ahora cierta la proposición para n y la demostraremos para $n+1$. Sean $P_1,P_2,\ldots, P_{n+1}$ los elementos de P; y por inducción el siguiente conjunto:

$$D=\{d(P_l,P_s) / l\neq s \quad l=1,2,\ldots,n;\ s=1,2,\ldots,n\}$$

ene a lo más $n-1$ valores diferentes.

a $D_{n+1}=\{ d(P_{n+1},P_l) \ / \ l=1,2,\ldots,n\}$ y ó bien $D_{n+1} \subset D$
n lo cual $Car(D \cup D_{n+1})=Car(D) \leq n-1 < n$; ó bien existe P_{lo} de forma
l que $d(P_{n+1},P_{lo}) \notin D$; en este caso como

$$d(P_{n+1},P_{lo}) \neq d(P_{lo},P_l) \qquad l \leq n$$

valor $d(P_{n+1},P_l)$ queda definido por la fórmula

$$d(P_{n+1},P_l)= \max \left[d(P_{n+1},P_{lo}),\ d(P_{lo},P_l)\right]$$

como consecuencia el conjunto $D \cup D_{n+1}$ que coincide con la ima-
n de d, tiene como máximo n elementos diferentes.

f. Diremos que un valor $v_1=d(P_{i1},P_{j1})$ de la ultramétrica d, es
ducible a partir del conocimiento de los valores $V_2=d(P_{i2},P_{j2})$ y
$=d(P_{i3},P_{j3})$ si satisface las siguientes condiciones:

$$V_2 \neq V_3$$

En los indices $i2,j2,i3,j3$, , hay dos de ellos que son iguales:
endo los otros dos iguales a i_1 y j_1 respectivamente.
suponemos que $i_2 = i_1$, $j_2 = j_3$ y $j_3 = j_1$ tenemos:

$$d(P_{i1},P_{j1}) = \max \left[d(P_{i2},P_{j2}),\ d(P_{i3},P_{j3})\right]$$

$$\text{o bien} \quad V_1 = \max \left[V_2,V_3\right]$$

a ahora $D \subset$ Imagen (d); siendo d una ultramétrica.
finimos a continuación los siguientes conjuntos:

$S_1=\{V/V$ es un valor de d, deducido a partir de $D\}$
endo $D_1= D_o \cup S_1$ con $D_o = D \neq \emptyset$; y por recurrencia en k obtene-
s;:

$S_k= \{V/V \in$ Imagen (d); y V es deducido a partir de un elemen-
$V_1 \in D_{k-1}$ y un $V_2 \in S_{k-1}\}$ y $D_k=D_{k-1} \cup S_k$

Es evidente que poniendo $S_o = D \neq \emptyset$; para todo
$\neq \emptyset$ existe un k_o tal que $S_{ko} \neq \emptyset$ y $S_{ko+1} = \emptyset$

Llamaremos al conjunto D_{ko} conjunto deducido de

Son propiedades evidentes de los conjuntos $\mathcal{D}(D)$ las siguientes:

1ª Si $D_1 \subset D$ entonces $\mathcal{D}(D_1) \subset \mathcal{D}(D)$

2ª Si V es deducido de D, entonces $\mathcal{D}(D\cup\{V\}) = \mathcal{D}(D)$

3ª Si D contiene n-1 elementos distintos entonces $\mathcal{D}(D) = P$.

Llamaremos $d_{ij} = d(P_i, P_j)$; y vamos a definir la operación max (máximo) entre los elementos de la imagen de d como sigue:

$$d_{il} \max d_{lj} = \max \left[d_{il}, d_{jl} \right]$$

siendo evidente que esta operación es asociativa y commutativa.

<u>LEMA</u>. Si $d_{iojo} \notin D$ y $d_{iojo} \in \mathcal{D}(D)$ entonces existe una cadena de èlementos de D, d_{ioi1}, $d_{i1i2}, \ldots, d_{imjo}$ tal que:

$$d_{iojo} = d_{ioi1} \max d_{i1i2} \max \ldots d_{im-1im} \max d_{imjo} .$$

<u>Demostración Por inducción.</u>

Como $\mathcal{D}(D) = D \cup (\bigcup_{k=1}^{ko} S_k)$ y $k_o \geq 1$ por hipótesis, vamos a suponer en primer lugar que $d_{iojo} \in S_1$; y en este caso existen d_{iol}, $d_{ljo} \in D$ con $d_{iojo} = d_{iol} \max d_{ljo}$; por definición de S_1

Suponemos por inducción que el lema es cierto para los elementos de $\bigcup_{k=1}^{l} S_k$, y le probamos para los elementos de S_{l+1}.

Sea $d_{iojo} \in S_{l+1}$; por definición existen d_{iol}, d_{ljo} tales que d_{iol} ó $d_{ljo} \in S_l$; y el otro a D_l. Sin pérdida de generalidad podemos suponer que $d_{iol} \in S_l$ y $d_{ljo} \in D_l$; por la hipótesis de inducción existen dos cadenas $d_{ioi1}\, d_{i1i2} \ldots d_{iml}$ y $d_{lj1}\, d_{j1j2} \ldots d_{jsjo}$ tales que:

$$d_{iol} = d_{ioi1} \max d_{i1i2} \max \ldots d_{im-1im} \max d_{iml}$$

$$d_{ljo} = d_{lj1} \max d_{j1j2} \max \ldots \max d_{jsjo}$$

y si unimos las dos cadenas obtenemos:

$$d_{iojo} = d_{ioi1} \max \ldots \max d_{iml} \max d_{lj1} \max \ldots \max d_{jsjo}$$

erminando aquí la demostración del Lema.

TEOREMA. Sea $d_{ij} \notin D \neq \emptyset$; $d_{ij} \in \mathcal{D}(D)$, entonces existe un lemento $d_{1s} \in D$ de tal forma que:

$$\mathcal{D}(D \cup \{d_{ij}\} - \{d_{1s}\}) = \mathcal{D}(D)$$

DEMOSTRACION. De las propiedades 1^a y 2^a tenemos:

$$\mathcal{D}(D \cup \{d_{ij}\} - \{d_{1s}\}) \subset \mathcal{D}(D \cup \{d_{ij}\}) = \mathcal{D}(D)$$

ara demostrar la otra inclusión sea $d_{ii1}d_{i1i2} \ldots d_{imj}$ una cadea tal que:

$$d_{ij} = d_{ii1} \max d_{i1i2} \max \ldots d_{im-1im} d_{i1i2} \ldots d_{imj} = d_{1s}$$

sea $d_{1s} = d_{ipip+1}$; entonces d_{ipip+1} queda definido por la siguiene fórmula:

$$d_{ipip+1} = d_{ipip+1} = d_{ip+1ip+2} \max \ldots d_{imj} \max d_{ji} \max d_{ii1} \ldots \max d_{ip-1ip}$$

or tanto $d_{1s} = d_{ipip+1} \in \mathcal{D}(D \cup \{d_{ij}\} - \{d_{ipip+1}\})$ y de aquí:

$$\mathcal{D}(D) \subset \mathcal{D}(D \cup \{d_{ij}\}) = \mathcal{D}(D \cup \{d_{ij}\} - \{d_{ipip+1}\})$$

c.q.d.

Este teorema nos justifica que si al elegir el eleento mínimo en el siguiente algoritmo; este no es único, podemos ncorporar al conjunto d_1 todos los elementos que sean iguales al ínimo.

lgoritmo para construir la máxima ultramétrica d_1 dominada por un oeficiente de desemejanza d' cualquiera.

Es evidente que el conjunto $\mathcal{D}(P)$ de ultramétricas s cerrado; y por lo tanto dado un d, existe d_o^1 tal que

$d_o^1 = \sup Y = \sup\{d'/d'$ dominado por d y $d' \in \mathcal{D}(P)\}$

el algoritmo siguiente construye d_1 a partir de d, siendo D=iman (d).

LGORITMO.

Paso 1 Se eligen los dos menores elementos del conjunto D; si estos son d_{i1j1} y d_{i2j2} colocamos

$d_o^1(P_{j1},P_{i1}) = d_o^1(P_{i1},P_{j1}) = d_{i1j1}$ $\quad d_o^1(P_{i2},P_{j2}) = d^1(P_{j2},P_{i2}) = d_{i2j2}$

$D=D-\{d_{i1j1},d_{i2j2}\}$

$D_1=\{d_{i1j1},d_{i2j2}\} = \{d_1(P_{i1},P_{j1}),d_1(P_{i2},P_{j2})\}$

<u>Paso 2</u> Si existe un $d^1_{ij}\notin D_1$; deducible a partir de D_1 se va al paso 3; en otro caso se va al paso 4.

<u>Paso 3</u> Si d^1_{ij} se deduce a partir de d^1_{il}, d^1_{lj} tenemos:

$d^1_{ij}=\max\{d^1_{il},d^1_{lj}\}$

$D_1= D_1 U \{d^1_{ij}\}$

$D=D- \{d_{ij}\}$

Si $D=\emptyset$ el àlgoritmo para; si no se va al paso 2

<u>Paso 4.</u> Se eligen los $d_{ij}\varepsilon$ D, tales que

$d_{ij}=\min \{d_{lk}/ d_{lk}\varepsilon D\}$ y colocamos

$d^1_{ji}=d^1_{ij} = d_{ij}$; $D_1 =D_1 \cup \{d^1_{ij}\}$ y $D=D-\{ d_{ij} \}$ para cada elemento que cumpla la anterior igualdad.

Si $D=\emptyset$ el algoritmo para en otro caso se va al paso 2.

<u>Proposición.</u> d_1 es la máxima ultramétrica dominada por d.

<u>Demostración</u> Sea d' ultramétrica dominada por d; y suponemos que existen un valor.

$d'(P_i,P_j) < d^1(P_i,P_j)$

de la construcción de d^1, es claro que $d^1(P_i,P_j)$ no es igual a d_{ij}; pues si fuera esto cierto d' no estaría dominada por d; por tanto $d^1(P_i,P_j)$ es un valor deducido y como consecuencia $d'(P_i,P_j)\leq d^1(P_i,P_j)$; luego d^1 domina a toda ultramétrica dominada por d.

<u>d_1 es una ultramétrica.</u>

Sea $P_i,P_j\varepsilon P$ y $P_l\varepsilon P$; $P_l \neq P_i$; $P_l\neq P_j$ y debemos probar que

$$d^1(P_i,P_j)\leq\max\{d^1_o(P_i,P_l), d^1(P_l,P_j)\}$$

La igualdad anterior es evidente en el caso de que ó bien $d^1(P_i,P_l)$ ó $d^1(P_l,P_j)$ fuese añadido a D_1 con posterioridad a $d^1(P_i,P_j)$.

Suponemos que $d^1(P_i,P_j)$ fuese añadido a D_1, después de que lo fueran $d^1(P_i,P_l)$ y $d^1(P_l,P_j)$; en este caso de la construcción del algoritmo es claro que el máximo valor del conjunto D_1 al añadir a el $d^1(P_i,P_j)$ coincide con el

$$\text{máximo}\left[d^1(P_i,P_1),d^1(P_1,P_j)\right]$$

y por tanto

$$d^1(P_i,P_j)=\max\left[d^1(P_i,P_1),d^1(P_1,P_j)\right]$$

y así d^1 es una ultramétrica.

4. METODOS SUBDOMINANTES K-ULTRAMETRICOS FUERTES.

Def. Una k-ultramétrica fuerte ó k-ultramétrica es un coeficiente de desemejanza que cumple la siguiente propiedad:

Para todo conjunto $S\subset P$ de k elementos y todo $a,b\in P$; con $a,b\notin S$ se tiene

$$d(a,b)\leq \max\{d(a,s),d(b,s) \;/\; \bar{s}\in S\}$$

El método subdominante k-ultramétrico consiste por tanto, en construir a partir de un coeficiente de desemejanza d ; la máxima k-ultramétrica fuerte d_1 dominada por d; y tiene como objetivo evitar que dos elementos formen conglomerados con k elementos separadamente; sin estar los dos elementos en un mismos conglomerado.

Si definimos $\mathcal{D}_k(P)$ como el conjunto de todas las k-ultramétricas fuertes, es evidente que estos conjuntos son cerrados, de ahí que tenga sentido este método subdominante.

Def! Diremos que el valor $d_1(P_i,P_j)$ de la k-ultramétrica fuerte d_1; construida con el método subdominante a partir de un coeficiente de desemejanza d, es deducible a partir de los valores

$$d_1(P_i,P_{i1}),\ d_1(P_i,P_{i2}),\dots,d_1(P_i,P_{ik});d_1(P_j,P_{i1}),\dots,d_1(P_j,P_{ik})$$

cuando

$$d(P_i,P_j)\geq\max\{d_1(P_i,P_{i1}),\dots,d_1(P_i,P_{ik}),d_1(P_j,P_{i1}),\dots,d_1(P_j,P_{ik})\}=V$$

y en este caso se define $d_1(P_i,P_j)=V$

ALGORITMO.

Sean $P=\{P_1,P_2,\dots,P_n\};d_{ij}=d(P_i,P_j);D=\{d_{ij}/\ i=1,2,\dots,n-1;j=i+1,\dots,n\}$

Paso 1 Se eligen 2k elementos; $d_{i1j1},d_{i2j2},\dots,d_{i2kj2k}$ de D cumpliendo la siguiente condición:

Para todo $d_{ij} \varepsilon D-\{d_{i1j1},\ldots,d_{i2kj2k}\}$, se debe tener:

$$d_{ij} \geq \max \{d_{i1j1},d_{i2j2},\ldots,d_{i2kj2k}\}$$

y colocamos:

$$d_1(P_{j1},P_{i1})=d_1(P_{i1},P_{j1})= d_{1_{i1j1}}$$

$$.D=D-\{d_{i1j1},d_{i2j2},\ldots,d_{i2kj2k}\}$$

$$D_1=\{d_{1i1j1},d_{1i2j2},\ldots,d_{1\ i2kj2k}\}$$

Paso 2. Si existe un d_{ij} ε D, de tal forma que

$$d_{ij}=\max\{d_{1s}\}$$

$$d_{1s} \ \varepsilon \ D_1$$

Colocamos $D=D-\{d_{ij}\}$; $D_1U\{d_{ij}\}$; $d_1(P_i,P_j)=d_{ij}$

y se vuelve al paso 2 si $D\neq\emptyset$, en otro caso se para. Si no existe ningún d_{ij} cumpliendo la condición anterior se va al paso 3.

Paso 3. Si existe un elemento d_{ij} ε D, de tal forma que existan k elementos $P_{s1},P_{s2},\ldots,P_{sk}$ con

$d_1(P_i,P_{s1})$ ε D_1; $d_1(P_j,P_{s1})$ ε D_1 para $1= 1,2,..,k$ se va al paso 4;

si no existe ningún d_{ij} cumpliendo la propiedad anterior se va al paso 5.

Paso 4. Sea d_{iojo} el elemento encontrado en el paso 3, colocamos entonces:

$$d_1(P_{io},P_{jo})=d_1(P_{jo},P_{io})=d_{1iojo}=\max\{d_1(P_i,P_{s1}),d_1(P_j,P_{s1});1=1,\ldots,k\}$$

$$D_1=D_1U\{d_{1iojo}\}$$

Si $D \neq \emptyset$ se vuelve al paso 2; si $D=\emptyset$ el algoritmo termina.

Paso 5. Se elige $d_{iojo}=\min_{d_{ij}\varepsilon D}\{d_{ij}\}$ y se coloca:

$=D-\{d_{iojo}\}$; $d_1(P_{io},P_{jo})= d_1(P_{jo},P_{io})=d_{1iojo}=d_{iojo}$

$$D_1= D_1 \cup \{d_{1iojo}\}$$

i $D \neq \emptyset$ se va al paso 2; si $D=\emptyset$ el algoritmo termina.

Probaremos a continuación que d_1 es la máxima -ultramétrica fuerte dominada por d.

De la construcción de d_1 es evidente, que todo valor construido en los pasos 2 y 4; es igual al máximo valor le D_1 en ese instante; y este conocimiento nos servirá para probar que d_1 es una k-ultramétrica fuerte.

Sean $P_i,P_j \in P$ y $\{P_{i1},P_{i2},\ldots,P_{ik}\} \varepsilon P-\{P_i,P_j\}$ cualquiera; pudiendo suceder los dos casos siguientes

1º Cuando se añadió $d_1(P_i,P_j)$ al conjunto D_1; este aún no contenía todos los elementos $\{d_1(P_i,P_{il}),d_1(P_j,P_{il});l=1,2,..,k\}$ siendo clara en este caso la siguiente desigualdad.

$$d_1(P_i,P_j) \leq \max\{d_1(P_i,P_{il}),d_1(P_j,P_{jl});l=1,2,\ldots,k\}$$

2º Cuando se añadió $d_1(P_i,P_j)$ al conjunto D_1; este ya contenia a los elementos $\{d_1(P_i,P_{il});d_1(P_j,P_{il});l=1,2,\ldots,k\}$ y en este caso es evidente que se definió $d_1(P_i,P_j)$ como

$$d_1(P_i,P_j)=\max\{d_1(P_i,P_{il}),\ d_1(P_j,P_{jl});l=1,2,..,k\}=\max_{d_{ij}\varepsilon D}\{d_{ij}\}$$

quedando así probado que d_1 es una k-ultramétrica fuerte.

Que es la máxima dominada por d; se deduce evidentemente de la construcción de d_1.

5.METODOS SUBDOMINANTES K-ULTRAMETRICOS FINOS.

Def. Un coeficiente de desemejanza es una k-ultramétrica fina si para todo $P_i,P_j \varepsilon P$ y todo $S \subset P-\{P_i,P_j\}$ de k elementos se cumple la siguiente desigualdad:

$$d(P_i,P_j) \leq \max\{d(P_l,P_s) \ / \ P_s \in S;\ P_l \ \varepsilon\ S \cup (P_i,P_j)\}$$

Llamaremos $\mathcal{B}^f_k(P)$, al conjunto de las k-ultramétricas finas definidas sobre P; y nuestro problema consiste ahora en construir la máxima k-ultramétrica fina dominada por un coeficiente de desemejanza d, que es evidente que existe debido al hecho de ser

$\mathcal{D}_k^{\ell}(P)$ cerrado.

El objetivo de construir los conglomerados mediante d_1, se debe a que la intersección entre dos de ellos tienen como máximo k-1 elementos; de ahí que se obtenga una reducción en su número.

Algoritmo para construir d_1 a partir de d.

Paso 1. Sea D=imagen{d} ; y se seleccionan $l=\frac{(k+2)(k+1)}{2}-1$ elementos de D cumpliendo la propiedad siguiente:

Si $d_{ij}\varepsilon D-\{d_{i1j1},d_{i2j2},\dots,d_{iljl}\}$; donde

$\{d_{i1j1},d_{i2j2},\dots,d_{iljl}\}$ son los elementos seleccionados entonces:

$d_{ij} \geq \max\{d_{i1j1},d_{i2j2},\dots,d_{iljl}\}$ y colocamos

$D=D-\{d_{i1j1},d_{i2j2},\dots,d_{iljl}\}$

$d_1(P_{is},P_{js})=d_1(P_{js},P_{is}) = d_{1isjs}=d_{isjs}$ $S=1,2,\dots,l$

$D_1=\{d_{1i1j1},d_{1i2j2},\dots,d_{1iljl}\}$

Paso 2. Se busca un elemento $d_{iojo}\in D$, de tal forma que

$$d_{iojo}=\max_{d_{1ij}\varepsilon D_1}\{d_{1ij}\}$$

si no existe ningún elemento de esta clase se va al paso 3; si por el contrario existe un d_{iojo} se colocan:

$D=D-\{d_{iojo}\}$; $d_1(P_{io},P_{jo})= d_1(P_{jo},P_{io})=d_{1iojo}=d_{iojo}$

$D_1=D_1 \cup\{d_{1iojo}\}$

y se vuelve al paso 2 si $D \neq \emptyset$; si por el contrario $D = \emptyset$ entonces el algoritmo termina.

Paso 3. Se busca un elemento d_{ij} ε D, de forma tal que existan k elementos $\{P_{s1},P_{s2},\dots, P_{sk}\}$ con

$d_1(P_{sl},P_i)\ \varepsilon\ D_1$; $d_1(P_{sl},P_j)\ \varepsilon\ D_1$ $l=1,2,\dots, K$ y

$d_1(P_{sl},P_{st})\ \varepsilon\ D_1$; $l=1,2,\dots,k-1$; $t=l+1,\dots, k$

Si existe este elemento se va al paso 4; si ningún elemento cumple las anteriores condiciones se va al paso 5.

<u>Paso 4.</u> Sea d_{iojo} el elemento hallado en el paso 3.

Colocamos $d_1(P_{io};P_{jo}) = d_1(P_{jo},P_{io}) = d_{1iojo}$

$$D = D - \{d_{iojo}\} \qquad d_1 = D_1 \cup \{d_{1iojo}\}$$

Si $D = \emptyset$ se acaba el algoritmo; si $D \neq \emptyset$ se va al paso 3.

<u>Paso 5.</u> Se elige un d_{iojo}; de tal forma que

$$d_{iojo} = \min_{d_{ij} \in D} \{d_{ij}\} \text{ y colocamos}$$

$$d_1(P_{io},P_{jo}) = d_{iojo};\ D = D - \{d_{iojo}\} \ \text{ y } \ D_1 \cup \{d_{iojo}\}$$

Si $D \neq \emptyset$ se va al paso 2; si por el contrario $D = \emptyset$ el algoritmo se termina.

<u>d_1 es la máxima k-ultramétrica fina dominada por d.</u>

Sean cualquier par de elementos $P_i, P_j \in P$; y cualquier conjunto $S \subset P - \{P_i,P_j\}$ de k elementos. Si algún elemento $d_1(P_l,P_s)$ con $P_l \in S \cup \{P_i,P_j\}$, $P_s \in S$ se añadió a D_1, después de $d_1(P_i,P_j)$ entonces es claro que:

$$d_1(P_i,P_j) \leq \max\{d_1(P_l,P_s);\ P_l \in S \cup \{P_i,P_j\}\ ;\ P_s \in S\}$$

Si por el contrario $d_1(P_i,P_j)$ se añadió a D_1 cuando

$D_1 \supset \{d_1(P_l,P_s);\ P_l \in S \cup \{P_i,P_j\}\ ,\ P_s \in S\}$; entonces de la construcción de d_1 es evidente que:

$$d_1(P_i,P_j) = \max\{d_1(P_l,P_s); P_l \in S \cup \{P_i,P_j\}\ ;\ P_s \in S\}$$

y así queda probado que d_1 es una k-ultramétrica fina dominada por d.

Que d_1 es la máxima k-ultramétrica fina dominada por d, es una clara evidencia de su construcción.

<u>NOTA.</u> La busca del elemento d_{iojo} en el paso 3 del algoritmo anterior; se hace mediante la siguiente relación binaria

$$P_l \curlyvee P_s \Leftrightarrow d_1(P_l,P_s) \in D_1$$

y se debe encontrar un conjunto S de k elementos de tal forma que:

$$\{P_i\}xSUSxSUSx\ \{P_j\} \subset Y$$

y si un tal conjunto existe se define:

$$d_1(P_i,P_j)=\max\{d_1(P_l,P_s)\ /\ P_l \in SU\{P_i,P_j\}\ ;\ P_s \in S\}$$

6. METODOS SUBDOMINANTES U-DIAMETRICOS.

Def. Un coeficiente de desemejanza d es u-diametrico si para todo P_i,P_j,P_s,P_l se cumple la siguiente proposición:

$$d(P_s,P_l)> um \Rightarrow d(P_i,P_j) \leq m$$

donde $m=\max\{d(P_s,P_l),\ d(P_s,P_i),\ d(P_s,P_j),\ d(P_l,P_i),\ d(P_l,P_j)\}$

El problema que nos proponemos en este apartado consiste en construir a partir de un coeficiente de desemejanza cualquiera d; otro d_1 que sea u-diamétrico; el objetivo que se consigue con los coeficientes de desemejanza u-diamétricos es que la máxima desemejanza de la intersección de dos conglomerados a nivel h, es menor ó igual que uh; y de aquí que debe ser $0 < u < 1$. Para la construcción de d_1 necesitamos la siguiente;

Déf. Diremos que el valor $d_1(P_i,P_j)$, es deducido del conocimiento de los valores ya construidos de d_1; si existen P_l,P_s tales que conocemos

$$d_1(P_l,P_s),\ d_1(P_l,P_i),\ d_1(P_l,P_j),\ d_1(P_s,P_i),\ d_1(P_s,P_j)$$

y que además si m es el máximo de los anteriores valores

$$d_1(P_l,P_s) > um \ \text{ y } \ d(P_i,P_j) > m$$

entonces se define $d_1(P_i,P_j)=m$

ALGORITMO:

Paso 1 Llamando D=imagen d , se eligen 5 elementos de D,

$\{d_{i1j1},d_{i2j2},\ldots,\ d_{i5j5}\}$, tal que para todo

$d_{ij} \in D-\{d_{i1j1},d_{i2j2},d_{i3j3},d_{i4j4},d_{i5j5}\}$ se tenga

$d_{ij} \geq \max\{d_{i1j1},d_{i2j2},d_{i3j3},d_{i4j4},d_{i5j5}\}$

y se hacen las siguientes transformaciones:

$D=D-\{d_{i1j1},d_{i2j2},d_{i3j3},d_{i4j4},d_{i5j5}\}$

$d_1\ (P_{il},P_{jl}) = d_1(P_{jl},P_{il})= d_{1iljl}= d_{iljl}$ $\quad l=1,2,\ldots,5$

$D_1=\{d_{i1j1},d_{i2j2},\ldots,d_{i5j5}\}$

Paso 2. Si existe un elemento d_{iojo} ε D, de forma que

$d_{iojo}=\max\limits_{d_{ij}\varepsilon D_1} \{d_{ij}\}$; se coloca:

$D=D-\{d_{iojo}\}$; $d_1(P_{io},P_{jo})=d_{1iojo}=d_{iojo}$; $D_1=D_1\cup\{d_{1iojo}\}$

Si $D \neq \emptyset$ se vuelve al paso 2; si $D = \emptyset$ el algoritmo termina.
Si $\nexists$ diojo cumpliendo la anterior condición se va al paso 3

Paso 3. Si no existe ningún elemento d_{ij} ε D, tal que d_{1ij} sea deducido de D_1 se va al paso 5; si por el contrario existe d_{ij} ε D cuyo valor correspondiente d_{1ij} sea deducido a partir de D_1 se va al paso 4.

Paso 4. Sea d_{iojo} el elemento hallado en 3.
Colocamos $d_1(P_{io},P_{jo}) = d_{1iojo} = d_{iojo}$

$D=D-\{d_{iojo}\}$ $\qquad D_1=D_1\cup\{d_{1iojo}\}$

Si $D \neq \emptyset$ se vuelve al paso 3; si $D = \emptyset$ el algoritmo termina.

Paso 5. Se elige un d_{iojo} tal que

$d_{iojo} = \min\limits_{d_{ij}\ \varepsilon\ D} \{d_{ij}\}$ y colocamos

$d_1(P_{io},P_{jo})= d_{1iojo}$; $D=D-\{d_{iojo}\}$ y $D_1=D_1\cup\{d_{iojo}\}$

Si $D \neq \emptyset$ se va al paso 2; si $D = \emptyset$ se acaba el algoritmo.

d_1 es la máxima u-diamétrica dominada por d.

Sean P_i,P_j,P_l,P_s arbitrarios y

$m= \max\{d_1(P_l,P_s),d_1(P_l,P_i),d_1(P_l,P_j),d_1(P_s,P_i),d_1(P_s,P_j)\}$

pudiendo darse dos casos

1º $d_1(P_i,P_j)$ fué añadido a D_1 antes de que lo fueran

$\{d_1(P_1,P_s),d_1(P_1,P_i),d_1(P_1,P_j),d_1(P_s,P_i),d_1(P_s,P_j)\}$

en este caso es evidente que $d_1(P_i,P_j) \leq m$

2º . Fue añadido después y pueden darse dos subcasos

a) $d_1(P_1,P_s) \leq um$, y entonces es cierta la proposición

$$d_1(P_1,P_s) > um \Rightarrow d_1(P_i,P_j) \leq m$$

por ser falsa la hipótesis

b) $d_1(P_1,P_s) > um$ en este caso se definió

$d_1(P_i,P_j) = m$; y en este caso también es cierta la proposición anterior.

NOTA. Para comprobar en el algoritmo anterior si es deducible el valor d_{1ij} , se observa los indices i_1 tales que . d_{1ii1} y $d_{1ji1} \in D_1$ si existen dos indices en el conjunto anterior s,t tales que $d_{1st} \in D_1$ y $d_{1st} > um$ donde

$$m=max\{d_{1is},d_{1js},d_{1it},d_{1jt},d_{1st}\} \quad \text{entonces se define } d_{1ij} = m$$

7. OBTENCION DE CONGLOMERADOS.

Como hemos visto anteriormente, dado un coeficiente de desemejanza d y un número real $h \geq 0$; se construia la relación binaria, reflexiva y simétrica

$$\gamma_{d(h)} = \{(a,b) \; / \; a,b \in P \text{ y } d(a,b) \leq h\}$$

siendo los conglomerados a nivel h, los conjuntos enlazados maximales de la relación anterior; y entre las propiedades de estos conjuntos están las siguientes:

Propiedad 1. Sea P_1 un conjunto enlazado maximal de la relación $\gamma_d(h)$; y $P_2 \neq \emptyset$, otro conjunto enlazado maximal, (C.E.M.) de la relación $\gamma_{\bar{d}}(h)$ restringida a $P-P_1$; entonces existe un subconjunto propio de P_1 que unido a P_2 forma un nuevo (C.E.M.) de $\gamma_d(h)$

Demostración Sea $P^*=\{a/a \in P_1$ y $(a,b) \in \gamma_d(h)$, para todo $b \in P_2\}$

Si $P^*=\{\emptyset\}$, es claro que P_2 es un nuevo CEM de $\gamma_d(h)$; si por el contrario $P^* \neq \emptyset$; tenemos que P^* es un subconjunto propio de P_1 y $P^* U P_2$ es un nuevo CEM; pues si $a \notin P^* U P_2$; pueden suceder dos casos

a) $a \notin P_1$ $a \in P-P_1$ y por ser P_2 un CEM en $P-P_1$, $\exists\, b \in P_2$ tal que $(a,b) \notin \gamma_d(h)$.

b) $a \notin P_1$; por $a \notin P^*$, $\exists\, b \in P_2$ tal que $(a,b) \notin \gamma_d(h)$

así queda probada la 2^{a} propiedad de los CEM.

La primera es evidente por construcción.

La propiedad anterior nos induce a realizar la siguiente partición en P.

Elegimos $P_1 \subset P$ como un CEM para $\gamma_d(h)$; P_2 un CEM para $\gamma_d(h)$ restringuido a $P-P_1$; y en general, P_i un CEM para $\gamma_d(h)$ restringido a $P-\bigcup_{j=1}^{i-1} P_i$; siendo evidente que:

$$P=P_1 U P_2 U \ldots\ldots U P_k \quad \text{con } P_i \cap P_j = \emptyset \text{ si } i \neq j$$

Propiedad 2 Para todo P_i con $i \neq 1$ existe un P^* subconjunto propio de $\bigcup_{j=1}^{i-1} P_j$ de tal forma que $P_i \cup P^*$ es un C.E.M. para $\gamma_d(h)$

Demostración Sea $P_1^*=\{a/a \in P_1$ y $(a,b) \in \gamma_d(h)$ para todo $b \in P_i\}$ y en general $P_l^*=\{a/a \in P_l$ y $(a,b) \in \gamma_d(h)$ para todo $b \in P_i U \left(\bigcup_{j=1}^{l-1} P_l\right)\}$ entonces es claro que $P_i U \left(\bigcup_{j=1}^{i-1} P_j^*\right)$ es un CEM que cumple las condiciones de la propiedad 2.

Def. Llamaremos

$$\mathcal{C}_{P_o}(P')=\{a/a \in P_o \text{ y } (a,b) \in \gamma_d(h) \text{ para todo } b \in P'\}$$

Algoritmo

Se construye en primer lugar una partición de P; $\{P_1,P_2,\ldots,P_k\}$ donde P_i es un CEM para la relación $\gamma_d(h)$

restringida a $P-\bigcup_{j=1}^{i-1} P_j$; y como consecuencia P_1 es un CEM.

Para determinar los C.E.M. que solo contienen elementos de $P_1 \cup P_2$ se procede de la forma siguiente:

Se elige un subconjunto $P_o \subset P_2$, de tal forma que no exista $x \in P_2$ y $x \notin P_o$ con $\mathcal{C}_{P_1}(P_o) \subset \mathcal{C}_{P_1}(x)$; y entonces ó bien $P_o \cup \mathcal{C}_{P_1}(P_o)$ es un CEM; ó bien existe $y \in \bigcup_{j=3}^{k} P_j$ con $\mathcal{C}_p(y) \supset P_o \cup \mathcal{C}_{P_1}(P_o)$.

Para demostrar la anterior afirmación suponemos que no existe el $y \in \bigcup_{j=3}^{k} P_j$ con $\mathcal{C}_p(y) \supset P_o \cup \mathcal{C}_{P_1}(P_o)$; es claro de la construcción que $P_o \cup \mathcal{C}_{P_1}(P_o)$ cumple la primera propiedad de C.E.M.

Para demostrar la segunda, sea $x \notin P_o \cup \mathcal{C}_{P_1}(P_o)$; si $x \in P_1$; por definición de $\mathcal{C}_{P_1}(P_o)$ existe $z \in P_o$ con $(x,z) \notin \gamma_d(h)$; si $x \in P_2$; $x \notin P_o$ por elección de P_o **existe** $z \in \mathcal{C}_{P_1}(P_o)$ de tal forma que $(x,z) \notin \gamma_d(h)$; finalmente si $x \in \bigcup_{j=3}^{k} P_j$; entonces al ser falso que $\mathcal{C}_p(x) \supset P_o \cup \mathcal{C}_{P_1}(P_o)$, existe $z \in P_o \cup \mathcal{C}_{P_1}(P_o)$ con $(x,z) \notin \gamma_d(h)$.

Como consecuencia los CEM de $P_1 \cup P_2$ se forman eligiendo todos los subconjuntos P_o de P_2, tales que no exista $x \in P_2$, $x \notin P_o$ con $\mathcal{C}_{P_1}(P_o) \subset \mathcal{C}_{P_1}(x)$; y si no existe $y \in \bigcup_{j=3}^{k} P_j$ con $P_o \cup \mathcal{C}_{P_1}(P_o) \subset \mathcal{C}_p(y)$ entonces $P_o \cup \mathcal{C}_{P_1}(P_o)$ es un C.E.M.

Procedemos por inducción. Suponemos que tenemos hallados todos los CEM que unicamente tienen elementos de $\bigcup_{j=1}^{i} P_j$, y queremos hallar los contenidos en $\bigcup_{j=1}^{i+1} P_j$ que tienen algún elemento de P_{i+1}.

Elegimos los subconjuntos $P_o \subset P_{i+1}$ tal que no exista $x \in P_{i+1}$, $x \notin P_o$ con $\mathcal{C}_{\bigcup_{j=1}^{i} P_j}(P_o) \subset \mathcal{C}_{\bigcup_{j=1}^{i} P_j}(x)$; y para cada $l < i+1$ se forman los conjuntos $P'_e=\{z/z \in P_l$ y $(z,b) \in \gamma_d(h)$ para todo $b \in P_0$ }; y sea $P'=\bigcup_{l=1}^{i} P'_l$; hallandose todos los CEM de $\gamma_d(h)$ restringido a P' ; y uniendo cada uno de ellos a P_o puede darse dos casos:

1º Que hayamos obtenido un nuevo C.E.M.

2º Que exista un $x \in \bigcup_{l=i+2}^{k} P_l$ que esté relacionado con todos los elementos del conjunto así obtenido.

En el caso 2º el C.E.M. correspondiente se hallaría en un paso posterior.

<u>Demostración de que el algoritmo obtiene todos los C.E.M para $\gamma_d(h)$.</u>

Sea P_o un C.E.M. para $\gamma_d(h)$; $P''_l=P_o \cap P_l$ y $l_o=\max\{i/\ P''_i \neq \emptyset\}$

Sean $P'_1= \mathcal{C}_{P_1}(P''_{lo})$;$P'_2= \mathcal{C}_{P_2}(P''_{lo}),\ldots,P'_{lo-1}= \mathcal{C}_{P_{lo-1}}(P''_{lo})$;

claramente $P''_i \subset P'_i$; $i = 1,2,\ldots,l_o-1$ y $P'=\bigcup_{i=1}^{lo-1} P''_i$ es un C.E.M. para $\gamma_d(h)$ restringuido a $A'=\bigcup_{i=1}^{l_o-1} P'_i$; pues si así no fuera existiria $x \in A'$, $x \notin P'$, tal que $(x,z) \in \gamma_d(h)$ para todo $z \in P' \cup P'_{lo} = P_o$ de donde se deduciria que $P_o \cup \{x\}$ es un conjunto enlazado, que contrario a la hipótesis de que P_o es un conjunto enlazado maximal; y es evidente que P_o es hallado por el algoritmo cuando une $P'_{lo} \subset P_{lo}$ con P'!

<u>OBTENCION DE CONGLOMERADOS A DISTINTOS NIVELES.</u>

Suponemos que tenemos construidos los conglomerados para un coeficiente de desemejanza d a nivel h; y los queremos construir a otro nivel h' con $h' > h$; es claro entonces que $\gamma_d(h') \supset \gamma_d(h)$; siendo el contenido estricto; pues en otro caso tendríamos ya hallados los conglomerados.

Podemos suponer sin pérdida de generalidad que $(a,b) \notin \gamma_d(h)$ y que $\gamma_d(h') = \gamma_d(h) \cup (a,b)$; pues si así no fuera, como P es finito, mediante un número finito de pasos como el anterior obtendriamos $\gamma_d(h')$; quedando así nuestro problema reducido a encontrar los C.E.M. de $\gamma_d(h') = \gamma_d(h) \cup (a,b)$.

Para ello sea $P' = \{z/z \in P \text{ y } (z,a) \in \gamma_d(h), (z,b) \notin \gamma_d(h)\}$ y hallaremos los C.E.M. de $\gamma_d(h')$ restringidos a P'; si alguno de ellos contiene a algún C.E.M. de $\gamma_d(h)$ este último se elimina; obteniendo los C.E.M. de $\gamma_d(h')$ como unión de los C.E.M. de $\gamma_d(h)$ no eliminados con los hallados en conjunto $P' \cup \{a,b\}$.

8. EJEMPLO PRACTICO.

Nos proponemos en este ejemplo, construir una 2-ultramétrica fina d_1 a partir de un coeficiente de desemejanza d; para después hallar conglomerados a distintos niveles partiendo de d.

Suponemos que P tiene 9 elementos $\{P_1, P_2, \ldots, P_9\}$ y que d viene dado por la siguiente matriz de desemejanzas

1	2	3	4	5	6	7	8	9
	4	6	9	21	12	41	6	21
		7	11	8	18	40	4	3
			13	9	32	28	10	7
				14	36	32	16	19
					32	12	12	11
						8	14	13
							18	15
								14

La 2-ultramátrica fina construida a partir de d viene dada por la siguiente matriz:

1	2	3	4	5	8	9	6	7
	0	0	0	0	0	0	1	1
		0	0	0	0	0	1	1
			0	0	0	0	1	1
				0	0	0	1	1
					0	0	1	1
						0	1	1
							1	1
								0

La partición de $P=\{1,2,3,4,5,6,7,8,9\}$ es $P_1=\{1,2,3,4,5,8,9\}$ y $P_2=\{6,7\}$ que son los conglomerados a nivel 9.

Los conglomerados a nivel 12 se obtienen de la matriz

1	2	3	4	5	8	9	6	7
	0	0	0	0	0	0	0	1
		0	0	0	0	0	1	1
			0	0	0	0	1	1
				0	0	0	1	0
					0	0	1	1
						0	1	1
							1	1
								0

Obteniendose como nuevos conglomerados el $P_3=\{1,6\}$ y $P_4=\{5,7\}$ que añadidos a P_1,P_2 forman los conglomerados a nivel 12.

El unico conglomerado a nivel 13 es el total.

que se construye como sigue:

1	2	3	4	5	6	7	8	9	
	4	6	9	9	12	13	6	7	1
		7	9	8	13	13	4	3	2
			9	9	13	13	7	7	3
				9	13	13	9	9	4
					13	12	9	9	5
						8	13	13	6
							13	13	7
								7	8
									9

Se eligen los 5 elementos minimos que son el {3,4,4,6,6} ; y como no se deduce ninguno pues intervienen 6 indices se añade el siguiente minimo que es el 7, que se alcanza en dos lugares $d(2,3)$ y $d(3,9)$; deduciendose $d(3,8)=7$, $d(1,9)=7$ y $d(8,9)=7$.

Como no se deduce ninguno más se añade el minimo que es el 8 que se alcanza en dos lugares $d(2,5)$ y $d(6,7)$; se sigue añadiendo el minimo que es el 9, que está en 3 lugares $d(1,4)$, $d(3,5)$, y $d(4,9)$; deduciendose

$d(2,4)=9$, $d(3,4)=9$, $d(4,8)=9$, $d(1,5)=9$, $d(4,5)=9$, $d(5,8)=9$, y $d(5,9)=9$

Como no se deducen mas se añade el minimo que es 12, que se alcanza en $d(1,6)$ y en $d(5,7)$ y se sigue añadiendo el minimo que es el 13 que se alcanza en $d(6,9)$; deduciendose

$d(2,6)=13$, $d(3,6)=13$, $d(4,6)=13$, $d(5,6)=13$, $d(6,8)=13$, $d(1,7)=13$

$d(2,7)=13$, $d(3,7)=13$, $d(4,7)=13$, $d(7,8)=13$, y $d(7,9)=13$

construyendo así la 2-ultramétrica fina.

Hallamos los conglomerados para d_1 a nivel 9; poniendo

$$f(a,b)=\begin{cases} 0 & \text{si} \quad (a,b) \in \gamma_{d_1}(9) \\ 1 & \text{si} \quad (a,b) \notin \gamma_{d_1}(9) \end{cases}$$

Tenemos así la matriz;:

C O M E N T A R I O   F I N A L

¿Qué sentido tienen los métodos subdominantes en conglomerados, cuando la máxima información se obtiene a partir de los conglomerados de desemejanza inicial?

Creo que son varias las razones que justifican estos métodos; algunas de las cuales expondremos brevemente a con= tinuación.

a) La dificultad de cálculo de todos los conglomerados para un coe- ficiente de desemejanza cualquiera.

b) La dificil interpretación de conglomerados que no cumplen unas determinadas propiedades.

Estas objeciones quedan salvadas mediante ciertos métodos subdominantes.

BIBLIOGRAFIA;

JARDINE, N. & ROBIN, S.

" *Mathematical Taxonomy* "

1971. New York, John Wiely.

SANCHEZ, M.

" *Determinación de las palabras maximales de un alfabeto* ". *Trabajos de Estadística e Investigación Operativa.*

Vol. XXVI - pág. 319 - 1975.