

# GENERALIZACIÓN DE LA CLASE DE ALGORITMOS DE FORMACIÓN DE CONGLOMERADOS JERÁRQUICOS CON ESTRATEGIA AGLOMERATIVA

M.Sc. Ramón Alberto Mollineda Cárdenas  
Lic. Boris Jesús Mederos Madrazo\*

## 1. INTRODUCCIÓN

Los métodos de formación de conglomerados jerárquicos vistos hasta ahora en la literatura, se basan en criterios de enlaces que sólo toman en cuenta la distancia absoluta entre dos conglomerados. Esta característica, obviamente daría una medida de la cercanía entre los dos conglomerados, ¿pero con respecto a qué?. Además estos métodos no tienen una condición de parada natural, o sea, no tienen criterio para determinar cuándo se ha alcanzado un agrupamiento óptimo. En algunos casos su objetivo es formar un número de conglomerados previamente establecido por el usuario. En otros casos van realizando progresivos agrupamientos hasta que el usuario determina que se ha obtenido uno que es satisfactorio para él. En ninguno de los dos casos se puede garantizar haber alcanzado un agrupamiento óptimo.

En este trabajo se expone una generalización de la clase de algoritmos de formación de conglomerados jerárquicos en su estrategia aglomerativa, que para realizar la unión de dos conglomerados no sólo toma en cuenta la distancia absoluta que los separa, sino que también considera la distancia relativa entre

ellos con respecto al promedio de las distancias que los separan de los demás conglomerados. Es decir no basta que ambos conglomerados sean los más cercanos o semejantes para que sean unidos, sino que además deben ser suficientemente distantes respecto de los demás conglomerados del agrupamiento actual. Este enfoque responde mejor al concepto de grupo o clase, es decir, no basta con que exista una pequeña variación intraclase, sino que además debe existir gran separación interclase.

Los algoritmos tendrán un criterio de parada natural, que ocurrirá cuando en un agrupamiento no es posible encontrar dos conglomerados que estén lo suficientemente cercanos con respecto al criterio de semejanza relativa establecido por el algoritmo y determinado por un parámetro suministrado por el usuario. Este parámetro representará un nivel de exigencia para unir dos conglomerados. De esta forma el algoritmo encuentra el agrupamiento óptimo con respecto al valor del parámetro suministrado por el usuario.

El menor agrupamiento que se obtiene usando estos algoritmos estará constituido

---

\* Facultad de Matemática, Física y Computación Universidad Central de Las Villas. CUBA. Email: mfc@ucentral.quantum.inf.cu

por dos conglomerados y no por uno como suele ocurrir con los criterios de enlace clásicos. Esto es debido a que la semejanza o cercanía entre dos conglomerados, siempre se establece con relación al contexto o agrupamiento en que se encuentran.

## 2. FUNDAMENTACIÓN TEÓRICA DE LA CLASE DE ALGORITMOS DE FORMACIÓN DE CONGLOMERADOS JERÁRQUICOS.

Dado un conjunto  $U$  de  $n$  objetos, se quiere agrupar sus elementos en diferentes clases, es decir, se quiere realizar una partición de  $U$  tal que de la pertenencia de un elemento de  $U$  a una clase se obtenga determinada información.

Se denomina esquema jerárquico de formación de conglomerados a la generación de una sucesión de particiones (agrupamientos) del conjunto  $U$ :

$$P = \{P_j\}_{j=0, \dots, m} \quad m < n$$

con cierto criterio de enlace y estrategia de aglomeración, según se define a continuación.

Una vez obtenida  $P$ , corresponde a un especialista seleccionar la partición que mejor clasifique los elementos de  $U$ .

Para realizar un agrupamiento jerárquico es necesario tomar en cuenta dos aspectos: el criterio de enlace y la estrategia, la cual puede ser aglomerativa o divisiva.

**Estrategia Aglomerativa:** Esta estrategia comienza con la partición trivial de  $U$ , que estará formada por  $n$  clases, cada una integrada por solo un elemento del conjunto  $U$ . Utilizando un criterio de enlace se van obteniendo diferentes particiones hasta

llegar a la partición final, que estará constituida por una sola clase, o se llega a una determinada partición por medio de un criterio de parada. En cada paso los subconjuntos de la partición surgen de unir dos o más subconjuntos de la partición precedente.

$$P = \{P_j\}_{j=0, \dots, m} \quad m < n, \quad P_0 \text{ partición trivial}$$

$$\text{donde: } P_j = \{C_{jk}\}_{k=1, \dots, m_j} \quad m_j = |P_j|$$

$$C_{jk} = \cup C_{j-1,r}, \quad r \in R_k, \quad R_k \subseteq \{1, \dots, m_{j-1}\}$$

**Estrategia Divisiva:** Tiene el sentido contrario de la anterior, es decir, parte de una partición de  $U$  formada por una sola clase que contiene todos los elementos del conjunto  $U$  y mediante un criterio de separación llega a una determinada partición.

### 2.1 Criterios de Enlace

Son las reglas que utilizan los algoritmos para agrupar entidades en la formación de conglomerados. Hay varios criterios de enlace, cada uno de los cuales determinan un método de aglomeración jerárquico. A continuación se explicarán cuatro de las formas clásicas de enlace más populares, así como dos nuevos criterio de enlaces propuestos por los autores que determinan los algoritmos de este trabajo.

**Enlace Simple:** Sea  $d$  una distancia definida sobre los elementos del conjunto  $U$ . Este criterio une dos conglomerados  $C_1$  y  $C_2$  si la "distancia":

$$\text{dist}(C_1, C_2) = \min d(x, y), \quad x \in C_1, \quad y \in C_2 \quad (2.1)$$

es la distancia más pequeña, entre todos los pares de conglomerados de cierto agrupamiento.

**Enlace Completo:** Sea  $d$  una distancia definida sobre los elementos del conjunto  $U$ . Este criterio une dos conglomerados  $C_1$  y  $C_2$  si la "distancia":

$$\text{dist}(C_1, C_2) = \max_{x \in C_1, y \in C_2} d(x, y) \quad (2.2)$$

es la distancia más pequeña, entre todos los pares de conglomerados de cierto agrupamiento.

**Enlace Promedio:** Sea  $d$  una distancia definida sobre los elementos del conjunto  $U$ . Este criterio une dos conglomerados  $C_1$  y  $C_2$  si la "distancia":

$$\text{dist}(C_1, C_2) = \frac{1}{|C_1| + |C_2|} \sum_{(x,y) \in C_1 \cdot C_2} d(x, y), \quad x \in C_1, y \in C_2 \quad (2.3)$$

es la distancia más pequeña, entre todos los pares de conglomerados de cierto agrupamiento.

**Enlace Centroide:** Sea  $d$  una distancia definida sobre los elementos del conjunto  $U$ . Este criterio une dos conglomerados  $C_1$  y  $C_2$  si la "distancia":

$$\text{dist}(C_1, C_2) = d(\bar{C}_1, \bar{C}_2) \quad (2.4)$$

donde  $\bar{C}_1$  es el centroide de  $C_1$  y  $\bar{C}_2$  es el centroide de  $C_2$ , es la distancia más pequeña, entre todos los pares de conglomerados de cierto agrupamiento.

### 3. GENERALIZACIÓN DE LOS CRITERIOS DE ENLACES.

Hasta aquí hemos visto algunos de los criterios de enlaces más populares. En todos ellos para realizar la unión de dos conglomerados, sólo toman en cuenta la distancia absoluta que los separa, sin considerar cuan distintos o separados están de los demás conglomerados.

A continuación aparecen los nuevos criterios de enlace, los cuales se basan en la filosofía de trabajo expuesta en la introducción. El primero de ellos representa una generalización de los criterios de enlace clásicos vistos anteriormente, y de otros que no se han referido aquí. Esto significa una generalización de la filosofía de trabajo de los algoritmos de formación de conglomerados jerárquicos. El segundo es más específico, pero también se basa en la misma idea.

**Criterio de Enlace Generalizado:** Sea  $d$  una distancia definida sobre los elementos del conjunto  $U$ . Sea  $\text{dist}$  cualquiera de las distancias entre conglomerados definidas en (2.1), (2.2), (2.3) y (2.4). Sea  $\alpha$  un parámetro entre 0 y 1. Sean los conglomerados  $C_1, C_2$  los que tuvieron la distancia  $\text{dist}$  mínima entre todos los pares de conglomerados. Sea  $d_1$  el promedio de las distancias entre el conglomerado  $C_1$  y los demás conglomerados del agrupamiento actual excluyendo a  $C_2$ . Sea  $d_2$  el promedio de las distancias entre el conglomerado  $C_2$  y los demás conglomerados del agrupamiento actual excluyendo a  $C_1$ . Este criterio une los conglomerados  $C_1$  y  $C_2$  si se cumple la siguiente condición:

$$\text{dist}(C_1, C_2) < \alpha \cdot \min\{d_1, d_2\}$$

De esta forma se está condicionando la unión de los dos conglomerados más cercanos, al cumplimiento de la desigualdad, que representa el criterio de semejanza relativa entre estos conglomerados, respecto al resto de los conglomerados del agrupamiento actual. Mientras más pequeño sea el parámetro  $\alpha$ , éste constituirá una condición más exigente para unir dos conglomerados. Si  $\alpha$  fuera 1, esta generalización se reduciría a la filosofía clásica, es decir funcionaría de la misma forma que los criterios de enlace clásicos.

**Otro Criterio de Enlace:** Sea  $d$  una distancia definida sobre los elementos del conjunto  $U$ . Sea  $\text{dist}$  la distancia entre conglomerados definida en (2.3). Sea  $\alpha$  un parámetro entre 0 y 1. Sean los conglome-

rados  $C_1, C_2$  los que tuvieron la distancia (2.3) mínima entre todos los pares de conglomerados. Este criterio une los conglomerados  $C_1$  y  $C_2$  si se cumple la siguiente condición:

$$\frac{\sum_{(x,y) \in C_1 \times C_2} d(x,y) + \sum_{(x,y) \in C_1 \times C_1} d(x,y) + \sum_{(x,y) \in C_2 \times C_2} d(x,y)}{|C_1| \cdot |C_2| + \binom{|C_1|}{2} + \binom{|C_2|}{2}} < \alpha \cdot \frac{\sum_{\substack{(x,y) \in C_1 \times C_k \vee (x,y) \in C_2 \times C_k \\ k \neq 1, k \neq 2}} d(x,y)}{(|C_1| + |C_2|) \cdot |U \setminus (C_1 \cup C_2)|}$$

Esta condición, aunque larga de escribir, es fácil de enunciar. El promedio de las distancias entre todos los pares no ordenados de elementos del nuevo conglomerado, no puede exceder el promedio de las distancias entre todos los pares no ordenados compuestos por un elemento del nuevo conglomerado y otro de su complemento multiplicado por un parámetro  $\alpha$  entre 0 y 1. El parámetro  $\alpha$  tiene el mismo significado que en el caso anterior. De no cumplirse esta condición los dos conglomerados más cercanos no podrán ser unidos, pues no se garantiza que estén, para ese  $\alpha$ , suficientemente separados del resto de los conglomerados, o sea no se garantiza que constituyan un grupo dentro de la muestra.

central consiste en unir los dos conglomerados más cercanos, si y sólo si, ambos están de forma relativa a la distancia entre ellos, lo suficientemente alejados del resto de los conglomerados.

Formalicemos la idea: sea  $\text{dist}$  una distancia entre conglomerados que podría ser cualquiera de las ya definidas en la sección de fundamentación teórica ya vista. Sea  $\alpha$  un parámetro entre 0 y 1, con el mismo significado ya explicado en la sección anterior. Sean  $C_p$  y  $C_q$  los dos conglomerados que por la distancia anterior resultaron ser los más cercanos, entonces la condición necesaria y suficiente que además de su cercanía necesitarían para ser unidos es:

$$\text{dist}(C_p, C_q) < \alpha \cdot \min\{\overline{d_p}, \overline{d_q}\}$$

Como estos dos criterios de enlace son aplicables a la estrategia aglomerativa, generalmente parten de la partición trivial, o sea, de la partición formada por  $n$  conglomerados, cada uno integrado por un solo elemento, y van creando sucesivas particiones hasta que en cierta partición, ningún par de conglomerados cumple con la condición.

donde  $\overline{d_p}$  sería el promedio de las distancias entre el conglomerado  $C_p$  y los demás conglomerados excluyendo a  $C_q$ , y  $\overline{d_q}$  sería el promedio de las distancias del conglomerado  $C_q$  a los demás conglomerados excluyendo a  $C_p$ .

#### 4. FORMALIZACIÓN DEL ALGORITMO DE FORMACIÓN DE CONGLOMERADOS JERÁRQUICOS BASADO EN EL CRITERIO DE ENLACE GENERALIZADO.

El criterio de enlace generalizado conduce a un algoritmo de agrupamiento, cuya idea

Por lo tanto no basta con la cercanía entre dos conglomerados para que éstos puedan ser unidos, sino que además ellos deben estar suficientemente separados de los demás conglomerados.

#### 4.1 Algoritmo

1. Formar un agrupamiento inicial. Dada una muestra  $U = \{O_1, O_2, \dots, O_n\}$ , generalmente se comienza por el agrupamiento trivial  $P_1 = \{\{O_1\}, \{O_2\}, \dots, \{O_n\}\}$ .

Especificar un valor para  $\alpha$  entre 0 y 1.

Inicializar el nivel de agrupamiento  $i = 1$ .

Sea  $dist$  una distancia entre conglomerados que puede ser cualquiera de las definidas en la sección de fundamentación teórica.

Se almacenará además la siguiente información que representa una relación entre cada par no ordenado de conglomerados distintos. En los pasos siguientes quedará aclarado su papel en el algoritmo.

$R(C_p, C_q) = NO\_CONSIDERADA$ , para todo  $p, q, p \neq q$ .

2. Seleccionar los dos conglomerados  $C_p, C_q, p \neq q$  del agrupamiento actual  $P_i$  con distancia  $dist(C_p, C_q)$  mínima, entre todos los pares no ordenados de conglomerados que cumplen que  $R(C_p, C_q) = NO\_CONSIDERADA$ .

Si no es posible encontrar dos conglomerados para ser seleccionados ir al paso 7.

Sean  $C_a, C_b$  los conglomerados seleccionados.

3. Calcular en  $\overline{d}_a$  el promedio de las distancias de  $C_a$  al resto de los conglomerados excluyendo a  $C_b$ .

Calcular en  $\overline{d}_b$  el promedio de las distancias de  $C_b$  al resto de los conglomerados excluyendo a  $C_a$ .

4. Verificar el cumplimiento de la desigualdad:

$$dist(C_a, C_b) < \alpha \cdot \min\{\overline{d}_a, \overline{d}_b\}$$

Si la desigualdad se cumple ir al paso 5; de lo contrario ir al paso 6.

5. Unir los conglomerados  $C_a$  y  $C_b$  de la siguiente forma:

$C_a = C_a \cup C_b$ ; posteriormente actualizar el agrupamiento  $P_{i+1} = P_i \setminus C_b$ , e incrementar  $i$ . Realizar las siguientes actualizaciones:

Recalcular las distancias del nuevo conglomerado  $C_a$  al resto de los conglomerados.

$R(C_p, C_q) = NO\_CONSIDERADA$  para todo par no ordenado  $(C_p, C_q)$  del agrupamiento actual

Ir al paso 2.

6. Marcar la relación entre los conglomerados  $C_a$  y  $C_b$  de la siguiente forma:

$R(C_a, C_b) = CONSIDERADA$

Ir al paso 2.

7. Dar como resultado el agrupamiento actual  $P_i$ .

FIN.

#### 5. COMPARACIÓN CON MÉTODOS TRADICIONALES.

Para realizar la comparación se usaron dos ficheros de datos de un problema caracterizado por 13 variables, y además con información de clases. Estos ficheros contenían 250 y 136 casos respectivamente separados en dos clases bien definidas previamente por un especialista. Cada fichero trata en cuestión individuos enfermos y sanos de disfunción cerebral y las variables constituyen posibles factores de riesgo de esa enfermedad con un nivel medición continuo. El primer fichero contiene 85 casos de individuos enfermos y 165 de individuos sanos y el segundo contiene 45 casos de individuos enfermos y 91 de sanos, es decir cada fichero contiene un conjunto de elementos sepa-

rados en dos grupos bien definidos por un experto.

Para realizar la validación del nuevo enfoque de agrupamiento se seleccionaron dos métodos pertenecientes a la clase de los algoritmos de formación de conglomerados jerárquicos. Estos son los determinados por los criterios de enlace Promedio y Centroide ya vistos en la sección de Fundamentación Teórica. Estos métodos clásicos fueron comparados con sus métodos homólogos respectivamente, es decir los determinados por el nuevo criterio de enlace generalizado para las distancias entre conglomerados presentes en los enlaces Promedio y Centroides. En todos los casos se usó como distancia entre elementos de la muestra la distancia euclidiana.

Las comparaciones fueron realizadas analizando los agrupamientos de cada uno

de los métodos para 2, 3 y 4 conglomerados. En el caso de los métodos clásicos que generan una única secuencia de agrupamientos se escogieron los que contenían el número de conglomerados especificados. Con respecto a los nuevos criterios de enlace, fueron ejecutados para los a más pequeños, que dieron como resultados agrupamientos con ese número de conglomerados respectivamente. El objetivo fundamental fue detectar si a esos niveles de agrupamientos los algoritmos eran capaces de reconocer la existencia de los dos grandes grupos. El criterio que tomamos para verificar la eficiencia de los agrupamientos fue comparar éstos con la clasificación dada por el especialista en enfermos y sanos.

Comenzaremos la comparación con los resultados obtenidos del fichero de datos formado por 250 elementos de la muestra.

### **Criterio de Enlace: Promedio. Distancia entre elementos: Euclidiana**

#### **Formación de dos conglomerados**

	<b>Método Clásico</b>		<b>Método Generalizado (<math>\alpha = 0.78</math>)</b>	
	Clase 1	Clase 2	Clase 1	Clase 2
Conglomerado 1	84	165	84	165
Conglomerado 2	1	0	1	0

#### **Formación de tres conglomerados**

	<b>Método Clásico</b>		<b>Método Generalizado (<math>\alpha = 0.75</math>)</b>	
	Clase 1	Clase 2	Clase 1	Clase 2
Conglomerado 1	84	162	3	163
Conglomerado 2	1	0	81	2
Conglomerado 3	0	3	1	0

#### **Formación de cuatro conglomerados**

	<b>Método Clásico</b>		<b>Método Generalizado (<math>\alpha = 0.70</math>)</b>	
	Clase 1	Clase 2	Clase 1	Clase 2
Conglomerado 1	10	157	3	162
Conglomerado 2	74	5	81	2
Conglomerado 3	1	0	1	0
Conglomerado 4	0	3	0	1

En el análisis anterior se destacan los siguientes casos: en la formación de tres conglomerados usando el criterio de enlace Promedio, el algoritmo generalizado logra reconocer los dos grandes

grupos y el método clásico no; en la formación de cuatro conglomerados el método generalizado logra reconocer con mayor eficiencia que el método clásico ambos grupos.

### Criterio de Enlace: Centroide. Distancia entre elementos: Euclidiana

#### Formación de dos conglomerados

	Método Clásico		Método Generalizado ( $\alpha = 0.74$ )	
	Clase 1	Clase 2	Clase 1	Clase 2
Conglomerado 1	84	165	84	165
Conglomerado 2	1	0	1	0

#### Formación de tres conglomerados

	Método Clásico		Método Generalizado ( $\alpha = 0.70$ )	
	Clase 1	Clase 2	Clase 1	Clase 2
Conglomerado 1	84	164	84	164
Conglomerado 2	1	0	1	0
Conglomerado 3	0	1	0	1

#### Formación de cuatro conglomerados

	Método Clásico		Método Generalizado ( $\alpha = 0.68$ )	
	Clase 1	Clase 2	Clase 1	Clase 2
Conglomerado 1	6	156	3	160
Conglomerado 2	78	8	81	4
Conglomerado 3	1	0	1	0
Conglomerado 4	0	1	0	1

En la tabla anterior en la comparación de los métodos con el criterio de enlace centroide, no se aprecia superioridad en la formación de 2 y 3 conglomerados, sin embargo en la formación de 4 aunque ambos reconocen los

dos grandes grupos, el método generalizado lo hace con mayor eficiencia.

A continuación mostramos los resultados comparativos para el fichero de datos de 136 elementos.

### Criterio de Enlace: Promedio. Distancia entre elementos: Euclidiana

#### Formación de dos conglomerados

	Método Clásico		Método Generalizado ( $\alpha = 0.95$ )	
	Clase 1	Clase 2	Clase 1	Clase 2
Conglomerado 1	32	2	38	2
Conglomerado 2	9	93	3	93

#### Formación de tres conglomerados

	Método Clásico		Método Generalizado ( $\alpha = 0.925$ )	
	Clase 1	Clase 2	Clase 1	Clase 2
Conglomerado 1	32	2	38	2
Conglomerado 2	7	90	1	90
Conglomerado 3	2	3	2	3

#### Formación de cuatro conglomerados

	Método Clásico		Método Generalizado ( $\alpha = 0.85$ )	
	Clase 1	Clase 2	Clase 1	Clase 2
Conglomerado 1	7	89	1	90
Conglomerado 2	7	1	7	0
Conglomerado 3	25	2	31	2
Conglomerado 4	2	3	2	3

En este caso específico, en todos los análisis fueron reconocidos ambos grupos, pero

siempre la eficiencia del método generalizado fue significativamente superior.

**Criterio de Enlace: Centroide. Distancia entre elementos: Euclidiana**  
**Formación de dos conglomerados**

	<b>Método Clásico</b>		<b>Método Generalizado (<math>\alpha = 0.77</math>)</b>	
	Clase 1	Clase 2	Clase 1	Clase 2
Conglomerado 1	39	95	39	95
Conglomerado 2	2	0	2	0

**Formación de tres conglomerados**

	<b>Método Clásico</b>		<b>Método Generalizado (<math>\alpha = 0.75</math>)</b>	
	Clase 1	Clase 2	Clase 1	Clase 2
Conglomerado 1	38	92	2	93
Conglomerado 2	2	0	38	0
Conglomerado 3	1	3	1	2

**Formación de cuatro conglomerados**

	<b>Método Clásico</b>		<b>Método Generalizado (<math>\alpha = 0.68</math>)</b>	
	Clase 1	Clase 2	Clase 1	Clase 2
Conglomerado 1	33	9	34	0
Conglomerado 2	5	84	2	93
Conglomerado 3	2	0	4	0
Conglomerado 4	1	2	1	2

En este último análisis tenemos que destacar el caso en que se formaron tres conglomerados y en el cual el método generalizado reconoció los dos grandes grupos y el método clásico no. En la formación de cuatro conglomerados se reflejó una mayor eficiencia en el agrupamiento del método clásico.

Como hemos podido apreciar los resultados en todos los casos, han sido mejores o iguales por parte del método generalizado. En la mayoría de los agrupamientos el método generalizado para el  $\alpha$  indicado ha reconocido mejor los dos grupos definidos que el método clásico y por supuesto ha cometido menos errores en el momento de ubicar los elementos de ambas clases definidas por los expertos y comprobado por la práctica en los grupos formados.

**6. Conclusiones**

Fue realizada una generalización a la clase de algoritmos de formación de conglomerados jerárquicos con estrategia aglomerativa. Esta se basa en una idea que obedece a las propiedades que debe cumplir cierta muestra que realmente esté distribuida formando grupos o clases bien definidas. Esta idea es la de que los grupos o clases no sólo posean una pequeña variación intraclase, sino que además mantengan suficiente separación interclase.

Los algoritmos basados en este enfoque unen los dos conglomerados más cercanos, si y sólo si cumplen cierta condición, en la que la distancia que los separa debe ser menor que el promedio de las distancias de cada uno de ellos a los demás conglomerados de

la muestra, multiplicado por cierto factor  $\alpha$  entre 0 y 1. En el caso de  $\alpha = 1$ , el algoritmo se comporta exactamente como los métodos clásicos. Estos métodos generalizados tienen un criterio de parada natural, que ocurrirá cuando en un agrupamiento no es posible encontrar dos conglomerados que estén lo suficientemente cercanos con respecto al criterio de semejanza relativa establecido por el algoritmo y determinado por el parámetro  $\alpha$  suministrado por el usuario. Este parámetro representa un nivel de exigencia para unir dos conglomerados.

Esta idea puede extenderse para realizar la unión de dos conglomerados utilizando la dispersión de los elementos que pertenezcan a ellos. Los dos conglomerados candidatos a unirse serán los que cumplan que al ser unidos forman un conglomerado con mínima dispersión de sus elementos, comparado con los que puedan formarse por la unión de todos los pares no ordenados de conglomerados del agrupamiento actual. La condición que debe satisfacerse para que sean definitivamente unidos es que el futuro conglomerado cumpla que la dispersión de

sus elementos no supere la dispersión de la muestra multiplicada por un factor  $\alpha$  entre 0 y 1.

Para validar el nuevo enfoque de agrupamiento se seleccionaron dos métodos pertenecientes a la clase de los algoritmos de formación de conglomerados jerárquicos. Estos métodos fueron determinados por dos criterios de enlace clásicos. Los métodos clásicos fueron comparados con sus métodos homólogos respectivamente, es decir los determinados por los mismos criterios de enlace con la condición generalizadora. Se usaron dos ficheros de datos de un problema caracterizado por 13 variables con información de clase. Estos ficheros contenían 250 y 136 casos respectivamente separados en dos clases. El criterio que tomamos para verificar la eficiencia de los agrupamientos, fue comparar éstos con las clases previamente determinadas por el especialista y comprobadas en la práctica. En todos los casos se obtuvieron resultados mejores o iguales por parte de los algoritmos generalizados en relación con los clásicos.

## 7. BIBLIOGRAFÍA

- ANDERSON, T. W. *An introduction to multivariate statistical analysis*. New York: Wiley, 1958.
- GRAU, R. *Estadística Aplicada con Ayuda de Paquetes de Software*. Curso de Postgrado, Imprenta UdeG. Guadalajara. México, 1994.
- GRUVAEUS, G., & Wainer, H. *Two additions to hierarchical cluster analysis*. The British Journal of Mathematical and Statistical Psychology, 1972, p. 25, 200-206.
- HARTIGAN, J. A. *Clustering algorithms*. New York: Wiley, 1975.
- JOHNSON, S. C. *Hierarchical clustering schemes*. Psychometrika, 1967, p. 32, 241-254.
- TRYON, R. C. *Cluster Analysis*. Ann Arbor, MI: Edwards Brothers, 1939.
- ZUPAN, J. *Clustering of large data sets*. New York: Research Studies Press, 1982.