CENATAV
Centro de Aplicaciones de Tecnologías de Avanzada

RNPS No. 2142
ISSN 2072-6287
Versión Digital

REPORTE TÉCNICO
Reconocimiento de Patrones

SERIE AZUL

Estructuraciones libres con datos mezclados e incompletos

José Ruiz-Shulcloper, Jesús Ariel Carrasco-Ochoa y José Francisco Martínez-Trinidad

RT_079 febrero 2016

7ma. No. 21812 e/218 y 222,
Rpto. Siboney, Playa;
La Habana.
Cuba. C.P. 12200
www.cenatav.co.cu

RNPS No. 2142
ISSN 2072-6287
Versión Digital

CENATAV
Centro de Aplicaciones de Tecnologías de Avanzada

REPORTE TÉCNICO
Reconocimiento de Patrones

SERIE AZUL

Estructuraciones libres con datos mezclados e incompletos

José Ruiz-Shulcloper, Jesús Ariel Carrasco-Ochoa y José Francisco Martínez-Trinidad

RT_079 febrero 2016

7ma. No. 21812 e/218 y 222,
Rpto. Siboney, Playa;
La Habana.
Cuba. C.P. 12200
www.cenatav.co.cu

Estructuraciones libres con datos mezclados e incompletos

José Ruiz-Shulcloper[1], Jesús Ariel Carrasco-Ochoa[2] y José Francisco Martínez-Trinidad[2]

[1] Equipo de Reconocimiento de Patrones, CENATAV - DATYS, La Habana, Cuba
jshulcloper@cenatav.co.cu
[2] Coordinación de Ciencias Computacionales, Instituto Nacional de Astrofísica, Óptica y Electrónica (INAOE), Puebla, México
{ariel, fmartine}@ccc.inaoep.mx

RT_079, Serie Azul, CENATAV - DATYS
Aceptado: 25 de noviembre de 2015

Resumen. En el trabajo se expone un conjunto de criterios de agrupamiento para la estructuración de un conjunto de objetos descritos en términos de datos mezclados e incompletos, en los cuales el número de agrupamientos no es predeterminado. Sobre la base de estas conceptualizaciones se establecen las relaciones conjuntuales entre cada uno de estos agrupamientos. Además, se exponen los algoritmos para obtener tales estructuraciones y se hace un análisis de las modificaciones que se producen en las estructuraciones cuando se efectúan diferentes cambios en los datos y se muestran los procedimientos para obtener los nuevos agrupamientos sin necesidad de recalcular desde el inicio la nueva estructuración.

Palabras clave: clasificación no supervisada, reconocimiento lógico combinatorio de patrones.

Abstract. In this work, several clustering criteria for structuring a mixed and incomplete dataset are presented. In these clustering criteria the number of clusters to be built is not specified a priori as a parameter. Over this conceptualization of clustering, the relations on clusters built by the different clustering criteria are enunciated. Additionally, the algorithms for structuring a dataset through the clustering criteria are presented. The analysis of changes in the clusters built by the clustering criteria when different types of changes in the dataset occur is performed. Based on this analysis, algorithms to build the new clusters, without re-computing the clusters from the beginning, are introduced.

Keywords: unsupervised classification, logical combinatorial pattern recognition.

1 Introducción

Uno de los problemas fundamentales en el Reconocimiento de Patrones es la *clasificación no supervisada* (*sin entrenamiento*) y esto obedece entre otras cosas a que esta problemática ha estado presente desde el inicio en prácticamente todas las áreas del conocimiento. Baste mencionar las investigaciones taxonómicas en la Biología, la Psicología, las Geociencias, las Ciencias Sociales y la Mercadotecnia, entre muchas otras. Es por otro lado, una de las bases fundamentales en la cada vez más necesaria línea de trabajos en Descubrimiento de Conocimiento y Minería de Datos. En todos ellos está presente el problema de la clasificación no supervisada. Este problema consiste, en esencia, en hallar la *estructura interna* de un conjunto de descripciones de objetos en su espacio de representación. Esta estructura interna depende, en una primera instancia, de la selección del propio espacio de representación y de la forma en que los objetos se comparen, es decir, del *concepto de similaridad* (*semejanza*) *que se utilice y de la forma en que éste se emplee*, lo que equivale a decir, que la estructura interna de un conjunto de datos depende del *criterio de agrupamiento* que se aplique.

De dicha estructuración, en términos generales se pudiera decir que:

a.- se sabe o se desea que se realice en un número dado de agrupaciones;

b.- se desconoce en cuántas agrupaciones se estructurará el conjunto de objetos una vez definidos el espacio de representación, los conceptos de similaridad y el criterio de agrupamiento.

Consecuentemente, los algoritmos de agrupamiento de objetos pueden tipificarse en dos grandes familias:

- *algoritmos de agrupamiento libre*, cuando el número de agrupamientos a obtener es desconocido; y
- *algoritmos de agrupamiento restringido*, cuando se exige que el universo sea estructurado en un número dado de agrupamientos.

Es importante resaltar que, en ambos casos, existen *paradigmas de la estructuración de universos*, que abordan esos problemas desde una óptica particular. Así podemos mencionar los paradigmas:

- *del conjunto cociente,*
- *del solapamiento,*
- *de estructuración difusa,*
- *de estructuración conceptual.*

En todos estos casos está presente también la problemática de los datos mezclados e incompletos y las funciones de similaridad (que no tienen que ser ni el opuesto ni el inverso de una función distancia, e incluso pudieran no ser funciones simétricas) y en todos ellos en la actualidad se cuenta con algoritmos que permiten resolver, y se han resuelto, problemas en diferentes áreas del conocimiento bajo las circunstancias mencionadas.

El *paradigma del conjunto cociente* consiste en la formación de una partición del conjunto de descripciones de los objetos bajo el presupuesto de que los mismos serán conjuntos en el sentido de la Teoría Clásica de Conjuntos. En otras palabras, de lo que se trata es de hallar el conjunto cociente del espacio de representación dado. Esto presupone que los agrupamientos serán disjuntos. Aquí las propiedades que caracterizan a un agrupamiento (*cluster* en inglés, *cúmulo*, *agrupación*) dado, contradicen las propiedades que caracterizan a cualquier otro de los restantes agrupamientos obtenidos. Esto es, son propiedades exclusivas.

El *paradigma del solapamiento* permite que los agrupamientos tengan elementos comunes, es decir, se trata de hallar un cubrimiento del conjunto de descripciones de objetos dado por subconjuntos (también en el sentido clásico) no necesariamente disjuntos. Las propiedades que caracterizan a un agrupamiento dado pudieran ser satisfechas por otros objetos de los agrupamientos restantes. Esto es, son propiedades inclusivas.

El *paradigma de estructuración difusa* consiste en que las agrupaciones son subconjuntos difusos del universo en cuestión. En este caso, las estructuraciones pueden ser cubrimientos o particiones difusas. Las propiedades que caracterizan a los diferentes agrupamientos son satisfechas por todos los objetos en un cierto grado, como es característico en la Teoría de Subconjuntos Difusos.

El *paradigma de estructuración conceptual* consiste en hallar la determinación intencional de las agrupaciones que conformarán la estructuración, es decir, dar las propiedades que cumplen los objetos que pertenecen a cada uno de los agrupamientos. Estas estructuraciones pueden estar constituidas por conjuntos duros o difusos formando particiones o cubrimientos.

En todos estos paradigmas hay factores comunes. Uno de ellos, esencial para la solución del problema de clasificación no supervisada, es la selección del *criterio de agrupamiento*.

La selección de un criterio de agrupamiento (que más adelante se define con precisión) puede realizarse de maneras diferentes:

- se llega al criterio de agrupamiento mediante la modelación matemática del problema y por la misma vía se llega al paradigma de realización de la estructuración del conjunto de objetos;
- se supone el paradigma, es decir, se impone, y se condiciona el criterio de agrupamiento de modo tal que resulte una estructuración acorde con el paradigma seleccionado.

El estudio de todos estos paradigmas se puede realizar bajo dos ópticas que aunque muy relacionadas poseen diferencias, en apariencia sutiles, que aquí se consideran de suma importancia para los análisis posteriores. Se trata de lo que pudiera denominarse un *enfoque clasificatorio* y un *enfoque conjuntual*.

En el *enfoque clasificatorio* se tiene un universo de objetos y se necesita agruparlos de modo tal que <u>los objetos del mismo agrupamiento se parezcan más entre sí que con objetos de otros agrupamientos</u>.

En el *enfoque conjuntual* se tiene un universo de objetos y se necesita agruparlos de modo tal que <u>los objetos que estén en el mismo agrupamiento cumplan (en un cierto grado) la propiedad que caracteriza al agrupamiento</u> (como conjunto en su determinación intencional).

Como ya se ha referido de alguna manera, el objetivo fundamental del problema de clasificación no supervisada es el de conocer la estructura interna de una población de objetos dada. Esa población pudiera ser, por ejemplo, una clase de objetos en un problema de clasificación supervisada o parcialmente supervisada. El interés en lograr esa estructuración puede ser porque se desea posteriormente clasificar nuevos objetos ya que la *población* a la que se está haciendo referencia no es todo el universo de objetos de un problema en cuestión. El interés puede ser, por ejemplo, la necesidad de analizar los datos con fines de depuración, de seleccionar la mejor representación de la clase en cuestión, entre otros.

Así, se puede afirmar que las causas intrínsecas de la agrupación responden a la *similitud* en el enfoque clasificatorio, al *cumplimiento de una propiedad* en el enfoque conjuntual.

Un concepto que sigue siendo tema de discusión es justamente el de *algoritmo de agrupamiento* (*clustering*). Hay muchas conceptualizaciones pero definiciones formales con un nivel general de aceptación no se han adoptado. En [Ruiz-Shulcloper and Montellano-Ballesteros (1995)] se introducen los siguientes conceptos:

Sea U un universo de objetos, $M\subseteq U$, Γ una función de semejanza definida sobre U, y β_0 un umbral de semejanza.

Como *criterio de agrupamiento duro* $\Pi(M,\Gamma,\beta_0)$ sobre $\Phi=(U,\Gamma)$ entenderemos un conjunto de predicados con parámetros M, Γ y β_0 tal que:

a) Genera una familia $\tau=\{NU_1,\ldots,NU_c\}$ de subconjuntos de M (*agrupamientos duros, clusters* en inglés) que cumplen:

i) $\forall NU\in\tau\,[NU\neq\varnothing]$,

ii) $\bigcup_{NU\in\tau} NU = M$,

iii) $\neg\exists NU_r, NU_{j_1},\ldots, NU_{j_k}\in\tau\,[NU_r\subseteq \bigcup_{\substack{t=1\\ j_t\neq r}}^{k} NU_{j_t}]$.

b) Define una relación $R_\Pi\subseteq M\times M\times\wp(M)$ (donde $\wp(M)$ es el conjunto potencia de M) que cumple:

iv) $\forall O_i,O_j\in M[\exists NU\in\tau\,\exists S\subseteq M\,[O_i,O_j\in NU\Leftrightarrow(O_i,O_j,S)\in R_\Pi]]$).

Cada NU_r elemento de la familia τ se denomina *núcleo*.

La relación R_Π es, en sí, la representación matemática de los predicados de Π, y por ello R_Π genera a la familia τ en el mismo sentido que una relación de equivalencia genera una partición, es decir, no define cuándo un objeto O_i pertenece a un agrupamiento (cluster), sino que nos dice cuándo dos objetos O_i y O_j pertenecen a un mismo núcleo (agrupamiento).

Los argumentos de R_Π son tales que la razón por la cual dos objetos pertenecen o no a un mismo agrupamiento puede depender del comportamiento global de las semejanzas entre otros objetos además del comportamiento de la semejanza entre ellos mismos. Ese conjunto de otros objetos es S, y por esto, R_Π determina de acuerdo a:

- el comportamiento entre las semejanzas de los objetos de $S \subseteq M$ entre sí,
- la semejanza entre O_i y O_j,
- y las semejanzas de éstos con los elementos de S,

cuándo O_i y O_j pertenecen a un mismo agrupamiento o no.

Como se puede constatar, dependiendo del criterio de agrupamiento, la función de semejanza Γ utilizada y la existencia de otros objetos, es lo que determina cuándo un objeto pertenece a un agrupamiento o porqué dos objetos pertenecen a un mismo agrupamiento. De esta manera se puede apreciar que la selección del criterio de agrupamiento a usar es crucial en la *calidad* de la solución del problema de clasificación no supervisada.

Con todo esto se observa que la definición del criterio de agrupamiento debe estar basada en el conocimiento que se tenga al respecto del problema en concreto que se está tratando, para poder definir así el tipo de comportamiento entre los objetos a partir de sus semejanzas, que resulte significativo, según el problema en particular.

Nótese que, al seleccionar algún criterio de agrupamiento, dado un conjunto de objetos y la función de semejanza, se ha definido indirectamente, la familia de agrupamientos, es decir, la estructura del universo ha sido conformada. En otras palabras, cada criterio de agrupamiento impone una estructuración sobre el conjunto que se pretende estructurar.

El planteamiento formal de la estructuración de universos, de la clasificación no supervisada, consiste en encontrar un criterio de agrupamiento que responda a los intereses del problema en cuestión.

2 Criterios de agrupamientos libres con datos mezclados e incompletos

Sea $M=\{O_1,...,O_m\} \subseteq U$, U un universo de objetos descritos en términos de un conjunto de rasgos $R=\{x_1,\ldots,x_n\}$ donde el rasgo $x_i: U \rightarrow M_i$, $x_i(O) \in M_i$; el conjunto de valores admisibles del rasgo x_i, i=1,...,n. Se considerará $M_1 \mathrm{x} \ldots \mathrm{x} M_n$, el producto cartesiano de los conjuntos de valores admisibles de los rasgos de R. Para cada rasgo x_i (i=1,...,n), se asocia un criterio de comparación C_i: $M_i \times M_i \longrightarrow L_i$. Para cualquier par de descripciones (subdescripciones) de objetos de U, se define una función de semejanza Γ: $\bigcup\limits_{T \subseteq R} (M_{i_1} x..xM_{i_p})^2 \rightarrow L$, donde L es un conjunto totalmente ordenado; $T=\{x_{i_1},..,x_{i_p}\} \subseteq R$, $p \geq 1$; $M_{i_1} x..xM_{i_p}$ el producto Cartesiano de sus respectivos conjuntos de valores admisibles.

A partir de M y Γ se puede construir una matriz que refleje las relaciones de semejanza entre todos los objetos sujetos a estudio. Esta matriz se denomina *matriz de similaridad* y será denotada como:

$$MS = \left\| \Gamma(O_i, O_j) \right\|_{mxm}.$$

MS es simétrica si lo es la función de semejanza. Por comodidad asumiremos en lo sucesivo que la función de semejanza es simétrica.

Sea $\beta_0 \in L$, un umbral de semejanza. Diremos que dos objetos O_i, O_j son β_0*-semejantes* si y sólo si $\Gamma(O_i,O_j) \geq \beta_0$. Análogamente, un objeto O_i es β_0*-ailsado* si y sólo si $\forall O_j \neq O_i \in M$ $\Gamma(O_i,O_j) < \beta_0$.

La magnitud $\beta_0 \in L$, es un parámetro que puede ser dado por expertos del área de la aplicación o calculado, por ejemplo, de alguna de las siguientes maneras:

a) $\beta_0 = \dfrac{2}{m(m-1)} \sum\limits_{i=1}^{m-1} \sum\limits_{j=i+1}^{m} \Gamma(O_i, O_j)$,

b) $\beta_0 = \max_{\substack{i=1,\dots,m\ 1 \\ i\neq j}} \{ \min_{j=i+1,\dots,m} \{\Gamma(O_i, O_j)\}\}$,

c) $\beta_0 = \min_{\substack{i=1,\dots,m\ 1 \\ i\neq j}} \{ \max_{j=i+1,\dots,m} \{\Gamma(O_i, O_j)\}\}$.

En todo caso, este parámetro debe reflejar la realidad que se pretende modelar, estando en concordancia con el modelo del especialista del área con el que se realiza la solución del problema en cuestión.

Dada una matriz de similaridad y un umbral de semejanza β_0, se puede construir un grafo no dirigido de β_0-similaridad $G_{\beta_0}=(M, A_{\beta_0})$, cuyos vértices son los elementos de M y las aristas son los elementos del conjunto $A_{\beta_0}=\{(O_i,O_j) / \Gamma(O_i,O_j)=\Gamma_{ij}\geq\beta_0\}$. Entonces, siguiendo la definición de criterio de agrupamiento, podemos definir un conjunto de proposiciones que generen una familia de subgrafos de G_{β_0} los cuales pueden ser interpretados como agrupamientos de M. A continuación enunciamos algunos ejemplos de criterios de agrupamiento siguiendo este enfoque.

Se dirá que $C\subseteq M$, $C\neq\emptyset$ es una *componente β_0-conexa* si y sólo si:

a) $\forall O_iO_j\in C\ \exists\ O_{i_1},\dots,O_{i_q}\in C\ [O_i=O_{i_1}\wedge O_j=O_{i_q}\wedge\forall p\in\{1,\dots,q\text{-}1\}\ \Gamma(O_{i_p},O_{i_{p+1}})\geq\beta_0]$,

b) $\forall O_i\in M[(O_j\in C,\ \Gamma(O_i,O_j)\geq\beta_0)\Rightarrow O_i\in C]$,

c) Todo elemento β_0-*aislado* es una *componente* β_0-conexa (*degenerada*).

La condición *a*) significa que para cualquier par de elementos de C existe una sucesión de elementos en C, que empieza en O_i y termina en O_j tales que cada uno es β_0-semejante al siguiente; *b*) significa que no existe fuera de C un elemento β_0-semejante a un elemento de C. También lo denominaremos *núcleo β_0-conexo*.

Se dirá que $B\subseteq M$, $B\neq\emptyset$es una *componente β_0-compacta* si y sólo si:

a)$\forall O_j\in M[O_i\in B\wedge(\max_{\substack{O_t\in M \\ O_t\neq O_i}}\{\Gamma(O_i,O_t)\}=\Gamma(O_i,O_j)\geq\beta_0 \vee \max_{\substack{O_t\in M \\ O_t\neq O_j}}\{\Gamma(O_j,O_t)\}=\Gamma(O_j,O_i)\geq\beta_0)]\Rightarrow O_j\in B$,

b) $\forall O_iO_j\in B\exists\ O_{i_1},\dots,O_{i_q}\in B[O_i=O_{i_1}\wedge O_j=O_{i_q}\wedge\forall p\in\{1,\dots,q\text{-}1\}$

$[\max_{\substack{O_t\in M \\ O_t\neq O_{i_p}}}\{\Gamma(O_{i_p},O_t)\}=\Gamma(O_{i_p},O_{i_{p+1}})\geq\beta_0\vee\max_{\substack{O_t\in M \\ O_t\neq O_{i_{p+1}}}}\{\Gamma(O_{i_{p+1}},O_t)\}=\Gamma(O_{i_p},O_{i_{p+1}})\geq\beta_0]]$,

c) Todo elemento β_0-aislado constituye una componente β_0-compacta (*degenerada*).

La condición *a*) dice que todo elemento de B tiene en B al elemento que más se le parece que es β_0-semejante con él y que no existe fuera de B un elemento cuyo elemento más parecido que sea β_0-semejante esté en B; la condición *b*) significa que para cualquier par de elementos O_i y O_j de B existe una sucesión de elementos en B, que empieza en O_i y termina en O_j tales que cada uno es el elemento que más se le parece y que es β_0-semejante al siguiente o que el siguiente es el que más se le parece y que es β_0-semejante al anterior. Esta condición garantiza además que el conjunto B es de cardinalidad mínima. También lo denominaremos *núcleo β_0-compacto*.

Se dice que B (una componente β_0-compacta) es una *componente β_0-homogénea* si y sólo si:

$$\forall O_i,O_j\in B\ \Gamma(O_i,O_j)\geq\beta_0.$$

Por definición todo elemento β_0-aislado es β_0-*homogéneo*.

Se dice que $F\subseteq M$, $F\neq\emptyset$ es una *componente β_0-fuertemente compacta* si y sólo si:

a) $\forall O_j \in M[O_i \in F \wedge \max\limits_{\substack{O \in M \\ O \neq O_i}} \{\Gamma(O_i,O)\} = \Gamma(O_i,O_j) \geq \beta_0] \Rightarrow O_j \in F$.

b) $\exists O_i \in F\ \forall O_j \in F \exists\, O_{i_1},\ldots,O_{i_q} \in F\ \ [O_i = O_{i_1} \wedge O_j = O_{i_q} \wedge \forall p<q[\max\limits_{\substack{O \in M \\ O \neq O_{i_p}}} \{\Gamma(O_{i_p},O)\} = \Gamma(O_{i_p},O_{i_{p+1}}) \geq \beta_0]]$.

c) No existe F' que cumpla a) y b) tal que $F \subset F'$.

d) Todo elemento β_0-aislado constituye una componente β_0- fuertemente compacta (*degenerada*).

La condición a) expresa que todo elemento de F tiene en F al elemento que más se le parece que es β_0-semejante con él. La condición b) significa que en F existe un elemento tal que para cualquier otro que pertenezca a F existe una sucesión de elementos de F tales que uno es el más parecido β_0-semejante al siguiente y por último la condición c) dice que F es el más grande que cumple las condiciones a) y b).

Un conjunto $S \subset M$, $S \neq \emptyset$ es un *conjunto β_0-completo maximal* si y sólo si:

a) $\forall O_i, O_j \in S\ \Gamma(O_i,O_j) \geq \beta_0$.

b) $\forall O_i \in M\ [(\forall O_j \in S\ \ \Gamma(O_i,O_j) \geq \beta_0) \Rightarrow O_i \in S]$.

c) Todo elemento β_0-aislado constituye un conjunto β_0-completo maximal (*degenerado*).

La condición a) dice que para cualquier par de elementos de S que se tomen, éstos deben ser β_0-semejantes y la condición b) dice que S debe ser el conjunto más grande que cumple la condición a).

No es difícil verificar que las estructuraciones de un universo en componentes β_0-fuertemente compactas y conjuntos β_0-completos maximales forman cubrimientos.

Dado un *grafo de β_0-similaridad* $G_{\beta_0}=(U, A_{\beta_0})$. Por $P(G_{\beta_0})$ se denotará el conjunto de todos los caminos en G_{β_0}. $p \in P(G_{\beta_0})$ entre los objetos O y O' se denotará por $p(O,O')$. Análogamente, un *grafo de β_0-máxima similaridad* $G_{\beta_0}^{\max}=(U, A_{\beta_0}^{\max})$, es un grafo dirigido cuyos nodos están en U y

$$A_{\beta_0}^{\max} = \{(O_i,O_j) / \max_{\substack{O_t \in U \\ O_t \neq O_i}} \{\Gamma(O_i,O_t)\} = \Gamma_{ij}^* \geq \beta_0\}.$$

Este grafo tiene asociado un grafo no dirigido, que se denotará $\overline{G_{\beta_0}^{\max}}$, el cual se obtiene de $G_{\beta_0}^{\max}$ desestimando la orientación de las aristas. Se denotará por $DP(G_{\beta_0}^{\max})$y por $P(\overline{G_{\beta_0}^{\max}})$ al conjunto de todos los caminos en $G_{\beta_0}^{\max}$ y en $\overline{G_{\beta_0}^{\max}}$, respectivamente. En los grafos dirigidos, se denotarán por $o(p)$ y $t(p)$ los *nodos inicial* y *terminal* de un camino p, respectivamente. Estos nodos pueden identificar un camino o un camino dirigido en los correspondientes grafos. Se asumirá que el conjunto de caminos en ambos casos es finito.

A partir de la formulación anterior podemos definir las siguientes relaciones sobre el universo de objetos U:

$R_1 = \{(O,O') \in U\text{x}U | O=O' \vee \exists p(O,O') \in P(G_{\beta_0})\}$,

$R_2 = \{(O,O') \in U\text{x}U | O=O' \vee \exists p(O,O') \in P(\overline{G_{\beta_0}^{\max}})\}$,

$R_3 = \{(O,O') \in U\text{x}U | O=O' \vee \exists p,p' \in DP(G_{\beta_0}^{\max})\ [(o(p)=O \wedge t(p)=O') \wedge (o(p')=O' \wedge t(p')=O)]\}$.

No es difícil verificar que todas son relaciones de equivalencia y por ende, cada una genera sobre U una partición. Esas clases de equivalencias de las respectivas relaciones antes descritas se denominan *componentes β_0-conexas, componentes β_0-compactas y componentes β_0-fuertemente conexas*.

Por comodidad en las notaciones se escribirá $(O_i,O_j) \in R_t = O_i R_t O_j$. Por $[O]_{R_t}$ se denotará la clase de equivalencia generada por la relación R_t, y U/R_t denotará al conjunto cociente de U por dicha relación, t=1,2,3.

Dichas clases de equivalencia son conjuntos maximales para la inclusión. Esto se expresa mediante la *condición de maximalidad* que se traduce a:

Dada una relación de equivalencia R sobre U y $M\subseteq U$ tal que $\forall O_i,O_j\in M(O_iRO_j)$, entonces existe una clase de equivalencia $C\in U/R$, inducida por R, tal que $M\subseteq C$.

La posibilidad de contar con más de un criterio de agrupamiento da lugar a la pregunta: ¿Qué relaciones conjuntuales existen entre estos criterios de agrupamientos?

3 Relaciones conjuntuales entre estructuraciones libres

Sea $P\subset M$, se define D_P y T_P como sigue:

$$D_P=\{O_i\in M \mid \max_{\substack{O_j\in M\\ O_j\neq O}}\{\Gamma(O,O_j)\}=\Gamma(O,O_i)\geq\beta_0;\ O\in P\},$$

$$T_P=\{O_i\in M \mid \max_{\substack{O_j\in M\\ O_j\neq O_i}}\{\Gamma(O_j,O_i)\}=\Gamma(O_i,O)\geq\beta_0;\ O\in P\}.$$

D_P representa al conjunto de los elementos de M que son los más parecidos y β_0-semejantes a los objetos que pertenecen a P y T_P contiene a aquellos elementos cuyos objetos más parecidos β_0-semejantes pertenecen a P.

Lema.- Sea C una componente β_0-conexa en M, $P\subset C$ entonces se tiene que $D_P\subset C$ y $T_P\subset C$.

Demostración.- La demostración es inmediata a partir de la definición de componente β_0-conexa ya que en D_P están los elementos que son los más parecidos y β_0-semejantes a $O\in P\subset C$; y sabemos que si un elemento es β_0-semejante a otro que está en la componente β_0-conexa, él también está en la componente β_0-conexa por lo que $D_P\subset C$. Por otro lado en T_P están los elementos que tienen como elementos más parecidos β_0-semejantes a $O\in P\subset C$ también por definición tenemos que estos elementos están en la componente β_0-conexa por lo que $T_P\subset C$. ■

Proposición.- En toda componente β_0-conexa existe al menos un subconjunto β_0-compacto.

Demostración.- Sea C una componente β_0-conexa, construyamos un conjunto β_0-compacto en C.

Si C es degenerada la demostración es inmediata. En otro caso sea $O_1\in C$, recursivamente se tiene lo siguiente; $P_0=\{O_1\}$ calcular D_{P_0} y T_{P_0} y formar $P_1=P_0\cup D_{P_0}\cup T_{P_0}$ ($P_1\subset C$ por el lema anterior). Puede suceder que $P_0\neq P_1$; en cuyo caso calcular D_{P_1} y T_{P_1} y formar $P_2=P_1\cup D_{P_1}\cup T_{P_1}$ ($P_2\subset C$ por el lema anterior). Nuevamente puede suceder $P_1\neq P_2$ y se debe calcular D_{P_2} y T_{P_2}. Como C es finita, en algún momento sucederá que $P_{s-1}=P_s$ y el conjunto así construido es un conjunto β_0-compacto. ■

Proposición.- Toda componente β_0-conexa es la unión finita de conjuntos β_0-compactos.

Demostración.- Sea C una componente β_0-conexa, por la proposición anterior podemos construir un conjunto β_0-compacto en C. Si C es degenerada la demostración es inmediata. En otro caso denótese como $B(O_i)$ al conjunto β_0-compacto generado a partir de O_i. Puede suceder que $B(O_i)=B(O_j)$ para $O_i\neq O_j$, lo que significa que O_i y O_j generan el mismo conjunto β_0-compacto además como $|C|<\infty$, entonces se puede afirmar que el número de conjuntos β_0-compactos generados a partir de cada elemento de C es finito. Por el modo que se construyeron los conjuntos β_0-compactos se concluye que $C=\bigcup_{O_i\in C} B(O_i)$, es decir, que toda componente β_0-conexa es la unión finita de conjuntos β_0-compactos. ■

Proposición.- Si B es un conjunto β_0-compacto. Entonces $\forall O_i \in B$ $B\backslash\{O_i\}$ no es un conjunto β_0-compacto.

Demostración.- Es evidente a partir de la condición *b*) de minimalidad de la definición de β_0-compacto. ■

Proposición.- Si B es un conjunto β_0-compacto en M, entonces B no necesariamente es una componente β_0-conexa de M.

Demostración.- Es inmediata la demostración ya que de la definición de β_0-compacidad lo que garantiza es la existencia en B dado O_i de un O_j tal que $\max_{\substack{O\in M\\ O\neq O_i}}\{\Gamma(O_i,O)\}=\Gamma(O_i,O_j)\geq\beta_0$ y no se garantiza que otros elementos, dígase O_q tal que $\Gamma(O_i,O_q)\geq\beta_0 \Rightarrow O_q\in B$. ■

Corolario.- Una condición necesaria y suficiente para que una componente β_0-conexa C sea un conjunto β_0-compacto B es que exista $O\in C$ tal que $B(O)=C$.

La demostración es inmediata. ■

Proposición.- En toda componente β_0-conexa existe al menos un conjunto β_0-fuertemente compacto.

Demostración.- Sea C una componente β_0-conexa, construyamos un conjunto fuertemente β_0-compacto en C.

Si C es degenerada la demostración es inmediata. En otro caso sea $O_1\in C$, tomemos $P_0=\{O_1\}$, calcular D_{P_0} y formar $P_1=P_0\cup D_{P_0}$ ($P_1\subset C$ por el lema anterior). Puede suceder que $P_0\neq P_1$, en este caso se calcula D_{P_1} y se forma $P_2=P_1\cup D_{P_1}$ ($P_2\subset C$ por el lema anterior). Como C es finito en algún momento se tendrá que $P_r=P_{r-1}$. Hágase $F_0=P_{r-1}$.

Ahora se calcula $Q_0=T_{P_0}$ y se forma $R_0=Q_0\backslash F_0$, puede suceder que $R_0\neq\varnothing$ entonces se toma $O_t\in R_0$. Hágase $P'_0=\{O_t\}$ y aplíquese el mismo proceso que para O_1 hasta tener $P'_r=P'_{r-1}$, sólo que ahora

$P'_i=P'_{i-1}\cup(D_{P'_{i-1}}\backslash F_0)$ ($P'_{r-1}\subset C$ por el lema anterior). Se forma $F_1=F_0\cup P'_{r-1}$ y se calcula $Q_1=T_{P'_0}$. Sea $R_1=Q_1\backslash F_1$. Nuevamente puede suceder que $R_1\neq\varnothing$, entonces se repite el mismo procedimiento. Como C es finito en algún momento se tendrá que $R_s=\varnothing$ y entonces el conjunto F_s es un conjunto β_0-fuertemente compacto ya que para cualquier elemento de él, éste tiene al (o los) elemento(s) más parecido(s) β_0-semejante(s) y además existe un elemento $O_i\in R_{s-1}$ (u $O_i\in P_0$ si $Q_0=\varnothing$) tal que:

$\forall O_j\in F_{s-1}\ \exists\, O_{i_1},\ldots,O_{i_q}\in F_{s-1}\ [\,O_{i_1}=O_i,\ O_{i_q}=O_j \wedge \forall p<q\ [\max_{\substack{O\in M\\ O\neq O_i}}\{\Gamma(O_{i_p},O)\}=\Gamma(O_{i_p},O_{i_{p+1}})\geq\beta_0]]$,

y por construcción F_{s-1} es maximal. ■

Proposición.- Toda componente β_0-conexa es la unión finita de conjuntos β_0-fuertemente compactos.

Demostración.- Sea C una componente β_0-conexa, por la proposición anterior se puede construir un conjunto fuertemente β_0-compacto en C. Si C es degenerada la demostración es inmediata. En otro caso se denota como $F(O_i)$ al conjunto β_0-fuertemente compacto generado a partir de O_i. Puede suceder que $F(O_i)=F(O_j)$ $O_i\neq O_j$ lo que significa que O_i y O_j generan el mismo conjunto β_0-fuertemente compacto. Además, como $|C|<\infty$, entonces se puede afirmar que el número de conjuntos β_0-fuertemente compactos generados a partir de cada elemento de C es finito, y por el modo en que se construyeron los conjuntos fuertemente β_0-compactos se puede concluir que $C=\bigcup_{O_i\in C}F(O_i)$, es decir, que toda componente β_0-conexa es la unión finita de conjuntos β_0-fuertemente compactos. ■

Lema.- Sea B un conjunto β_0-compacto, $P\subset B$, sean D_P y T_P como antes, entonces se tiene que $D_P\subset B$ y $T_P\subset B$.

Demostración.- Es inmediata a partir de la definición de conjunto β_0-compacto, ya que en D_P están los elementos que son los más parecidos β_0-semejantes a $O \in P \subset B$ y por definición, éstos están en B, por lo tanto $D_P \subset B$. Por otra parte, en T_P están los elementos que tienen como elementos más parecidos β_0-semejantes a los $O \in P \subset B$, también por definición se tiene que estos elementos están en B, por lo tanto $T_P \subset B$. ■

Proposición.- En todo conjunto β_0-compacto existe al menos un subconjunto β_0-fuertemente compacto.

Demostración.- Sea B un conjunto β_0-compacto. Se construirá un conjunto β_0-fuertemente compacto en B.

Si B es degenerado la demostración es inmediata. En otro caso sea $O_1 \in B$. Tómese $P_0=\{O_1\}$, calcúlese D_{P_0} y fórmese $P_1=P_0 \cup D_{P_0}$ ($P_1 \subset C$ por el lema anterior). Puede suceder que $P_0 \neq P_1$. En este caso se calcula D_{P_1} y se forma $P_2=P_1 \cup D_{P_1}$ ($P_2 \subset B$ por el lema anterior). Como B es finito en algún momento se tendrá que $P_r=P_{r\text{-}1}$. Hágase $F_0=P_{r\text{-}1}$. Ahora se calculará $Q_0=T_{P_0}$ y se formará $R_0=Q_0 \backslash F_0$. Puede suceder que $R_1 \neq \varnothing$, entonces se toma $O_t \in R_0$, se hace $P'_0=\{O_t\}$ y se aplica el mismo proceso que para O_1 hasta tener $P'_r=P'_{r\text{-}1}$, sólo que ahora $P'_i=P'_{i\text{-}1} \cup (D_{P'_{i\text{-}1}} \backslash F_0)$ ($P'_{r\text{-}1} \subset B$ por el lema anterior). Fórmese $F_1=F_0 \cup P'_{r\text{-}1}$ y calcúlese $Q_1=T_{P'_0}$. Se toma $R_1=Q_1 \backslash F_1$ y nuevamente puede suceder que $R_1 \neq \varnothing$, entonces se repite el mismo procedimiento. Como B es finito, en algún momento se tendrá que $R_s=\varnothing$ y entonces el conjunto F_s es un conjunto β_0-fuertemente compacto, ya que para cualquier elemento de él, éste tiene al (o los) elemento(s) más parecido(s) β_0-semejante(s) y además existe un elemento $O_i \in R_{s\text{-}1}$ (u $O_i \in P_0$, si $Q_0=\varnothing$) tal que:

$$\forall O_j \in F_{s\text{-}1}\ \exists\, O_{i_1},\ldots,O_{i_q} \in F_{s\text{-}1}\ [\,O_{i_1}=O_i,\ O_{i_q}=O_j \wedge \forall p<q\ [\max_{\substack{O \square M \\ O \neq O_i}} \{\Gamma(O_{i_p},O)\}=\Gamma(O_{i_p},O_{i_{p+1}}) \geq \beta_0]],$$

y por construcción $F_{s\text{-}1}$ es maximal. ■

Proposición.- Todo conjunto β_0-compacto es la unión finita de conjuntos β_0-fuertemente compactos.

Demostración.- Sea B un conjunto β_0-compacto. En la proposición anterior se da el proceso para construir un conjunto β_0-fuertemente compacto F en un conjunto β_0-compacto a partir de un elemento $O_i \in B$.

Si B es degenerado la demostración es inmediata. En otro caso, sea $F(O_i)$ el conjunto β_0-fuertemente compacto generado a partir de O_i. Se puede aplicar este proceso para cada elemento de B. Puede ocurrir que $F(O_i)=F(O_j)$ para $O_i \neq O_j$, lo que significa que O_i y O_j generan el mismo conjunto β_0-fuertemente compacto. Además, como $|C|<\infty$, entonces se puede afirmar que el número de conjuntos β_0-fuertemente compactos generados a partir de cada elemento de B es finito, y por el modo que se construyeron los conjuntos fuertemente β_0-compactos se tiene que $B=\bigcup_{O_i \in B} F(O_i)$, es decir, que un conjunto β_0-compacto es la unión finita de conjuntos β_0-fuertemente compactos. ■

Proposición.- En toda componente β_0-conexa existe al menos un conjunto β_0-completo maximal.

Demostración.- Sea C una componente β_0-conexa, construyamos un conjunto β_0-completo maximal en C.

Si C es degenerada la demostración es inmediata. En otro caso sea $O_1 \in C$, entonces por la definición de componente β_0-conexa existe $O_2 \in C$ $O_2 \neq O_1$ tal que $\Gamma(O_1,O_2) \geq \beta_0$. Pueden suceder dos casos:

1) $\{O_1,O_2\}$ es un conjunto β_0-completo maximal.

2) $\{O_1,O_2\}$ no es un conjunto β_0-completo maximal.

Para el primer caso, se ha encontrado un conjunto β_0-completo maximal. En el segundo caso, como el conjunto no lo es, entonces existe $O_3 \in C$, $O_3 \neq O_1$, $O_3 \neq O_2$ tal que $\Gamma(O_2,O_3) \geq \beta_0$ y $\Gamma(O_1,O_3) \geq \beta_0$. Ahora nuevamente pueden suceder dos casos:

1) $\{O_1,O_2,O_3\}$ es un conjunto β_0-completo maximal.

2) $\{O_1,O_2,O_3\}$ no es un conjunto β_0-completo maximal.

Como C es finita en algún momento se termina el proceso y se habrá encontrado un conjunto β_0-completo maximal. ■

Proposición.- Toda componente β_0-conexa es la unión finita de conjuntos β_0-completos maximales.

Demostración.- Sea C una componente β_0-conexa, en la proposición anterior se da el proceso para construir un conjunto β_0-completo maximal en una componente β_0-conexa a partir de un elemento $O_i \in C$.

Si C es degenerada la demostración es inmediata. En otro caso sea $S(O_1)$ el conjunto β_0-completo maximal generado por el procedimiento anterior a partir de O_1. El proceso anterior puede aplicarse a partir de cualquier $O \in C$, entonces se puede dar el caso de que $S(O_i)=S(O_j)$ para $O_i \neq O_j$, es decir, que dos objetos diferentes generen el mismo conjunto β_0-completo maximal y además como C es finito podemos afirmar que el número de conjuntos β_0-completos maximales es finito, y por construcción se tiene que $C=\bigcup_{O_i \in C} S(O_i)$, es decir, que toda componente β_0-conexa es la unión finita de conjuntos β_0-completos maximales. ■

Las proposiciones anteriores asumen que los conjuntos a estructurar son finitos, que en términos prácticos no es una restricción fuerte. Sin embargo, en [Martínez-Trinidad, et al. (2000)] se encuentran demostraciones de estos mismos resultados donde la hipótesis de la finitud de los conjuntos se obvia, por lo que la demostración es más general. La que a continuación se expone, no tiene la restricción antes mencionada.

Proposición.- Toda componente β_0-compacta se puede particionar en componentes β_0-fuertemente conexas.

Demostración.- Sea $O \in B$ y $\aleph = [O]_{R_2}$, es decir, la componente β_0-compacta generada por O. Se probará que para cualquier

$$O' \in [O]_{R_2},\ [O']_{R_3} \subseteq [O]_{R_2} \quad (*).$$

Probando esta relación, se ve fácilmente que en el conjunto $\{[O']_{R_3} \mid O' \in [O]_{R_2}\}$ existe un único subconjunto que satisface la definición de partición.

Es claro que $[O']_{R_3} \neq \varnothing$ y que existe un camino $p(O',O) \in P(\overline{G_{\beta_0}^{\max}})$. Por lo que se pueden dar dos casos: $[O']_{R_3} = \{O'\}$ ó $[O']_{R_3}$ tiene más de un elemento. En el primer caso la condición (∗) se cumple de inmediato. En el segundo caso, existe $O'' \in [O']_{R_3}$ tal que $O'' \neq O'$. Entonces, por la definición de la relación R_3, para dos objetos cualesquiera de $[O']_{R_3}$ existen caminos $p(O'',O')$, $p'(O',O'') \in D_P(G_{\beta_0}^{\max})$ que unen dichos objetos de modo tal que $o(p)=O''$, $t(p)=O'$ y $o(p')=O'$, $t(p')=O''$. No es difícil probar que también ambos caminos $p(O'',O')$, $p'(O',O'') \in P(\overline{G_{\beta_0}^{\max}})$. Por lo que existe un camino $p(O'',O) \in P(\overline{G_{\beta_0}^{\max}})$, lo cual implica que $O'' \in [O]_{R_2}$, luego se cumple la condición (∗). ■

Corolario.- Toda componente β_0-conexa C se puede particionar en componentes β_0-fuertemente conexas.

Puesto que la estructura de un universo está conformada por los agrupamientos encontrados al aplicar un criterio de agrupamiento, entonces se tienen diferentes maneras de estructurar a los objetos y obviamente cada criterio dará una estructuración diferente, pero si se hace uso de las proposiciones demostradas anteriormente se puede verificar que agrupamientos formados por un criterio Π resultan ser unión de agrupamientos formados por otro criterio Π' lo cual da la posibilidad de hallar una estructuración más fina del universo y establecer comparaciones del cumplimiento o no de propiedades diferentes entre los objetos. Así por ejemplo, las estructuraciones en componentes β_0-conexas, β_0-compactas y β_0-fuertemente conexas forman un refinamiento de particiones. Una propiedad que se cumple de manera laxa en una componente conexa se cumple con más exigencia en una componente compacta y más aún en una fuertemente conexa, lo que permite profundizar el análisis del universo de estudio.

De igual manera, por ejemplo, una estructuración en componentes β_0-conexas pudiera refinarse mediante cubrimientos con una estructuración en componentes β_0-fuertemente compactas. En este caso el refinamiento no se hace con particiones pero en ciertas aplicaciones es lo que corresponde pues las propiedades que se estudian son compartidas en diferentes agrupamientos.

Estos últimos resultados [Martínez-Trinidad, et al. (2000); Martínez-Trinidad and Guzmán-Arenas (2001); Martínez-Trinidad (1995); Montellano-Ballesteros (1994); Ruiz-Shulcloper and Montellano-Ballesteros (1995)] como decíamos, tienen importancia práctica ya que permiten realizar diferentes agrupamientos de objetos conociendo el nivel de exigencia con el cual se forman dichos grupos. Esos resultados, sometidos a análisis, por ejemplo, en el contexto de un modelo geólogo-geofísico con vistas a la determinación de perspectividad de cualquier materia prima mineral, arroja información valiosa para los objetivos mencionados [Gómez-Herrera, et al. (1994)].

4 Algoritmos para la obtención de estructuraciones libres

En la literatura existen varios algoritmos que permiten la determinación de todas las componentes β_0-conexas, β_0-compactas y β_0-fuertemente conexas. Sin embargo, en los algoritmos que a continuación se exponen hay un elemento novedoso: las características peculiares de los grafos de β_0-similaridad y máxima β_0-similaridad.

A continuación se exponen resultados que permiten caracterizar las componentes β_0-fuertemente conexas de un grafo de máxima β_0-similaridad, $G_{\beta_0}^{max}$.

Para hacer más claro el resultado se expondrá un ejemplo sencillo. Sea $O_1,O_2,O_3\in M$ objetos que forman un ciclo en el grafo de máxima similaridad asociado a M. Sin pérdida de generalidad se puede asumir que ese ciclo es O_1,O_2,O_3,O_1. Esto significa que $\Gamma(O_1,O_3)\leq\Gamma(O_1,O_2)$; $\Gamma(O_2,O_1)\leq\Gamma(O_2,O_3)$; $\Gamma(O_3,O_2)\leq\Gamma(O_3,O_1)$; pero si Γ es una función simétrica (lo que aquí se asume), entonces $\Gamma(O_1,O_2)\leq\Gamma(O_2,O_3)\leq\Gamma(O_3,O_1)$; y además $\Gamma(O_1,O_3)\leq\Gamma(O_3,O_2)\leq\Gamma(O_2,O_1)$. Esto significa que O_1,O_3,O_2,O_1 es también un ciclo, y por ello se cumple que $\Gamma(O_1,O_2)=\Gamma(O_2,O_3)=\Gamma(O_3,O_1)$.

Proposición.- Para cualquier ciclo $O_{i_1},\ldots,O_{i_q}$ en $G_{\beta_0}^{\max}$, $O_{i_q},O_{i_{q-1}},\ldots,O_{i_1}$ es también un ciclo en $G_{\beta_0}^{max}$.

Demostración.- Por ser $O_{i_1},\ldots,O_{i_q}$ un ciclo en $G_{\beta_0}^{max}$, se tiene que:

$$\forall p=1,\ldots,q\text{-}1 \text{ las aristas } (O_{i_p},O_{i_{p+1}}),\ (O_{i_q},O_{i_1})\in A_{\beta_0}^{max},$$

lo que implica que:

$\Gamma(O_{i_1},O_{i_2})\leq\Gamma(O_{i_2},O_{i_3})\leq\ldots\leq\Gamma(O_{i_q},O_{i_1})$, y en particular $\Gamma(O_{i_1},O_{i_2})\leq\Gamma(O_{i_q},O_{i_1})$.

Como $(O_{i_1},O_{i_2})\in A_{\beta_0}^{max}$ se tiene que $\max\limits_{\substack{O_t\in M\\ O_t\neq O_1}}\{\Gamma(O_{i_t},O_{i_1})\}=\Gamma(O_{i_1},O_{i_2})$ por lo que se concluye que:

$\Gamma(O_{i_1}, O_{i_2}) \geq \Gamma(O_{i_q}, O_{i_1})$ y por tanto $\Gamma(O_{i_1}, O_{i_2}) = \Gamma(O_{i_q}, O_{i_1})$.

Consecuentemente, $\Gamma(O_{i_1}, O_{i_2}) = \Gamma(O_{i_2}, O_{i_3}) = \ldots = \Gamma(O_{i_q}, O_{i_1})$. Además,

$$\forall p=1,\ldots,q\text{-}1 \ \max_{\substack{O_t \in M \\ O_t \neq O_p}} \{\Gamma(O_{i_p}, O_{i_t})\} = \Gamma(O_{i_p}, O_{i_{p+1}}) = \Gamma(O_{i_{p-1}}, O_{i_p}),$$

por lo que la arista $(O_{i_{p-1}}, O_{i_p}) \in A_{\beta_0}^{max}$. ■

Corolario.- Sea N una componente β_0-fuertemente conexa en $G_{\beta_0}^{max}$ y O, $O' \in N$ tal que $(O,O') \in A_{\beta_0}^{max}$, entonces $(O',O) \in A_{\beta_0}^{max}$.

Demostración.- Como N es una componente β_0-fuertemente conexa, entonces existen caminos $p,p' \in DP(G_{\beta_0}^{max})$ tales que $o(p)=t(p')=O$ y $o(p')=t(p)=O'$. Así, se obtiene un ciclo en O y también en O'. en virtud de la proposición anterior, para todas las aristas de este ciclo la arista con orientación inversa pertenece a $A_{\beta_0}^{max}$, en particular (O',O). ■

Este último resultado asegura que si N es una componente β_0-fuertemente conexa de $G_{\beta_0}^{max}$, para todos los O, $O' \in N$ tales que $(O,O') \in A_{\beta_0}^{max}$, entonces $\Gamma(O,O')$ toma los mismos valores. A este valor común se le denomina *nivel de similaridad* de la componente β_0-fuertemente conexa.

Apoyándose en los resultados anteriores se pueden determinar todas las componentes β_0-conexas, β_0-compactas y β_0-fuertemente conexas del siguiente modo:

Se denotará al conjunto de todos los objetos β_0-similares a O por $D_S(O)$.

Se denominará *predecesores directos* de O en M a los elementos del conjunto:

$$DPred_M(O) = \{O' \in M \mid O' \neq O \wedge \max_{\substack{O_t \in M \\ O_t \neq O}} \{\Gamma(O_t, O)\} = \Gamma(O', O)\}.$$

Se denominará *predecesores* de O en M a los elementos del conjunto:

$$Pred_M(O) = \{O' \in M \mid \exists p \in DP(G_{\beta_0}^{max})[o(p)=O' \wedge t(p)=O]\}.$$

El conjunto $DPred_M(O)$ puede ser calculado buscando en la fila del objeto O en la matriz de similaridad el valor máximo y tomando todos los objetos que la similaridad con O coincida con ese valor.

En el caso del conjunto $Pred_M(O)$ la idea es formar los caminos directos p tales que $o(p)=O$.

Análogamente, se denominará *sucesores directos* de O en M a los elementos del conjunto:

$$DSuc_M(O) = \{O' \in M \mid O' \neq O \wedge \max_{\substack{O_t \in M \\ O_t \neq O}} \{\Gamma(O, O_t)\} = \Gamma(O, O')\}.$$

Se denominará *sucesores* de O en M a los elementos del conjunto:

$$Suc_M(O) = \{O' \in M \mid \exists p \in DP(G_{\beta_0}^{max})[o(p)=O \wedge t(p)=O']\}.$$

4.1 Algoritmo para el cálculo de todas las componentes β_0-conexas

Para el caso de las componentes β_0-conexas el procedimiento es el mismo que se sigue en otros algoritmos diseñados con este propósito. La idea es construir el grafo de β_0-similaridad usando para ello la matriz de función de similaridad MS, y a partir de ese grafo, para cada objeto $O \in M$, calcular la clase de equivalencia generada por la relación R_1, es decir, calcular la clase de equivalencia por dicha relación de cada objeto. Lo que significa, para cada objeto $O \in M$, calcular todos los objetos que son β_0-similares con O y repetir ese proceso con los nuevos elementos incorporados hasta que no exista en M elemento alguno que sea similar a uno del conjunto que se está construyendo, esto es, la clase de equivalencia del objeto O. Este proceso se repite con los otros elementos de M que no están en el conjunto construido.

4.2 Algoritmo para el cálculo de todas las componentes β_0-compactas

Para el caso de las componentes β_0-compactas, lo primero que se hace es construir el grafo de máxima β_0-similaridad usando para ello la función de similaridad Γ y el umbral de similaridad β_0; a partir de ese grafo, para cada objeto $O \in M$, calcular la clase de equivalencia generada por la relación R_2, es decir, calcular la clase de equivalencia por dicha relación de cada objeto (la componente β_0-compacta del objeto O).

Find $M/R_2(M,\Gamma,\beta_0)$

Entradas: M (Matriz de datos), Γ (función de similaridad), β_0 (umbral de similaridad)

Salida: M/R_2=El conjunto de todas las componentes β_0-compactas de M.

begin
    1.- $Visited=\emptyset$;
    2.- $M/R_2=\emptyset$;
    3.- for each $O \in M$ do
        4.- if $O \notin Visited$ then
            begin
                5.- $Visited=Visited \cup \{O\}$
                6.- $M/R_2=M/R_2 \cup \{Find[O]_{R_2}(M,O,Visited,\Gamma,\beta_0)\}$;
            end
    end
    7.- $return=M/R_2$;
end.

$Find[O]_{R_2}(M,O,Visited,\Gamma,\beta_0)$

Entradas: M (Matriz de datos), O (objeto semilla), *Visited* (objetos ya visitados), Γ (función de similaridad), β_0 (umbral de similaridad)

Salida: $[O]_{R_2}$ =El conjunto de todas la componentes β_0-compacta de M construida a partir de O.

begin
    1.- $DPred_M(O)=\{O\}$;
    2.- for each $O' \in M$ do
        3.- if $\Gamma(O,O')=\max\limits_{\substack{O_t \in M \\ O_t \neq O}}\{\Gamma(O,O_t)\} \geq \beta_0$ then
            4.- $DPred_M(O)=DPred_M(O) \cup \{O'\}$;
    end
    5.- $DSuc_M(O)=\{O\}$;
    6.- for each $O' \in M$ do

7.- if $\Gamma(O,O')=\max_{\substack{O_t\in M\\ O_t\neq O'}}\{\Gamma(O',O_t)\}\geq\beta_0$ then

8.- $DSuc_M(O)=DSuc_M(O)\cup\{O'\}$;

end

9.- $Visited=Visited\cup\{O\}$;

10.- $[O]_{R_2}=DPred_M(O)\cup DSuc_M(O)$

11.- if $DPred_M(O)\cup DSuc_M(O)\neq\emptyset$ then

12.- for each $O'\in Dpred_M(O)\cup Dsuc_M(O)$ do

13.- if $O'\notin Visited$ then

begin

14.- $Visited=Visited\cup\{O'\}$

15.- $[O]_{R_2}=[O]_{R_2}\cup\{Find[O]_{R_2}(M,O',Visited,\Gamma,\beta_0)\}$;

end

end

16.- $return=[O]_{R_2}$;

end

4.3 Algoritmo para el cálculo de todas las componentes β_0-fuertemente conexas

Para el caso de las componentes β_0-fuertemente conexas se procede como en el caso anterior: se construye el grafo de máxima β_0-similaridad usando para ello la de similaridad Γ y el umbral de similaridad β_0; a partir de ese grafo, para cada objeto $O\in M$, calcular la clase de equivalencia generada por la relación R_3, es decir, calcular la clase de equivalencia por dicha relación de cada objeto (la componente β_0-fuertemente conexa del objeto O).

Find $M/R_3(M\ \Gamma,\beta_0)$

Entradas: M (Matriz de datos), Γ (función de similaridad), β_0 (umbral de similaridad)

Salida: M/R_3=El conjunto de todas las componentes β_0-fuertemente conexas de M.

begin

1.- $Visited=\emptyset$;

2.- $M/R_3=\emptyset$;

3.- for each $O\in M$ do

4.- if $O\notin Visited$ then

begin

5.- $Visited=Visited\cup\{O\}$;

6.- $M/R_3=M/R_3\cup\{Find[O]_{R_3}(M,O,MS)\}$;

end

end

7.- $return=M/R_3$;

end.

$Find[O]_{R_3}(M,O,Visited,\Gamma,\beta_0)$

Entradas: M (Matriz de datos), O (objeto semilla), *Visited* (objetos ya visitados), Γ (función de similaridad), β_0 (umbral de similaridad)

Salida: $[O]_{R_3}$ =El conjunto de todas la componentes β_0-fuertemente conexas de M construida a partir de O.

begin

1.- $DPred_M(O)=\{O\}$;

2.- for each $O' \in M$ do

3.- if $\Gamma(O,O') = \max\limits_{\substack{O_t \in M \\ O_t \neq O}} \{\Gamma(O,O_t)\} \geq \beta_0$ then

4.- $DPred_M(O)=DPred_M(O) \cup \{O'\}$;

end

5.- $DSuc_M(O)=\{O\}$;

6.- for each $O' \in M$ do

7.- if $\Gamma(O,O') = \max\limits_{\substack{O_t \in M \\ O_t \neq O'}} \{\Gamma(O',O_t)\} \geq \beta_0$ then

8.- $DSuc_M(O)=DSuc_M(O) \cup \{O'\}$;

end

9.- $Visited=Visited \cup \{O\}$;

10.- if $DPred_M(O) \cap DSuc_M(O) \neq \emptyset$ then

11.- for each $O' \in DPred_M(O) \cap DSuc_M(O)$ do

12.- if $O' \notin Visited$ then

begin

13.- $Visited=Visited \cup \{O'\}$;

14.- $[O]_{R_3} = (DPred_M(O) \cap DSuc_M(O)) \cup Find[O]_{R_3}(M,O',MS)$;

end

end

15.-*return*=$[O]_{R_3}$;

end

5 Métodos de sensibilidad en el cálculo de estructuraciones libres

En problemas de clasificación no supervisada, la adición o eliminación de objetos, rasgos o valores de algún rasgo en algún objeto es frecuente en la práctica. Estos cambios obedecen a diferentes factores, como por ejemplo: errores en la recopilación de los datos; aparición o desaparición de algunos objetos de estudio, nuevos criterios para la evaluación de los objetos de estudio, etc. Estas modificaciones en el espacio de representación, pueden inducir modificaciones en la forma en que se agrupan los objetos. Por lo que la solución ante tales situaciones es recalcular los agrupamientos a partir del nuevo conjunto de objetos. Esto claramente tiene un costo computacional, que en algunos casos puede ser sustancial.

Sean dados: un algoritmo de reconocimiento de patrones que depende de los parámetros $p_1,\ldots,p_s$, lo cual se denotará $A(\{p_1,\ldots,p_s\})$, donde el conjunto de parámetros puede ser vacío; un conjunto de datos M sobre los que el algoritmo trabajará y $\mathfrak{R}$ el resultado que se obtiene de la aplicación de $A(\{p_1,\ldots,p_s\})$ sobre M. Es decir, $A(\{p_1,\ldots,p_s\})(M)=\mathfrak{R}$.

Por Sensibilidad de $\mathfrak{R}$ se entenderá las alteraciones que se producen en dicho conjunto, bien a partir de modificaciones de M, manteniendo inalterable $A(\{p_1,\ldots,p_s\})$ o bien manteniendo fijo M y variando algunos de los parámetros del algoritmo A. Aquí sólo se considerará el primer caso.

Aquí se denomina *Análisis de Sensibilidad* [Carrasco-Ochoa and Ruiz-Shulcloper (1999); Carrasco-Ochoa, et al. (2001)] al problema de estudiar la sensibilidad de $\Re$, con el fin de encontrar mecanismos que permitan obtener la nueva solución con un costo computacional menor que el de la aplicación del algoritmo original, con las nuevas condiciones.

El análisis de la sensibilidad para las componente β_0-conexas se resolverá utilizando árboles de búsqueda en profundidad [Aho, et al. (1988)]. Éstos son utilizados para definir un orden de recorrido en un grafo, en particular en grafos no dirigidos.

Como antes, consideraremos un conjunto de n objetos $M=\{O_0,...,O_{n-1}\}\subseteq U$, siendo U un universo de objetos, y una función de semejanza Γ, que permite comparar objetos en U. Además, dado un $\beta_0\in L$ consideraremos el grafo $G_{\beta_0}=(M, A_{\beta_0})$ de β_0-similaridad en M.

El orden de recorrido de búsqueda en profundidad (*ordenBP*) en G_{β_0}, iniciando en O_0, se obtiene con los siguientes algoritmos iniciando con *actual*=0:

Orden_bp

begin

1.- for each $O_i\in M$ $(i=0,..,n-1)$ do *marcar* O_i *como no visitado*;

2.- for each $O_i\in M$ $(i=0,..,n-1)$ do if O_i *no ha sido visitado* then *bpf*(i);

end

bpf(x)

Entrada: x el objeto que se está analizando

begin

1.- *marcar* O_x *como visitado*;

2.- *ordenBP*[x] = *actual*;

3.- *actual* = *actual*+1;

4.- for each $O_i\in M$ $(i=0,..,n-1)$ do if $(O_x,O_i)\in A_{\beta_0}$ y O_i *no ha sido visitado* then *bpf*(i)

end.

El árbol se obtiene marcando como aristas de árbol aquellas que llegan a vértices no visitados y como aristas de retroceso aquellas que van a vértices ya visitados. En la figura 1, en (a) se muestra un grafo y en la parte (b), sus árboles de búsqueda en profundidad. Las aristas de retroceso están marcadas con líneas punteadas.

Cada nodo tiene asociado, aparte del número correspondiente a su nombre, un índice indicando el orden de recorrido, y un valor denominado *bajo*, el cual, para las hojas del árbol (nodos sin hijos), se obtiene como el mínimo valor entre el índice de recorrido del nodo y los índices de recorrido de los nodos a los cuales se puede llegar con sus aristas de retroceso. Para el resto de los nodos, se obtiene como el mínimo valor entre el índice de recorrido del nodo, los índices de recorrido de los nodos a los cuales se puede llegar con sus aristas de retroceso y los valores bajo de sus hijos.

Como se puede apreciar en la Figura 1, dado un grafo, el recorrido en profundidad generará un árbol independiente por cada componente conexa.

Se dice que un vértice es un *punto de articulación* [Grimaldi (1989)], si al eliminarlo junto con todas las aristas que inciden en él, la componente conexa a la que pertenece se divide.

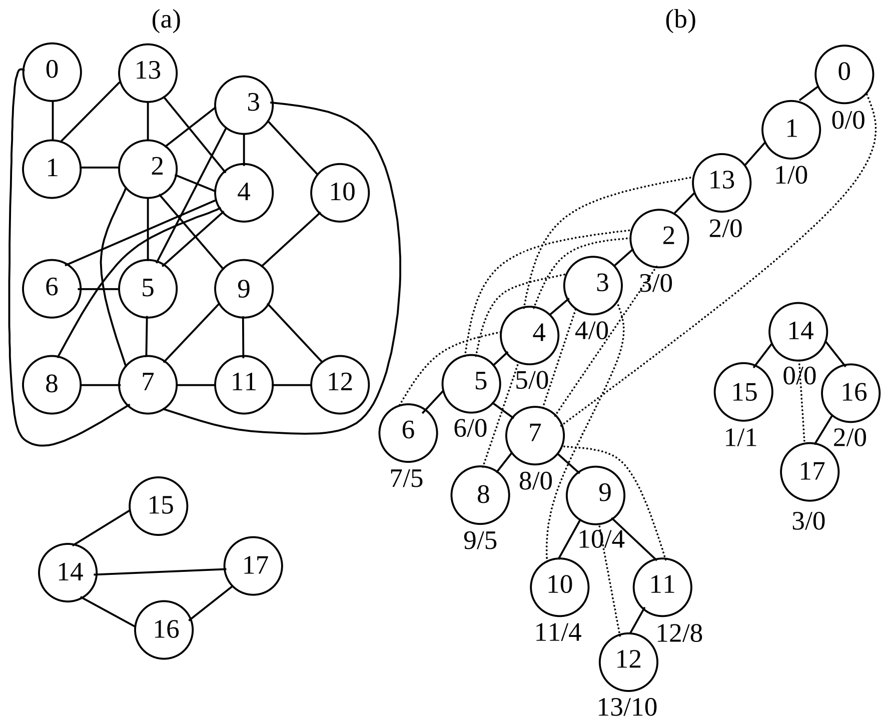

Fig. 1. [Carrasco-Ochoa and Ruiz-Shulcloper (1999)] Un grafo con dos componentes conexas y uno de los posibles árboles de búsqueda en profundidad asociados.

Cuando un nodo es eliminado del grafo, para determinar si es un punto de articulación, o no, se analizan los hijos del nodo eliminado y si el valor bajo de un hijo es mayor o igual al índice de recorrido del nodo eliminado, entonces este hijo junto con todos sus descendientes forman una nueva componente conexa. Por el contrario, si el valor bajo de un hijo es menor que el índice de recorrido del nodo eliminado, este hijo, junto con todos sus descendientes, permanece unido a la componente conexa original.

En [Grimaldi (1989); Carrasco-Ochoa (2001)] se dice que un conjunto de aristas de un grafo es llamado *conjunto de corte*, si al eliminar dichas aristas, sin eliminar ningún nodo, se incrementa el número de componentes conexas, y esto no se logra con ningún subconjunto propio no vacío del mismo.

Sea $G=(V,A)$ un grafo no dirigido y AG uno de los posibles conjuntos de árboles de búsqueda en profundidad asociado a G; sean $x_1,x_2 \in V$ y $(x_1,x_2) \in A$.

Proposición.- El conjunto de aristas $\{(x_1,x_2)\}$ es un conjunto de corte si y sólo si no existe un conjunto de vértices $\{v_1,v_2,\ldots,v_n\} \subset V$, con $n \geq 1$ y $\forall i=1,\ldots,n$ $x_1 \neq v_i \neq x_2$, tal que:

$$(x_1,v_1), (v_1,v_2),\ldots,(v_n,x_2) \in A \quad (*).$$

Demostración.- Primero, si $\{(x_1,x_2)\}$ es un conjunto de corte, entonces si se elimina la arista (x_1,x_2), la componente se divide y los nodos x_1 y x_2 quedan en diferentes componentes. Por lo que no hay un camino entre ellos y como no se ha eliminado ninguna otra arista. El único camino en el grafo original era el camino directo dado por la arista (x_1,x_2) y por lo tanto no existe un conjunto de vértices $\{v_1,v_2,\ldots,v_n\} \subset V$, que cumpla $(*)$.

Por otro lado, si no existe un conjunto de vértices que cumpla la condición $(*)$ entonces, el único camino entre x_1 y x_2 en G es el camino directo dado por la arista $\{(x_1,x_2)\}$. Además, $\{(x_1,x_2)\}$ no tiene subconjunto propios no vacíos, por lo cual $\{(x_1,x_2)\}$ es un conjunto de corte. Luego se concluye la validez de la proposición. ■

Proposición.- Si el conjunto $\{(x_1,x_2)\}$ es un conjunto de corte, entonces (x_1,x_2) no es una arista de retroceso en *AG*.

Demostración.- Supongamos que $\{(x_1,x_2)\}$ es un conjunto de corte y (x_1,x_2) es una arista de retroceso en *AG*, entonces sucede una de las siguientes cosas:

1) x_1 es descendiente de x_2 en *AG*, y entonces la rama que conduce de x_2 a x_1 contiene a los vértices $\{x_1,v_1,v_2,\ldots,v_n,x_2\}$ y a las aristas (x_1,v_1), (v_1,v_2),….,(v_n,x_2); y como x_2 no es hijo de x_1, puesto que (x_1,x_2) es una arista de retroceso en *AG*, entonces $n\geq 1$, con esto, por la proposición anterior se tiene que $\{(x_1,x_2)\}$ no es un conjunto de corte, lo cual es una contradicción. (1)
2) x_2 es descendiente de x_1 en *AG*, entonces la rama que conduce de x_1 a x_2 contiene a los vértices $\{x_2,v_1,v_2,\ldots,v_n,x_1\}$ y a las aristas (x_2,v_1), (v_1,v_2),….,(v_n,x_1); y como x_1 no es hijo de x_2, puesto que (x_1,x_2) es una arista de retroceso en *AG*, entonces n≥1, con esto, por la proposición anterior se tiene que $\{(x_1,x_2)\}$ no es un conjunto de corte, lo cual es una contradicción. (2)

De (1) y (2) se concluye que (x_1,x_2) no es una arista de retroceso en *AG*. ■

Proposición.- El conjunto $\{(x_1,x_2)\}$ es un conjunto de corte si y sólo si se cumple alguna de las siguientes condiciones:

c1) x_1 es hijo de x_2 en *AG* y el valor bajo de x_1 es mayor o igual al índice de recorrido de x_2.

c2) x_2 es hijo de x_1 en *AG* y el valor bajo de x_2 es mayor o igual al índice de recorrido de x_1.

Demostración.- Primero, si el conjunto $\{(x_1,x_2)\}$ es un conjunto de corte entonces, por la proposición anterior, (x_1,x_2) no es una arista de retroceso en *AG*, por lo tanto se cumple alguna de las siguientes dos condiciones:

a) x_1 es hijo de x_2 en *AG*, y como $\{(x_1,x_2)\}$ es un conjunto de corte entonces, si se elimina la arista (x_1,x_2), x_2 y todos sus descendientes se separan de la componente conexa, lo cual implica que no hay aristas de retroceso desde x_2 o cualquiera de sus descendientes, hacia ninguno de los ascendientes de x_2. Lo cual significa que el valor bajo de x_1 es mayor o igual al índice de recorrido de x_2. (1)
b) x_2 es hijo de x_1 en *AG*, y como $\{(x_1,x_2)\}$ es un conjunto de corte entonces, si se elimina la arista (x_1,x_2), x_1 y todos sus descendientes se separan de la componente conexa. Lo cual implica que no hay aristas de retroceso desde x_1 o cualquiera de sus descendientes, hacia ninguno de los ascendientes de x_1. Lo cual significa que el valor bajo de x_2 es mayor o igual al índice de recorrido de x_1. (2)

De (1) y (2) se tiene que si el conjunto $\{(x_1,x_2)\}$ es un conjunto de corte entonces se cumple alguna de las condiciones c1 ó c2. (3)

Por otra parte, si se cumple alguna de las condiciones c1 ó c2, se tiene que (analizando por separado el caso de que se cumpla cada una):

c1) si x_1 es hijo de x_2 en *AG*, y valor bajo de x_1 es mayor o igual al índice de recorrido de x_2, entonces no hay aristas de retroceso desde x_2 o cualquiera de sus descendientes, hacia ninguno de los ascendientes de x_2. Lo cual implica que la única arista que une al subárbol formado por x_2 y sus descendientes es (x_1,x_2) y por lo tanto, si se eliminara, x_2 y todos sus descendientes se separarían de la componente, es decir, $\{(x_1,x_2)\}$ es un conjunto de corte. (4)

c2) si x_2 es hijo de x_1 en *AG*, y valor bajo de x_2 es mayor o igual al índice de recorrido de x_1 entonces, no hay aristas de retroceso desde x_1 o cualquiera de sus descendientes, hacia ninguno de los ascendientes de x_1, lo cual implica que la única arista que une al subárbol formado por x_1 y sus descendientes es (x_1,x_2) y por lo tanto, si se eliminara, x_1 y todos sus descendientes se separarían de la componente, es decir, $\{(x_1,x_2)\}$ es un conjunto de corte. (5)

De (4) y (5) se tiene que si se cumple c1 ó c2 entonces $\{(x_1,x_2)\}$ es un conjunto de corte. (6)

Finalmente de (3) y (6) se concluye la validez de la proposición. ■

Todas estas proposiciones, en particular la última, se han formulado con el fin de tener un criterio simple para decidir si una arista cualquiera forma por si sola un conjunto de corte, lo cual es necesario para el método que se propondrá a continuación.

5.1 Sensibilidad en el cálculo de las componentes β_0-conexas

Si se elimina un objeto O que está en la componente β_0-conexa N_j, solo es necesario recalcular las componentes β_0-conexas con los objetos restantes de N_j. Cuando se inserta un nuevo objeto O, se realiza la unión de las componentes β_0-conexas a las que pertenezcan aquellos objetos que son β_0-semejantes con O.

Esta estrategia, aunque parece natural, es eficiente solamente para el caso de agregar objetos, porque en el caso de eliminación su complejidad es el cuadrado del número de elementos de la componente β_0-conexa considerada, y si este número es una fracción considerable de n, el número de objetos originalmente agrupados, la complejidad sigue siendo de orden n^2. Por esto es necesario buscar un método alternativo que sea más eficiente.

La idea del algoritmo es encontrar procedimientos para reorganizar los árboles de búsqueda en profundidad cuando se eliminan o agregan objetos (nodos). Los procedimientos propuestos son los siguientes:

Cuando un nodo es eliminado, si dicho nodo constituía una componente β_0-conexa degenerada, entonces simplemente se elimina dicha componente. Si el nodo eliminado es un nodo hoja en AG, solamente se elimina dicho nodo junto con todas sus aristas. En otro caso se procede de la siguiente manera:

If el nodo eliminado es un nodo raíz
  then
    Cada uno de sus hijos con sus descendientes forman un nuevo árbol de búsqueda en profundidad (componente β_0-conexa); se ajustan los índices de recorrido y los valores *bajo*
  else
    for each hijo del nodo eliminado,
      if valor *bajo* del hijo ≥ índice de recorrido del nodo eliminado
        then (no hay conexión entre este hijo o sus descendientes con los ascendientes del nodo eliminado)
          Este hijo junto con todos sus descendientes forman un nuevo árbol de búsqueda en profundidad (componente β_0-conexa); se ajustan los índices de recorrido y los valores *bajo*
        else (si hay conexión entre este hijo o sus descendientes con los ascendientes del nodo eliminado)
          Aplicar el procedimiento de reorganización a partir de este hijo.
    Ajustar los índices de recorrido y los valores *bajo* del árbol original.

El procedimiento de reorganización a partir de un nodo t es el siguiente:

1.- Buscar la arista de retroceso, desde t o de alguno de sus descendientes, cuyo par de nodos (α,β) cumplan las siguientes condiciones:

α=*max*{índices de recorrido de los ascendientes del padre de t posteriores a t' que pueden ser alcanzados por t o alguno de sus descendientes} (este conjunto no es vacío pues este procedimiento sólo se aplica cuando hay conexión entre t o sus descendientes con los ascendientes del padre de t)

β=*min*{índices de recorrido de t y de sus descendientes que alcanzan al nodo α}

2.- Al padre de β se le quita β como hijo

3.- β se convierte en hijo de α

4.- Se reconstruye el árbol de búsqueda en profundidad con los nodos del subárbol con raíz en t, pero utilizando β como raíz

- El resultado es un árbol de búsqueda en profundidad pues no hay conexiones entre ningún nodo de este subárbol y los descendientes de α, pues α es el nodo más lejano de la raíz que puede ser alcanzado por aristas de retroceso de t o sus descendientes)
- Este subárbol se reconstruye a partir de β para reducir el número de ajustes necesarios, pues β es el nodo más cercano a t que puede retroceder hacia α

Cuando un nodo O es agregado se procede de la siguiente manera:

1.- *for each* α β_0-semejante con O
  2.- *if* componente β_0-conexa a la que pertenece α ya fue agregada
    then
      3.- Agregar un retroceso de α a O
    else
      4.- Marcar la componente β_0-conexa a la que pertenece α como ya agregada
      5.- *if* α no es un nodo raíz
        then
          6.- Reconstruir el árbol de búsqueda en profundidad de la componente β_0-conexa a la que pertenece α, pero utilizando α como raíz
      7.- Agregar α como hijo de O
8.- Ajustar los índices de recorrido y los valores bajo del nuevo núcleo conexo.

5.2 Sensibilidad en el cálculo de las componentes β_0-compactas

Considere $\overline{G_{\beta_0}^{\max}}$ =(M,A) el grafo de β_0-*máxima similaridad* sobre M (ver sección 2).

Proposición.- El conjunto de componentes β_0-compactas de M coincide con el conjunto de componentes β_0-conexas del grafo $\overline{G_{\beta_0}^{\max}}$.

Demostración.- Primero, sea C una componente β_0-compacta de M, entonces cumple las siguientes dos condiciones:

a) $\forall O_j \in M\ [O_i \in C \wedge (\max_{\substack{O_t \in M \\ O_t \neq O_i}} \{\Gamma(O_i,O_t)\} = \Gamma(O_i,O_j) \geq \beta_0 \ \square \ \max_{\substack{O_t \in M \\ O_t \neq O_j}} \{\Gamma(O_j,O_t)\} = \Gamma(O_j,O_i) \geq \beta_0)] \Rightarrow O_j \in C,$

pero por definición del grafo $\overline{G_{\beta_0}^{\max}}$, esto sólo se cumple si y sólo si:

$\forall O_j \in M\ [O_i \in C \wedge (O_i,O_j) \in A] \Rightarrow O_j \in C.$ (1)

b) $\forall O_i, O_j \in C\ \exists\ O_{i_1}, \ldots, O_{i_q} \in C$

$$[O_i = \mathrm{O}_{\mathrm{i}_1} \wedge O_j = O_{i_q} \wedge \forall p=1,\ldots,q\text{-}1\ (\max_{\substack{O_t \square M \\ O_t \neq O_{i_p}}} \{\Gamma(O_{i_p}, O_t)\} = \Gamma(O_{i_p}, O_{i_{p+1}}) \geq \beta_0 \square$$

$$\max_{\substack{O_t \square M \\ O_t \neq O_{i_{p+1}}}} \{\Gamma(O_{i_{p+1}}, O_t)\} = \Gamma(O_{i_{p+1}}, O_{i_p}) \geq \beta_0)],$$

pero por definición del grafo $\overline{G_{\beta_0}^{\max}}$, esto sólo se cumple si y sólo si:

$$\forall O_i, O_j \in C\ \exists\ O_{i_1}, \ldots, O_{i_q} \in C\ [O_i = O_{i_1} \wedge O_j = O_{i_q} \wedge \forall p=1,\ldots,q\text{-}1\ (O_{i_p}, O_{i_{p+1}}) \in A]. \quad (2)$$

De (1) y (2) se concluye que C es una componente β_0-conexa de $\overline{G_{\beta_0}^{\max}}$. (3)

Por otro lado, sea C una componente β_0-conexa de $\overline{G_{\beta_0}^{\max}}$, entonces cumple las siguientes dos condiciones:

a) $\forall O_i, O_j \in M\ [O_i \in C \wedge (O_i, O_j) \in A] \Rightarrow O_j \in C$;

pero por definición del grafo $\overline{G_{\beta_0}^{\max}}$, esto se cumple si y sólo si:

$$\forall O_j \in M [O_i \in C \wedge (\max_{\substack{O_t \square M \\ O_t \neq O_i}} \{\Gamma(O_i, O_t)\} = \Gamma(O_i, O_j) \geq \beta_0 \square \max_{\substack{O_t \square M \\ O_t \neq O_j}} \{\Gamma(O_j, O_t)\} = \Gamma(O_j, O_i) \geq \beta_0)] \Rightarrow O_j \in C; \quad (4)$$

b) $\forall O_i, O_j \in C\ \exists\ O_{i_1}, \ldots, O_{i_q} \in C [O_i = O_{i_1} \wedge O_j = O_{i_q} \wedge \forall p=1,\ldots,q\text{-}1\ (O_{i_p}, O_{i_{p+1}}) \in A]$;

pero por definición del grafo $\overline{G_{\beta_0}^{\max}}$, esto se cumple si y sólo si:

$$\forall O_i, O_j \in C\ \exists\ O_{i_1}, \ldots, O_{i_q} \in C\ [O_i = O_{i_1} \wedge O_j = O_{i_q} \wedge \forall p=1,\ldots,q\text{-}1$$

$$(\max_{\substack{O_t \square M \\ O_t \neq O_{i_p}}} \{\Gamma(O_{i_p}, O_t)\} = \Gamma(O_{i_p}, O_{i_{p+1}}) \geq \beta_0 \square \max_{\substack{O_t \square M \\ O_t \neq O_{i_{p+1}}}} \{\Gamma(O_{i_{p+1}}, O_t)\} = \Gamma(O_{i_{p+1}}, O_{i_p}) \geq \beta_0)]. \quad (5)$$

De (4) y (5) se concluye que C es una componente β_0-compacta de M. (6).

De (3) y (6) se concluye la validez de la proposición. ■

Como resultado de la proposición anterior se tiene que el problema del análisis de la sensibilidad para componentes β_0-compactos puede resolverse sobre las componentes conexas del grafo $\overline{G_{\beta_0}^{\max}}$.

A continuación se analizará qué sucede en el grafo $\overline{G_{\beta_0}^{\max}}$, cuando se agrega o elimina un objeto de M.

Considérese la siguiente función:

$$F_M(O_i) = \{O_{i_1}, \ldots, O_{i_q}\} \text{ si y sólo si } \forall p=1,\ldots,q\ \ \Gamma(O_i, O_{i_p}) = \max_{\substack{O_t \square M \\ O_t \neq O_i}} \{\Gamma(O_i, O_t)\}.$$

Esta función nos da el conjunto de objetos con el máximo parecido a O_i.

- **Si agregamos un objeto *O* a *M***, entonces en $\overline{G_{\beta_0}^{\max}}$ naturalmente se agrega el nodo O, pero además se eliminarán algunas aristas y se agregarán otras de la siguiente manera:

  Para cada nodo O_i, i=1,…,n, con $F_M(O_i)$={ O_{i_1} ,..., O_{i_q} }.

  Verificar si $\Gamma(O_i, O_{i_1}) < \Gamma(O_i, O)$, en caso que:

  Sí) Para cada p=1,...,q se determina si $O_i \notin F_M(O_{i_p})$, en caso de que:

  Sí) se elimina la arista (O_i, O_{i_p}). Se agrega la arista (O_i,O) y queda $F_M(O_i)$={O};

  No) verificar si $\Gamma(O_i, O_{i_1}) = \Gamma(O_i, O)$, en caso de que:

  Sí) agrega la arista (O_i,O) y queda $F_M(O_i)$={ O_{i_1} ,..., O_{i_q} ,O}.

- **Si eliminamos un objeto O_k de *M***, 1≤k≤n, entonces en $MAX(M,\beta_0)$ naturalmente se elimina el nodo O_k, con todas las aristas que inciden en él, pero además se eliminarán algunas aristas adicionales y se agregarán otras de la siguiente manera:

  Para cada nodo O_i, i=1,…,n, $i \neq k$, con $F_M(O_i)$={ O_{i_1} ,..., O_{i_q} } verificar si $O_k \in F_M(O_i)$:

  Sí) verificar si $F_M(O_i)$={O_k}.

  Sí) Calcular el nuevo $F_M(O_i)$={ O'_{i_1} ,..., O'_{i_q} } después de eliminar O_k.

  Para p=1,…,q verificar si (O_i, O'_{i_p})∉A:

  Sí) agregar la arista (O_i, O'_{i_p}).

Como puede apreciarse, se necesita resolver el problema de análisis de la sensibilidad para componentes β_0-conexas en el caso de agregar y eliminar nodos y aristas. Una solución al problema de ajustar las componentes β_0-conexas después de agregar o eliminar nodos está dada en el epígrafe 5.1.

Para el problema de ajustar las componentes β_0-conexas después de agregar o eliminar aristas se propone el siguiente método:

Se supondrá que las componentes β_0-conexas se encuentran representadas por un conjunto de árboles de búsqueda en profundidad:

- **Si se agrega una arista entre los nodos O_i y O_j**, las componentes β_0-conexas a las que pertenecen se unen.
- **Si se elimina una arista entre los nodos O_i y O_j,** se determina, utilizando las proposiciones del epígrafe 5, si {(O_i,O_j)} es un conjunto de corte, en caso de que:

  Sí) se determina si O_i es padre de O_j.

  Sí) se elimina la arista y O_j junto con todos sus descendientes forman una nueva componente β_0-conexa y por lo tanto un nuevo árbol de búsqueda en profundidad, para esto último se coloca O_j como nodo raíz y se ajustan los índices de recorrido y valores *bajo* de los nodos de este árbol.

  No) se elimina la arista y O_i junto con todos sus descendientes forman una nueva componente β_0-conexa y por lo tanto un nuevo árbol de búsqueda en profundidad, para esto último se coloca O_i como nodo raíz y se ajustan los índices de recorrido y valores *bajo* de los nodos de este árbol.

  No) se determina si (O_i,O_j) es una arista de retroceso:

Sí) se elimina la arista;

No) se elimina la arista y se aplica el procedimiento de reorganización dado en el epígrafe 5.1 a partir de O_j;

Ajustar los índices de recorrido y los valores *bajo* del árbol donde están O_i y O_j.

5.3 Sensibilidad en el cálculo de los conjuntos β_0-completos maximales

Para el análisis de la sensibilidad de los conjuntos β_0-completos maximales existe un método intuitivo que consiste en:

- **Si se agrega un nuevo objeto *O***, el cual es β_0-semejante a $O_{i_1},...,O_{i_q}$, se verifica para cada conjunto β_0-completo maximal C, si $C\subseteq\{O_{i_1},...,O_{i_q}\}$. De ser así se agrega O a este conjunto, y en caso contrario $\{O\}\cup(C\cap\{O_{i_1},...,O_{i_q}\})$ es un conjunto β_0-completo maximal.

En este caso $O_{i_1},...,O_{i_q}$ forman la nueva fila y columna de la matriz de similaridad, y hay que intersectar esta nueva fila con los vectores característicos de los conjuntos β_0-completos maximales existentes, para decidir si se agrega el objeto o se forma un nuevo conjunto β_0-completo maximal. En este caso hay que verificar que no se haya generado ya a partir de otro objeto. Todo esto es muy sencillo de programar y la complejidad está dada por el número original de conjuntos β_0-completos maximales y por el tiempo que se tarde en la intersección de los vectores característicos.

Para el caso de eliminación de un objeto, éste simplemente se elimina y los conjuntos a los que pertenece no se ven afectados, excepto si él solo forma un núcleo β_0-completo maximal degenerado, en cuyo caso se reduce en 1 el número de conjuntos.

En este caso la complejidad depende del número de conjuntos β_0-completos maximales a los que pertenezca el objeto eliminado y del tiempo que se tarde en eliminar el objeto de un conjunto.

De la baja complejidad de este método intuitivo se puede concluir que en el caso de los conjuntos β_0-completos maximales no es necesario buscar una estrategia alternativa.

6 Conclusiones

Los problemas de estructuración de objetos descritos en términos de datos mezclados e incompletos aparecen con frecuencia en problemas de la práctica real de muchas áreas del conocimiento. La estructuración de esos objetos y la relación entre esas estructuraciones es un instrumento que apoya el desarrollo de dichas disciplinas. Sin embargo, a esta problemática no se le ha dedicado suficiente atención, es un área de trabajo que requiere más desarrollos. Las herramientas que aquí se exponen constituyen pasos importantes en esta dirección que merecen ser continuados. La introducción de las mismas en la solución de problemas reales es una tarea que amerita la atención de los especialistas del área.

Referencias bibliográficas

1. Aho, A.V., Hopcroft, J.E., Ullman, J.D. (1988). Estructuras de datos y algoritmos, Addison-Wesley Iberoamericana.
2. Carrasco-Ochoa, J.A. (2001). Sensibilidad en Reconocimiento Lógico Combinatorio de Patrones. Tesis en opción del grado de Doctor en Ciencias de la Computación. Centro de Investigación en Computación, Instituto Politécnico Nacional, México, D.F.
3. Carrasco-Ochoa, J.A., Ruiz-Shulcloper, J. (1999). Análisis de Sensibilidad para Agrupamientos Duros. IV Simposio Iberoamericano de Reconocimiento de Patrones SIARP99, Cuba, pp223-233.
4. Gómez-Herrera, J.E., Rodríguez-Morán, O., Valladares-Amaro, S., Ruiz-Shulcloper, J., Pico-Peña, R., Echevarría-Rodríguez, G., Tenreyro-Pérez, R., Otero-Marrero, R., Cheremisina, E.N., Cruz-Toledo, R., Barceló-Carol, G., Álvarez-Castro, J., Barea-Centeno, M., García-Sánchez, R. (1994). Pronóstico Gasopetrolífero en la Asociación Ofeolítica Aplicando la Modelación Matemática. Revista Geofísica Internacional, Volumen 33, No. 3, July-Sept., pp 447-467. México (In Spanish).
5. Grimaldi, R.P. (1989). Matemáticas discreta y combinatoria, Addison-Wesley Iberoamericana.
6. Martínez-Trinidad, J.F. (1995). Sistema automatizado para el análisis de datos y la estructuración conceptual de universos. Tesis en opción del grado de Maestro en Ciencias, Benemérita Universidad Autónoma de Puebla, México.
7. Martínez-Trinidad, J.F., Guzmán-Arenas, A. (2001). The logical combinatorial approach to Pattern Recognition, an overview through selected works. Pattern Recognition, Vol. 34, Issue 4, pp.1-13.
8. Martínez-Trinidad, J.F., Ruiz-Shulcloper, J., Lazo-Cortés, M. (2000). Structuralization of Universes. Fuzzy Sets and Systems, Vol. 112, No. 3, pp. 485-500.
9. Montellano-Ballesteros, J.J. (1994). Agrupaciones en Gráficas Difusas. Tesis en opción del grado de Maestro en Ciencias, Instituto de Matemáticas, Universidad Nacional Autónoma de México, México, DF.
10. Ruiz-Shulcloper, J., Montellano-Ballesteros, J.J. (1995). A new model of fuzzy clustering algorithms. Proceedings of European Congress on Intelligent Techniques and Soft Computing, EUFIT'95, 1484.

RT_079, febrero 2016
Aprobado por el Consejo Científico CENATAV

Editor: Lic. Lucía González Bayona
Diseño de Portada: Di. Alejandro Pérez Abraham
RNPS No. 2142
ISSN 2072-6287
Indicaciones para los Autores:
Seguir la plantilla que aparece en www.cenatav.co.cu
C E N A T A V
7ma. A No. 21406 e/214 y 216, Rpto. Siboney, Playa;
La Habana. Cuba. C.P. 12200
Impreso en Cuba