Bases conceptuales para una teoría de objetos simbólicos

José Ruiz Shulcloper ***-*
Martín G. Chac Kantún-*
José F. Martínez Trinidad-*

-*Laboratorio de Procesamiento Digital de Imágenes
Instituto Politécnico Nacional
Centro de Investigación en Computación
E-mail cenacjr@pollux.cenac.ipn.mx

***Instituto de Cibernética, Matemática y Física, CITMA, Cuba.
E-mail recpat@cidet.icmf.edu.cu

Resumen

En este trabajo presentamos una nueva formalización del concepto de objeto simbólico de modo tal que constituye una base sólida para una Teoría de Objetos Simbólicos, la que contendrá como caso particular la Teoría del Análisis Simbólico de Datos de Diday y la Teoría Clásica del Análisis de Datos.

Palabras clave

Agrupamiento conceptual; objeto simbólico; análisis de datos; variables simbólicas; clasificación no supervisada.

Introducción

Uno de los problemas centrales en muchas disciplinas de carácter aplicado lo constituye el análisis de datos. Se ha desarrollado un conjunto grande de herramientas para la solución de distintas manifestaciones de este problema a partir de diferentes enfoques. En la medida en que los datos se hacen más complejos, por ejemplo dejan de ser exclusivamente numéricos para presentarse mezclados con datos de naturaleza cualitativa, con subjetividad, imprecisión y otros elementos de esta índole, mayor es la dificultad de extraer información útil de los mismos. Por ejemplo, en la práctica profesional, en empresas comerciales, en industrias, en investigaciones sociales, antropológicas o económicas y en muchas otras más, surge la necesidad de hacer análisis comparativos entre conjuntos de objetos. Nos interesan las similaridades y diferencias entre grupos poblacionales atendiendo a factores sociales, económicos, culturales, de salud y otros; o se hace necesario analizar el comportamiento de las ventas en una cadena de tiendas de ropas que tiene sucursales en diferentes estados de un país, atendiendo a las características de la publicidad empleada, las peculiaridades socioculturales de las diferentes poblaciones, las edades de las mismas y otras. Tradicionalmente procedemos sobre la base de la estrategia de encontrar *prototipos* de cada conjunto de objetos sujeto a estudio y comparar dichos prototipos. Para lograr esto, en muchos casos empezamos por homogeneizar el espacio de representación de los objetos, considerando sólo las variables numéricas o excluyentemente las cualitativas y asumiendo que nunca vamos a tener *datos perdidos*. En caso de que éstos aparecieran, se estiman, empleando para ello herramientas que en la mayoría de las ocasiones asumen comportamientos en el conjunto de datos que ni tan siquiera se verifican y casi nunca se satisfacen en la práctica. No nos vamos a referir a aquellas investigaciones en las que se consideran codificaciones (alto=5, mediano=4, bajo=3) como números y realizan operaciones aritméticas entre los mismos. Otra peculiaridad de esta tradicional estrategia es que en muchas ocasiones dichos prototipos son *monstruos* inexistentes en nuestra población de objetos, por lo que además es difícil confiar que el resultado de la comparación entre dichos monstruos nos pueda ofrecer una información valiosa al estudio que se realiza.

Con un sentido pragmático, a finales de los años 70, R.S.Michalski [1], introdujo un conjunto de ideas que han dado en llamarse *agrupamiento conceptual*. En esa ocasión lo que se pretendía era aportar una información adicional a la estructuración de un espacio; no se pretendía sólo decir quiénes formaban un agrupamiento sino además decir en alguna medida por qué. Esto es, se pretendía arrojar más información acerca de los agrupamientos, caracterizarlos a partir de las propiedades que éstos cumplían, definidas sobre la base de los rasgos en términos de los cuales se describen los objetos en estudio. A partir de aquí E. Diday y un grupo de científicos han estado desarrollando una serie de trabajos en torno al concepto de *objeto simbólico*, que constituyen la base de este artículo.

Génesis del concepto de objeto simbólico

Los objetos simbólicos tienen origen precisamente en los trabajos desarrollados sobre agrupamiento conceptual por R.S. Michalski [1], cuya finalidad era, dada una colección de objetos, construr agrupamientos o subcategorías de los mismos, caracterizados por descripciones conjuntivas de propiedades formadas a partir de los atributos de los objetos, teniendo estas últimas una interpretación conceptual simple. Siguiendo a Michalski [2] tenemos:

Sea $x_1,...,x_n$ variables discretas, las cuales son seleccionadas para describir objetos de la muestra a ser estructurada. Para cada variable se ha definido un conjunto de valores admisibles, el cual contiene todos los posibles valores que esta variable puede tomar para cualquier objeto de la muestra.

Se asume que los conjuntos de valores de las variables son finitos, y que además pueden ser representados como: $D_i=\{0,1,...,d_{i-1}\}$, i=1,2,...,n. En general, los conjuntos de valores pueden diferir no sólo con respecto a su tamaño, sino además en la estructura relacionada con sus elementos.

En los trabajos desarrollados por Michalski y sus colaboradores, se distingue solamente entre variables nominales y lineales, cuyos dominios son no ordenados y conjuntos linealmente ordenados respectivamente.

Un **evento** es definido como cualquier secuencia de valores de las variables $x_1,...,x_n$, y se denota $e=(r_1,r_2,...,r_n)$ donde $r_i \in D_i$, i=1,2,...,n. El conjunto de todos los posibles eventos S, es llamado el espacio de eventos denotado por $S=\{e_1,...,e_d\}$ donde $d=d_1 x d_2 x ... x d_n$ (el tamaño del espacio de eventos).

Una proposición relacional, del cálculo proposicional clásico, $[x_i \# R_i]$ donde R_i es un conjunto de uno o más elementos del dominio de x_i, y # simboliza los operadores relacionales $\geq$, $>,<,\leq,=,\in$, o sus negaciones es llamada ***selector***.

El selector $[x_i=R_i]$ ($[x_i\neq R_i]$) es interpretado como "el valor de $x_i \in R_i$" ("el valor de $x_i \notin R_i$"). En el caso de variables lineales, el operador "=" en $[x_i=R_i]$ puede ser reemplazado por los operadores relacionales $\geq$, $>$, $<$, $\leq$ para un R_i apropiado .

Ejemplos de selectores:

[altura ≥ 1,75]	(la altura es mayor o igual a 1.75)
[color=azul,rojo]	(el color es azul o rojo)
[tamaño ≠ mediano]	(el tamaño no es el mediano)
[peso=2...5]	(el peso está entre 2 y 5)

Como notaremos más adelante, a los selectores E. Diday les llama eventos. Nosotros preservaremos la terminología de Diday.

Un producto lógico de selectores es llamado un ***complejo lógico*** (*l-complejo*); si los valores de las variables en *e* satisfacen todos los selectores en el complejo se dice que *e* satisface al *l*-complejo. Por ejemplo, el evento $e=(2,7,0,1,5,4,6)$ satisface al *l*-complejo $[x_1=2,3]$ $[x_3\leq 3]$ $[x_5=3..8]$ (la concatenación de selectores significa conjunción). Aquí estamos preservando las notaciones de Michalski que posteriormente modificaremos.

Un *l*-complejo puede ser visto como una representación simbólica exacta de los eventos a los cuales satisface. Por ejemplo, el *l*-complejo antes mencionado es la representación simbólica de todos los eventos para los cuales x_1 es 2 ó 3, x_3 es menor o igual que 3, y x_5 está entre 3 y 8.

Cualquier conjunto de eventos para los cuales exista un *l*-complejo (propiedad) que estos eventos satisfagan y sólo ellos, es llamado un ***conjunto complejo*** (*s-complejo*).

Los primeros conceptos de objeto simbólico

En el análisis de datos clásico, los objetos son t-uplos numéricos. La agrupación de tales objetos es lograda por la minimización de la disimilaridad dentro de cada agrupamiento y maximizando la disimilaridad entre agrupamientos. Con los objetos simbólicos Diday pretende lograr extensiones de tipos de datos clásicos. En conjuntos de datos clásicos, los objetos son *individualizados*, mientras que en los conjuntos de datos simbólicos éstos son *unificados* por medio de vínculos, propiedades. Estos últimos son más complejos que los datos convencionales de la siguiente manera:

1) Todos los objetos de un conjunto de datos simbólicos pueden no estar definidos en términos de las mismas variables.
2) Cada variable, en la descripción de un objeto simbólico, puede tomar más de un valor, incluso un conjunto infinito de valores.
3) En objetos simbólicos más complejos, los valores que toman las variables pueden incluir uno o más objetos *simbólicos elementales*.
4) La descripción de un objeto simbólico puede depender de las relaciones que existan entre otros objetos simbólicos.
5) Los valores que las variables toman pueden estar tipificados, indicando la frecuencia de ocurrencia, probabilidad relativa, nivel de importancia de los valores, etc.

En sus trabajos [3-6], Diday nos brinda varias definiciones y descripciones que son necesarias recordar:

"Un ***evento*** es una pareja (variable-valor) que enlaza las variables y valores de las mismas en los objetos". Ejemplos de eventos:

e_1=[color={verde, azul, rojo}]
e_2=[altura=156]
e_3=[tiempo=[1.2...3.1]].

Los ***objetos simbólicos*** son definidos por una conjunción lógica de eventos (en la terminología de Diday).

Como puede observarse en la formulación de Michalski, un *l*-complejo es un objeto simbólico.

Basados en la complejidad de las expresiones que los describen, Diday introduce diferentes tipos de objetos simbólicos:

aseverativos (assertion), acumulativos (hordes) o de tipo síntesis (synthesis).

Un *objeto aseverativo* es una conjunción de eventos pertenecientes a las descripciones (individuales) de objetos reales. Ejemplo:

$a=e_1 \wedge e_2 \wedge e_3$ =[color={verde, azul,rojo}]∧[altura=156]∧[tiempo=[1.2,...,3.1]]

Aquí a es un objeto simbólico aseverativo que tiene las siguientes propiedades:

1) el color es verde, azul o rojo.
2) la altura es igual a 156.
3) el tiempo está entre 1.2 y 3.1.

Un *objeto simbólico acumulativo* es una conjunción de 2 ó más objetos simbólicos aseverativos y eventos. Ejemplo:

h=[VDU(Comp1)=color]∧[RAM(Comp1)=64k]∧[Keys(Comp1)=[57,...,63]]∧
[VDU(Comp2)=[B&N]∧[RAM(Comp2)=48k]∧[Keys(Comp2)=[64,...,73]]

Esto significa que el objeto **h** consiste de 2 objetos elementales (aseverativos):

1) Computadora 1 con VDU a color, RAM de 64k y Keys entre 57 y 63.
2) Computadora 2 con VDU en B&N, RAM de 48k y Keys entre 64 y 73.

Un *objeto simbólico tipo síntesis* es una conjunción de 2 ó más objetos simbólicos acumulativos y eventos. Ejemplo:

s=$\mathbf{h}_1$&$\mathbf{h}_2$= [Tipo(r_1)=autopista][Vehículos(r_1)=2]∧[Tipo(r_2)=carretera][Vehículos(r_2)=1]∧
[Tipo(v_1)=carro][Color(v_1)=azul][Movimiento(v_1)=r_1]∧
[Tipo(v_2)=trailer][Color(v_2)=rojo][Movimiento(v_2)=r_1]∧
[Tipo(v_3)=autobús][Color(v_3)=verde][Movimiento(v_3)=r_2]

Definición formal de E. Diday

A partir de todas estas ideas y resultados iniciales, Diday en [7] propone una formalización del concepto de objeto simbólico. Sea dado el diagrama de la fig 1.

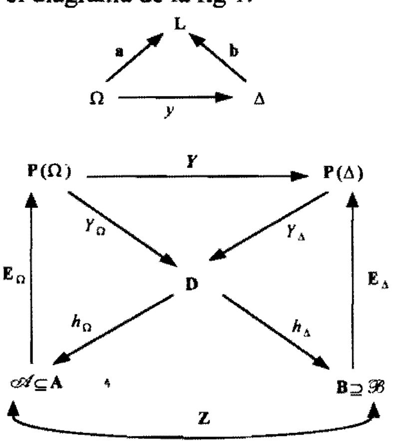


Figura 1. Diagrama de E. Diday

Donde:

Ω es un conjunto de entes elementales llamados "objetos individuales";

Δ es un conjunto de posibles descripciones de los elementos de Ω;

y: Ω →Δ, la cual asocia a cualquier w∈Ω su descripción δ=$y(w)$;

D es un conjunto de descripción de subcunjuntos de Ω;

Y_Ω: **P**(Ω)→**D**, donde **P**(Ω) es el conjunto potencia de Ω el cual asocia a cualquier Ω'⊆Ω su descripción **d**∈**D**;

Y: **P**(Ω)→**P**(Δ) tal que Y(Ω')=Δ' si y sólo si (ssi) Δ'={$y(w)$ | w∈Ω'};

Y_Δ: **P**(Δ)→**D** la cual asocia a cualquier Δ'⊆Δ una descripción **d**∈**D** que satisface al menos la siguiente propiedad: Y_Δ(Δ')⊆**D**;

A es un conjunto de mapeos Ω→**L** donde **L**={verdadero, falso} (más general **L**=[0, 1]);

h_Ω: **D**→**A** tal que h_Ω(**d**)=**a**, **a**:Ω→**L** tal que: **a**(w)=verdadero ssi y (w)=δ∈**d**;

B es el conjunto de mapeos Δ→**L** tal que h_Δ(**d**)=**b**, **b**:Δ→**L** tal que **b**(δ)=verdadero ssi δ∈**d**;

$\mathcal{A}$= h_Ω(**D**) y $\mathcal{B}$=h_Δ(**D**);

Z: $\mathcal{B}$→$\mathcal{A}$ tal que **Z**(**b**)=**a** ssi **a**=**b**°y.

Una "intención" de un conjunto de objetos individuales Ω'⊆Ω puede ser definida por **d**=Y_Ω (Ω'), **a**=h_Ω (Y_Ω (Ω')),ó **b**=$h\Delta$(Y_Ω(Ω')). La "extensión" de **a** en Ω es un subconjunto de Ω denotado Ext(**a**/Ω) y definido por:
Ext(**a**/Ω)={w∈Ω| **a**(w)=verdadero}; la "extensión" de **b** es un subconjunto de Δ definido por:
Ext(**b**/Δ)={δ∈Δ | **b**(δ)=verdadero}; la "extensión" de **d**∈**D** en Q=Δ,es denotado Ext(**d**/Q); por definición, tomamos Ext(**d**/Ω)=Ext(**a**/Ω) y Ext(**d**/Δ)=Ext(**b**/Δ).
$\mathbf{E}_\Delta$ es el mapeo $\mathcal{B}$→**P**(Δ) tal que $\mathbf{E}_\Delta$(**b**)=Ext(**b**/Ω); $\mathbf{E}_\Omega$:$\mathcal{A}$→**P**(Ω) tal que $\mathbf{E}_\Omega$:(**a**)=Ext(**a**/Ω).

De tal manera que, para Diday, un "objeto simbólico" *es un conjunto de propiedades concernientes a un subconjunto de* Ω. Cualquier elemento de **D**, **B**, o **A** puede ser considerado como un "objeto simbólico".

Elementos críticos

En el trabajo de E. Diday [3], en particular en la definición formal de objeto simbólico, la notación que utiliza no hace una distinción precisa entre un objeto individual y un objeto simbólico.

También tiene algunas imprecisiones:

1) En la definición de Y_Δ tiene la imprecisión siguiente : para Δ⊆ Y_Δ(Δ')=**d**∈**D**, por definición; sin embargo Y_Δ(Δ')⊆ **D** por la propiedad de satisfacción, entonces **d**=Y_Δ (Δ')⊆ **D**, lo cual es una inconsistencia.

2) E. Diday da un ejemplo (Sección 2.2 [3]), en el cual Δ= $O_1xO_2x,...,xO_p$ y **D**= **P**(O_1)x**P**(O_2)x,...,x**P**(O_p), en particular Δ= O_1xO_2 , **D**= **P**(O_1)x**P**(O_2) y define **d**=$\mathbf{V}_1x\mathbf{V}_2$ con $\mathbf{V}_1$⊆O_1

y $\mathbf{V}_2 \subseteq O_2$, aquí existe la incongruencia de que $\mathbf{d} \in \mathbf{D}$, ya que **d** es un conjunto de t-uplos, cuyas componentes son valores simples, y **D** es un conjunto de t-uplos pero cuyas componentes son a la vez conjuntos de valores. De la misma manera comete el error de considerar a={a}, cuando hace que $\mathbf{D} \subseteq \mathbf{P}(\Delta)$ (Sección 2.4 [3]).

3)En la práctica existen objetos individuales que sus variables aceptan un conjunto de valores, a diferencia de los dados por E. Diday, que sólo aceptan un solo valor por la función *y*.
Es el caso, por ejemplo, en el que estamos describiendo una persona y decimos que su estatura está entre 1,70 m y 1,80 m; o cuando afirmamos que el color del pelo era entre rubio y castaño claro, esto es, cuando nuestra imprecisión está acotada y sabemos que en definitiva la persona tiene una única estatura y un único color de pelo.

4)La formalización del concepto de objeto simbólico, no es general, ya que al considerar objetos de tipo "hordes" y de "synthesis" se modifica su diagrama particularmente en lo que se refiere al universo de objetos Ω al considerar Ω^p y $\Omega_1,\ldots,\Omega_p$ respectivamente y por consiguiente todas las funciones que utilizan Ω.

Redefinición de la formalización anterior

Debido a las anteriores irregularidades presentes en la formalización anterior, se han analizado y discutido varias soluciones para mejorar lo hecho por Diday; antes de desarrollar una nueva formalización es preciso enmendar lo establecido con miras a generalizar estos conceptos.

Sea Ω un conjunto de objetos del universo de estudio **E** ($\Omega \subseteq \mathbf{E}$); Δ un conjunto descripciones admisibles de Ω (**F** el conjunto de todas las descripciones posibles, $\Delta \subseteq \mathbf{F}$).

Sean O_i el conjunto de valores admisibles de la variable (rasgo) x_i i=1,...,p, $\Delta = O_1 x \ldots x O_p$, $\Delta_1 = \mathbf{P}(O_1) x \ldots x \mathbf{P}(O_p)$, con $\mathbf{P}(O_i)$ conjunto potencia de O_i, i=1,...,p y $\mathbf{D} \subseteq \Delta_1$. Sea **B** un conjunto de fórmulas bien formadas (fbf) en un cálculo proposicional *Cp*.
Consideremos el siguiente diagrama :

$$\begin{tikzcd} \Omega \arrow[r, "y"] \arrow[dr, "y_1"'] & \Delta \\ & \Delta_1 \end{tikzcd}$$

$$\begin{tikzcd} \mathbf{P}(\Omega \arrow[r, "Y"] & \mathbf{P}(\Delta_1) \arrow[d, "X"] \\ \mathbf{B} \arrow[u, "Z"] & \mathbf{D} \arrow[l, "S"] \end{tikzcd}$$

Figura 2. Diagrama modificado

Donde :

y: $\Omega \rightarrow \Delta$ es una función de descripción de objetos de Ω en p-uplos cuyas componentes tienen asignado un sólo valor. $y(w)=\delta$ con $w \in \Omega$, $\delta=(x_1(w),\ldots,x_p(w))$.
y_1: $\Omega \rightarrow \Delta_1$ función de descripción de objetos de Ω en p-uplos cuyas componentes tienen asignado un conjunto de valores. Esto permite a un rasgo x_i tomar uno o más valores admisibles en O_i, se da comúnmente cuando hay imprecisión del valor de un rasgo, es decir, el valor se encuentra entre ciertos valores o un intervalo de valores (finitos o infinitos).
Y: $\mathbf{P}(\Omega) \rightarrow \mathbf{P}(\Delta_1)$ función de descripción de subconjuntos de Ω en un conjunto de p-uplos cuyas componentes tienen asignado un conjunto de valores, es decir, dado $\Omega' \subseteq \Omega$, , $Y(\Omega')=\{(V_1^1,\ldots,V_p^1),\ldots,(V_1^s,\ldots,V_p^s)\}$ $0 \le s \le p$.
X: $\mathbf{P}(\Delta_1) \rightarrow \mathbf{D}$ la cual asigna a un conjunto de p-uplos, cuyas componentes son conjuntos de valores, un conjunto de un solo p-uplo, cuyas componentes son conjuntos de valores, al cual denominaremos objeto simbólico (Os). En símbolos:
$X(\{(V_1^1,\ldots,V_p^1),\ldots,(V_1^s,\ldots,V_p^s)\})=\{(V_1,\ldots,V_p)\}=Os$
En particular, restringiendo el dominio $\mathbf{P}(\Delta_1)$ a $Y(\mathbf{P}(\Omega))$, tenemos que:

$$X(Y(\Omega'))=\{(V_1,\ldots,V_p) \mid V_i=\Phi_i(Y(\Omega')_{[xi]})\}$$

donde:

$$\Phi_i: \mathbf{P}(O_i)^s \rightarrow \mathbf{P}(O_i) \quad 0 \le s \le p$$

$$Y_{[xi]}: \mathbf{P}(\Delta_1) \rightarrow \mathbf{P}(O_i)^s$$

$$Y(\Omega')_{[xi]}=(V_i^s,\ldots,V_i^s)$$

$$\Phi_i((V_i^s,\ldots,V_i^s))=V_i.$$

S: $\mathbf{D} \rightarrow \mathbf{B}$, función de asignación (determinación intencional) de un Os en una fórmula lógica bien formada, es decir: $S(Os)=fbf(\{(V_1,\ldots,V_p)\})$, como por ejemplo la conjunción de proposiciones de la forma $[x_i \in V_i]$ (l-complejos de Michalski ver [1]).
Z: $\mathbf{B} \rightarrow \mathbf{P}(\Omega)$, función de asignación (determinación extensional) de una fbf(Os), a un subconjunto $\Omega'' \subseteq \Omega$.
No necesariamente $\Omega''=\Omega'$, aunque si $\Omega' \subseteq \Omega''$ (s-complejos de Michalski ver [1]).

Del diagrama modificado debemos puntualizar, que si bien se introducen una serie de elementos que salvan las deficiencias señaladas al diagrama de Diday, sin embargo, no se supera el inciso 4) del apartado anterior. De aquí que a continuación se propone una nueva formalización del concepto de objeto simbólico, la cual tiene como casos particulares, la propuesta por E. Diday y la modificada.

Nueva formalización del concepto de objeto simbólico

Daremos una serie de definiciones sobre la base de las cuales se construirá todo el andamiaje de nuestra Teoría de Objetos Simbólicos, que tiene como objetivo central, la comparación entre conjuntos de datos.

Definición 1.- Sean $U_1,...,U_r$ universos, no necesariamente diferentes, de objetos reales, una *variable simbólica sobre* $U_1,...,U_r$ es una función parcialmente definida sobre $2^{*U_1} x \ldots x 2^{*U_r}$ como sigue

$$X: 2^{*U_1} x \ldots x 2^{*U_r} \rightarrow 2^M$$
$$X((A_1,...,A_r))=V$$

donde $A_i \subseteq U_i$, para $i=1,...,r$; $V \subseteq M$; $X((A_1,...,A_r))$ denota el valor de la variable X en el r-uplo $(A_1,...,A_r)$, $2^{*U_1} x \ldots x 2^{*U_r}$ es el producto cartesiano de los conjuntos potencia de los respectivos U_i, con $i=1,...,r$, excluyendo al vacío. A M lo denominamos *conjunto generador de los valores admisibles de la variable* X y a 2^M, el conjunto potencia de M, lo denominaremos *conjunto de valores admisibles de la variable* X. El valor $X((A_1,...,A_r))=\varnothing$ denota la ausencia de información en cuanto al valor que toma el t-uplo $(A_1,...,A_r)$ en la variable X.

Ejemplo: Sea U={Juan, Pedro}, M=N (el conjunto de los números naturales), X: $2^{*U} \rightarrow 2^M$ la variable simbólica (aquí r=1) que denota la edad de un elemento de 2^{*U}.

Así:

X({Juan})={17, ...,22}
X({Pedro})={44}
X({Juan, Pedro})={17,...,44}

donde {17,...,44} en este caso significa que la edad del conjunto {Juan, Pedro} está entre 17 y 44.

Definición 2.- Una *variable simbólica* se denomina *relacional heterogénea* ssi $U_i \neq U_j$ para $i \neq j=1,...,r$; $r>1$.

Ejemplo: Sea X (porcentaje de venta) una variable que relaciona al porcentaje de venta de un conjunto de vendedores U_1={Pedro, Juan, María} y las marcas de automóviles U_2={VW, Datsun} y M=[0,100].
X({Juan},{VW})={42}
X({Pedro, María},{VW})={100}
X({Pedro},{Datsun})= {45,...,55}
X({Juan, Pedro},{Datsun})= {50,...,55}

Definición 3.- Una *variable simbólica* se denomina *relacional homogénea* ssi $U_i=U_j$ para $i \neq j=1,...,r$; $r>1$.

Ejemplo: Sea X la variable vecino_de, U_1={Pedro, Juan, María}, U_2={Pedro, Juan, María} y M={si, no, no_sé}.
X({Juan},{María})={si}
X({Juan},{María, Pedro})={no_sé}

Nota : Este es un ejemplo particular de una variable simbólica relacional homogénea *binaria*, ya que r=2.

Definición 4.- Una *variable simbólica* se denomina *conjuntual* ssi r=1.

Ejemplo: Sea X la variable bebida predilecta, U={Pedro, Juan, María} M={Coca-Cola, Pepsi-Cola, Otra}
X({Juan})={Coca-Cola}
X({Juan, Pedro})={Coca-Cola, Pepsi-Cola}
X({Pedro})={Pepsi-Cola, Coca-Cola}

Las variables relacionales de acuerdo al cardinal de los conjuntos A_i pueden ser : a) *unitarias* $|A_i|=1$, $i=1,...,s$; siendo s el orden de la relación, b) parcialmente unitarias si $\exists j_1,...,j_p\ |A_t|>1$, para $t=j_1,...,j_p$; $\exists i_1,...,i_q\ |A_t|=1$, para $t=i_1,...,i_q$, p+q=s.
De manera análoga las variables conjuntuales se denominan unitarias si $|A|=1$.

Tanto las variables relacionales como conjuntuales se denominan *univaluadas* si $|V|=1$. Hay que hacer notar que estos atributos no son excluyentes, es decir, puede haber variables simbólicas *unitarias univaluadas*. A continuación ejemplificamos una *variable simbólica conjuntual unitaria*.

Ejemplo: Sea X la variable consumo_de_energía_eléctrica, U={Pedro, Juan, María} y $M=R^+$ el conjunto de los números reales positivos.
X({Pedro})={1350,23}
X({María})={880,99}

Nota : Este ejemplo también puede ilustrar el caso univaluado.

Atendiendo a la naturaleza del conjunto M las variables simbólicas pueden ser reales, enteras, de intervalos, booleanas, k-valentes, difusas, lingüísticas, etc.

Es importante observar que las variables simbólicas conjuntuales unitarias se pueden poner en correspondencia biunívoca con las variables tradicionales, esto es, variables cuantitativas (reales, enteras, de intervalos, etc.) y cualitativas (booleanas, k-valentes, difusas, lingüísticas, etc.). En particular tenemos que X({Juan})={22}, se corresponde con X(Juan)=22. Ponemos especial atención en las variables simbólicas *difusas* y *lingüísticas*.

Definición 5.- Una *variable simbólica* se denomina *difusa* si es una función parcialmente definida sobre $2^{*U_1} x \ldots x 2^{*U_r}$ como sigue

$$\tilde{X}: 2^{*U_1} x \ldots x 2^{*U_r} \rightarrow 2^{\tilde{M}}$$
$$\tilde{X}((A_1,...,A_r))=\tilde{V}$$

siendo

$$\mathrm{Img}\,\tilde{X}(2^{*U_1} x \ldots x 2^{*U_r}) = \varnothing + \sum_{i=1}^{n} X((A_1,...,A_r)) \,|\, \mu_{\mathrm{Img}X}(X((A_1,...,A_r)))$$

el conjunto de todos los subconjuntos difusos en 2^M por X, donde $\mu_{\mathrm{Img}X}(X((A_1,...,A_r)))$ describe el grado con el que la variable simbólica X toma el valor V en el t-uplo de conjunto de objetos reales $(A_1,...,A_r)$, donde $A_i \subseteq U_i$, para $i=1,...,r$; $\tilde{V} \subseteq 2^M$.
Denotaremos $2^{\tilde{M}}$ al conjunto de todos los subconjuntos difusos de 2^M y la denominaremos *potencia difusa* de 2^M.

Definición 6.- Una *variable simbólica* se denomina *lingüística* ssi toma como valores variables simbólicas difusas.

$$X: 2^{*U_1} x \ldots x 2^{*U_r} \rightarrow (2^{\tilde{M}})^{2^{*U_1} x \ldots x 2^{*U_r}}$$
$$X((A_1,\ldots,A_r)) = \tilde{X}((A_1,\ldots,A_r))$$

Definición 7.- Sea dada una variable simbólica cualquiera X_i, $i=1,\ldots,n$. Un *operador asociado a* X_i es una función parcialmente definida sobre $2^{*U_1} x \ldots x 2^{*U_r}$

$$\varphi|_{X_i}: 2^{*U_1} x \ldots x 2^{*U_r} \rightarrow (2^{M_i})^{2^{*U_1} x \ldots x 2^{*U_r}}$$

tal que $\varphi|_{X_i}((A_1,\ldots,A_r)) = X_i((A_{i_1},\ldots,A_{i_s})) = V_i$, donde $V_i \subseteq M_i$, $U_{i_1},\ldots,U_{i_s}$, son los universos sobre los que está definida la variable X_i; $i=1,\ldots,n$; $1 \leq s \leq r$.

Definición 8.- Sean dados $U_1,\ldots,U_r$ universos de objetos reales no necesariamente diferentes, $r \geq 1$; $X_1, \ldots, X_n$ variables simbólicas, $\varphi|_{X_i}$ los operadores asociados a X_i; $i=1,\ldots,n$; $A_j \subseteq U_j$; $j=1,\ldots,r$. Diremos que un *objeto simbólico* (Os) es un elemento cualquiera de la imagen del operador

$$D_c: 2^{*U_1} x \ldots x 2^{*U_r} \rightarrow 2^{M_1} x \ldots x\, 2^{M_n}$$
$$(A_1,\ldots,A_r) \rightarrow D_c(A_1,\ldots,A_r) = (V_1,\ldots,V_n) = Os$$

tal que $\varphi|_{X_i}(A_1,\ldots,A_r) = V_i$, siendo $V_{i_j} = \emptyset$, para $j=1,\ldots,p$, $0 \leq p < n$. Tomamos la convención

$Os = (\emptyset,\ldots,\emptyset, Vi_1,\ldots,Vi_j,\ldots,\emptyset,\ldots, Vi_{j+1},\ldots, Vi_p, \emptyset, \ldots,\emptyset) = (Vi_1,\ldots,Vi_p)$

Definición 9.- Sea dado un cálculo proposicional Cp. Sobre el mismo se considera el concepto de fórmula bien formada (fbf). Una *determinación intencional* de un objeto simbólico Os está dada por una imagen cualquiera del operador:

$$D_I: 2^{M_1} x \ldots x\, 2^{M_n} \rightarrow Cp$$
$$D_c(A_1,\ldots,A_r) = (V_1,\ldots,V_n) = Os \rightarrow D_I(Os) = fbf(Vi_1, \ldots, Vi_P)$$

Ejemplo: $D_I(Os) = \bigwedge_{j=1}^{p} [X_{i_j}(Os) \in V_{i_j}]$

Definición 10.- Una *determinación extensional* de un objeto simbólico Os es una imagen cualquiera del operador:

$$D_E: Img\varphi|_{X_1} x \ldots x Img\varphi|_{X_n} \rightarrow 2^M$$
$$D_E(Os) = \{(o, \mu_{D_E(Os)}(o)) | o \in U = \bigcup_{j=1}^{r} U_{ij} \wedge \mu_{D_E(Os)}(o) = v(fbf(Vi_1,\ldots,Vi_p))\}$$

donde $\mu_{D_E(Os)}(o)$ describe el grado de pertenencia de o al subconjunto difuso $D_E(Os) \subseteq 2^M$ y v es la función valor veritativo definida sobre el cálculo proposicional seleccionado Cp.

Corolario **a)** $SopD_E(Os) = \{o | \{o\} \in 2^{*U}$ objeto simbólico conjuntual unitario asociado a $o \wedge$

$D_I(\{o\}) = fbf(X'_1(\{o\}),\ldots, X'_n(\{o\})) \wedge \forall\, i=1,\ldots,n\ X'_i(\{o\}) \subseteq X_i(Os)\}$

donde X'_i es la variable simbólica conjuntual asociada a X_i.

$$\textbf{b)}\ |SopD_E(Os)| \geq \prod_{i=1}^{n} \left(2^{|X'_i(\{o\})|} - 1 \right)$$

Definición 11.- De acuerdo a su descripción intencional los objetos simbólicos pueden ser *booleanos*, *k-valentes*, *difusos*, *de creencia*, *probabilísticos*, *posibilísticos*, *modales*, etc. en dependencia del cálculo proposicional Cp que se emplee.

Un caso particular de objeto simbólico, el cual nos ocuparemos inmediatamente, es cuando r=1, es decir, cuando trabajamos con objetos de un solo universo.

Definición 12.- Un *objeto simbólico individual* está definido por $D_c(\{a\}) = (x_{i_1}(\{a\}),\ldots, x_{i_p}(\{a\}))$, donde, x_{i_j} $j=1,\ldots,p$ son variables simbólicas conjuntuales. Cuando las x_{i_j} sean variables simbólicas conjuntuales unitarias entonces $D_c(\{a\})$ será denominado *objeto simbólico individual unitario*.

Definición 13.- Un *objeto simbólico individual con imprecisión acotada* en las variables x_{i_j} $j=1,\ldots,s \leq p$ es un objeto simbólico individual tal que $|V_{i_t}(w(a))| > 1$, $t=1,\ldots,s$

Definición 14.- Un *evento* es un objeto simbólico $Os(A) = D_c(A) = (V_1,\ldots,V_p)$ tal que $\exists!\, i=1,\ldots,p$ $Vi \neq \emptyset$. Los eventos serán denotados $[X_i(Os(A)) \in V_i]$, con $A \subseteq U$.

Independientemente del cálculo proposicional, empleado los objetos simbólicos pueden ser: *Aseverativos*, *Acumulativos* y *Síntesis*. De acuerdo a su descripción extensional los objetos simbólicos pueden ser *duros* si $\mu_{D_E(Os)}(o) \in \{0, 1\}$; *difusos* si $\mu_{D_E(Os)}(o) \in [0, 1]$; *L-difuso* si $\mu_{D_E(Os)}(o) \in L$, con L conjunto totalmente ordenado. A cada uno de los posibles tipos de objetos simbólicos le corresponde una expresión para las determinaciones extensional e intencional respectivamente.

CONCLUSIONES

En primer lugar es importante destacar que los objetos simbólicos individuales unitarios pueden ponerse en correspondencia biunívoca con las descripciones de los objetos reales con las que habitualmente trabajamos en el Análisis de Datos y el Reconocimiento de Patrones. No es difícil ver que las operaciones entre descripciones de objetos reales (tradicionales) y entre variables (tradicionales) pueden ser fácilmente extensibles al caso de objetos simbólicos individuales unitarios y de variables simbólicas conjuntuales unitarias respectivamente. Esto nos permite aseverar que todo lo que hasta ahora se ha podido hacer (y se hará) con estos entes tradicionales, resulta un caso particular del tratamiento con nuestros objetos simbólicos.

En segundo lugar, que este conjunto de definiciones constituye un lenguaje que nos permitirá elaborar, sobre bases sólidas, las herramientas necesarias para el desarrollo de algoritmos de estructuración de espacios simbólicos, la clasificación supervisada de conjuntos de objetos y de individuos, a partir de una muestra de aprendizaje formada con conjuntos de objetos, la clasificación parcialmente supervisada en condiciones análogas a las descritas para la clasificación supervisada, la selección de variables mediante la extensión de herramientas de la Teoría de Testores [8], sin tener que fabricar *monstruos* como los mencionados en la introducción.

En tercer lugar, las definiciones de determinación intencional de los objetos simbólicos, en términos del cálculo proposicional conveniente (clásico, de Lukasiewicz, posibilístico, de creencia, etc.) crea un marco amplio de trabajo, que aunque complejo, nos permitirá extraer la información que hoy no podemos obtener de los conjuntos de datos disponibles.

En la actualidad ya hemos definido los conceptos de funciones de semejanza entre objetos simbólicos, entre los valores de una misma variable y un conjunto de expresiones que nos permiten establecer estructuraciones de los espacios de representación de objetos simbólicos atendiendo a diferentes formas de realización, como componentes conexas, conjuntos compactos, clanes y otros criterios de agrupamientos duros y difusos [9,10]. Estos resultados serán objeto de una publicación posterior.

Teniendo en cuenta las ventajas que permiten todos estos conceptos para el tratamiento de conjuntos de objetos reales, grandes volúmenes de información, tratamiento de objetos de universos diferentes, descritos en términos de variables no necesariamente iguales, puede afirmarse que estas formalizaciones constituyen un paso importante en la elaboración de una Teoría de Objetos Simbólicos, que sin lugar a dudas ocupará un lugar en el Reconocimiento de Patrones y el Análisis de Datos.

Algunos de los resultados aquí presentados fueron expuestos en eventos internacionales [11,12]. Esta investigación estuvo parcialmente financiada por la Dirección de Estudios de Posgrado y de Investigación del Instituto Politécnico Nacional de México.

Referencias

[1] Michalski, R.S., (1981). Revealing conceptual structure in data by inductive inference. *Machine Intelligence 10*, eds. Hayes-Michic, D. Michie, and Y.H. Pao, Chichister: Ellis Horwood, New York: Halsted Press (John Wiley).

[2] Michalski, R.S., E. Stepp and E. Diday, (1981). A recent advanced in data analysis: clustering objects into classes characterizad by conjuntive concepts. *Progress in Pattern Recognition*, Vol. 1, L. Kanal and A Rosenfeld, eds.

[3] Chidananda Gowda, K. and E. Diday, (1992). Symbolic Clustering Using a New Similarity Measure. *IEEE Transactions on Systems, Man, and Cybernetics*, Vol 22, No. 2, pp. 368-378.

[4] Chidananda Gowda, K. and E. Diday, (1992). Symbolic Clustering Using a New Dissimilarity Measure. *IEEE Transactions on Systems, Man, and Cybernetics*, Vol 22, No. 2, pp. 567-578.

[5] Chidananda Gowda, K. and T.V. Ravi, (1995). Agglomerative clustering of symbolic objects using the concepts of both similarity and dissimilarity. *Pattern Recognition Letters* 16, pp. 647-652.

[6] Chidananda Gowda, K. and E. Diday, (1991). Unsupervised Learning through Symbolic Clustering. *Pattern Recognition Letters* 12, North-Holland, pp. 259-264.

[7] Diday, E. (1995). Probabilist, possibilist and belief objects for knowledge analysis. *Annals of Operations Research* 55, pp. 227-276.

[8] Lazo-Cortés, M. and J. Ruiz-Shulcloper, (1995) Determining the feature relevance for non-clssically described objects and a new algorithm to compute typical fuzzy testors. *Pattern Recognition Letters* 16, pp. 1259-1265.

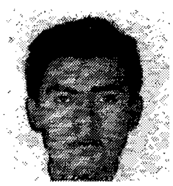
[9] Ruiz-Shulcloper, J. and J.J. Montellano-Ballesteros, (1995). A new model of fuzzy clustering algorithms. *Proceedings of the EUFIT'95, august 28-31, Aachen*, pp. 1484-1488. Germany.

[10] Martínez-Trinidad, J.F. and J. Ruiz-Shulcloper. Fuzzy Semantic Clustering. *Proceedings of the EUFIT'96, sept. 2-5, Aachen,1996*, pp. 1397-1401. Germany.

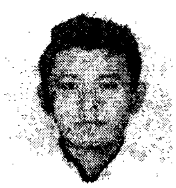
[11] Ruiz-Shulcloper, J.; M. Chac-Kantún, J.F. Martínez-Trinidad, (1996). Data Analysis Between Sets of Objects. In *Advances in Artificial Intelligence and Ingineering Cybernetics, Volume III*, Ed. George E. Lasker (*Proceedings of the 8th International Conference on Systems Research, Informatics and Cybernetics, Baden Baden, Germany, August 14-18,1996*) pp. 81-85.

[12] Ruiz-Shulcloper, J.; M. Chac-Kantún, J.F. Martínez Trinidad, (1996). Acerca de una Teoría de Objetos Simbólicos. *Memorias del Simposium Internacional de Computación*, pp. 137-144. México.

José Ruiz Shulcloper. *Nacionalidad cubana, obtuvo el Doctorado en Matemáticas por la Universidad Estatal de Moscú y el Centro de Cálculo de la Academia de Ciencias de la URSS. Sus áreas de interés son: el enfoque lógico combinatorio del Reconocimiento de Patrones, la Teoría de Testores, la Teoría de Objetos Simbólicos y los Modelos Difusos.*
Actualmente es investigador titular del Instituto de Cibernética, Matemática y Física de Cuba, donde es jefe del grupo de Reconocimiento de Patrones. Además es profesor titular del Centro de Investigación en Computación del IPN, México.



José Francisco Martínez Trinidad. *Nacionalidad mexicana, obtuvo la Maestría en Ciencias de la Computación, por la Benemérita Universidad Autónoma de Puebla, México. Sus áreas de interés son: el enfoque lógico combinatorio del Reconocimiento de Patrones, el Agrupamiento Conceptual y la Teoría de Objetos Simbólicos.*
Actualmente es profesor Investigador en el Centro de Investigación en Computación del IPN, México.



Martín Chac Kantún. *Nacionalidad mexicana, obtuvo el grado de Licenciado en Matemáticas por la Escuela de Matemáticas de la Universidad Autónoma de Yucatán. Sus áreas de interés son: la Teoría de Objetos Simbólicos, el enfoque lógico combinatorio del Reconocimiento de Patrones y el Análisis de Datos. Actualmente es estudiante en el laboratorio de Procesamiento de Imágenes del Centro de Investigación en Computación del IPN, México.*

FE DE ERRATAS

DICE	DEBE DECIR
PORTADA : Revista Iberoamerica de Computación	 Revista Iberoamericana de Computación
CONTRAPORTADA : Volumen. I Número.1 Volume. I. Number .1	 Volumen I Número1 Volume I Number 1
006 Bruce H. Mc.Cormick, David A. Batte and Andrew T. Duchwski	5 Bruce H. McCormick, David A. Batte and Andrew T. Duchowski
014 José Ruiz. Shulcloper, Martín G. Chac y José F. Martínez Bases Conceptuales para una Teoría de Objetos Simbólicos	13 José Ruiz Shulcloper, Martín G. Chac y José F. Martínez Bases Conceptuales para una Teoría de Objetos Simbólicos
022 Gustavo Núñez and Matías Alvarado Contextual Knowledge Reasoning : Representation and Automatizaton	21 Gustavo Núñez and Matías Alvarado Contextual Knowledge and Belief : Representation and Reasoning
028 Igor A. Bolshakov El Modelo Morfológico Formal para la Síntesis de Sustantivos y Adjetivos en Español	27 Igor A. Bolshakov El Modelo Morfológico Formal para Sustantivos y Adjetivos en el Español
038 J. Humberto Sossa and J. Luis Díaz de León Recognizing Planar Riding and Non-Riding Wire-Shapes	37 J. Humberto Sossa and J. Luis Díaz de León Recognizing Planar Rigid and Non-Rigid Wire-Shapes
046 Sergio R. Sandoval Revisión de libros	45 Sergio Sandoval Reyes Revisión de libros
048 Anuncios y Eventos Announcements and call for papers	47 Anuncios y Eventos Announcements and call for papers
PRIMERA DE FORROS Juan Manuel Ibarra Centro de Ivestigación en Computación-IPN México	 Juan Manuel Ibarra Zannatha Centro de Investigación en Computación-IPN México
PÁG. 1 Volumen. I Número.1 Volume. I. Number .1	 Volumen I Número 1 Volume I Number 1
006 Bruce H. Mc.Cormick, David A. Batte and Andrew T. Duchwski	5 Bruce H. McCormick, David A. Batte and Andrew T. Duchowski
014 José Ruiz. Shulcloper, Martín G. Chac y José F. Martínez Bases Conceptuales para una Teoría de Objetos Simbólicos	13 José Ruiz Shulcloper, Martín G. Chac y José F. Martínez Bases Conceptuales para una Teoría de Objetos Simbólicos
022 Gustavo Núñez and Matías Alvarado Contextual Knowledge Reasoning : Representation and Automatizaton	21 Gustavo Núñez and Matías Alvarado Contextual Knowledge and Belief : Representation and Reasoning
028 Igor A. Bolshakov El Modelo Morfológico Formal para la Síntesis de Sustantivos y Adjetivos en Español	27 Igor A. Bolshakov El Modelo Morfológico Formal para Sustantivos y Adjetivos en el Español
038 J. Humberto Sossa and J. Luis Díaz de León Recognizing Planar Riding and Non-Riding Wire-Shapes	37 J. Humberto Sossa and J. Luis Díaz de León Recognizing Planar Rigid and Non-Rigid Wire-Shapes
046 Sergio R. Sandoval Revisión de libros	45 Sergio Sandoval Reyes Revisión de libros
048 Anuncios y Eventos Announcements and call for papers	47 Anuncios y Eventos Announcements and call for papers
PÁG. 2 Línea 11 invited several well know Mexican and foreign researchers in Computer Science and related fields,	 invited several well known Mexican and foreign researchers in Computer Science and related fields,
PÁG. 3 Línea 9 Correo electrónico : Revista@pollux,cenac.ipn.mx.	 Correo electrónico : Revista@pollux.cenac.ipn.mx.
PÁG. 26 **Matías Alvarado Mentado**. Nacionalidad mexicana, es licenciado en Matemáticas por la Facultad de Ciencias de la Universidad Nacional Autonoma de México.	 **Matías Alvarado Mentado**. Nacionalidad mexicana, es licenciado en Matemáticas por la Facultad de Ciencias de la Universidad Nacional Autónoma de México.
PÁG. 44 **Juan Luis Díaz de León Santiago**. Nacionalidad mexicana, obtuvo el doctorado en el CINVESTAV del I.P.N, México con especialidad con ingenieria eléctrica. Profesor titular en el Centro de Investigación en el I.P.N.	PÁG. 44 **Juan Luis Díaz de León Santiago**. Nacionalidad mexicana, obtuvo el doctorado en el CINVESTAV del I.P.N, México con especialidad en ingeniería eléctrica. Profesor titular en el Centro de Investigación en Computación en el I.P.N.
PÁG. 52 Línea 33 [7]...	 Línea 33 [7] A. Rosenfeld. Digital Topology. *Amer.Math. Montly, 86 :621-630, 1979.* [7] G. Matheron. *Random Sets and Integral Geometry.* John Wile and Sons, New York, 1995. [7] B.Vazquez, J.H. Sossa and J.L. Díaz de León Autoguided Vehicle Control using Expended Time B-splines. In *IEEE International Conference on Systems, Man and Cybernetics,* pages 2786 -2791, San Antonio Texas, October, 1994. [7] P. Perner, P. Wang and A. Rosefeld (Eds). Advances in Structural and Syntactical Pattern Recognition, Springer 1996.
PÁG. 53 Línea 14 Loa artículos deben ajustarse a las siguientes especificaciones :	 Línea 14 Los artículos deben ajustarse a las siguientes especificaciones :