Objeto Simmbólico: una forma de tomar en cuenta la variabilidad de los datos.

Mallea Adriana-Herrera Myriam

LXV Reunión Anual de Comunicaciones Científicas. Bahia Blanca, Octubre de 2016

Introducción

En el Análisis de Datos clásico, los datos proceden de observaciones únicas, de determinadas variables sobre individuos únicos. Pero cuando los elementos de interés son clases o grupos de algún tipo , como los ciudadanos que viven en una ciudad determinada por ejemplo, hay variabilidad inherente en los datos.

Los datos simbólicos, introducidos por E. Diday en los ochenta, proporcionan un marco que permite representar datos con variabilidad mediante nuevos tipos de variables.

En el Análisis de Datos clásico, los datos proceden de observaciones únicas, de determinadas variables sobre individuos únicos. Pero cuando los elementos de interés son clases o grupos de algún tipo , como los ciudadanos que viven en una ciudad determinada por ejemplo, hay variabilidad inherente en los datos.

Los datos simbólicos, introducidos por E. Diday en los ochenta, proporcionan un marco que permite representar datos con variabilidad mediante nuevos tipos de variables.

En los últimos años surgió el término Big Data , refiriéndose a conjuntos de datos tan grandes y complejos que se vuelven difíciles de procesar en un tiempo razonable con aplicaciones tradicionales de análisis de datos. El análisis de objetos simbólicos, al ofrecer la posibilidad de agregación de datos al nivel de granularidad elegido por el usuario, mientras se mantiene la información sobre la variabilidad intrínseca, desempeña un papel importante en este contexto.

Los datos simbólicos pueden provenir de individuos considerando clases o grupos de los mismos y, en tal caso, representan las propiedades o descripciones de un elemento genérico de la clase que describen. Desde otro punto de vista,los datos simbólicos pueden establecerse por el conocimiento del experto sin necesidad de datos individuales; o venir dados con incertidumbre.

Un poco de Historia

El Análisis de Datos Simbólicos permite la extensión de la Estadística a la Estadística de las intenciones o conceptos y más concretamente la extensión de problemas, métodos y algoritmos de Análisis de Datos a datos simbólicos. Según Diday el Análisis de Datos Simbólicos crea un puente entre la Estadística y el Aprendizaje Automático.

Diday (Diday (1987,1988)) formaliza los conceptos de intención y extensión, debidos a Arnauld y Nicole (Arnauld y Nicole, 1662). La intención de un concepto constituye su descripción, mientras que la extensión es el conjunto de individuos cuya descripción es acorde a la del concepto.

Un poco de Historia

El Análisis de Datos Simbólicos permite la extensión de la Estadística a la Estadística de las intenciones o conceptos y más concretamente la extensión de problemas, métodos y algoritmos de Análisis de Datos a datos simbólicos. Según Diday el Análisis de Datos Simbólicos crea un puente entre la Estadística y el Aprendizaje Automático.

Diday (Diday (1987,1988)) formaliza los conceptos de intención y extensión, debidos a Arnauld y Nicole (Arnauld y Nicole, 1662). La intención de un concepto constituye su descripción, mientras que la extensión es el conjunto de individuos cuya descripción es acorde a la del concepto.

Según Diday (Diday, 2000), el Análisis de Datos Simbólicos nace influido por tres campos:

- El Análisis Exploratorio de Datos,
- La Inteligencia Artificial, donde gran esfuerzo se realiza por el desarrollo de lenguajes de representación de conocimiento,
- la Taxonomía Numérica usada en las Ciencias Biológicas.

Según Diday (Diday, 2000), el Análisis de Datos Simbólicos nace influido por tres campos:

- El Análisis Exploratorio de Datos,
- La Inteligencia Artificial, donde gran esfuerzo se realiza por el desarrollo de lenguajes de representación de conocimiento,
- la Taxonomía Numérica usada en las Ciencias Biológicas.

Según Diday (Diday, 2000), el Análisis de Datos Simbólicos nace influido por tres campos:

- El Análisis Exploratorio de Datos,
- La Inteligencia Artificial, donde gran esfuerzo se realiza por el desarrollo de lenguajes de representación de conocimiento,
- la Taxonomía Numérica usada en las Ciencias Biológicas.

Según Diday (Diday, 2000), el Análisis de Datos Simbólicos nace influido por tres campos:

- El Análisis Exploratorio de Datos,
- La Inteligencia Artificial, donde gran esfuerzo se realiza por el desarrollo de lenguajes de representación de conocimiento,
- la Taxonomía Numérica usada en las Ciencias Biológicas.

En el análisis de datos clásico, los datos se presentan en una matriz X de orden $n \times p$, donde cada uno de los n individuos (fila) toma un único valor para cada una de las p variables (columnas).
En una encuesta de hogares, por ejemplo, cada hogar se encuentra caracterizado por la región en la que se encuentra, el número de habitaciones, el número de baños y su categoría socio profesional.

ID	Región	Nro de Habit.	Nro de baños	CSP
11404	Sur	2	1	Baja
11405	Sur	2	3	Media
11406	Sur	1	1	Media
...	...	...	...	...
12112	Capital	3	1	Baja
12113	Capital	2	2	Media
12114	Capital	1	3	Alta

En el análisis de datos clásico, los datos se presentan en una matriz X de orden $n \times p$, donde cada uno de los n individuos (fila) toma un único valor para cada una de las p variables (columnas).
En una encuesta de hogares, por ejemplo, cada hogar se encuentra caracterizado por la región en la que se encuentra, el número de habitaciones, el número de baños y su categoría socio profesional.

ID	Región	Nro de Habit.	Nro de baños	CSP
11404	Sur	2	1	Baja
11405	Sur	2	3	Media
11406	Sur	1	1	Media
...	...	...	...	...
12112	Capital	3	1	Baja
12113	Capital	2	2	Media
12114	Capital	1	3	Alta

Las variables simbólicas tienen en cuenta la variabilidad inherente a los datos. Esta variabilidad ocurre cuando:

- Los datos son de pacientes de un centro médico, pero, analizamos el centro médico, no el paciente.
- Los datos son acerca de vuelos, pero analizamos los aeropuertos, no cada vuelo individual
- Los datos son acerca de personas, pero analizamos las ciudades, regiones, países, no cada persona
- Como en el ejemplo, los datos son acerca de hogares, pero nos interesa analizar las regiones, no cada hogar.

En estos casos los valores de las variables son conjuntos, distribución de intervalos o un conjunto subyacente de subintervalos o categorías.

Micro-dato $\Rightarrow$ Macro-dato

Las variables simbólicas tienen en cuenta la variabilidad inherente a los datos. Esta variabilidad ocurre cuando:

- Los datos son de pacientes de un centro médico, pero, analizamos el centro médico, no el paciente.
- Los datos son acerca de vuelos, pero analizamos los aeropuertos, no cada vuelo individual
- Los datos son acerca de personas, pero analizamos las ciudades, regiones, países, no cada persona
- Como en el ejemplo, los datos son acerca de hogares, pero nos interesa analizar las regiones, no cada hogar.

En estos casos los valores de las variables son conjuntos, distribución de intervalos o un conjunto subyacente de subintervalos o categorías.

$$\text{Micro-dato} \Rightarrow \text{Macro-dato}$$

Región	Nro de Habit.	Nro de baños	CSP
Sur	$2(2/3), 1(1/3)$	$1(2/3), 3(1/3)$	Baja$(1/3)$, Media$(2/3)$
...	...	...	...
Capital	$1(1/3), 2(1/3), 3(1/3)$	$1(1/3), 2(1/3), 3(1/3)$	Media$(2/3)$, Alta$(1/3)$

Fuentes de Datos Simbólicos

- Agregación de microdatos: contemporánea, temporal.
- Descripción de conceptos abstractos.

Región	Nro de Habit.	Nro de baños	CSP
Sur	$2(2/3), 1(1/3)$	$1(2/3), 3(1/3)$	Baja$(1/3)$, Media$(2/3)$
...	...	...	...
Capital	$1(1/3), 2(1/3), 3(1/3)$	$1(1/3), 2(1/3), 3(1/3)$	Media$(2/3)$, Alta$(1/3)$

Fuentes de Datos Simbólicos

- Agregación de microdatos: contemporánea, temporal.
- Descripción de conceptos abstractos.

Tipos de Variables Simbólicas

- Variables Numéricas o Cuantitativas
  - Variables Numéricas monovaluadas.
  - Variables Numéricas multivaluadas
  - Variables de Intervalos
  - Variables de Histograma
- Variables Categóricas o Cualitativas
  - Variables categóricas monovaluadas
  - Variables categóricas multivaluadas
  - Variables Categóricas modales

Sea $\Omega = \{\omega_1, ..., \omega_m\}$ un conjunto de individuos y sea $\mathcal{Y}$ un conjunto o dominio de posibles valores observados.

Sea $E = \{e_1, ..., e_n\}$ un conjunto de objetos. Como casos particulares más frecuentes se tiene que E es un subconjunto de Ω o un subconjunto de las clases de Ω, es decir, $E \subseteq \Omega$ o $E \subseteq P(\Omega)$. En el segundo caso, los datos simbólicos correspondientes, describen clases de individuos de Ω.

Definición

Se dice que una variable conjunto valuada es una aplicación:

$$X : E \to P(\mathcal{Y})$$
$$e \to X(e)$$

- $X(e)$ es la descripción de un elemento $e \in E$ en $P(\mathcal{Y})$ dada por la variable X
- $P(\mathcal{Y})$ es el conjunto de descripciones de los elementos de E.

En caso que $|X(e)| = 1$ para todo $e \in E$ se denomina monovaluada. Se llama multivaluada se $1 < |X(e)| < \infty$ para todo $e \in E$. Se puede tratar de una variable categórica o cuantitativa.

Definición

Se dice que una variable conjunto valuada es una aplicación:

$$\begin{aligned} X : E &\rightarrow P(\mathcal{Y}) \\ e &\rightarrow X(e) \end{aligned}$$

- $X(e)$ es la descripción de un elemento $e \in E$ en $P(\mathcal{Y})$ dada por la variable X
- $P(\mathcal{Y})$ es el conjunto de descripciones de los elementos de E.

En caso que $|X(e)| = 1$ para todo $e \in E$ se denomina monovaluada. Se llama multivaluada se $1 < |X(e)| < \infty$ para todo $e \in E$. Se puede tratar de una variable categórica o cuantitativa.

Se puede extender la definición anterior al caso **multivariante**.

Definición

Sean $X_1, \ldots, X_p$, p variables conjunto valuadas definidas en E, con dominios respectivos $\mathcal{Y}_j$.
Sea $P(\mathcal{Y}) = P(\mathcal{Y}_1) \times \ldots \times P(\mathcal{Y}_p)$ el producto cartesiano de las partes de dichos dominios. El vector de variables **X** *definido en E es:*

$$\mathbf{X} : E \rightarrow P(\mathcal{Y})$$
$$e \rightarrow \mathbf{X}(e) = (X_1(e), \ldots, X_p(e))$$

En el caso en que $E \subseteq P(\Omega)$, la variable X o el **X** se llama **descriptor** de clases de individuos de Ω y $P(\mathcal{Y})$ conjunto de las descripciones de clases de Ω, o de los elementos de $P(\Omega)$

Se puede extender la definición anterior al caso **multivariante**.

Definición

Sean $X_1, ..., X_p$, p variables conjunto valuadas definidas en E, con dominios respectivos $\mathcal{Y}_j$.
Sea $P(\mathcal{Y}) = P(\mathcal{Y}_1) \times ... \times P(\mathcal{Y}_p)$ el producto cartesiano de las partes de dichos dominios. El vector de variables $\mathbf{X}$ definido en E es:

$$\mathbf{X} : E \to P(\mathcal{Y})$$
$$e \to \mathbf{X}(e) = (X_1(e), ..., X_p(e))$$

En el caso en que $E \subseteq P(\Omega)$, la variable X o el $\mathbf{X}$ se llama **descriptor** de clases de individuos de Ω y $P(\mathcal{Y})$ conjunto de las descripciones de clases de Ω, o de los elementos de $P(\Omega)$

Descripción de clase de individuos a partir de descripciones de individuos

Sea el conjunto $E = \{S_1, ..., S_m\} \subseteq P(\Omega)$ .Se presenta la forma más habitual de descripción de una clase por generalización de las descripciones de los individuos que la componen. Sea $\widetilde{X} = \left(\widetilde{X}_1, ..., \widetilde{X}_p\right)$ un vector de variables definidas en Ω con dominios respectivos $\mathcal{Y}_j$:

A partir del vector de variables monoevaluadas $\widetilde{X}$ de descripción de individuos se define el vector de variables $X = (X_1, ..., X_p)$ en E de descripción de clases de individuos como:

$$X : E \to P(\mathcal{Y})$$
$$S_i \to X(S_i) = (X_1(S_1), ..., X_p(S_p)) =$$
$$\left(\left\{\widetilde{X}_1(\omega) : \omega \in S_i\right\}, ..., \left\{\widetilde{X}_p(\omega) : \omega \in S_i\right\}\right)$$

La descripción de la clase S_i de individuos se obtiene a partir de los valores de $\mathcal{Y}$ que son observadas por el vector $\widetilde{X}$ en los individuos $\omega \in S_i$.

Descripción de clase de individuos a partir de descripciones de individuos

Sea el conjunto $E = \{S_1, ..., S_m\} \subseteq P(\Omega)$ .Se presenta la forma más habitual de descripción de una clase por generalización de las descripciones de los individuos que la componen. Sea $\widetilde{X} = \left(\widetilde{X}_1, ..., \widetilde{X}_p\right)$ un vector de variables definidas en Ω con dominios respectivos $\mathcal{Y}_j$:
A partir del vector de variables monoevaluadas $\widetilde{X}$ de descripción de individuos se define el vector de variables $X = (X_1, ..., X_p)$ en E de descripción de clases de individuos como:

$$X : E \to P(\mathcal{Y})$$
$$S_i \to X(S_i) = (X_1(S_1), ..., X_p(S_p)) =$$
$$\left(\left\{\widetilde{X}_1(\omega) : \omega \in S_i\right\}, ..., \left\{\widetilde{X}_p(\omega) : \omega \in S_i\right\}\right)$$

La descripción de la clase S_i de individuos se obtiene a partir de los valores de $\mathcal{Y}$ que son observadas por el vector $\widetilde{X}$ en los individuos $\omega \in S_i$.

Descripción de clase de individuos a partir de descripciones de individuos

Sea el conjunto $E = \{S_1, ..., S_m\} \subseteq P(\Omega)$ .Se presenta la forma más habitual de descripción de una clase por generalización de las descripciones de los individuos que la componen. Sea $\widetilde{X} = \left(\widetilde{X}_1, ..., \widetilde{X}_p\right)$ un vector de variables definidas en Ω con dominios respectivos $\mathcal{Y}_j$:
A partir del vector de variables monoevaluadas $\widetilde{X}$ de descripción de individuos se define el vector de variables $X = (X_1, ..., X_p)$ en E de descripción de clases de individuos como:

$$X : E \to P(\mathcal{Y})$$
$$S_i \to X(S_i) = (X_1(S_1), ..., X_p(S_p)) =$$
$$\left(\left\{\widetilde{X}_1(\omega) : \omega \in S_i\right\}, ..., \left\{\widetilde{X}_p(\omega) : \omega \in S_i\right\}\right)$$

La descripción de la clase S_i de individuos se obtiene a partir de los valores de $\mathcal{Y}$ que son observadas por el vector $\widetilde{X}$ en los individuos $\omega \in S_i$.

Ejemplo

Sea el conjunto de individuos $\Omega = \{\omega_1, ..., \omega_7\}$ *descrito por las variables categóricas monoevaluadas* $\widetilde{X}_1 = \widetilde{sexo}$ *e* $\widetilde{X}_2 = \widetilde{estado\ civil}$ *con dominios respectivos* $\mathcal{Y}_1 = \{masculino, femenino\}$ *e* $\mathcal{Y}_2 = \{soltero, casado, viudo\}$.
La matriz de datos se representa por:

$$\begin{pmatrix} id & \widetilde{sexo} & \widetilde{estado\ civil} \\ \omega_1 & femenino & casado \\ \omega_2 & femenino & soltero \\ \omega_3 & femenino & soltero \\ \omega_4 & masculino & casado \\ \omega_5 & masculino & viudo \\ \omega_6 & masculino & casado \\ \omega_7 & masculino & viudo \end{pmatrix}$$

Sea $E \subset P(\Omega)$, $E = \{S_1, S_2\} = \{\{\omega_1, \omega_2, \omega_3\}, \{\omega_4, \omega_5, \omega_6, \omega_7\}\}$. A partir de los descriptores de individuos $\widetilde{sexo}$ y $\widetilde{estado\ civil}$ se pueden definir los descriptores de clase de individuos sexo y profesión. Las variables multievaluadas sexo y profesión se definen como:

$$sexo : E \longrightarrow P(\{masculino, femenino\})$$
$$S_i \longrightarrow sexo(S_i) = \{\widetilde{sexo}(\omega) : \omega \in S_i\}$$

$$estado\ civil : E \longrightarrow P(\{soltero, casado, viudo\})$$
$$S_i \longrightarrow estado\ civil(S_i) = \left\{\widetilde{estado\ civil}(\omega) : \omega \in S_i\right\}$$

Así, por ejemplo para la clase de individuos S_1, se tiene que $sexo(S_1) = \{femenino\}$ y $profesión(S_1) = \{soltero, casado\}$ y el vector de descripciones multievaluadas para la clase S_1 es

$$(sexo, estado\ civil)\,(S_1) = (sexo(S_1), estado\ civil(S_1)) = \\ (\{femenino\}\,, \{soltero, casado\})$$

En este caso, la matriz de datos simbólicos que representa E es:

$$\begin{pmatrix} ID & sexo & estado\ civil \\ S_1 & \{femenino\} & \{soltero, casado\} \\ S_2 & \{masculino\} & \{casado, viudo\} \end{pmatrix}$$

Variables modales probabilistas

Una variable modal es aquella que describe un elemento del conjunto E no sólo por un subconjunto de elementos del dominio sino también por modos o pesos de cada uno de ellos. Las variables modales probabilistas asocian a cada elemento de E una distribución de probabilidad o de frecuencias.

Sea $\mathcal{Y} = \{z_1, ..., z_x\}$ y sea $\mathcal{M}(\mathcal{Y}) = \{q : q \text{ es una distribución de probabilidad definida en } \mathcal{Y}\}$, el **conjunto de descripciones modales probabilistas** de elementos de E. Una descripción $q \in \mathcal{M}(\mathcal{Y})$ se define como:

Definición

$$q : \mathcal{Y} \rightarrow [0, 1]$$
$$z_i \rightarrow q(z_i) \quad con \sum_{i=1,...,x} q(z_i) = 1$$

Variables modales probabilistas

Una variable modal es aquella que describe un elemento del conjunto E no sólo por un subconjunto de elementos del dominio sino también por modos o pesos de cada uno de ellos. Las variables modales probabilistas asocian a cada elemento de E una distribución de probabilidad o de frecuencias.
Sea $\mathcal{Y} = \{z_1, ..., z_x\}$ y sea
$\mathcal{M}(\mathcal{Y}) = \{q : q$ es una distribución de probabilidad definida en $\mathcal{Y}\}$, el **conjunto de descripciones modales probabilistas** de elementos de E.
Una descripción $q \in \mathcal{M}(\mathcal{Y})$ se define como:

Definición

$$q : \mathcal{Y} \rightarrow [0,1]$$
$$z_i \rightarrow q(z_i) \quad con \sum_{i=1,\dots,x} q(z_i) = 1$$

Variables modales probabilistas

Una variable modal es aquella que describe un elemento del conjunto E no sólo por un subconjunto de elementos del dominio sino también por modos o pesos de cada uno de ellos. Las variables modales probabilistas asocian a cada elemento de E una distribución de probabilidad o de frecuencias.
Sea $\mathcal{Y} = \{z_1, ..., z_x\}$ y sea
$\mathcal{M}(\mathcal{Y}) = \{q : q \text{ es una distribución de probabilidad definida en } \mathcal{Y}\}$, el **conjunto de descripciones modales probabilistas** de elementos de E.
Una descripción $q \in \mathcal{M}(\mathcal{Y})$ se define como:

Definición

$$q : \mathcal{Y} \rightarrow [0, 1]$$
$$z_i \rightarrow q(z_i) \quad con \sum_{i=1,...,x} q(z_i) = 1$$

Se identifica el **dato simbólico** o descripción simbólica q con

$$q \equiv (z_1 q(z_1), ..., z_x q(z_x))$$

Definición

se dice que X es una variable modal probabilista definida en E, si es una aplicación

$$\begin{gathered} X : E \to \mathcal{M}(\mathcal{Y}) \\ e \to X(e) = q_e \end{gathered}$$

tal que dado $e \in E$ le asocia $X(e) = q_e$ donde q_e es una distribución de probabilidad en el conjunto $\mathcal{Y}$ de posibles valores de observación completado por una $\sigma -$ álgebra

$X(e)$ es la descripción modal probabilista (en $\mathcal{M}(\mathcal{Y})$) del elemento $e \in E$ dada por la variable modal probabilista X.

Se identifica el **dato simbólico** o descripción simbólica q con

$$q \equiv \left(z_1 q\left(z_1\right), ..., z_x q\left(z_x\right)\right)$$

Definición

se dice que X es una variable modal probabilista definida en E, si es una aplicación

$$\begin{aligned} X : E &\rightarrow \mathcal{M}(\mathcal{Y}) \\ e &\rightarrow X(e) = q_e \end{aligned}$$

tal que dado $e \in E$ le asocia $X(e) = q_e$ donde q_e es una distribución de probabilidad en el conjunto $\mathcal{Y}$ de posibles valores de observación completado por una $\sigma -$ álgebra

$X(e)$ es la descripción modal probabilista (en $\mathcal{M}(\mathcal{Y})$) del elemento $e \in E$ dada por la variable modal probabilista X.

Se identifica el **dato simbólico** o descripción simbólica q con

$$q \equiv \left(z_1 q\left(z_1\right), ..., z_x q\left(z_x\right)\right)$$

Definición

se dice que X es una variable modal probabilista definida en E, si es una aplicación

$$X : E \to \mathcal{M}(\mathcal{Y})$$
$$e \to X(e) = q_e$$

tal que dado $e \in E$ le asocia $X(e) = q_e$ donde q_e es una distribución de probabilidad en el conjunto $\mathcal{Y}$ de posibles valores de observación completado por una $\sigma -$ álgebra

$X(e)$ es la descripción modal probabilista (en $\mathcal{M}(\mathcal{Y})$) del elemento $e \in E$ dada por la variable modal probabilista X.

Descripción modal probabilista de clases de a partir de descripciones de individuos

Sea $E = \{S_1, ..., S_m\} \subseteq \mathcal{P}(\Omega)$, $\widetilde{X} = \left(\widetilde{X}_1, ..., \widetilde{X}_p\right)$ un vector de variables categóricas monoevaluadas definidas en Ω con dominios respectivos $\mathcal{Y}_j$, $\mathcal{M}(\mathcal{Y}) = \mathcal{M}(\mathcal{Y}_1) \times ... \times \mathcal{M}(\mathcal{Y}_p)$.

A partir del vector de variables monoevaluadas $\widetilde{X}$ de descripción de individuos se define el vector de variables modales probabilistas $X = (X_1, ..., X_p)$ en E de descripción de clase de individuos:

$$X : E \to \mathcal{M}(\mathcal{Y})$$
$$S_i \to X(S_i) = (X_1(S_i), ..., X_p(S_i)) = (q_{S_i,1}, ..., q_{S_i,p})$$

Descripción modal probabilista de clases de a partir de descripciones de individuos

Sea $E = \{S_1, ..., S_m\} \subseteq \mathcal{P}(\Omega)$, $\widetilde{X} = \left(\widetilde{X}_1, ..., \widetilde{X}_p\right)$ un vector de variables categóricas monoevaluadas definidas en Ω con dominios respectivos $\mathcal{Y}_j$, $\mathcal{M}(\mathcal{Y}) = \mathcal{M}(\mathcal{Y}_1) \times ... \times \mathcal{M}(\mathcal{Y}_p)$.
A partir del vector de variables monoevaluadas $\widetilde{X}$ de descripción de individuos se define el vector de variables modales probabilistas $X = (X_1, ..., X_p)$ en E de descripción de clase de individuos:

$$X : E \to \mathcal{M}(\mathcal{Y})$$
$$S_i \to X(S_i) = (X_1(S_i), ..., X_p(S_i)) = (q_{S_i,1}, ..., q_{S_i,p})$$

Descripción modal probabilista de clases de a partir de descripciones de individuos

Sea $E = \{S_1, ..., S_m\} \subseteq \mathcal{P}(\Omega)$, $\widetilde{X} = \left(\widetilde{X}_1, ..., \widetilde{X}_p\right)$ un vector de variables categóricas monoevaluadas definidas en Ω con dominios respectivos $\mathcal{Y}_j$, $\mathcal{M}(\mathcal{Y}) = \mathcal{M}(\mathcal{Y}_1) \times ... \times \mathcal{M}(\mathcal{Y}_p)$.

A partir del vector de variables monoevaluadas $\widetilde{X}$ de descripción de individuos se define el vector de variables modales probabilistas $X = (X_1, ..., X_p)$ en E de descripción de clase de individuos:

$$\begin{gathered} X : E \rightarrow \mathcal{M}(\mathcal{Y}) \\ S_i \rightarrow X(S_i) = (X_1(S_i), ..., X_p(S_i)) = (q_{S_i,1}, ..., q_{S_i,p}) \end{gathered}$$

donde la distribución de probabilidad $q_{S_i,j}$ para $j \in \{1, ..., p\}$ se define como:

$$q_{S_i,j} : \mathcal{Y}_j \to [0,1]$$

$$y_j \to q_{S_i,j}(y_j) = \frac{Card\left(\left\{\omega \in S_i : \widetilde{X}_j(\omega) = y_j\right\}\right)}{Card(S_i)}$$

$X(S_i)$ es la **descripción de la clase de individuos** $S_i \in \mathcal{P}(\Omega)$ el **vector de variables modales probabilistas** definido en $\mathcal{P}(\Omega)$, inducido por el vector $\widetilde{X}$ de **variables monoevaluadas** definido en Ω.

donde la distribución de probabilidad $q_{S_i,j}$ para $j \in \{1, ..., p\}$ se define como:

$$q_{S_i,j} : \mathcal{Y}_j \to [0, 1]$$

$$y_j \to q_{S_i,j}(y_j) = \frac{Card\left(\left\{\omega \in S_i : \widetilde{X}_j(\omega) = y_j\right\}\right)}{Card(S_i)}$$

$X(S_i)$ es la **descripción de la clase de individuos** $S_i \in \mathcal{P}(\Omega)$ el **vector de variables modales probabilistas** definido en $\mathcal{P}(\Omega)$, inducido por el vector $\widetilde{X}$ de **variables monoevaluadas** definido en Ω.

donde la distribución de probabilidad $q_{S_i,j}$ para $j \in \{1, ..., p\}$ se define como:

$$q_{S_i,j} : \mathcal{Y}_j \to [0,1]$$

$$y_j \to q_{S_i,j}(y_j) = \frac{Card\left(\left\{\omega \in S_i : \widetilde{X}_j(\omega) = y_j\right\}\right)}{Card(S_i)}$$

$X(S_i)$ es la **descripción de la clase de individuos** $S_i \in \mathcal{P}(\Omega)$ el **vector de variables modales probabilistas** definido en $\mathcal{P}(\Omega)$, inducido por el vector $\widetilde{X}$ de **variables monoevaluadas** definido en Ω.

Ejemplo

$$\begin{pmatrix} id & \widetilde{sexo} & \widetilde{estado\ civil} \\ \omega_1 & femenino & casado \\ \omega_2 & femenino & soltero \\ \omega_3 & femenino & soltero \\ \omega_4 & masculino & casado \\ \omega_5 & masculino & viudo \\ \omega_6 & masculino & casado \\ \omega_7 & masculino & viudo \end{pmatrix}$$

$E = \{S_1, S_2\} = \{\{\omega_1, \omega_2, \omega_3\}, \{\omega_4, \omega_5, \omega_6, \omega_7\}\}.$

$$\begin{pmatrix} ID & sexo & estado\ civil \\ S_1 & (femenino) & (casado\frac{1}{3}, soltero\frac{2}{3}) \\ S_2 & (masculino) & (casado\frac{1}{2}, viudo\frac{1}{2}) \end{pmatrix}$$

Ejemplo

$$\begin{pmatrix} id & \widetilde{sexo} & \widetilde{estado\ civil} \\ \omega_1 & femenino & casado \\ \omega_2 & femenino & soltero \\ \omega_3 & femenino & soltero \\ \omega_4 & masculino & casado \\ \omega_5 & masculino & viudo \\ \omega_6 & masculino & casado \\ \omega_7 & masculino & viudo \end{pmatrix}$$

$E = \{S_1, S_2\} = \{\{\omega_1, \omega_2, \omega_3\}, \{\omega_4, \omega_5, \omega_6, \omega_7\}\}$.

$$\begin{pmatrix} ID & sexo & estado\ civil \\ S_1 & (femenino) & (casado\frac{1}{3}, soltero\frac{2}{3}) \\ S_2 & (masculino) & (casado\frac{1}{2}, viudo\frac{1}{2}) \end{pmatrix}$$

Ejemplo

$$\begin{pmatrix} id & \widetilde{sexo} & \widetilde{estado\ civil} \\ \omega_1 & femenino & casado \\ \omega_2 & femenino & soltero \\ \omega_3 & femenino & soltero \\ \omega_4 & masculino & casado \\ \omega_5 & masculino & viudo \\ \omega_6 & masculino & casado \\ \omega_7 & masculino & viudo \end{pmatrix}$$

$E = \{S_1, S_2\} = \{\{\omega_1, \omega_2, \omega_3\}, \{\omega_4, \omega_5, \omega_6, \omega_7\}\}$.

$$\begin{pmatrix} ID & sexo & estado\ civil \\ S_1 & (femenino) & (casado\frac{1}{3}, soltero\frac{2}{3}) \\ S_2 & (masculino) & (casado\frac{1}{2}, viudo\frac{1}{2}) \end{pmatrix}$$

Las distribuciones de posibilidad y los conjuntos difusos son otras formas de representación de la incertidumbre y pueden encuadrarse en el marco de las variables y datos simbólicos. Según Diday (Diday (1991,1995a)) las variables modales posibilistas asocian a cada elemento de E una distribución de posibilidad sobre el conjunto de categorías.

Desde otro punto de vista, las categorías de una variable se pueden definir cómo conjuntos difusos y los elementos de E tienen unos grados de pertenencia a estos conjuntos difusos.

Las distribuciones de posibilidad y los conjuntos difusos son otras formas de representación de la incertidumbre y pueden encuadrarse en el marco de las variables y datos simbólicos. Según Diday (Diday (1991,1995a)) las variables modales posibilistas asocian a cada elemento de E una distribución de posibilidad sobre el conjunto de categorías.

Desde otro punto de vista, las categorías de una variable se pueden definir cómo conjuntos difusos y los elementos de E tienen unos grados de pertenencia a estos conjuntos difusos.

Conjunto de descripciones simbólicas

$$X : E \longrightarrow \mathcal{D}$$

Con $\mathcal{D}$ un conjunto de descripciones de elementos de E asociado al conjunto o dominio $\mathcal{Y}$.

- $\mathcal{D} = \mathcal{Y}$ en el caso de que X sea una variable monoevaluada
- $\mathcal{D} = \mathcal{P}(\mathcal{Y})$ en el caso de que X sea una variable simbólica multievaluada
- $\mathcal{D} = \mathcal{M}(\mathcal{Y})$ en el caso de que X sea una variable simbólica modal.

Un vector de variables simbólicas asocia a un elemento de E un vector de descripciones. En este caso

$$\mathcal{D} = \mathcal{D}_1 \times \ldots \times \mathcal{D}_p$$

Conjunto de descripciones simbólicas

$$X : E \rightarrow \mathcal{D}$$

Con $\mathcal{D}$ un conjunto de descripciones de elementos de E asociado al conjunto o dominio $\mathcal{Y}$.

- $\mathcal{D} = \mathcal{Y}$ en el caso de que X sea una variable monoevaluada
- $\mathcal{D} = \mathcal{P}(\mathcal{Y})$ en el caso de que X sea una variable simbólica multievaluada
- $\mathcal{D} = \mathcal{M}(\mathcal{Y})$ en el caso de que X sea una variable simbólica modal.

Un vector de variables simbólicas asocia a un elemento de E un vector de descripciones. En este caso

$$\mathcal{D} = \mathcal{D}_1 \times \ldots \times \mathcal{D}_p$$

Conjunto de descripciones simbólicas

$$X : E \rightarrow \mathcal{D}$$

Con $\mathcal{D}$ un conjunto de descripciones de elementos de E asociado al conjunto o dominio $\mathcal{Y}$.

- $\mathcal{D} = \mathcal{Y}$ en el caso de que X sea una variable monoevaluada
- $\mathcal{D} = \mathcal{P}(\mathcal{Y})$ en el caso de que X sea una variable simbólica multievaluada
- $\mathcal{D} = \mathcal{M}(\mathcal{Y})$ en el caso de que X sea una variable simbólica modal.

Un vector de variables simbólicas asocia a un elemento de E un vector de descripciones. En este caso

$$\mathcal{D} = \mathcal{D}_1 \times \ldots \times \mathcal{D}_p$$

Conjunto de descripciones simbólicas

$$X : E \to \mathcal{D}$$

Con $\mathcal{D}$ un conjunto de descripciones de elementos de E asociado al conjunto o dominio $\mathcal{Y}$.

- $\mathcal{D} = \mathcal{Y}$ en el caso de que X sea una variable monoevaluada
- $\mathcal{D} = \mathcal{P}(\mathcal{Y})$ en el caso de que X sea una variable simbólica multievaluada
- $\mathcal{D} = \mathcal{M}(\mathcal{Y})$ en el caso de que X sea una variable simbólica modal.

Un vector de variables simbólicas asocia a un elemento de E un vector de descripciones. En este caso

$$\mathcal{D} = \mathcal{D}_1 \times \ldots \times \mathcal{D}_p$$

Conjunto de descripciones simbólicas

$$X : E \rightarrow \mathcal{D}$$

Con $\mathcal{D}$ un conjunto de descripciones de elementos de E asociado al conjunto o dominio $\mathcal{Y}$.

- $\mathcal{D} = \mathcal{Y}$ en el caso de que X sea una variable monoevaluada
- $\mathcal{D} = \mathcal{P}(\mathcal{Y})$ en el caso de que X sea una variable simbólica multievaluada
- $\mathcal{D} = \mathcal{M}(\mathcal{Y})$ en el caso de que X sea una variable simbólica modal.

Un vector de variables simbólicas asocia a un elemento de E un vector de descripciones. En este caso

$$\mathcal{D} = \mathcal{D}_1 \times \ldots \times \mathcal{D}_p$$

Objetos Simbólicos

Un objeto simbólico se describe por variables simbólicas referidas a un conjunto E y relaciones de dominio.

Un objeto simbólico por una parte representa la intención de un concepto y por otra proporciona una herramienta para la obtener la extensión de ese concepto en un conjunto de individuos.

Sea el conjunto $E = \{e_1, ..., e_n\}$ de elementos descriptos por p simbólicas $X_1, ..., X_p$ definidas en E con dominios finitos $\mathcal{Y}_1, ..., \mathcal{Y}_p$.
y $(\mathcal{D}_1, ..., \mathcal{D}_p)$ y $\left(\mathcal{D}'_1, ..., \mathcal{D}'_p\right)$ dos colecciones de p conjuntos de descripciones asociados a los dominios $\mathcal{Y}_1, ..., \mathcal{Y}_p$.

Objetos Simbólicos

Un objeto simbólico se describe por variables simbólicas referidas a un conjunto E y relaciones de dominio.

Un objeto simbólico por una parte representa la intención de un concepto y por otra proporciona una herramienta para la obtener la extensión de ese concepto en un conjunto de individuos.

Sea el conjunto $E = \{e_1, ..., e_n\}$ de elementos descriptos por p simbólicas $X_1, ..., X_p$ definidas en E con dominios finitos $\mathcal{Y}_1, ..., \mathcal{Y}_p$.
y $(\mathcal{D}_1, ..., \mathcal{D}_p)$ y $\left(\mathcal{D}'_1, ..., \mathcal{D}'_p\right)$ dos colecciones de p conjuntos de descripciones asociados a los dominios $\mathcal{Y}_1, ..., \mathcal{Y}_p$.

Objetos Simbólicos

Un objeto simbólico se describe por variables simbólicas referidas a un conjunto E y relaciones de dominio.

Un objeto simbólico por una parte representa la intención de un concepto y por otra proporciona una herramienta para la obtener la extensión de ese concepto en un conjunto de individuos.

Sea el conjunto $E = \{e_1, ..., e_n\}$ de elementos descriptos por p simbólicas $X_1, ..., X_p$ definidas en E con dominios finitos $\mathcal{Y}_1, ..., \mathcal{Y}_p$.
y $(\mathcal{D}_1, ..., \mathcal{D}_p)$ y $\left(\mathcal{D}'_1, ..., \mathcal{D}'_p\right)$ dos colecciones de p conjuntos de descripciones asociados a los dominios $\mathcal{Y}_1, ..., \mathcal{Y}_p$.

Objetos Simbólicos

Un objeto simbólico se describe por variables simbólicas referidas a un conjunto E y relaciones de dominio.

Un objeto simbólico por una parte representa la intención de un concepto y por otra proporciona una herramienta para la obtener la extensión de ese concepto en un conjunto de individuos.

Sea el conjunto $E = \{e_1, ..., e_n\}$ de elementos descriptos por p simbólicas $X_1, ..., X_p$ definidas en E con dominios finitos $\mathcal{Y}_1, ..., \mathcal{Y}_p$.
y $(\mathcal{D}_1, ..., \mathcal{D}_p)$ y $\left(\mathcal{D}'_1, ..., \mathcal{D}'_p\right)$ dos colecciones de p conjuntos de descripciones asociados a los dominios $\mathcal{Y}_1, ..., \mathcal{Y}_p$.

Definición

Sean $\mathcal{D}$ y $\mathcal{D}'$ dos conjuntos de descripciones de clase asociados a un mismo dominio, $\mathcal{D} \times \mathcal{D}'$ su producto cartesiano, una relación de dominio $\mathcal{R}$ definida en $\mathcal{D} \times \mathcal{D}'$ es una aplicación:

$$\mathcal{R} : \mathcal{D} \times \mathcal{D}' \rightarrow \mathcal{L}$$
$$(d, d') \rightarrow \mathcal{R}\,(d, d') := [d\mathcal{R}d']$$

donde, $\mathcal{L} = \{0, 1\}$ o $\mathcal{L} = [0, 1]$.

Definición

Definición: $\mathcal{R}$ *es una relación de dominio booleana si el conjunto de comparación de descripciones es* $\mathcal{L} = \{0, 1\}$. *Dado un par de descripciones* $(d, d') \in \mathcal{D} \times \mathcal{D}'$,

- *cuando* $[d\mathcal{R}d'] = 1$ *entonces la relación entre* d *y* d' *es verdad, o* d *y* d' *se relacionan*
- *cuando* $[d\mathcal{R}d'] = 0$ *entonces la relación entre* d *y* d' *es falsa, o* d *y* d' *no se relacionan*

Definición

Sea una colección de relaciones de dominio $(\mathcal{R}_1, ..., \mathcal{R}_p)$ *, cada* $\mathcal{R}_j$ *definida en el producto cartesiano* $\mathcal{D}_j \times \mathcal{D}'_j$*.Sean* $\mathcal{D} := \mathcal{D}_1 \times ... \times \mathcal{D}_p$ *y* $\mathcal{D}' := \mathcal{D}'_1 \times ... \times \mathcal{D}'_p$ *los correspondientes productos cartesianos de los conjuntos de descripciones. La relación producto* $\mathcal{R} = \mathcal{R}_1 \times ... \times \mathcal{R}_p$ *definida en* $\mathcal{D} \times \mathcal{D}'$ *es la aplicación:*

$$\mathcal{R} : \mathcal{D} \times \mathcal{D}' \to \mathcal{L}$$

$$(d, d') \to \mathcal{R}(d, d') := [d\mathcal{R}d'] := g\left(\left\{\left[d_j \mathcal{R}_j d'_j\right], j = 1, ..., p\right\}\right) = \bigwedge_{j=1,...,p} \left[d_j \mathcal{R}_j d'_j\right]$$

Nota

La aplicación $g(.)$ es una aplicación simétrica, se denota por $\wedge$ si bien este operador no es siempre el operador conjuntivo lógico estándar.
La aplicación g, llamada aplicación de combinación de niveles de relación (o de adecuación), verifica

$$g(1, l_2, ..., l_p) = g(l_2, ..., l_p), \textit{para todo} \quad l_2, ..., l_p \in [0, 1]$$

$$g(0, l_2, ..., l_p) = 0, \textit{para todo} \quad l_2, ..., l_p \in [0, 1]$$

Definición

Un objeto simbólico de tipo evento en E es una t-upla $(a, \mathcal{R}, d)$ donde:

- *a es una función, denotada por $a = [X\mathcal{R}d]$, con X una variable simbólica con dominio $\mathcal{Y}$ definida por $X : E \rightarrow \mathcal{D}$ y $\mathcal{D}$ un conjunto de descripciones de elementos de E, asociado al conjunto $\mathcal{Y}$. La función a es:*

$$
\begin{gathered}
a : E \rightarrow \mathcal{L} \\
e \rightarrow a(e) = [X(e)\mathcal{R}d]
\end{gathered}
$$

- *$\mathcal{R}$ es una relación de dominio definida en $\mathcal{D} \times \{d\}$*
- *d es una descripción de un conjunto de descripciones asociado al conjunto $\mathcal{Y}$.*

Definición

Sea $(a, \mathcal{R}, d)$ con $a = [X\mathcal{R}d]$, un evento booleano definido en E. Se llama extensión del evento booleano a en E y se denota por $Ext_E(a)$, al subconjunto de elementos de E cuya descripción en $\mathcal{D}$ (dada por X) se relaciona con el evento a:

$$Ext_E(a) = \{e \in E : a(e) = [X(e)\mathcal{R}d] = 1\}$$

Nota

Sea $(a, \mathcal{R}, d)$, $a = [X\mathcal{R}d]$ *un evento no booleano definido en* E *y* $\delta \in [0,1]$. *Se llama extensión de nivel* δ *del evento a en* E *y se denota por* $Ext_{E,\delta}(a)$, *al subconjunto de elementos de* E *cuya descripción en* $\mathcal{D}$ *(dada por* X*) tiene un nivel de relación con el evento a igual o superior a* δ :

$$Ext_{E,\delta}(a) = \{e \in E : a(e) = [X(e)\mathcal{R}d] \geq \delta\}$$

Definición

Un objeto simbólico de tipo **aserción** *definido en E es una t-upla $(a, \mathcal{R}, d)$ donde:*

- *a es una función, denotada por $a = [X\mathcal{R}d]$ con $X = (X_1, ..., X_p)$ un vector de variables simbólicas con dominios $\mathcal{Y}_1, ..., \mathcal{Y}_p$ definido por $X : E \longrightarrow \mathcal{D}$, y $\mathcal{D} = \mathcal{D}_1 \times ... \times \mathcal{D}_p$ un conjunto de descripciones de elementos de E, $\mathcal{D}_j$ asociado al dominio $\mathcal{Y}_j$.*

$$a : E \longrightarrow \mathcal{L}$$

$$e \longrightarrow a(e) = [X(e)\mathcal{R}d] = g\left(\{[X_j(e)\mathcal{R}_j d_j], j = 1, ..., p\}\right) = \bigwedge_{j=1,...,p} [X_j(e)\mathcal{R}_j d_j]$$

- *$\mathcal{R} = \mathcal{R}_1 \times ... \times \mathcal{R}_p$ es un producto de relaciones de dominio definido en $\mathcal{D} \times \{d\}$ como $[d'\mathcal{R}d] = g\left(\left\{\left[d'_j\mathcal{R}_j d_j\right], j = 1, ..., p\right\}\right)$ para $d' = (d'_1, ..., d'_p) \in \mathcal{D}$.*

Definición

Un objeto simbólico de tipo aserción es booleano si el conjunto de comparación de descripciones es $\mathcal{L} = \{0, 1\}$ *. En este caso, dado* $e \in E$:

- *si* $a(e) = [X(e)\mathcal{R}d] = 1$ *entonces la descripción de e en* $\mathcal{D}$ *(dada por el vector* X*) se relaciona con la descripción* d*, o e se relaciona con la aserción a*
- *si* $a(e) = [X(e)\mathcal{R}d] = 0$ *entonces la descripción de e en* $\mathcal{D}$ *(dada por el vector* X*) no se relaciona con la descripción* d*, o e no se relaciona con la aserción a*

Ejemplo

$$\begin{pmatrix} id & \widetilde{sexo} & \widetilde{estado\ civil} \\ \omega_1 & femenino & casado \\ \omega_2 & femenino & soltero \\ \omega_3 & femenino & soltero \\ \omega_4 & masculino & casado \\ \omega_5 & masculino & viudo \\ \omega_6 & masculino & casado \\ \omega_7 & masculino & viudo \end{pmatrix}$$

$$a_1 = [\widetilde{sexo} \in \{femenino\}] \wedge \left[\widetilde{estado\ civil} \in \{casado, soltero\}\right]$$
$$a_2 = [\widetilde{sexo} \in \{masculino\}] \wedge \left[\widetilde{estado\ civil} \in \{casado, viudo\}\right]$$

$$a_1 : \Omega \to \{0, 1\}$$

$$\omega \to a_1(\omega) =$$
$$[\widetilde{sexo}(\omega) \in \{femenino\}] \wedge \left[\widetilde{estado\ civil}(\omega) \in \{casado, soltero\}\right]$$

Ejemplo

$$\begin{pmatrix} id & \widetilde{sexo} & \widetilde{estado\ civil} \\ \omega_1 & femenino & casado \\ \omega_2 & femenino & soltero \\ \omega_3 & femenino & soltero \\ \omega_4 & masculino & casado \\ \omega_5 & masculino & viudo \\ \omega_6 & masculino & casado \\ \omega_7 & masculino & viudo \end{pmatrix}$$

$$a_1 = [\widetilde{sexo} \in \{femenino\}] \wedge \left[\widetilde{estado\ civil} \in \{casado, soltero\}\right]$$

$$a_2 = [\widetilde{sexo} \in \{masculino\}] \wedge \left[\widetilde{estado\ civil} \in \{casado, viudo\}\right]$$

$$a_1 : \Omega \to \{0, 1\}$$

$$\omega \to a_1(\omega) =$$

$$[\widetilde{sexo}(\omega) \in \{femenino\}] \wedge \left[\widetilde{estado\ civil}(\omega) \in \{casado, soltero\}\right]$$

Ejemplo

$$\begin{pmatrix} id & \widetilde{sexo} & \widetilde{estado\ civil} \\ \omega_1 & femenino & casado \\ \omega_2 & femenino & soltero \\ \omega_3 & femenino & soltero \\ \omega_4 & masculino & casado \\ \omega_5 & masculino & viudo \\ \omega_6 & masculino & casado \\ \omega_7 & masculino & viudo \end{pmatrix}$$

$$a_1 = [\widetilde{sexo} \in \{femenino\}] \wedge \left[\widetilde{estado\ civil} \in \{casado, soltero\}\right]$$
$$a_2 = [\widetilde{sexo} \in \{masculino\}] \wedge \left[\widetilde{estado\ civil} \in \{casado, viudo\}\right]$$

$$a_1 : \Omega \rightarrow \{0, 1\}$$

$$\omega \rightarrow a_1(\omega) =$$
$$[\widetilde{sexo}(\omega) \in \{femenino\}] \wedge \left[\widetilde{estado\ civil}(\omega) \in \{casado, soltero\}\right]$$

Ejemplo

$$\begin{pmatrix} id & \widetilde{sexo} & \widetilde{estado\ civil} \\ \omega_1 & femenino & casado \\ \omega_2 & femenino & soltero \\ \omega_3 & femenino & soltero \\ \omega_4 & masculino & casado \\ \omega_5 & masculino & viudo \\ \omega_6 & masculino & casado \\ \omega_7 & masculino & viudo \end{pmatrix}$$

$$a_1 = [\widetilde{sexo} \in \{femenino\}] \wedge \left[\widetilde{estado\ civil} \in \{casado, soltero\}\right]$$
$$a_2 = [\widetilde{sexo} \in \{masculino\}] \wedge \left[\widetilde{estado\ civil} \in \{casado, viudo\}\right]$$

$$a_1 : \Omega \rightarrow \{0, 1\}$$

$$\omega \rightarrow a_1(\omega) =$$
$$[\widetilde{sexo}(\omega) \in \{femenino\}] \wedge \left[\widetilde{estado\ civil}(\omega) \in \{casado, soltero\}\right]$$

Ejemplo

Las aserciones a_1 y a_2 representan la intención de dos subconjuntos de individuos S_1 y S_2, respectivamente. Cada una de las aserciones a_1 y a_2 son también un medio de obtención de individuos que verifican dicha intención.

Las extensiones en Ω de las aserciones a_1 y a_2 son:

$$Ext_\Omega(a_1) = \{\omega \in \Omega : a_1(\omega) = 1\} = \{\omega_1, \omega_2, \omega_3\}$$

$$Ext_\Omega(a_2) = \{\omega \in \Omega : a_2(\omega) = 1\} = \{\omega_4, \omega_5, \omega_6, \omega_7\}$$

Ejemplo

Las aserciones a_1 *y* a_2 *representan la intención de dos subconjuntos de individuos* S_1 *y* S_2*, respectivamente. Cada una de las aserciones* a_1 *y* a_2 *son también un medio de obtención de individuos que verifican dicha intención.*
Las extensiones en Ω *de las aserciones* a_1 *y* a_2 *son:*

$$Ext_{\Omega}(a_1) = \{\omega \in \Omega : a_1(\omega) = 1\} = \{\omega_1, \omega_2, \omega_3\}$$

$$Ext_{\Omega}(a_2) = \{\omega \in \Omega : a_2(\omega) = 1\} = \{\omega_4, \omega_5, \omega_6, \omega_7\}$$

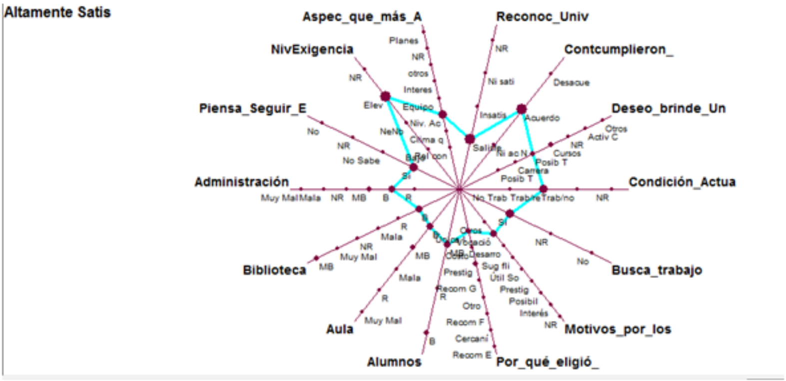
Ejemplo

Las aserciones a_1 *y* a_2 *representan la intención de dos subconjuntos de individuos* S_1 *y* S_2*, respectivamente. Cada una de las aserciones* a_1 *y* a_2 *son también un medio de obtención de individuos que verifican dicha intención.*
Las extensiones en Ω *de las aserciones* a_1 *y* a_2 *son:*

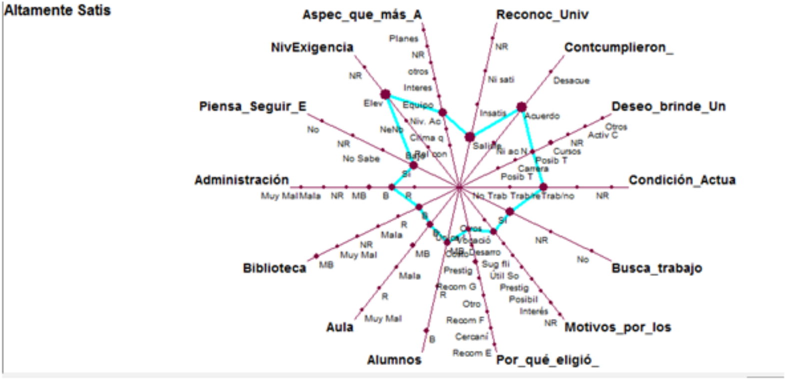
$$Ext_\Omega(a_1) = \{\omega \in \Omega : a_1(\omega) = 1\} = \{\omega_1, \omega_2, \omega_3\}$$

$$Ext_\Omega(a_2) = \{\omega \in \Omega : a_2(\omega) = 1\} = \{\omega_4, \omega_5, \omega_6, \omega_7\}$$

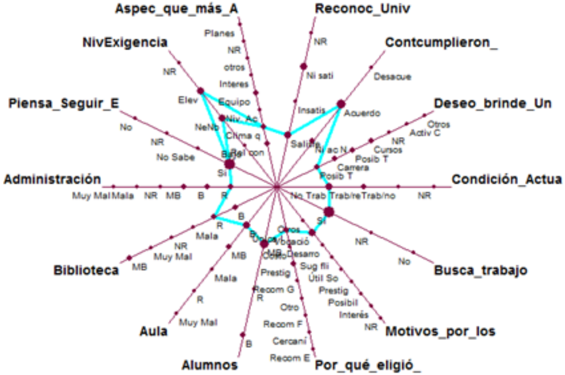
Con los datos de la encuesta Kolla al recién graduado, se han creado objetos simbólicos que representan a las tres clases de egresados 2014 de la FFHA-UNSJ que se encontraron utilizando clustering a partir de los factores principales de un análisis de correspondencias múltiples, obtenido usando el software estadístico SPAD_N.

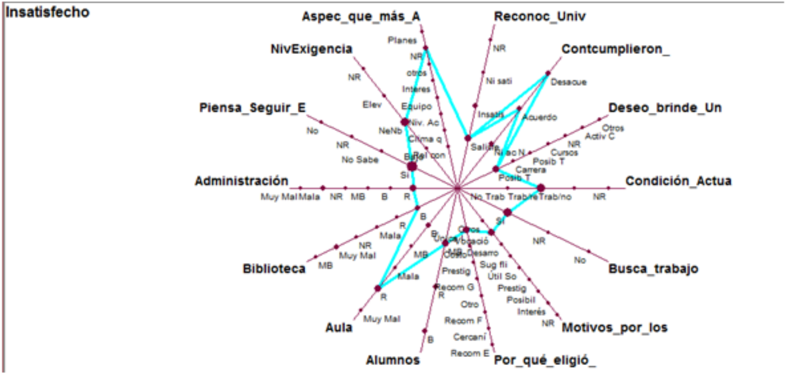


Con los datos de la encuesta Kolla al recién graduado, se han creado objetos simbólicos que representan a las tres clases de egresados 2014 de la FFHA-UNSJ que se encontraron utilizando clustering a partir de los factores principales de un análisis de correspondencias múltiples, obtenido usando el software estadístico SPAD_N.



Satisfecho


Aspec_que_más_A
Reconoc_Univ
NivExigencia
Contcumplieron_
Piensa_Seguir_E
Deseo_brinde_Un
Administración
Condición_Actua
Biblioteca
Busca_trabajo
Aula
Motivos_por_los
Alumnos
Por_qué_eligió_


Insatisfecho
Aspec_que_más_A
Reconoc_Univ
NivExigencia
Contcumplieron_
Piensa_Seguir_E
Deseo_brinde_Un
Administración
Condición_Actua
Biblioteca
Busca_trabajo
Aula
Motivos_por_los
Alumnos
Por_qué_eligió_
Planes
NR
otros
Interes
Equipo
Niv. Ac
Clima q
Rel con
Elev
NeNb
Bajo
No
No Sabe
Si
Ni sati
Desacue
Insatis
Acuerdo
Satisfe
Ni ac N
Otros
Activ C
Cursos
Posib T
Carrera
Muy Mal
Mala
MB
B
R
No Trab
Trab/no
Sug fli
Útil So
Prestig
Posibil
Interés
Vocació
Desarro
Recom G
Otro
Recom F
Cercaní
Recom E

Conclusiones

- Analizar macro-datos en lugar de micro-datos.
- Analizar datos en intervalos o valuados es distribución
- Tomar en cuenta la variabilidad en lo datos
- Se han propuesto varias metodologías que extienden el análisis clásico al análisis simbólico de datos.

Software Disponible

- SODAS: libre, http://www.info.fundp.ac.be/asso/sodaslink.htm
- Syrokko: comercial, http://syrokko.com/
- R: paquetes disponibles para datos en intervalo: RSDA ; Regression, MAINT.Data

Conclusiones

- Analizar macro-datos en lugar de micro-datos.
- Analizar datos en intervalos o valuados es distribución
- Tomar en cuenta la variabilidad en lo datos

- Se han propuesto varias metodologías que extienden el análisis clásico al análisis simbólico de datos.

Software Disponible

- SODAS: libre, http://www.info.fundp.ac.be/asso/sodaslink.htm
- Syrokko: comercial, http://syrokko.com/
- R: paquetes disponibles para datos en intervalo: RSDA ; Regression, MAINT.Data

Conclusiones

- Analizar macro-datos en lugar de micro-datos.
- Analizar datos en intervalos o valuados es distribución
- Tomar en cuenta la variabilidad en lo datos

- Se han propuesto varias metodologías que extienden el análisis clásico al análisis simbólico de datos.

Software Disponible

- SODAS: libre, http://www.info.fundp.ac.be/asso/sodaslink.htm
- Syrokko: comercial, http://syrokko.com/
- R: paquetes disponibles para datos en intervalo: RSDA ; Regression, MAINT.Data

Libros y Artículos Principales

- Bock, H.-H.; Diday, E. (2000): Analysis of Symbolic Data: Exploratory methods for extracting statistical information from complex data. Berlin-Heidelberg: Springer-Verlag.
- Billard, L., Diday, E. (2007): Symbolic Data Analysis: Conceptual Statistics and Data Mining. Wiley.
- Diday, E., Noirhomme-Fraiture, M. (2008): Symbolic Data Analysis and the SODAS Software. Wiley.
- Special Issue of Statistical Analysis and Data Mining. Vol. 4, Issue 2, April 2011. Wiley, on behalf of the American Statistical Association.
- Billard, L. and Diday, E. (2003). From the statistics of data to the statistics of knowledge: Symbolic Data Analysis. Journal of the American Statistical Association, 98 (462), pp. 470487.
- Noirhomme-Fraiture, M. and Brito, P. (2011). Far beyond the classical data models: Symbolic data analysis. Statistical Analysis and Data Mining, 4(2), 157-170.
- Brito, P. (2014). Symbolic Data Analysis: another look at the interaction of Data Mining and Statistics. WIREs Data Mining and Knowledge Discovery, Volume 4, Issue 4, July/August 2014, 281-295.

MUCHAS GRACIAS!!