FXA in Data Mining

Francisco J. Valverde-Albacete[1], Carmen Peláez-Moreno[1], Pablo Cordero-Ortega[2], Inma de las Peñas[2] and Manuel Ojeda-Aciego[2]

[2]Dep. Matemática Aplicada Universidad de Málaga

8-10/11/2019, Grazalema, Cádiz

Outline

Outline

Consideraciones Iniciales. . .

En un interesante "final remarks" de su artículo seminal sobre FCA, Wille renunció a cualquier atisbo de exhaustividad en su propuesta de reestructuración de la teoría de retículos y recomendó:

"Besides the interpretation by hierarchies of concepts, other basic interpretations of lattices should be introduced; . . . "[a]

[a] Rudolf Wille. "Restructuring lattice theory: an approach based on hierarchies of concepts". In: *Ordered sets (Banff, Alta., 1981).* Reidel, 1982, pp. 445–470.

Maslow acuño una metáfora imprescindible para el científico/ingeniero (matemáticos incluidos)

"... I suppose it is tempting, if the only tool you have is a hammer, to treat everything as if it were a nail. "[a].

[a] A H Maslow. *The Psychology of Science: A Reconnaissance.* The John Dewey Society lectureship series. Harper & Row, 1966, Ch 1.

Consideraciones Iniciales. . .

En un interesante "final remarks" de su artículo seminal sobre FCA, Wille renunció a cualquier atisbo de exhaustividad en su propuesta de reestructuración de la teoría de retículos y recomendó:

"Besides the interpretation by hierarchies of concepts, other basic interpretations of lattices should be introduced; . . . "[a]

[a] Rudolf Wille. "Restructuring lattice theory: an approach based on hierarchies of concepts". In: *Ordered sets (Banff, Alta., 1981)*. Reidel, 1982, pp. 445–470.

Maslow acuño una metáfora imprescindible para el científico/ingeniero (matemáticos incluidos)

"... I suppose it is tempting, if the only tool you have is a hammer, to treat everything as if it were a nail. "[a].

[a] A H Maslow. *The Psychology of Science: A Reconnaissance*. The John Dewey Society lectureship series. Harper & Row, 1966. Ch 1.

Consideraciones Iniciales. . .

A. Rényi es famoso por sus aforismos (a menudo atribuidos a P. Erdös, su colega y amigo.)

"Un matemático es una máquina de transformar café en teoremas."

Con esta perspectiva, debieramos considerar. . .

- Qué otra información porta un contexto formal.
- Qué pueden ser conceptualizaciones alternativas de esa información.
- ¡Esta es una consideración epistemológica fundamental para no limitar FCA!
- ¿Cómo utilizar productivamente el café que tenemos?

Consideraciones Iniciales...

A. Rényi es famoso por sus aforismos (a menudo atribuidos a P. Erdös, su colega y amigo.)

"Un matemático es una máquina de transformar café en teoremas."

Con esta perspectiva, debieramos considerar...

- Qué otra información porta un contexto formal.
- Qué pueden ser conceptualizaciones alternativas de esa información.
- ¡Esta es una consideración epistemológica fundamental para no limitar FCA!
- ¿Cómo utilizar productivamente el café que tenemos?

El Teorema Fundamental del FCA, Extendido (1/3)

Theorem

Sea el contexto formal (G, M, R) en donde,

- *G es un conjunto de objetos formales,*
- *M es un conjunto de attributos formales,*
- *$I \in 2^{g \times m}$ es una relación de incidencia.*

Entonces

(a) *La fase de análisis del contexto.*

  (i) *The polar operators $\cdot^{\uparrow} : 2^G \to 2^M$ and $\cdot^{\downarrow} : 2^M \to 2^G$.*

$$A^{\uparrow} = \{m \in M \mid \forall g \in A, gIm\} \qquad B^{\downarrow} = \{g \in G \mid \forall m \in B, gIm\}$$

form a Galois connection $(\cdot^{\uparrow}, \cdot^{\downarrow}) : 2^G \rightleftarrows 2^G$ whose formal concepts are the pairs (A, B) of closed elements such that $A^{\uparrow} = B \Leftrightarrow A = B^{\downarrow}$ whence

$$\mathfrak{B}(G, M, I) = \{(A, B) \in 2^G \times 2^M \mid A^{\uparrow} = B \Leftrightarrow A = B^{\downarrow}\}$$

El Teorema Fundamental del FCA, Extendido (2/3)

Theorem

(a) *Concepts are partially ordered with the hierarchical order as*

$$(A_1, B_1) \leq (A_2, B_2) \Leftrightarrow A_1 \subseteq A_2 \Leftrightarrow B_1 \supseteq B_2 .$$

and the set of formal concepts with the hierarchical order $\langle \mathfrak{B}(G, M, I), \leq \rangle$ is a complete lattice $\underline{\mathfrak{B}}(G, M, I)$ called the concept lattice of (G, M, I).

(b) *In $\underline{\mathfrak{B}}(G, M, I)$ infima and suprema are given by:*

$$\bigwedge_{t \in T} (A_t, B_t) = \left(\bigcap_{t \in T} A_t, \left(\bigcup_{t \in T} B_t \right)^{\uparrow} \right) \quad \bigvee_{t \in T} (A_t, B_t) = \left(\left(\bigcup_{t \in T} A_t \right)^{\downarrow}, \bigcap_{t \in T} B_t \right)$$

(c) *The mappings $\overline{\gamma} : G \to V$ and $\overline{\mu} : M \to V$*

$$g \mapsto \overline{\gamma}(g) = (\{g\}^{\uparrow\downarrow}, \{g\}^{\uparrow}) \qquad m \mapsto \overline{\mu}(m) = (\{m\}^{\downarrow}, \{m\}^{\downarrow\uparrow})$$

are such that $\overline{\gamma}(G)$ is supremum-dense in $\underline{\mathfrak{B}}(G, M, I)$, $\overline{\mu}(M)$ is

El Teorema Fundamental del FCA, Extendido (3/3)

Theorem

(b) *La fase de síntesis del contexto*

(i) *A complete lattice $\mathscr{V} = \langle V, \leq \rangle$ is isomorphic to (read can be built as) $\underline{\mathfrak{B}}(G, M, I)$ if and only if there are mappings $\overline{\gamma} : G \rightarrow V$ and $\overline{\mu} : M \rightarrow V$ such that*

- *$\overline{\gamma}(G)$ is supremum-dense in $\mathscr{V}$, $\overline{\mu}(M)$ is infimum-dense in $\mathscr{V}$, and*
- *gIm is equivalent to $\overline{\gamma}(g) \leq \overline{\mu}(m)$ for all $g \in G$ and all $m \in M$.*

(ii) *In particular, $\mathscr{V} \cong \underline{\mathfrak{B}}(V, V, \leq)$ using the assignments $G := J(\mathscr{V})$ and $M := M(\mathscr{V})$ where these are the sets of join- and meet-irreducibles, respectively, of $\mathscr{V}$.*

Una Especie de Transformada. . .

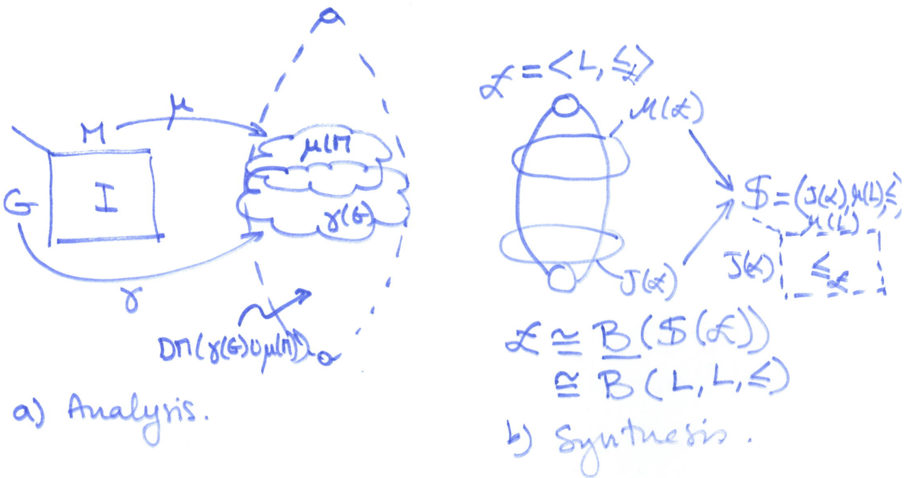


Figure: Las dos fases de FCA como una Transformada de Contextos: a) Análisis y b) Síntesis.

FCA tiene “Puntos Ciegos”

Los polares del contexto generan relaciones de equivalencia

Usando las funciones generadoras de contexto:

$$(g_1, g_2) \in \ker \overline{\gamma} \iff \overline{\gamma}(g_1) = \overline{\gamma}(g_2)$$

$$(m_1, m_2) \in \ker \overline{\mu} \iff \overline{\mu}(g_1) = \overline{\mu}(g_2)$$

El contexto purificado o reducido resintetizado por la parte de síntesis del Teorema Fundamental es

$$\mathbb{S}(\mathfrak{B}(G, M, I)) = (G/\ker \overline{\gamma}, M/\ker \overline{\mu}, I')$$

donde $([g]_{\ker \overline{\gamma}}, [m]_{\ker \overline{\mu}}) \in I' \iff gIm$.

Los elementos individuales de tales clases no están contenidos en $\mathscr{V}$ y no se pueden recuperar.

FCA tiene un “punto ciego” en

las clases de equivalencia de los polares.

FCA tiene "Puntos Ciegos"

Los polares del contexto generan relaciones de equivalencia

Usando las funciones generadoras de contexto:

$$(g_1, g_2) \in \ker \overline{\gamma} \iff \overline{\gamma}(g_1) = \overline{\gamma}(g_2)$$
$$(m_1, m_2) \in \ker \overline{\mu} \iff \overline{\mu}(g_1) = \overline{\mu}(g_2)$$

El contexto purificado o reducido resintetizado por la parte de síntesis del Teorema Fundamental es

$$\mathbb{S}(\mathfrak{B}(G, M, I)) = (G/\ker \overline{\gamma}, M/\ker \overline{\mu}, I')$$

donde $([g]_{\ker \overline{\gamma}}, [m]_{\ker \overline{\mu}}) \in I' \iff gIm$.

Los elementos individuales de tales clases no están contenidos en $\mathscr{V}$ y no se pueden recuperar.

FCA tiene un "punto ciego" en

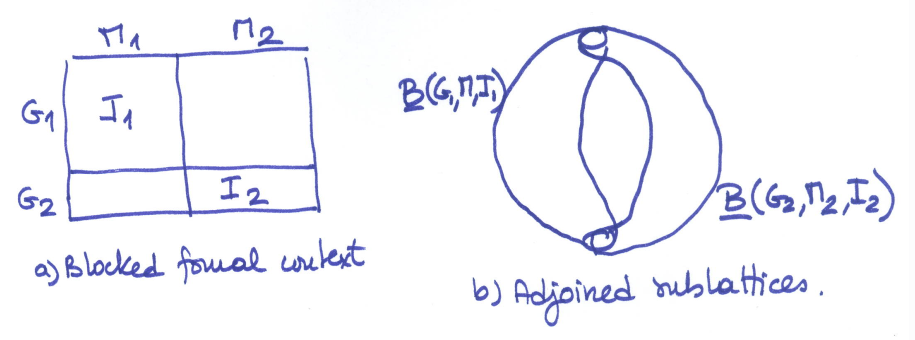
las clases de equivalencia de los polares.

FCA tiene “Espejismos” (I)

Los contextos formales diagonales por bloques son “especiales”.

- Cada bloque es independiente de los demás.
- Lo esperable es que cada bloque genere su propio retículo conceptual.

Pero FCA fuerza a que los retículos compartan $\top$ y $\bot$.

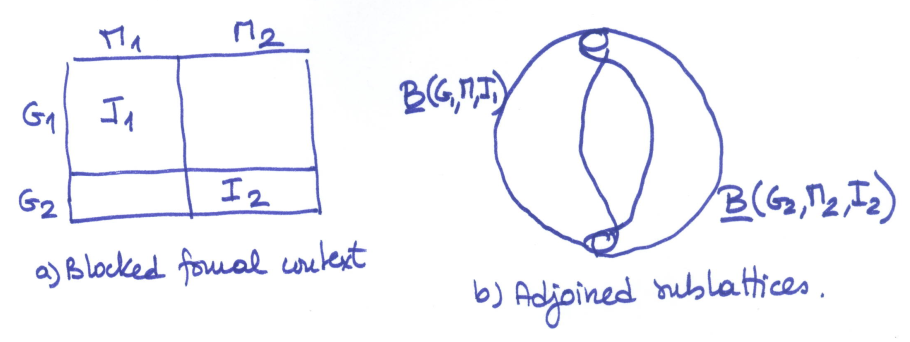


FCA tiene “Espejismos” (I)

Los contextos formales diagonales por bloques son “especiales”.

- Cada bloque es independiente de los demás.
- Lo esperable es que cada bloque genere su propio retículo conceptual.

Pero FCA fuerza a que los retículos compartan $\top$ y $\bot$.



En resumen. . .

FCA no extrae *toda* la información de un contexto formal.

- FCA ignora las equivalencias entre objetos y atributos inducidas por los polares.

FCA introduce cierta "información visual" en el retículo.

- FCA fuerza relaciones jerárquicas donde no las hay.

Necesitamos otras "lentes" para mirar en los "puntos ciegos" y deshacer los "espejismos" del FCA sobre un contexto formal.

Formal XX Analysis

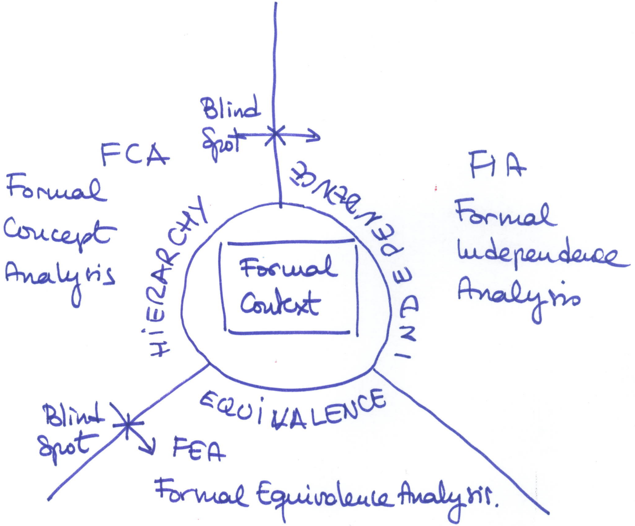


Figure: A variety of formal analysis around a formal context are conceivable: formal concept analysis around the notion of hierarchy as captured by upper

Outline

1. Motivación
2. Formal XXX Analysis
3. Aplicación: Clasificación multietiqueta

Componentes de un "Formal X Analysis"

Si analizamos el Teorema Fundamental, vemos que hay ciertos "ingredientes" fundamentales.

- Una conexión (adjunción) de Galois entre los conjuntos potencia de objetos y atributos, de forma que los pares de subconjuntos estables capturen un "sabor" del retículo.
  - FCA => Los conjuntos de cotas sup. e inf. capturan "jerarquía"
- Un "gránulo" que captura nuclearmente dicho "sabor".
  - FCA => Supremo- e ínfimo-irreducibles.
- Un par de funciones(no inyectivas!) de los objetos formales y los atributos en los gránulos:
  - FCA => $\overline{\gamma}$ y $\overline{\mu}$ generadores de conceptos.
- Un teorema de representación en función de los "gránulos" de determinada capacidad(universal, lo mejor).
  - FCA => Teorema de Dedekind-McNeille $\mathscr{L} \equiv DM(\mathscr{J}(\mathscr{L}) \cup \mathscr{M}(\mathscr{L}))$

Ejemplo: Formal Independence Analysis

Sea un orden parcial finito $\mathscr{P} = \langle P, \leq \rangle$

- Las anticadenas capturan la noción de no-relación.
- No conocemos un teorema de representación pero existe...

Theorem (Behrendt's representation theorem)

Let $\mathscr{L} = \langle L, \leq \rangle$ be a lattice. Then there exists a poset $\mathscr{P} = \langle P, \leq_P \rangle$ such that $|P| = 2|L|$, where any chain has at most 2 elements and such that $\mathscr{L}$ is isomorphic to the lattice of maximal antichains of $(P, \leq_P)$.

$$\mathscr{L} \cong MA(\mathscr{P})$$

- Esto sugiere que las anticadenas maximales (y minimales para el dual) son los “gránulos”.
- Existen conexiones de Galois entre los conjuntos de cotas superiores e inferiores y las anticadenas maximales y minimales.

El Teorema Fundamental del FIA (1/2)

Theorem (Basic theorem of formal independence analysis (Valverde, Peláez, Cabrera, Cordero, Ojeda, 2018))

(a) *The context analysis phase: Given a formal context (G,M,I),*

(i) *The context operators $\cdot^{\sim} : 2^G \to 2^M$ and $\cdot_{\sim} : 2^M \to 2^G$*

$$\alpha^{\sim} = M \setminus \uparrow \alpha \qquad \beta_{\sim} = G \setminus \downarrow \beta$$

form a right-Galois connection $(\cdot^{\sim}, \cdot_{\sim}) : (2^G, \subseteq) \rightharpoonup (2^M, \subseteq)$ whose formal tomoi are the pairs (α, β) such that $\alpha^{\sim} = \beta$ and $\alpha = \beta_{\sim}$.

(ii) *The set of formal tomoi $\mathfrak{A}(G,M,I)$ with the relation*

$$(\alpha_1, \beta_1) \leq (\alpha_2, \beta_2) \text{ iff } \alpha_1 \supseteq \alpha_2 \text{ iff } \beta_1 \subseteq \beta_2$$

is a complete lattice, which is called the tomoi lattice of (G,M,I) and denoted $\underline{\mathfrak{A}}(G,M,I)$, where infima and suprema are given by:

$$\bigwedge_{t\in T} (\alpha_t, \beta_t) = \left(\bigcup_{t\in T} \alpha_t, \left(\bigcap_{t\in T} \beta_t \right)_{\sim}^{\;\sim} \right) \quad \bigvee_{t\in T} (\alpha_t, \beta_t) = \left(\left(\bigcap_{t\in T} \alpha_t \right)^{\sim}_{\;\sim}, \bigcup_{t\in T} \beta_t \right)$$

El Teorema Fundamental del FIA (2/2)

Theorem

(a) *The mappings $\overline{\gamma}: G \to \underline{\mathfrak{A}}(G,M,I)$ and $\overline{\mu}: M \to \underline{\mathfrak{A}}(G,M,I)$*

$$g \mapsto \overline{\gamma}(g) = (\{g\}^{\sim}{}_{\sim}, \{g\}^{\sim}) \qquad m \mapsto \overline{\mu}(m) = (\{m\}_{\sim}, \{m\}_{\sim}{}^{\sim})$$

are such that $\overline{\gamma}(G)$ is infimum-dense in $\underline{\mathfrak{A}}(G,M,I)$, $\overline{\mu}(M)$ is supremum-dense in $\underline{\mathfrak{A}}(G,M,I)$.

(b) *The context synthesis phase: Given a complete lattice $\mathbb{L} = \langle L, \leq \rangle$*

(i) *$\mathbb{L}$ is isomorphic to[a] $\underline{\mathfrak{A}}(G,M,I)$ if and only if there are mappings $\overline{\gamma}: G \to L$ and $\overline{\mu}: M \to L$ such that*

- *$\overline{\gamma}(G)$ is infimum-dense in $\mathbb{L}$, $\overline{\mu}(M)$ is supremum-dense in $\mathbb{L}$, and*
- *$g\, I\, m$ is equivalent to $\overline{\gamma}(g) \not\geq \overline{\mu}(m)$ for all $g \in G$ and all $m \in M$.*

(ii) *In particular, $\mathbb{L} \cong \underline{\mathfrak{A}}(L, L, \not\geq)$ and, if L is finite, $\mathbb{L} \cong \underline{\mathfrak{A}}(M(\mathbb{L}), J(\mathbb{L}), \not\geq)$ where $M(\mathbb{L})$ and $J(\mathbb{L})$ are the sets of meet- and join-irreducibles, respectively, of $\mathbb{L}$.*

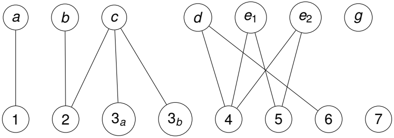
---

[a] Read *can be built as*.

Example context

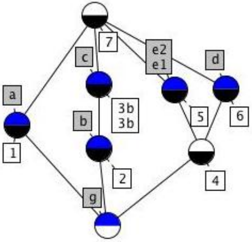
$\mathbb{K}_1$	a	b	c	d	e1	e2	g
1	×						
2		×	×				
3a			×				
3b			×				
4				×	×	×	
5					×	×	
6				×			
7							

(a) Tabular representation of $\mathbb{K}_1$

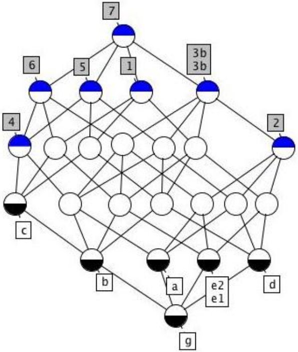


(b) Bipartite graph representation of $\mathbb{K}_1$

Figure: **Equivalent representations of an example context $\mathbb{K}_1 = (G, M, I)$.** (a)) tabular representation. (b)) bipartite graph representation.



(a) Concept lattice $\mathfrak{B}(G, M, I)$



(b) Tomoi lattice $\underline{\mathfrak{A}}(G, M, I)$

Figure: **Two different lattices for context** $\mathbb{K}_1 = (G, M, I)$ (a) lattice of formal concepts *showing* three adjoint sublattices, and (b) lattice of formal tomoi, *describing* the three adjoint sublattices.

Carrying formal analysis on partitions

We have seen that:

- There are partitions $\ker \overline{\gamma}$ and $\ker \overline{\mu}$ related to a formal context.
- The FCA procedure is blind to the classes in these partitions, that is, there is an information loss in the process of FCA.
- There is a notion of “mutual determination” of objects and attributes, whether from the qualitative point of view of Psychology and Cognition, or from the purely quantitative of Information Theory.

Define the *coarsening order*:

Let $\mathrm{Part}(G)$ denote the set of partitions over G.

$$\pi \leq \sigma \iff \forall g_1, g_2 \in G, g_1 \equiv g_2(\pi) \text{ implies that } g_1 \equiv g_2(\sigma)$$

The lattice of partitions of a set

Theorem

Let G be a set. Then $\mathrm{Part}(G) = \langle \Pi(G), \subseteq \rangle$ is a complete lattice, called the partition lattice (or equivalence lattice of G) where:

- *The bottom of $\mathrm{Part}(G)$ is $\iota_G = \{\{g\} \mid g \in G\}$ is the set of trivial blocks.*
- *The top of $\mathrm{Part}(G)$ is $\omega_G = \{G\}$.*
- *The meet of partitions $\{\pi_i \mid i \in I\}$ is defined, for all $g_1, g_2 \in G$ as:*

$$g_1 \equiv g_2(\wedge_{i \in I} \pi_i) \iff \forall i \in I, g_1 \equiv g_2(\pi_i)$$

- *The join of partitions $\{\pi_i \mid i \in I\}$ is defined, for all $g_1, g_2 \in G$ as:*

$$g \equiv d(\vee_{i \in I} \pi_i) \iff$$

there is a natural number n, a subset $J = \{i_0, \ldots, i_n\} \subseteq I$, and $g_0, \ldots, g_{n+1} \in G$ such that $g = g_0, \ldots, d = g_{n+1}$ and $g_j = g_{j+1}(\pi_{i_j})$, for $0 \leq j \leq n$

More partitions

Theorem (Continued)

- *The atoms of* $\mathrm{Part}(G)$ *are the partitions with exactly one non-trivial block and this block has two elements.*
- *The co-atoms of* $\mathrm{Part}(G)$ *are the partitions with exactly two blocks.*
- *The covering relation in* $\mathrm{Part}(G)$ *holds* $\pi \prec \sigma$ *iff* σ *is the result of replacing two blocks of* π *by their union.*

Proposition

Let G be a set and consider $\pi \in \Pi(G)$*. Then,*

- $\uparrow \pi \subseteq \mathrm{Part}(G)$ *is isomorphic to the partition lattice of the set* π*,* $\uparrow \pi \cong \mathrm{Part}(\pi)$*.*
- $\downarrow \pi \subseteq \mathrm{Part}(G)$ *is isomorphic to direct product of* $\mathrm{Part}(X)$ *where* X *ranges over the non-trivial blocks of* π*,* $\downarrow \pi \cong \bigtimes_{X \text{ non trivial in } \pi} \mathrm{Part}(X)$*.*

La “bombilla” de las particiones

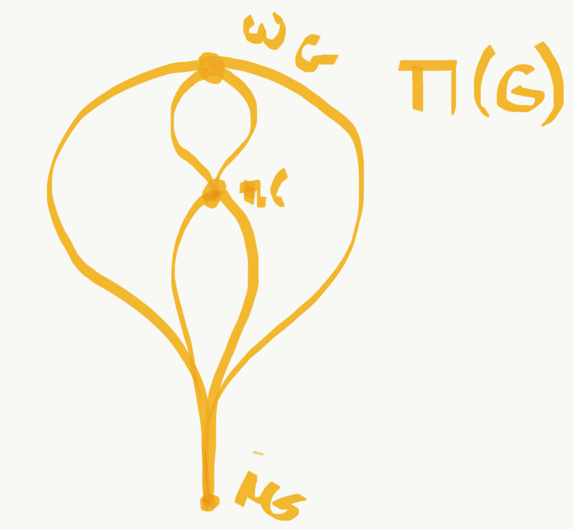


Figure: Sketch of the lattice of partitions with the filter and ideal of $\pi \in \mathrm{Part}(G)$ drawn.

El Teorema Fundamental del FEA (1/2)

Theorem (Basic theorem of formal equivalence analysis (Valverde, Peláez, Cordero, Ojeda, 2019))

(a) *The context analysis phase: Given a formal context* (G, M, I)*,*

(i) *The operators form a left adjunction whose formal equivalences or partitions are the pairs* (π, σ) *such that* $\pi^{\exists}_{\sqcap} = \sigma \iff \sigma^{\forall}_{\sqcap} = \pi$

(ii) *The set of formal partitions* $\mathfrak{P}(\mathbb{K})$ *with the relation* $(\pi_1, \sigma_1) \leq (\pi_2, \sigma_2)$ *iff* $\pi_1 \leq \pi_2$ *iff* $\sigma_1 \leq \sigma_2$ *is a complete lattice, which is called the partition lattice of* $\mathbb{K}$ *and denoted* $\underline{\mathfrak{P}}(\mathbb{K})$*, with infima and suprema given by:*

$$\bigwedge_{t \in T} (\pi_t, \sigma_t) = \left(\bigwedge_{t \in T} \pi_t, \left[\left(\bigwedge_{t \in T} \sigma_t \right)^{\forall}_{\sqcap} \right]^{\exists}_{\sqcap} \right)$$

$$\bigvee_{t \in T} (\pi_t, \sigma_t) = \left(\left[\left(\bigvee_{t \in T} \pi_t \right)^{\exists}_{\sqcap} \right]^{\forall}_{\sqcap}, \bigvee_{t \in T} \sigma_t \right)$$

El Teorema Fundamental del FEA (2/2)

Theorem

(a) *The mappings in* (**??**) $\overline{\gamma}_\Pi(\cdot) : G \to \mathfrak{P}(\mathbb{K})$ *and* $\overline{\mu}_\Pi(\cdot) : M \to \mathfrak{P}(\mathbb{K})$ *are such that* $\overline{\gamma}_\Pi(G)$ *is* $\vee$*-dense in* $\mathfrak{P}(\mathbb{K})$, $\overline{\mu}_\Pi(M)$ *is* $\wedge$*-dense in* $\mathfrak{P}(\mathbb{K})$.

(b) *The context synthesis phase: Given a complete lattice* $\mathbb{L} = \langle L, \leq \rangle$ *Given a complete lattice* $\mathbb{L} = \langle L, \leq \rangle$

(i) $\mathbb{L}$ *is isomorphic to*[a] $\mathfrak{P}(G, M, I)$ *if and only if there are mappings* $\overline{\gamma} : G \to L$ *and* $\overline{\mu} : M \to L$ *such that*

- $\overline{\gamma}(G)$ *is* $\vee$*-dense in* $\mathbb{L}$, $\overline{\mu}(M)$ *is* $\wedge$*-dense in* $\mathbb{L}$, *and*
- $g\, I\, m$ *is equivalent to* $\overline{\gamma}(g) \not\geq \overline{\mu}(m)$ *for all* $g \in G$ *and all* $m \in M$.

(ii) *In particular,* $\mathbb{L} \cong \mathfrak{P}(L, L, \not\geq)$ *and, if* L *is finite,* $\mathbb{L} \cong \mathfrak{P}(J(\mathbb{L}), M(\mathbb{L}), \not\geq)$ *where* $M(\mathbb{L})$ *and* $J(\mathbb{L})$ *are the sets of meet- and join-irreducibles, respectively, of* $\mathbb{L}$.

---

[a] Read *can be built as*.

Example context

$\mathbb{K}_0$	a	b	c	d
1	×	×	×	×
2			×	×
3			×	×
4	×	×		
5	×	×		
6	×	×		

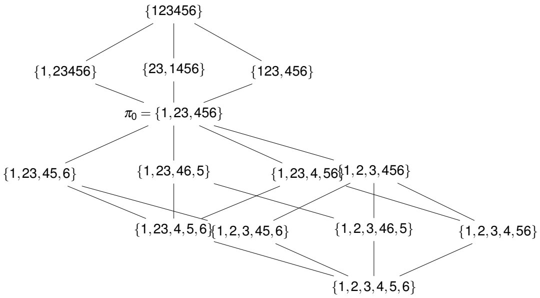
(a) Context $\mathbb{K}_0$

$\mathbb{K}_0^0$	a	c
1	×	×
2		×
4	×	

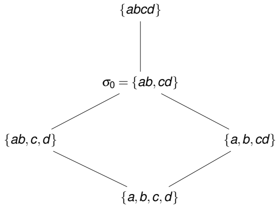
(b) Context $\mathbb{K}_0^0$

Figure: Context $\mathbb{K}_0$, its clarified context $\mathbb{K}_0^0$

The ambient partition lattices



(a) Neighbourhood of $\pi_0 = \{1, 23, 456\}$.



(b) Neighbourhood of $\sigma_0 = \{ab, cd\}$.

The FEA lattice

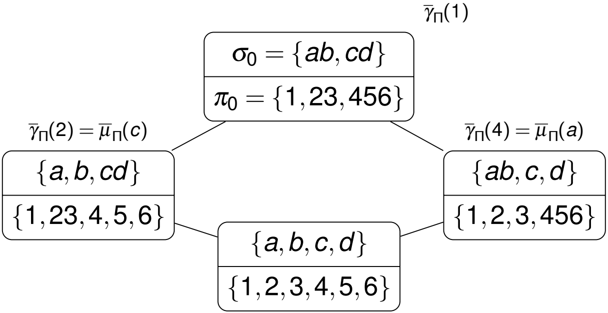


Figure: Partition lattice $\mathfrak{P}(\mathbb{K}_0)$.

Outline

Multi-Label Classification (MLC)

- Let $K \neq \varnothing$ be a set of *labels* and $\overline{K} = 2^L$ be a space of *labelsets*.
- Let $\overline{X} = \mathbb{R}^{mx}$ be a space of *observations*.
- Let $\{(\vec{k}^i, \vec{x}^i)\}_{i=1}^N$ be a collection of labelsets $\vec{k}^i$ assigned to observation vectors $\vec{x}^i$, that we call the *task dataset.*

The Multi-Label Classification task

- Divide the task datasets in two non-overlapping subsets
  - The *training dataset*, $\{(\vec{k}_T^i, \vec{x}_T^i)\}_{i=1}^{N_T}$
  - The *test dataset*, $\{(\vec{k}_E^i, \vec{x}_E^i)\}_{i=1}^{N_E}$

  So $N = N_T + N_E$
- Use the training dataset to build a model: $\hat{h} : \mathbb{R}^{mx} \to 2^L$
- Predict the outputs $\hat{k}_E^i = \hat{h}(\vec{x}_E^i)$ on the testing observation vectors.
- Measure the quality of approximating $\vec{k}_E^i$ by $\hat{k}_E^i$

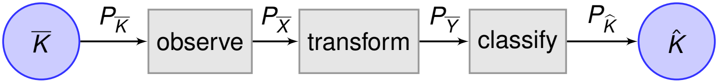
Some datasets...

Name	$N_E + N_T$	M	L	distinct
emotions	391+202	72	6	27
scene	1 211+1 196	294	6	15
yeast	1 500+917	103	14	198

Figure: A selection of multi-label classificaton databases (data from [**tso:kat:vla:10**]). “distinct” refers to distinct label sets ocurring in the training decision data.

Multi-label Classification as an Information Channel

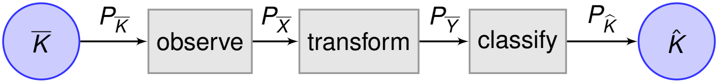
- Supervised Classification as an Information Channel



- It is possible to *assess* this system by entropic means. We have
  - modelling tools, the positive (dual pairs of) semifields (cfr. keynote from ESCIM'17, Entropy'19)
  - "prescription" tools, the balance equations and,
  - data exploratory tools, the entropy triangles and diagrams (PRL'11, ESWA'17,Entropy'18).
- Each tool provides different insight into the technologies implementing each subsystem.
- We want to add FXA tools to these!

Multi-label Classification as an Information Channel

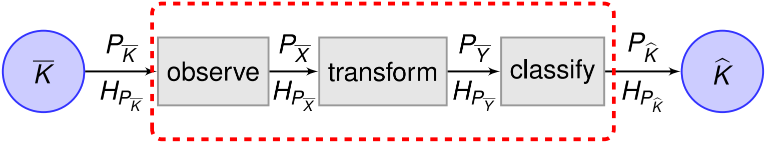
- Supervised Classification as an Information Channel



- It is possible to *assess* this system by entropic means. We have
  - modelling tools, the positive (dual pairs of) semifields (cfr. keynote from ESCIM'17, Entropy'19)
  - "prescription" tools, the balance equations and,
  - data exploratory tools, the entropy triangles and diagrams (PRL'11, ESWA'17,Entropy'18).
- Each tool provides different insight into the technologies implementing each subsystem.
- We want to add FXA tools to these!

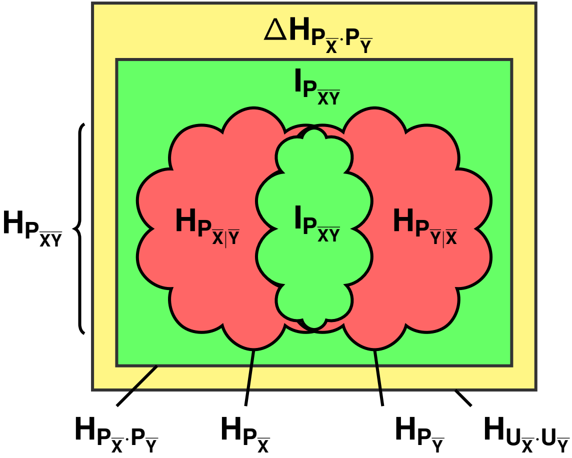
Sidetrack: End-to-End Evaluation of $P_{K\widehat{K}}$

- We consider classification as the transmission of $\overline{K}$.
- For end-to-end evaluation we ignore all details within the outer hatched block:



- The whole chain is considered here as a multiple-input multiple-output (MIMO) block with joint distribution $(\overline{K}, \hat{K}) \sim P_{\overline{K}\hat{K}}$

Our method is based in an exhaustive analysis of the information in $P_{\overline{XY}}$.



Context Modelling of an MLC task

We concentrate on the labelling part

- Call G the set of indices on observation pairs. We keep L as the name of the attribute set.
- Call $\mathbb{K} = (G, L, I)$ where $gIm \iff$ "observation g has label m"

Define training and testing parts, where G_T the index set on the training and G_E the index set on the test observation pairs.

- Call $\mathbb{K}_T = (G_T, L, I \cap G_T \times L)$ and Similarly define $\mathbb{K}_E = (G_E, L, \cap G_E \times L)$
- Clearly $G = G_T \cup G_E$, and for good statistic practice we need $G_T \cap G_E = \varnothing$ so

$$\mathbb{K} \cong \mathbb{K}_T / \mathbb{K}_E \quad (1)$$

  This is the subposition of contexts.
- In this situation, FIA is indicated for task simplification.

MLC Stratified N-fold validation is “solved” by FXA

Statistical validation of models

- This is a process of assessing the quality of statistical models issued from data.
- N-fold validation is quite used:
  - Divide *all* data in *n* parts or *folds*
  - Use $n-1$ folds as the training part and 1 fold as the test part.
  - Permute to carry out a total of *n* models and obtain assessment figures.
  - Assess the distribution of assessment figures.
- *Stratified* n-fold validation is very problematic for MLC

FXA can provide leverage to solve this:

- The Train-Test duality supposes (for instance) that $\mathfrak{B}\mathbb{K}_T \cong \mathfrak{B}\mathbb{K}_E$
- The solution is to carry out stratification modulo $\ker \tilde{\gamma}(\mathbb{K})$

Conclusiones

- Some MLC concepts can be naturally captured by FXA abstractions.
- It would seem that FXA can guide improve the modelling in MLC label space.
- To do likewise with observation space, we would need *numeric* FXA.

¡Gracias!
¿Alguna pregunta?
¿Alguna sugerencia?

CLA 2020

- The 15th International Conference on Concept Lattices and Their Applications
- Tallinn, Estonia
- Fechas: June 29–July 1, 202
- Abstract deadline: January 26, 2020
- Paper deadline: February 2, 2020

¡Enviadnos contribuciones!