Del Análisis de Conceptos Formales al 'co-clustering' idempotente

Francisco J. Valverde-Albacete

Dep. Lenguajes y Sistemas Informáticos
NLP & IR group, UNED, Spain

02/04/2013, Seminario MAVIR, Madrid, Spain

Outline

Biclustering, coclustering: a definition

Given:

- a set of samples (or objects or observations, etc.) G, with $|G| = g$,
- a set of features (or attributes, etc.) M, with $|M| = m$, and
- a data matrix $R \in \mathcal{K}^{g \times m}$, where $\mathcal{K}$ is generally any non-negative section of a field, say $\mathbb{R}_0^+$,

Direct clustering[Hartigan, 1972]: generate

- permutations for rows I and columns J ...
- so that $R(I, J)$ is block diagonal.

More generally[Mirkin, 1996], generate:

- biclusters, that is pairs (A, B) of sets of samples $A \subseteq G$ and features $B \subseteq M$...
- that are naturally related to each other.

Biclustering definition (II)

Different models of what a matrix is generate different concepts of "natural relations" and algorithms

- As a contingency matrix, find a non-negative factorization minimizing the reconstruction loss.
  - Iterative (direct clustering) techniques
  - Non-negative matrix factorization techniques
- As bipartite (weighted) graph, maximize/minimize measure on a cut
  - Graph-partitioning techniques
  - Spectral coclustering techniques
- As a product of RV's, minimize loss of mutual information in coclustering taken as a compression of the joint distribution.
  - Information-theoretic techniques

Outline

A model for batch ad-hoc tasks [Fuhr, 1992]

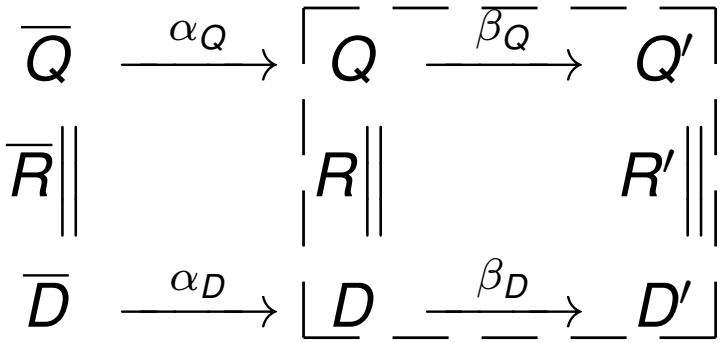


Figure: An adaptation of the conceptual model of Fuhr.

Given D, Q and R, the ideal IR system is $S_{D,Q}(R) =< \varrho_R > \ldots$

- with a relevance function

$$\varrho_R : Q \to 2^D \tag{1}$$
$$q_i \mapsto \varrho_R(q_i) = \{d_j \in D | d_j R q_i\} \ .$$

An IR model solving the batch ad-hoc task

- Given
  - a collection, $D_T \subseteq D$,
  - a set of topics, $Q_T \subseteq Q$, and
  - a set of relevance judgments, $R_T \subseteq D_T \times Q_T$,
- the implemented IR system $S_{D,Q}(\hat{R}) =< \varrho_{\hat{R}} >$ is what we can *actually* build, with approximated relevance $\hat{R} \neq R$
- using a retrieval function:

$$\varrho_{\hat{R}} : Q \to 2^D$$
$$q_i \mapsto \varrho_{\hat{R}}(q_i) = \{d_j \in D | d_j \hat{R} q_i\} \ .$$

- for each query $q \in Q$ we have precision $P_{\hat{R}}$ and recall $R_{\hat{R}}$

$$P_{\hat{R}}(q) = \frac{|\varrho_R(q) \cap \varrho_{\hat{R}}(q)|}{|\varrho_{\hat{R}}(q)|} \qquad R_{\hat{R}}(q) = \frac{|\varrho_R(q) \cap \varrho_{\hat{R}}(q)|}{|\varrho_R(q)|} \ .$$

Outline

Biclusters appear naturally for relevance relations...

Consider a set of queries $B \in 2^Q$:

- It is natural to think of a set of documents *relevant to all queries*:

$$B_R = \{d \in D \mid \forall q \in B, dRq\}$$

Dually, consider a set of documents $A \in 2^D$:

- And the set of queries *for which all documents are relevant*

$$A^R = \{q \in Q \mid \forall d \in A, dRq\}$$

Clearly the following is a bicluster

$$(A, B) \quad \text{such that} \quad A^R = B \wedge B_R = A$$

Q: What is the organization of D and Q implied by this coclustering?

The affordances of Formal Concept Analysis

Affordance 1

FCA implements the (conjunctive) Boolean model of IR[Godin et al., 1986, Valverde-Albacete, 2006].

- There exists a set of keywords T (after normalization, stoplisting, stemming)
- Queries are represented as sets of keywords $Q' \equiv 2^T$
- Documents are represented as set of keywords $D' \equiv 2^T$
- Retrieval (estimated relevance) $\hat{R}'$ is modelled as inclusion

$$d' \hat{R}' q' \Leftrightarrow q' \subseteq d'$$

- The retrieval function is the query polar,

$$\varrho_{\hat{R}'}(q') = q'_{\hat{R}'}$$

Outline

Formal contexts and their polars

A *Formal Context* (D, Q, R) is a triple of:

- A set of *objects* D
- A set of *attributes* Q
- A (boolean) *incidence* relation $R \in 2^{D \times Q}$

$$dRq \Leftrightarrow \quad \text{"object } d \text{ has attribute } q\text{"}$$

The polars of the formal context:

Given (D, Q, R) and subsets of objects A and attributes B

$$\varphi(\cdot) : 2^D \to 2^Q \qquad \psi(\cdot) : 2^Q \to 2^D$$

$$\varphi(A) = A^R \qquad \psi(B) = B_R$$

$$= \{q \in Q | \forall d \in A, dRq\} \qquad = \{d \in D | \forall q \in B, dRq\}$$

Formal contexts and polars (II)

The polars form an (antitone) Galois connection $(\varphi, \psi) : 2^D \leftharpoonup 2^Q$

$$\varphi(A) \geq_Q B \Leftrightarrow \psi(B) \geq_D A \tag{2}$$

The closures of the polars: monotone, expansive and idempotent

$$\gamma_D = \psi \circ \varphi \qquad \gamma_Q = \varphi \circ \psi$$

$$A_1 \leq A_2 \Rightarrow \gamma_D(A_1) \leq \gamma_D(A_2) \qquad B_1 \leq B_2 \Rightarrow \gamma_Q(B_1) \leq \gamma_Q(B_2)$$

$$\gamma_D(A) \geq A \qquad \gamma_Q(B) \geq B$$

$$\gamma_D(\gamma_D(A)) = \gamma_D(A) \qquad \gamma_Q(\gamma_Q(B)) = \gamma_Q(B)$$

Outline

Concepts

The set of concepts

$$(A, B) \in \mathfrak{B}(D, Q, R) \Leftrightarrow \varphi(A) = B \Leftrightarrow A = \psi(B)$$

Extents and intents: let $c = (A, B) \in \mathfrak{B}(G, M, I)$

$$\text{ext}(\cdot) : \mathfrak{B}(D, Q, R) \to \mathfrak{B}(D, Q, R) \qquad \text{int}(\cdot) : \mathfrak{B}(D, Q, R) \to \mathfrak{B}(D, Q, R)$$
$$c = (A, B) \mapsto \text{ext}(c) = A \qquad c = (A, B) \mapsto \text{int}(c) = B$$

The concept order $\underline{\mathfrak{B}}(D, Q, R) = \langle \mathfrak{B}(D, Q, R), \leq \rangle$

$$(A_1, B_1) \leq (A_2, B_2) \Leftrightarrow A_1 \leq A_2 \Leftrightarrow B_1 \geq B_2$$

The fundamental theorem of Formal Concept Analysis

The Concept Lattice $\underline{\mathfrak{B}}(D, Q, R)$ is a complete lattice in which infima and suprema are given by:

$$\bigwedge_{i\in I}(A_i, B_i) = \left(\bigcap_{i\in I} A_i, \left[\bigcup_{i\in I} B_i\right]^{R}_{R}\right) \quad \bigvee_{i\in I}(A_i, B_i) = \left(\left[\bigcup_{i\in I} A_i\right]^{R}_{R}, \bigcap_{i\in I} B_i,\right)$$

A complete latttice $\mathcal{V}$ is isomorphic to $\underline{\mathfrak{B}}(D, Q, R)$ if and only if there are mappings

$$\tilde{\gamma} : D \to V \qquad \tilde{\mu} : Q \to V$$

$$\tilde{\gamma}(D) \supseteq \mathcal{J}(\mathcal{V}) \qquad \tilde{\mu}(Q) \supseteq \mathcal{M}(\mathcal{V})$$

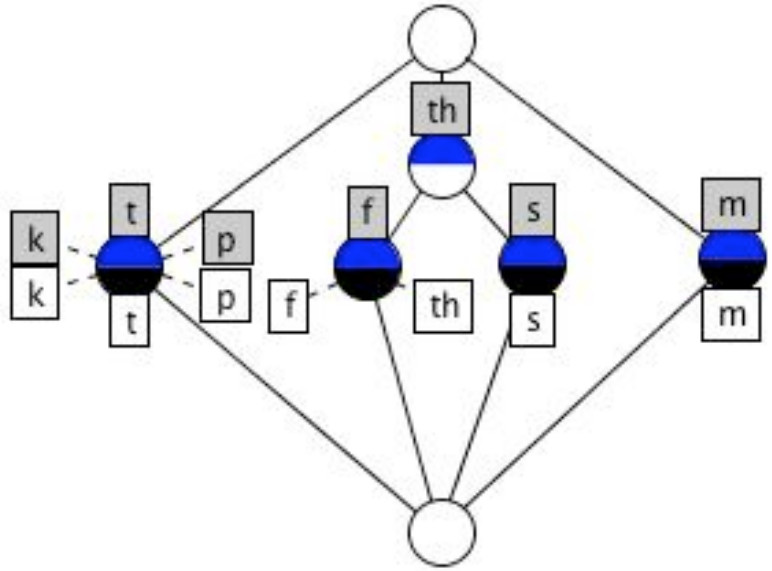
such that, $dRq \Leftrightarrow \tilde{\gamma}(d) \leq \tilde{\mu}(q)$

In particular $\mathcal{V} \cong \underline{\mathfrak{B}}(V, V, \leq)$.

A logarithmic connection...

By the fundamental theorem incidence and Concept Lattice are interchangeable:

	/p/	/m/	/t/	/f/	/th/	/k/	/s/
/p/	×		×			×	
/m/		×					
/t/	×		×			×	
/f/				×	×		
/th/				×	×		
/k/	×		×			×	
/s/					×		×



They are a pair of analysis and synthesis equations! Metaphor:

- The concept lattice is the *exponential* of the formal context.
- The formal concept is the *logarithm* of the concept lattice.

Outline

Examples: Confusion matrices of multiclass classifiers

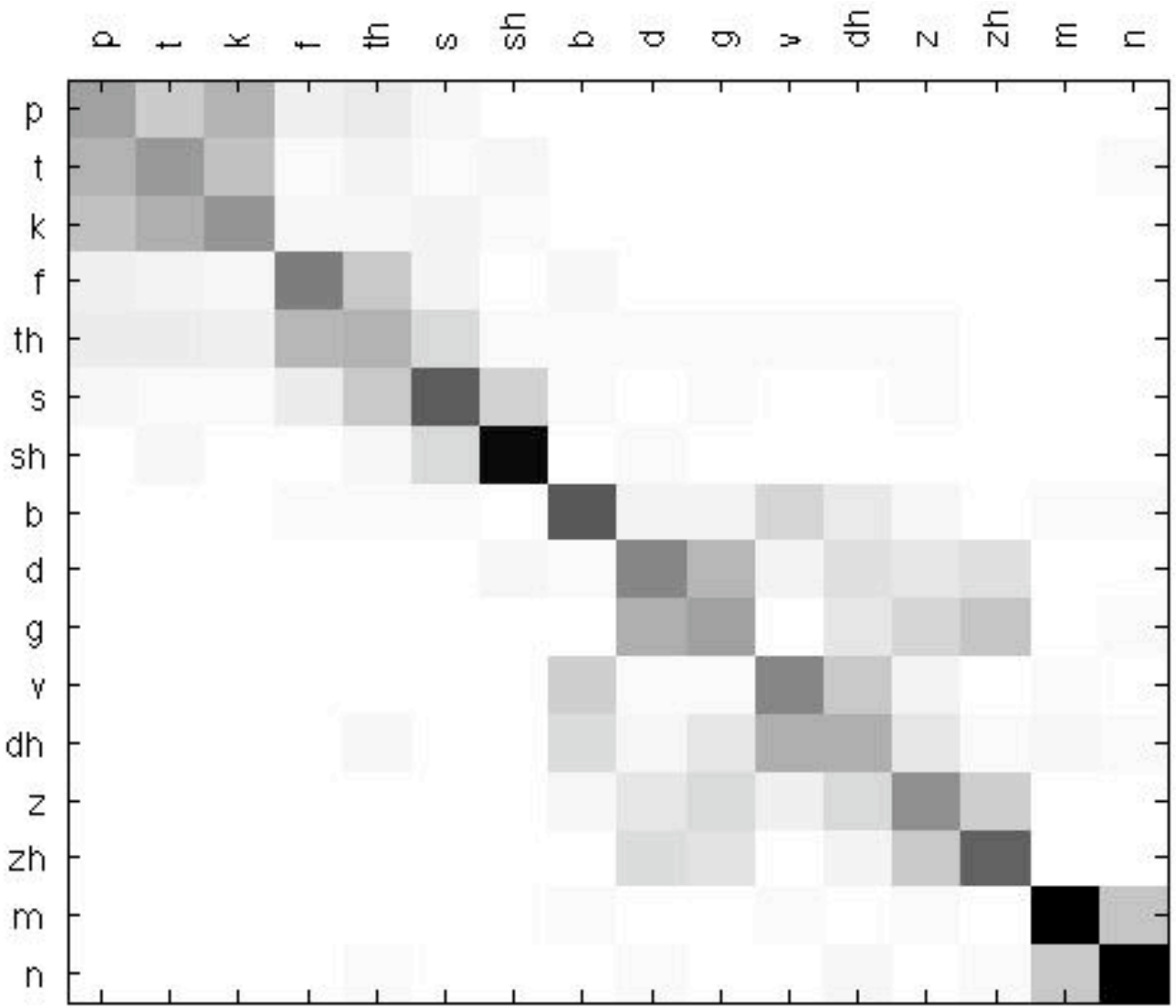
- The data matrix is in a semiring $N_{DQ} \in \mathbb{N}^{D \times Q}$

N_{DQ}	p	m	t	f	th	k	s
p	150	0	38	7	13	88	0
m	0	201	0	0	0	0	0
t	30	0	193	1	0	28	0
f	4	1	3	199	46	5	4
th	11	0	6	85	114	4	10
k	86	0	45	4	1	138	0
s	0	0	2	5	38	1	170

Figure: N_{DQ} at $SNR = 0$ dB

- Notice: no symmetry, certain sparsity.

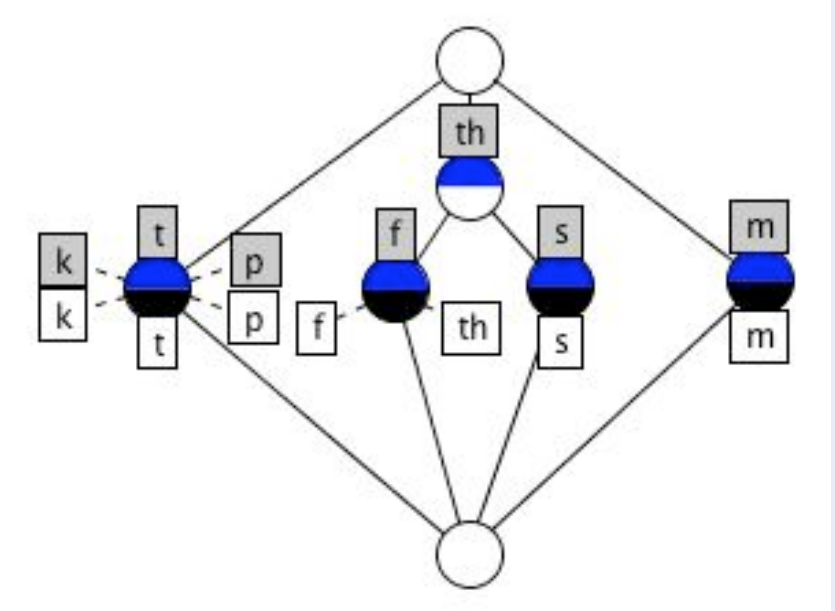
Usual representation: heatmaps



Boolean confusion matrix and lattice of confusions

Boolean confusion matrices can be subjected to FCA by simply thresholding counts:

	/p/	/m/	/t/	/f/	/th/	/k/	/s/
/p/	×		×			×	
/m/		×					
/t/	×		×			×	
/f/				×	×		
/th/				×	×		
/k/	×		×			×	
/s/					×		×



Confusion lattices represent *some* information about CM:

- Stimuli in white boxes; percepts in grey.
- The strength of the confusion is not clear.

Outline

Data preparation

We substitute $N_{DQ} \in \mathbb{N}^{D \times Q}$ by $MI_{DQ} \in \overline{\mathbb{R}}_{\min,+}^{D \times Q}$ by computing the **point-wise mutual information** of the count matrix N_{DQ}:

- The MLE of the joint probability is

$$\hat{P}_{DQ}(a_i, b_j) = \frac{n_{ij}}{\sum_{ij} n_{ij}} \ ,$$

- with marginals,

$$\hat{P}_D(a_i) = \sum_i n_{ij}/N \qquad \hat{P}_Q(b_j) = \sum_j n_{ij}/N \ .$$

- Then the mutual information matrix becomes,

$$\mathrm{MI}_{DQ}(a_i, b_j) = \log \left(\frac{\hat{P}_{DQ}(a_i, b_j)}{\hat{P}_D(a_i) \cdot \hat{P}_Q(b_j)} \right) \ .$$

Generalized Formal Concept Analysis

- The entries are now in the min-plus semiring: $MI \in \overline{\mathbb{R}}_{\min,+}^{D\times Q}$

R	p	m	t	f	th	k	s
p	2.851	$-\infty$	0.824	-1.717	-0.305	2.155	$-\infty$
m	$-\infty$	4.202	$-\infty$	$-\infty$	$-\infty$	$-\infty$	$-\infty$
t	0.761	$-\infty$	3.401	-4.292	$-\infty$	0.735	$-\infty$
f	-2.213	-3.793	-2.674	3.277	1.683	-1.817	-1.626
th	-0.567	$-\infty$	-1.487	2.236	3.179	-1.953	-0.117
k	2.149	$-\infty$	1.169	-2.424	-3.904	2.905	$-\infty$
s	$-\infty$	$-\infty$	-3.047	-1.826	1.619	-3.928	3.995

Figure: (pointwise) mutual information from N_{DQ}

Interpretations of $MI(i,j) = \lambda$

- "stimulus i is confused with percept j in degree λ"
- "percept j is taken for stimulus i to degree λ".

Generalized Formal Concept Analysis (cont) [Valverde-Albacete and Peláez-Moreno, 2011]

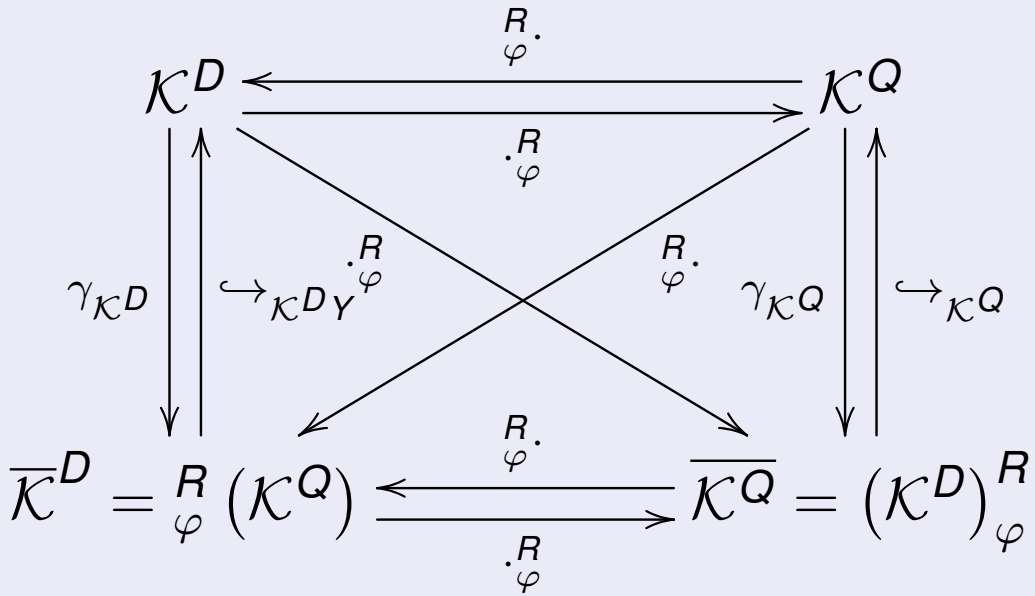
A $\mathcal{K}$-valued formal context is a triple $(D, Q, R)_{\mathcal{K}}$ with:

- $\mathcal{K}$, a complete, reflexive idempotent semifield
- two finite set of objects D and attributes Q,
- a $\mathcal{K}$-valued incidence between them, $R \in \mathcal{K}^{D \times Q}$, where $R(d, q) = \lambda$ reads as:
  - "object d has attribute q in degree λ" or
  - "attribute q is manifested in object d to degree λ",

φ-polars

Consider $(D, Q, R)_{\mathcal{K}}$, an invertible $\varphi \in K$ and the bracket $\langle y \mid x \rangle = y^{\mathsf{T}} \mathbin{\dot{\otimes}} R \mathbin{\dot{\otimes}} x$. Then the φ-polars are the maps of the GC $\left((\cdot)^{R}_{\varphi}, {}^{R}_{\varphi}(\cdot) \right) : \mathcal{K}^{D} \leftrightarrows \mathcal{K}^{Q}$:

$$(y)^{R}_{\varphi} = (y^{\mathsf{T}} \mathbin{\dot{\otimes}} R) \mathbin{\dot{\setminus}} \varphi \qquad {}^{R}_{\varphi}(x) = \varphi \mathbin{\dot{/}} (R \mathbin{\dot{\otimes}} x)$$



Formal φ-Concepts

A (formal) φ-concept of the formal context $(G, M, R)_{\mathcal{K}}$ is a pair $(a, b) \in \mathcal{Y} \times \mathcal{X}$ such that $(a)^R_\varphi = b$ and $^R_\varphi(b) = a$. We call:

- a the φ-*extent* and
- b the φ-*intent* of the concept (a, b), and
- φ its *(minimum) degree of existence*.

φ-concepts are pairs $(A, B)_\varphi$ with similar properties to those of standard Formal Concept Analysis.

$\varphi \in \mathbb{R}$ describes a minimum degree of existence required for pairs $(A, B) \in \mathbb{R}^D_{\text{min},+} \times \mathbb{R}^Q_{\text{min},+}$ to be considered as members of the φ-lattice $\underline{\mathfrak{B}}^\varphi(D, Q, MI_{DQ})_{\mathbb{R}_{\text{min},+}}$.

Basic theorem of $\mathcal{K}$-valued Formal Concept Analysis, finite version, 1^{st} half

The hierarchical order. If (a_1, b_1) (a_2, b_2) are φ-concepts,

$$(a_1, b_1) \leq (a_2, b_2) \iff a_1 \leq_{\mathcal{K}^D} a_2 \iff b_1 \leq^{op}_{\mathcal{K}^Q} b_2$$

Given a reflexive, idempotent semiring $(\mathcal{K}, \varphi)$, the φ-concept lattice $\underline{\mathfrak{B}}^{\varphi}(D, Q, R)_{\mathcal{K}}$ of a $\mathcal{K}$-valued formal context $(D, Q, R)_{\mathcal{K}}$ is a (finite, complete) lattice in which infimum and supremum are given by:

$$\bigwedge_{t \in T} (a_t, b_t) = \left(\sum\nolimits^{\bullet}_{t \in T} a_t, \left(\overset{R}{\varphi} \left(\sum\nolimits_{t \in T}^{\bullet} b_t \right) \right)^{R}_{\varphi} \right)$$

$$\bigvee_{t \in T} (a_t, b_t) = \left(\overset{R}{\varphi} \left(\left(\sum\nolimits_{t \in T}^{\bullet} a_t \right)^{R}_{\varphi} \right), \sum\nolimits^{\bullet}_{t \in T} b_t \right)$$

Challenges

Theoretical

- The relationship with the VSM in NLP and IR is very evident.
  - tfidf is related to mutual information (Roelleke, 2008)
  - (k)FCA is a VSM in a different algebraic setting.
- The entailments, very enticing:
  - There is a concept lattice structure underlying the VSM.
  - There is an actual topology of information that is 'finer' than the discrete topology.
  - kFCA actually shows how IR and IF **are** two sides of the same coin.
- The development of idempotent semiring algebra is way behind that of normal algebra (e.g. no known SVD, so idempotent LSI is unavailable).

Challenges

Practical

- The complexity of CL building algorithms is not good: $O(DQK)$ where K is the number of concepts in the lattice.
  - But Big Data techniques may be of great help.
- Most toolkits deal with the dense context case, which for us is less interesting.
- The theory is agnostic with respect to the interpretations of D and Q. This is a mixed blessing.

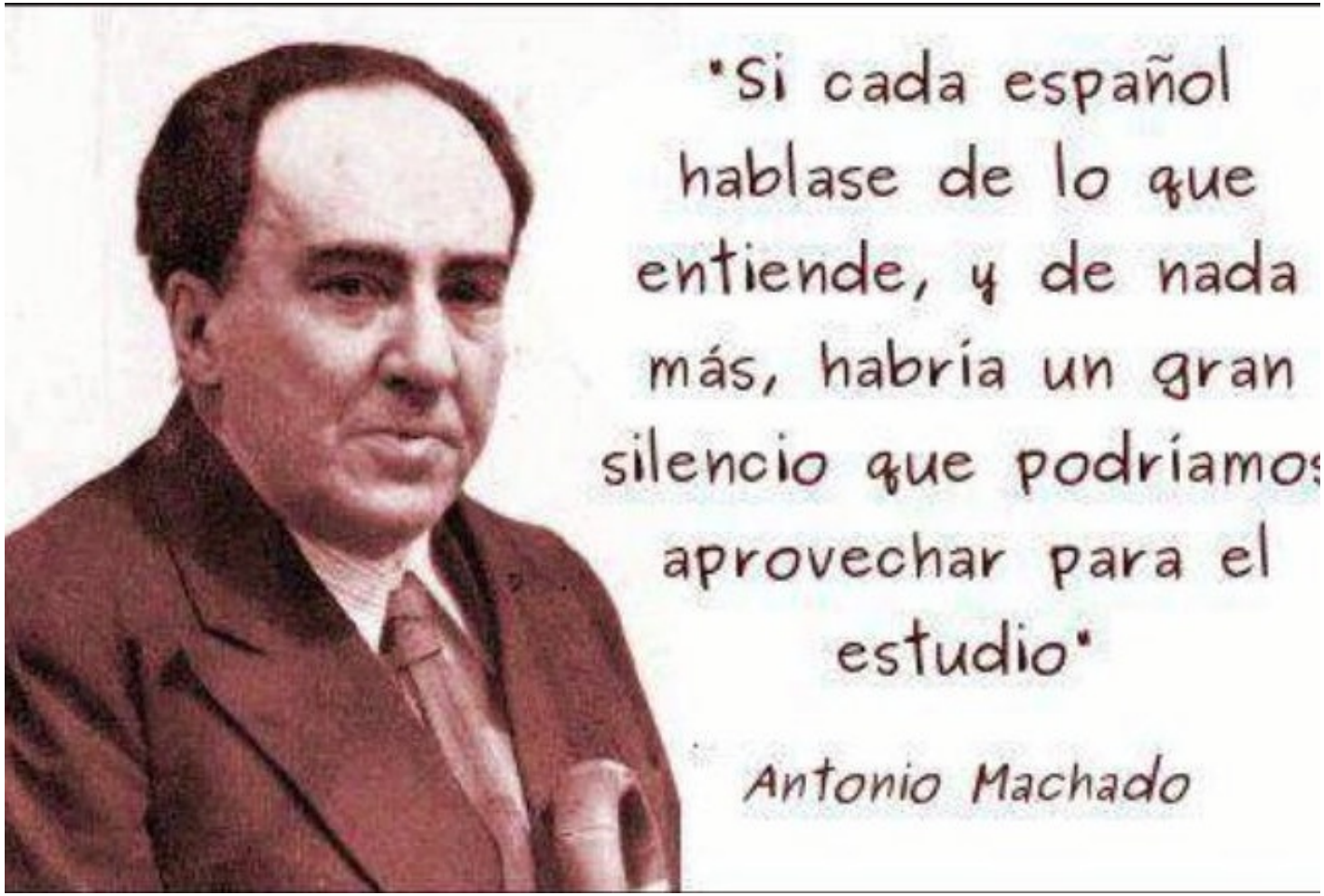
Summary

(K)FCA as a coclustering strategy...

- KFCA does not try to solve the original (direct clustering) task.
- But it provides an alternative look into the task that makes it more realistic and varied:
  - Deals naturally with lack of symmetry (confusion matrices)
  - Deals naturally with data with many objects/few attributes (GED data) or viceversa (itemset analysis).

Most of the advantages stem from:

- A very solid theory (FCA).
- A deep understanding of the maths behind (order lattice theory).
- Appropriateness of use: KFCA deals with counts, probabilities, concentrations: all positive quantities.


"Si cada español
hablase de lo que
entiende, y de nada
más, habría un gran
silencio que podríamos
aprovechar para el
estudio"
Antonio Machado

Thank you!

Norbert Fuhr. Probabilistic models or information retrieval. *The Computer Journal*, 35(3):243–255, 1992.

R Godin, E Saunders, and Jan Gecsei. Lattice model of browsable data spaces. *Information Sciences*, 40:89–116, 1986.

J Hartigan. Direct clustering of a data matrix. *Journal of the American Statistical Association*, Jan 1972.

Boris Mirkin. *Mathematical Classification and Clustering*, volume 11 of *Nonconvex Optimization and Its Applications*. Kluwer Academic Publishers, 1996.

Francisco J. Valverde-Albacete. Combining soft and hard techniques for the analysis of batch retrieval tasks. In Enrique Herrera-Viedma, Gabriella Pasi, and Fabio Crestani, editors, *Soft Computing for Information Retrieval on the Web. Models and Applications*, volume 197 of *Studies in Fuzziness and Soft Computing*, pages 239–258. Springer, 2006.

Francisco J. Valverde-Albacete and Carmen Peláez-Moreno. Extending conceptualisation modes for generalised Formal Concept Analysis. *Information Sciences*, 181:1888–1909, May 2011.