Universidad
Carlos III de Madrid

Detecting Features from Confusion Matrices using Generalized Formal Concept Analysis

F.J. Valverde-Albacete and Carmen Peláez-Moreno

Departamento de Teoría de la Senal y Comunicaciones
Universidad Carlos III de Madrid, Spain

24/06/2010 / San Sebastián

This talk: `www.tsc.uc3m.es/~carmen/HAIS_24_06_10.pdf`

Outline

1. Motivation
2. $(\overline{\mathbb{R}}_{\max,+})$-Formal Concept Analysis applied to confusion matrices
   - Formal Concept Analysis of boolean confusion matrices
   - Exploration of non-boolean confusion matrices
   - The analysis of articulation confusion matrices
3. Conclusions

Classification as Transmission

Consider

- A set of input labels (stimuli) G
- A set of output labels (percepts) M
- Events consisting of presenting one input label (stimulus) $G = g_i$ and returning an output label (percept) $M = m_j$ for it.

The Miller & Nicely [1955] Hypothesis:

Recognition of speech seems to be based in the independent transmission of *phonetic features*

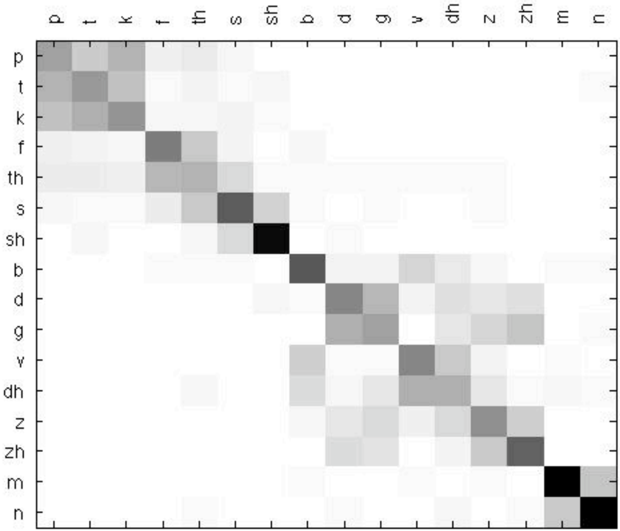
Confusion matrices of multiclass classifiers

N	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_{GM} at $SNR = 0$ dB

- The performance of classifier system can be measured post-operation by a *confusion matrix* counting the ocurrence of the events $N_{ij} = N_{GM}(g_i, m_j)$.
- Notice: no symmetry. Some nulls.

An example of a depiction of a confusion matrix



Outline

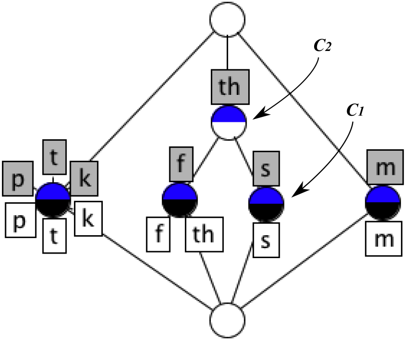
1. Motivation

2. $(\overline{\mathbb{R}}_{\text{max},+})$-Formal Concept Analysis applied to confusion matrices
   - Formal Concept Analysis of boolean confusion matrices
   - Exploration of non-boolean confusion matrices
   - The analysis of articulation confusion matrices

3. Conclusions

Formal context and Concept Lattice

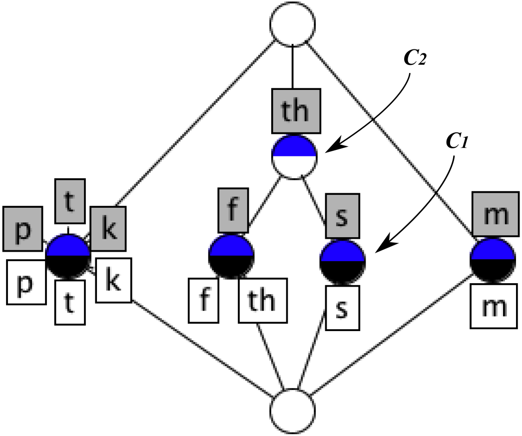
	p	m	t	f	th	k	s
p	×		×			×	
m		×					
t	×		×			×	
f				×	×		
th				×	×		
k	×		×			×	
s					×		×



Concepts: for context (G, M, I) and concept lattice $\mathfrak{B}(G, M, I)$

$$(A, B) \in \mathfrak{B}(G, M, I) \Leftrightarrow A^{+} = B \Leftrightarrow A = {}^{+}B$$

Confusion lattices (cont.)



join- and -meet irreducible concepts

$$\tilde{\gamma}(s) = (\{s\}, \{s, th\}) \qquad \tilde{\mu}(th) = (\{f, s, th\}, \{th\})$$

Outline

Generalized Formal Concept Analysis

- *R* is the *mutual information distribution*

$$R(i,j) = \log \frac{\hat{P}_{GM}(i,j)}{\hat{P}_G(i) \cdot \hat{P}(j)}$$

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

$(\mathbb{R}_{\max,+})$-Formal Concept Analysis [Valverde & Peláez, 2007]

$(G, M, R)_{\mathbb{R}_{\max,+}}$ is called a $\mathbb{R}_{\max,+}$*-valued formal context*

- $R(i, j) = \lambda$ reads as "stimuli g_i is confused with received stimuli m_j in degree λ"
- or dually "received stimuli m_j is taken for stimuli g_i to degree λ".
- We can construct φ-concepts as pairs $(A, B)_\varphi$ with similar properties to those of standard Formal Concept Analysis.

$$(A, B)_\varphi \Leftrightarrow (A)^{+}_{R,\varphi} = B \Leftrightarrow {}^{+}_{R,\varphi}(B) = A$$

- The set of φ-concepts, ordered componentwise is a complete lattice $\underline{\mathfrak{B}}^\varphi(G, M, R)_{\mathbb{R}_{\max,+}}$.
- $\varphi \in \mathbb{R}$ describes a *maximum degree of existence* allowed for pairs (A, B) to be considered members of $\underline{\mathfrak{B}}^\varphi(G, M, R)_{\mathbb{R}_{\max,+}}$.

Structural lattices

- The φ-concept lattice $\underline{\mathfrak{B}}^{\varphi}(G, M, R)_{\overline{\mathbb{R}}_{\max,+}}$ is huge!
- To simplify, for each choice of φ deemed interesting:
  1. Work out the concepts $\gamma(g_i)^{+}_{R,\varphi}$ and $\mu(m_j)^{+}_{R,\varphi}$ associated to singleton stimuli and responses, respectively.
  2. Build a binary incidence $I^{+}_{R,\varphi}$ associated to those concepts by adequately comparing them to create the binary context $(G, M, I^{+}_{R,\varphi})$ with the binary incidence
  $g_i \langle I^{+}_{R,\varphi} \rangle m_i \iff \gamma(g_i)^{+}_{R,\varphi} \leq \mu(m_j)^{+}_{R,\varphi}$.

The *structural lattice* $\underline{\mathfrak{B}}(G, M, I^{+}_{R,\varphi})$ is the (standard) confusion lattice of the binary incidence, $I^{+}_{R,\varphi}$

Valverde-Albacete, F. J. and Peláez-Moreno, C. (in press). *Extending conceptualisation modes for generalised Formal Concept Analysis*. Information Sciences, 2010.

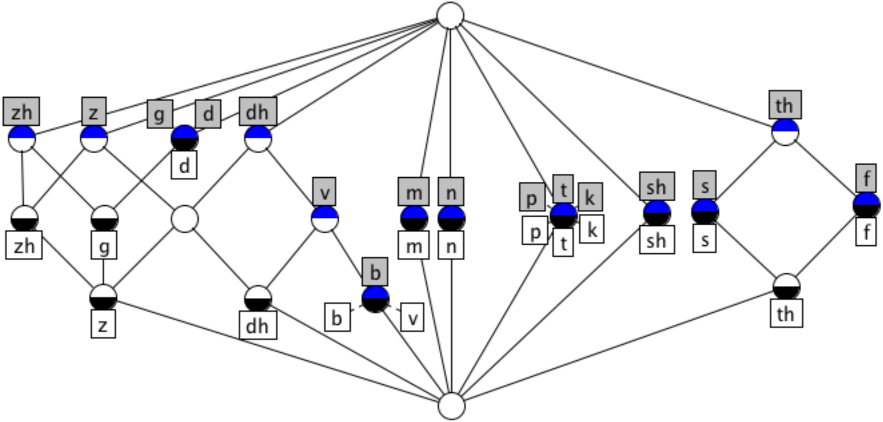
Outline

1. Motivation

2. $(\overline{\mathbb{R}}_{\max,+})$-Formal Concept Analysis applied to confusion matrices
   - Formal Concept Analysis of boolean confusion matrices
   - Exploration of non-boolean confusion matrices
   - The analysis of articulation confusion matrices

3. Conclusions

Phonetic confusion lattice ($\varphi = 0.11716$ and SNR$=$ 0 dB)



Adjoined sublattices establish separate groups or channels!

C. Peláez-Moreno, A. I. García-Moral, and F. J. Valverde-Albacete, *Analyzing phonetic confusions using formal concept analysis,* accepted for publication in Journal of the Acoustical Society of America, 2010.

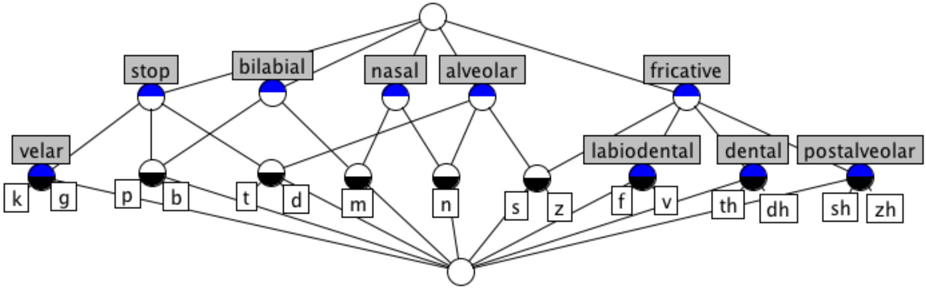
Classical phonetic theory: articulatory features

	labial	lab-dental	dental	alveolar	p-alv	velar
stops	/p/ /b/			/t/ /d/		/k/ /g/
fricative		/f/ /v/	/th/ /dh/	/s/ /z/	/sh/ /zh/	
nasal	/m/			/n/		

Table: Classification of consonant phonemes in English

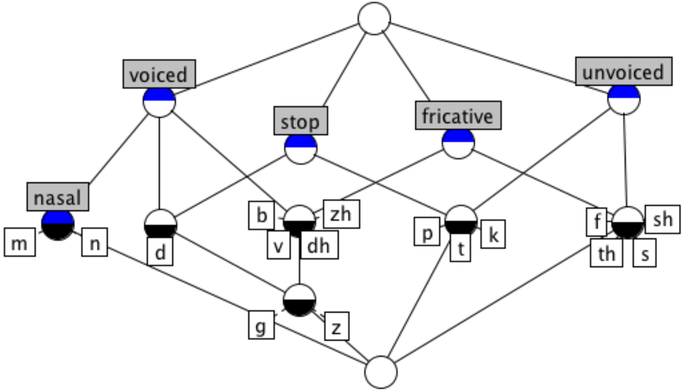
- Examples: /p/ pit, /b/ bit, /t/ tin, /d/ din, /k/ cut, /g/ gut, /f/ fat, /v/ vat, /th/ thin, /dh/ then, /s/ sap, /z/ zap, /sh/ she, /zh/ measure, /m/ map, /n/ nap.

Classical phonetic theory: a lattice representation



- Voiced sounds on the right of each node.

Estimated lattice (including voicedness)



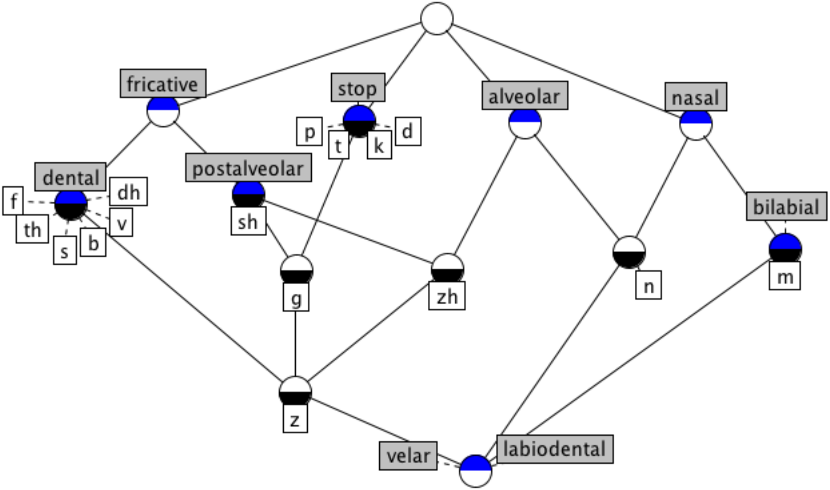
Conclusions

- $\mathcal{K}$-Formal Concept Analysis is a powerful tool for extracting information from matrices extending the capabilities of Formal Concept Analysis over matrices with entries in a semiring like $\overline{\mathbb{R}}_{\text{max},+}$
- $\varphi \in \mathbb{R}$ describes a *maximum degree of existence* required for pairs (A, B) to be considered as members of the φ-lattice $\underline{\mathfrak{B}}^{\varphi}(G, M, R)_{\mathbb{R}_{\text{max},+}}$
- We can explore a $(\mathbb{R}_{\text{max},+})$-valued confusion matrices to find channels "trasmitting" some phonetic features, but not *all* those suggested by the standard theory.

References

- G. A. Miller and P. E. Nicely, *An analysis of perceptual confusions among some english consonants,* Journal of the Acoustical Society of America, vol. 27, no. 2, pp. 338–352, 1955.
- F. J. Valverde-Albacete and C. Peláez-Moreno, *Galois connections between semimodules and applications in data mining,* Lecture Notes on Artificial Intelligence, S. Kusnetzov and S. Schmidt, Eds., vol. 4390. Berlin, Heidelberg: Springer, 2007, pp. 181–196.
- C. Peláez-Moreno, A. I. García-Moral, and F. J. Valverde-Albacete, *Analyzing phonetic confusions using formal concept analysis,* accepted for publication in Journal of the Acoustical Society of America, 2010.
- C. Peláez-Moreno, F. J. Valverde-Albacete and V. Esteban-Alonso, *KFCA on-line demo*, `http://www.tsc.uc3m.es/~carmen/KFCADemo.html`

Sequences of lattices: feature detection (cont.)



- For unvoiced groups we can conclude that friction is the distinctive feature.

$\mathcal{K}$-Formal Concept Analysis

$$\overline{\mathcal{K}} = \langle\, \overline{K}, \preccurlyeq \,\rangle \equiv \langle\, \overline{K}, \underset{\cdot}{\oplus} = \vee, \underset{\cdot}{\otimes}, \cdot^{-1}, \bot, e \,\rangle$$

$$\overline{\mathcal{K}}^{\mathrm{d}} = \langle\, \overline{K}, \preccurlyeq^{\mathrm{d}} \,\rangle \equiv \langle\, \overline{K}, \dot{\oplus} = \wedge, \dot{\otimes}, \cdot^{-1}, \top, e \,\rangle$$

$$\overline{\mathbb{R}}_{\max,\min,+} = \langle\, \mathbb{R} \cup \{\pm\infty\}, \max, \underset{\cdot}{+}, \min, \dot{+}, \cdot-, -\infty, 0, \infty \,\rangle$$

The concept forming operators:

$\langle x \mid R \mid y \rangle = y^{\mathsf{T}} \underset{\cdot}{\otimes} R^{-1} \underset{\cdot}{\otimes} x$, where $y \in \mathcal{Y}$ is a multivalued set of objects and $x \in \mathcal{X}$ a multivalued set of attributes is actually a binary residuated operation:

$$(y)^{+}_{R,\varphi} = \bigvee \{\, x \in \mathcal{X} \mid \langle x \mid R \mid y \rangle \leq \varphi \,\} \qquad {}^{+}_{R,\varphi}(x) = \bigvee \{\, y \in \mathcal{Y} \mid \langle x \mid R \mid y \rangle \leq$$

$$(y)^{+}_{R,\varphi} = R^{\mathsf{T}} \dot{\otimes}\, y^{-1} \dot{\otimes}\, \varphi \qquad {}^{+}_{R,\varphi}(x) = R \dot{\otimes}\, x^{-1} \dot{\otimes}\, \varphi$$