Evidence Theory [episode 2]
Implementation Issues and Applications

David Lesage <david@lrde.epita.fr>

LRDE seminar, May 28, 2003

Table of Contents

Fed up with TV series? **4**

Episode 1 5

Episode 2 6

Episode 3 8

Belief State and Decision after 3 episodes 10

Implementation **14**

Mass Functions representation 15

Focal Elements representation 19

Potentials representation 24

Combination & fusion issues 26

Evidenz: an evidence theory engine 29

Evidenz characteristics 30

Performances 31

Evidenz engine future improvements 35

Applications 38

Document classification 39

Cytology (Adhoc^2) 42

Satellite image classification 46

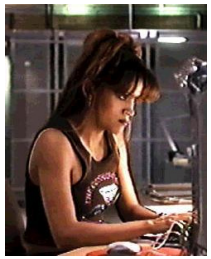
Conclusion 50

Questions? 51

References **52**

Fed up with TV series?

Someone is a traitor...



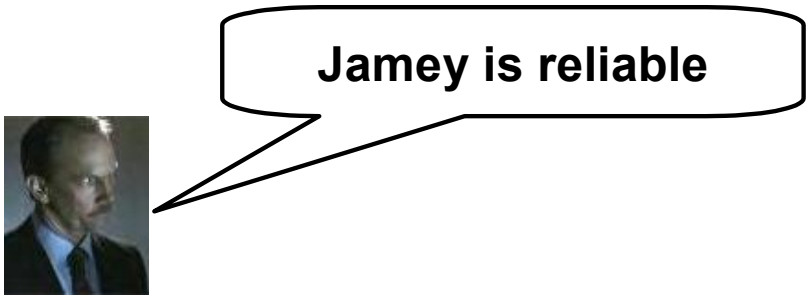
Jamey Farrell

Tony Almeida

Nina Myers

- 1 variable: traitor = { Jamey, Tony, Nina }
- belief mass on focal element X = belief that X contains the traitor
- potentials = episodes

Episode 1



Richard Walsh

▷ $m(\{Tony, Nina\}) = 0.7$

Richard is not omnipotent.

▷ $m(\{Jamey, Tony, Nina\}) = 0.3$

- Reminder:

$$\sum_{A \subseteq \Omega} m(A) = 1$$

$$ep_1 = \{\{Tony, Nina\}, \{Jamey, Tony, Nina\}\}$$

Episode 2

Jack Bauer

Statistically, people with a family are more sensitive to pressure or money temptation

- $m(\{Nina\}) = 0.4$
- $m(\{Jamey, Tony\}) = 0.6$

$$ep_2 = \{\{Nina\}, \{Jamey, Tony\}\}$$

➔ Statistical support

- Combination of ep_1 and ep_2

$$ep_1 = \{\{T, N\}[0.7], \{J, T, N\}[0.3]\}$$
$$ep_2 = \{\{N\}[0.4], \{J, T\}[0.6]\}$$

  - **Reminder**: Combination
    - focal elements intersections
    - $m_{D=A\cap B=A\cap C} = m_A \times m_B + m_A \times m_C$

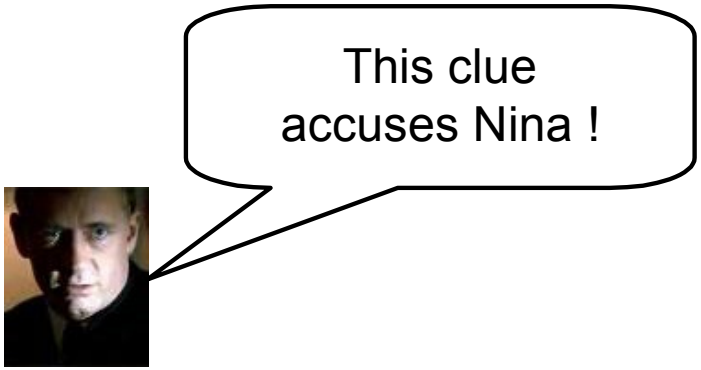
  - Generated focals elements:

$$ep_{1\oplus 2} = \{\{N\}, \{T\}, \{J, T\}\}$$

  - Belief Masses:

$$m(\{Nina\}) = 0.7 \times 0.4 + 0.3 \times 0.4 = 0.4$$
$$m(\{Tony\}) = 0.7 \times 0.6 = 0.42$$
$$m(\{Jamey, Tony\}) = 0.3 \times 0.6 = 0.18$$

Episode 3



Georges
Masson

- $m(\{Nina\}) = 0.8$
- $m(\{Jamey, Tony\}) = 0.2$

$$ep_3 = \{\{Nina\}, \{Jamey, Tony\}\}$$

- Combination of ep_3 and $ep_{1\oplus 2}$

$$ep_3 = \{\{N\}[0.8], \{J,T\}[0.2]\}$$
$$ep_{1\oplus 2} = \{\{N\}[0.4], \{T\}[0.42], \{J,T\}[0.18]\}$$

  - ▷ Generated focals elements:

$$ep_{1\oplus 2\oplus 3} = \{\{N\}, \{T\}, \{J,T\}\}$$

  - ▷ Conflict C:

$$\{N\} \cap \{T\} = \varnothing \text{ and } \{N\} \cap \{J,T\} = \varnothing$$
$$C = 0.8 \times 0.42 + 0.8 \times 0.18 + 0.2 \times 0.4 = 0.56$$

  - ▷ Belief Masses:

$$m(\{Nina\}) = \frac{0.8 \times 0.4}{1-C} \simeq 0.73$$
$$m(\{Tony\}) = \frac{0.2 \times 0.42}{1-C} \simeq 0.19$$
$$m(\{Jamey, Tony\}) = \frac{0.2 \times 0.18}{1-C} \simeq 0.08$$

Belief State and Decision after 3 episodes

- **Reminder:**

$$bel_A = \sum_{\emptyset \neq X \subseteq A} m(X)$$

$$pl_A = 1 - bel_{\overline{A}} = \sum_{X \cap A \neq \emptyset} m(X)$$

$$ep_{1\oplus 2\oplus 3} = \{\{N\}[0.73], \{T\}[0.19], \{J,T\}[0.08]\}$$

- **{Jamey}**

$bel_{\{Jamey\}} = 0$ $\quad$ $pls_{\{Jamey\}} = 0.08$

- **{Tony}**

$bel_{\{Tony\}} = 0.19$ $\quad$ $pls_{\{Tony\}} = 0.19 + 0.08 = 0.27$

- **{Nina}**

$bel_{\{Nina\}} = 0.73$ $\quad$ $pls_{\{Nina\}} = 0.73$

➔ Take care about interpretations !

- Decision ➔ probability construction (Smets et al., 1992)

$$BetP(x,m) = \sum_{x \in A \subseteq \Omega} \frac{m(A)}{|A|}$$

$$ep_{1 \oplus 2 \oplus 3} = \{\{N\}[0.73], \{T\}[0.19], \{J,T\}[0.08]\}$$

▷ **{Jamey}**

$$BetP(J) = \frac{0.08}{2} = 0.04$$

▷ **{Tony}**

$$BetP(T) = 0.19 + \frac{0.08}{2} = 0.24$$

▷ **{Nina}**

$$BetP(N) = 0.73$$

Is this theory right?
To be continued...

illustrations from “24” TV series: http://www.fox.com/24/

Implementation

- Mass Functions representation
- Focal Elements representation
- Potentials representation
- Combination & fusion issues

Mass Functions representation

- Mass function = distribution of belief on Ω

▷ **In theory**:
mass functions assign mass **values** to sets.

▷ **In reality**:
masses depend on **context** (events).

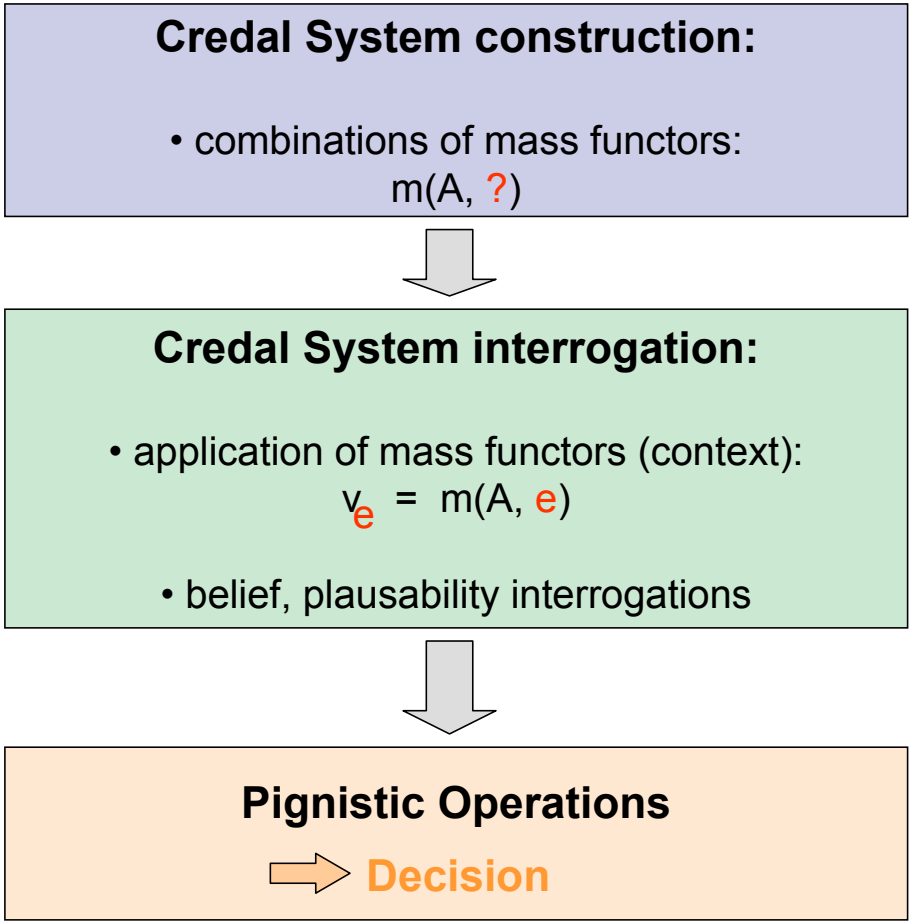
Example, in the field of **Medical Diagnosis**:

Combining “general rules” (potentials), and applying the system obtained on multiple cases, according to patient’s characteristics.

➔ Need for **functors**

$$\begin{array}{rccc} m: & 2^{\Omega} \times E & \rightarrow & [0,1]_{\Re} \\ & (A,e) & \rightarrow & v \end{array}$$

➔ Model principles:



- **Consequences of this model**
  - ▷ Different levels of evaluation
  - ▷ Approximations impossible

Focal Elements representation

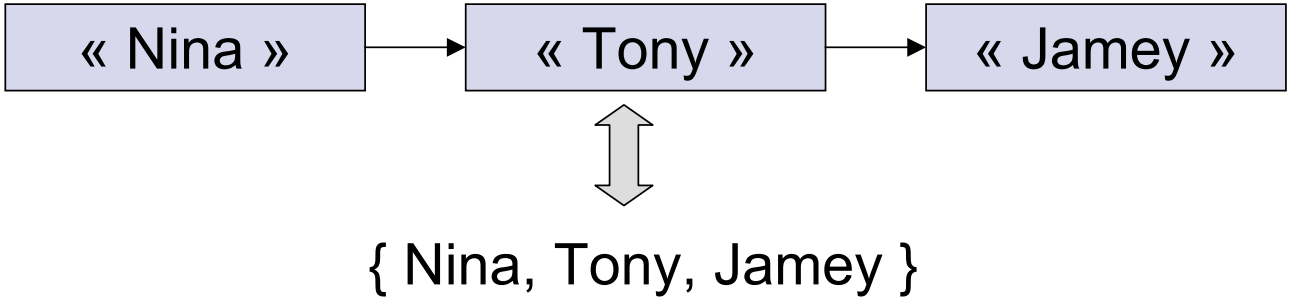
- Focal Element = set of variable configurations ➔ **hypothesis**

▷ Operations: **set intersections, projections, extensions**

▷ Size issues

▷ Efficiency issues

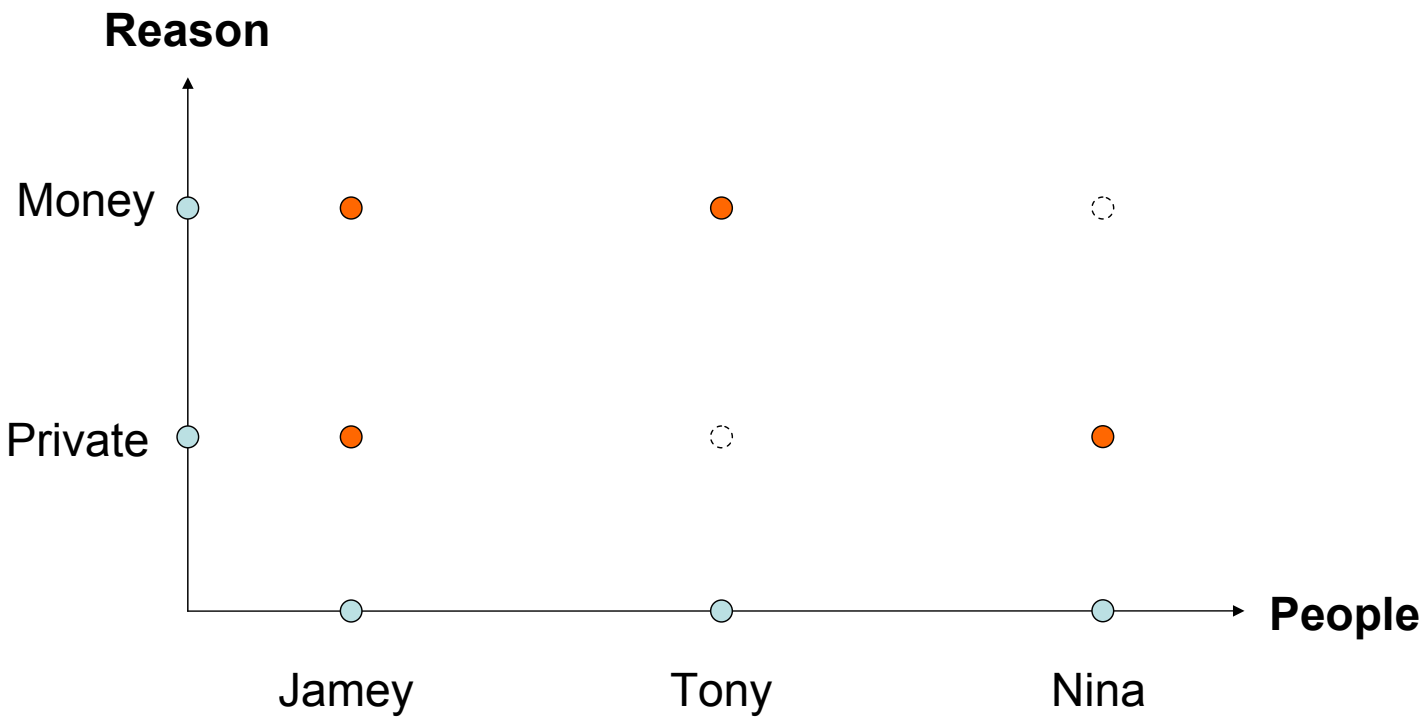
- **Lists?**



  - ▷ speed, size, ordering problems

- **Ordered sets?**
  - ▷ solve only ordering problems

- **Bitsets?**



{ (Jamey, Money), (Jamey, Private),
(Tony, Money),
(Nina, Private) }

⟺ **< 11 10 01 >**

Bitsets properties

- ▷ efficient operations on sets
- ▷ set bit number = configurations number
- ▷ constant size:

$$S = \prod_{i=1}^{nbvars} \#X_i bits$$

ex: for 23 binary variables, $size(FS) = 2^{23}$ bits $= 1$ Mo

- ▷ $Index \Leftrightarrow Configuration$ correspondancy formulas
- ▷ Independence towards original representation !

- **Other representations**
  - ▷ Normal forms

Disjunctive forms

$$r(A) = \bigcup_{i=1}^{h1} \bigcap_{j=1}^{n} S_j^i$$

$$S_j^i \in \Theta_{x_j}$$

Conjunctive forms

$$r(A) = \bigcap_{i=1}^{h1} \bigcup_{j=1}^{n} S_j^i$$

$$S_j^i \in \Theta_{x_j}$$

Potentials representation

- Potential = collection of pairs (F_i, m_i)
  with $F_i \in FS$ and m_i mass function ➔ **1 source of information**
  - ▷ Operations on potentials: **combinations, fusions**
  - ▷ Need for efficient research and combination ➔ **regroupments**
  - ▷ **Lists?** research: *O(n)* regroupment: *O(n^2)*
  - ▷ **Balanced Trees?** research: *O($log(n)$)* regroupment: *O($n \times log(n)$)*
  - ▷ **Hash Tables?** research: *O(1)* regroupment: *O(n^2/s)*

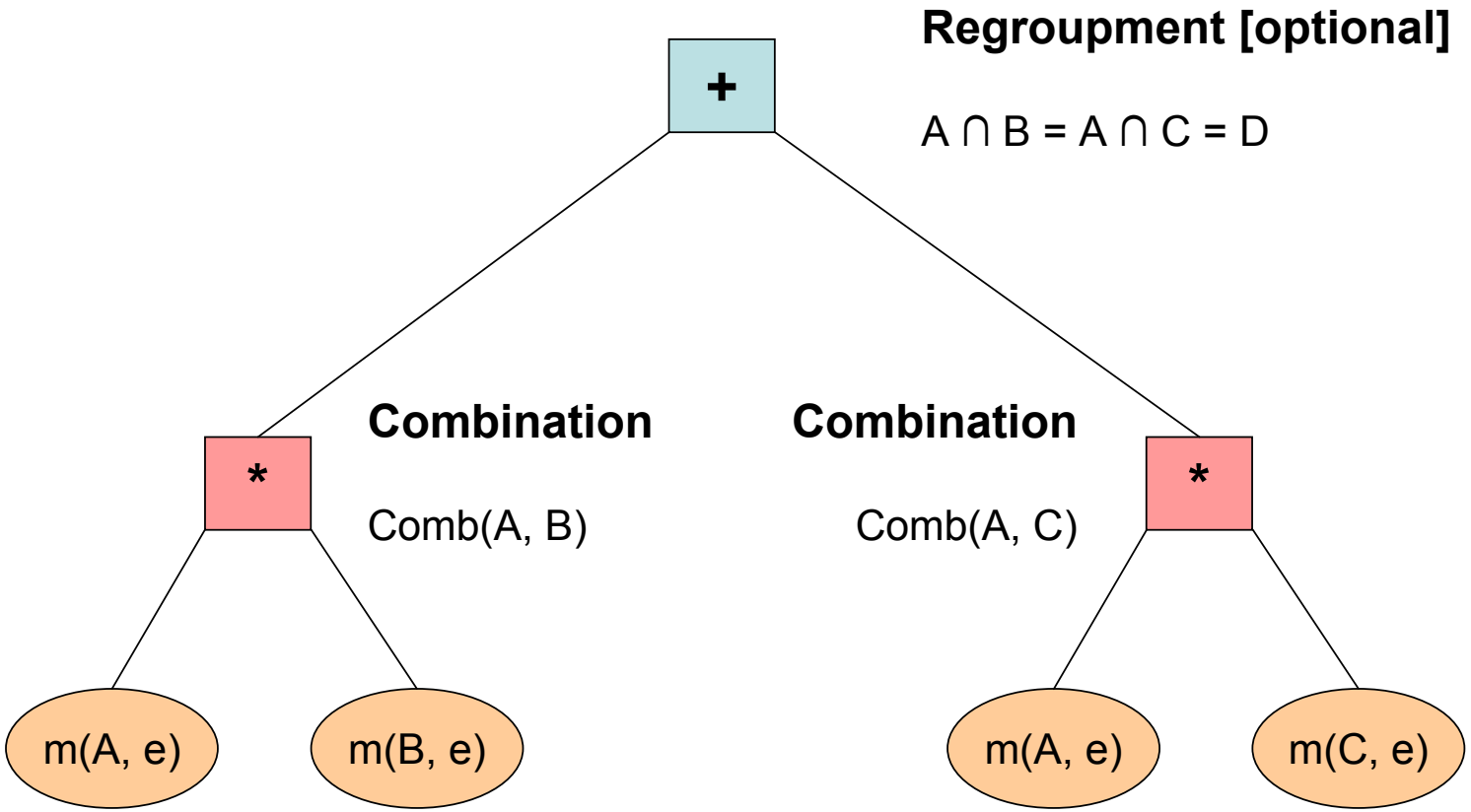
Example of a combination of 2 potentials of 1000 focal elements with 1000 configurations:

balanced trees (AVL)	6 seconds
lists	14 hours 22 minutes

Combination & fusion issues

- New potentials creation ➔ focal elements creation (by intersection)
- New mass functions creation ➔ functors creation
- Mass functions combinations and regroupment

Mass Function of D = A ∩ B = A ∩ C



▷ **indirections** due to functors !

▷ example, with 1 potential P, 10 focal elements of 10 configurations, on 1 variable with 10 realisations, after 5 combinations:

construction time	1 second
memory load	1 Mo
original focal elements	10
final focal elements	163
interrogation time (1 Focal element)	2 minutes
mass regroupments (plus indirections)	1.6059×10^9
mass combinations (times indirections)	3.2642×10^9

- optimization?

Evidenz: an evidence theory engine

- Existing engines: research prototypes, written in Lisp...
  (Haenni and Lehmann, 2001) (Saffiotti and Umkehrer, 1991)

  - Evidenz characteristics
  - Performances
  - Evidenz engine future improvements

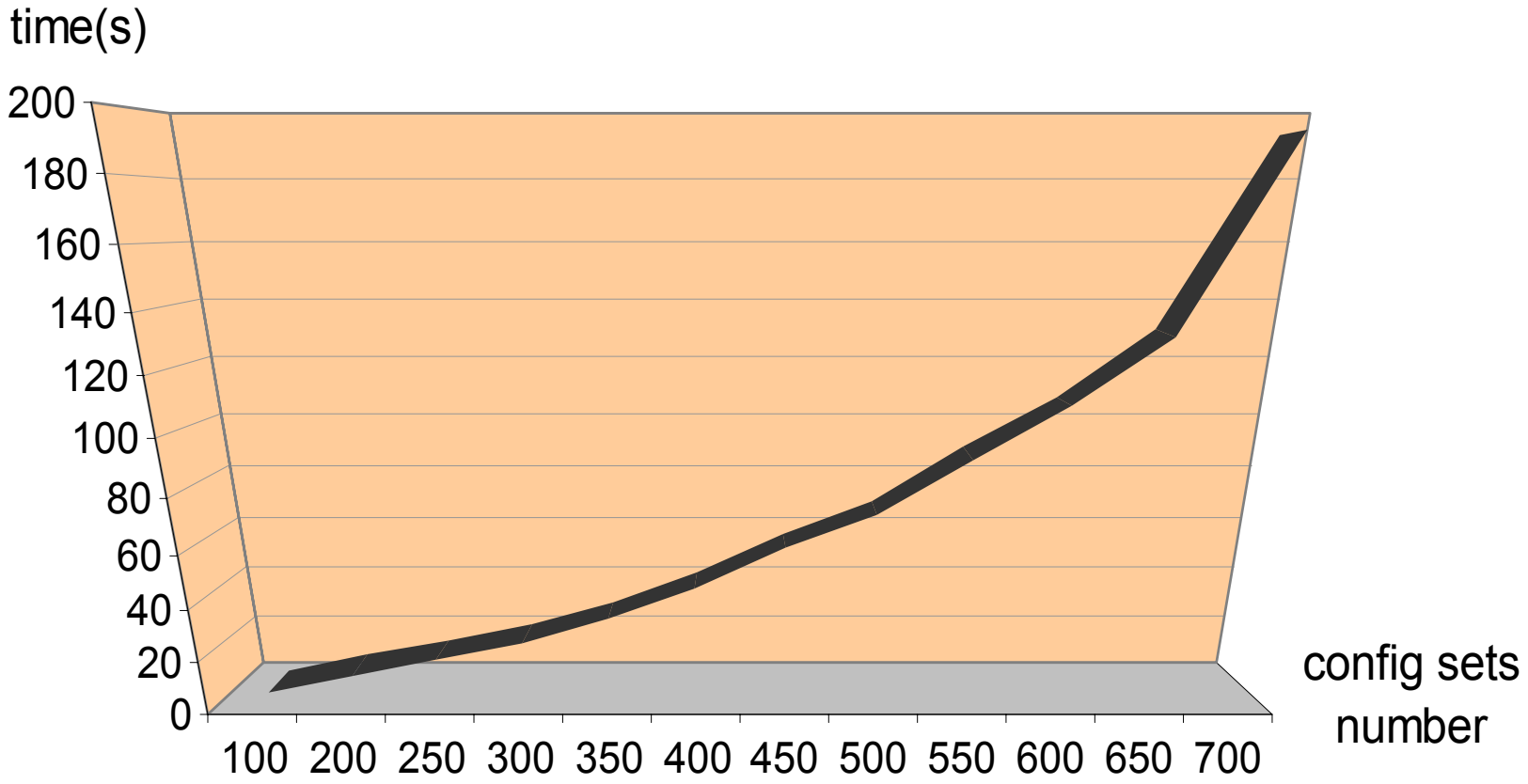
Evidenz characteristics

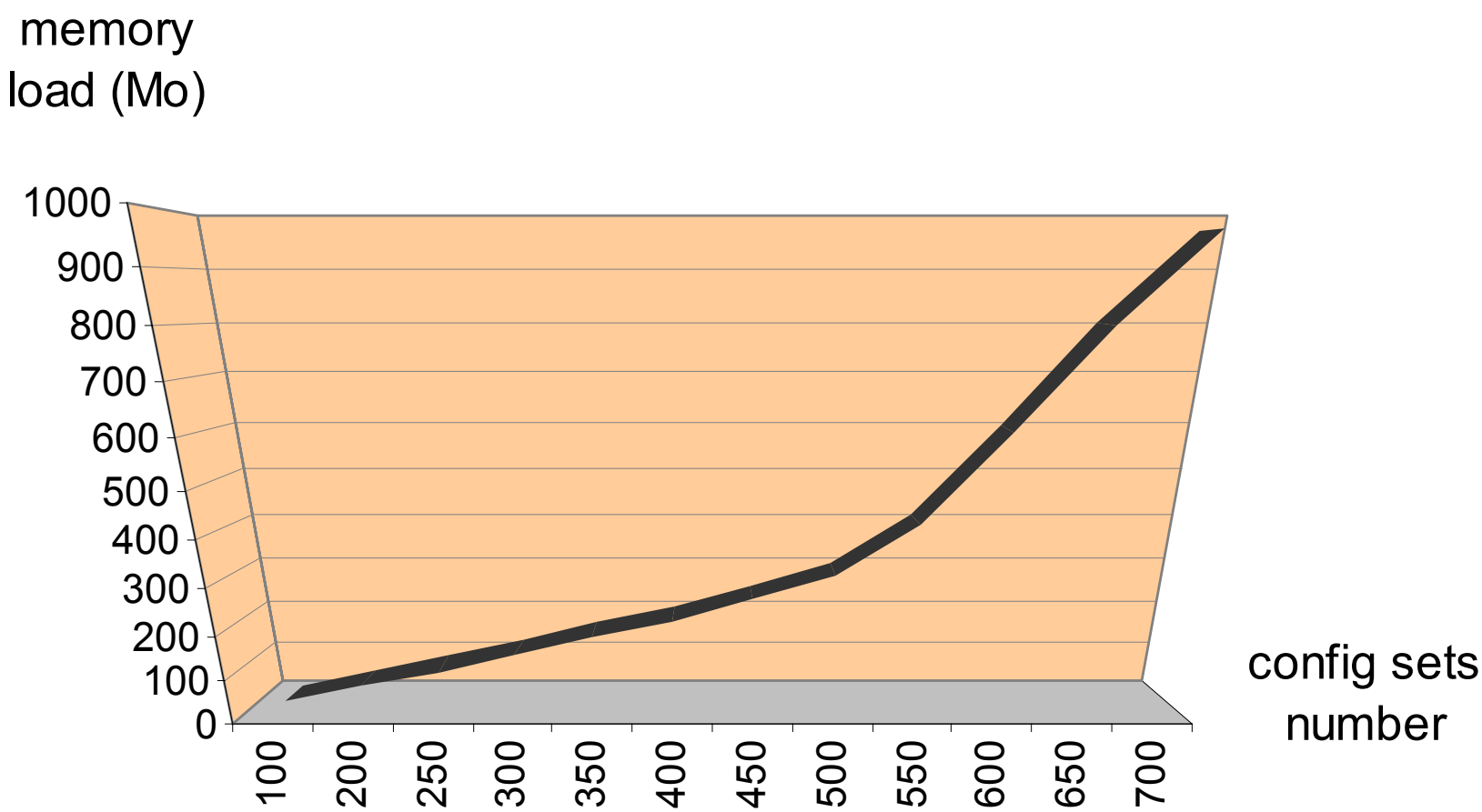
- TBM exact modeling (Smets et al., 1992)
- written in C++ (dynamic version)

- focal elements represented by bitsets (Boost)
- generic towards configurations representation
- potentials represented by balanced trees (STL maps)

Performances

- Study of $\neq$ factors influence on execution **time** and **memory load**:
  - ▷ variables
  - ▷ variable realisations
  - ▷ focal elements per potential
  - ▷ configurations per focal element
  - ▷ **one combination of 2 potentials randomly generated**
- example, on **focal elements per potential** influence:
  - ▷ 2 variables with 500 realisations
  - ▷ 100 configurations per focal element
  - ▷ 100 to 700 focal elements per potential


time(s)
200
180
160
140
120
100
80
60
40
20
0
100 200 250 300 350 400 450 500 550 600 650 700
config sets number

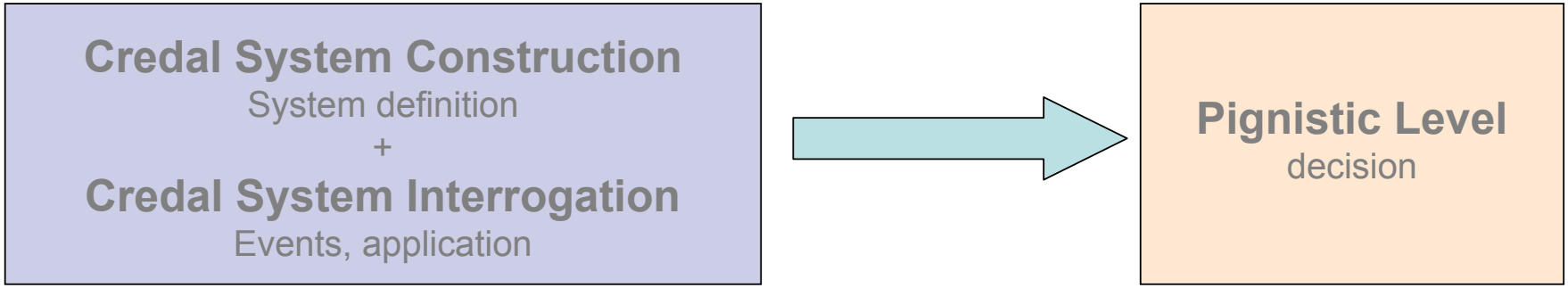

memory
load (Mo)
1000
900
800
700
600
500
400
300
200
100
0
100
200
250
300
350
400
450
500
550
600
650
700
config sets
number

- **Performance study conclusions**
  - ▷ Time and memory load evolutions are **exponential** towards **all** the factors
  - ▷ Problems become quickly infeasible !
  - ▷ In practice, applications are **much more** reasonable...

Evidenz engine future improvements

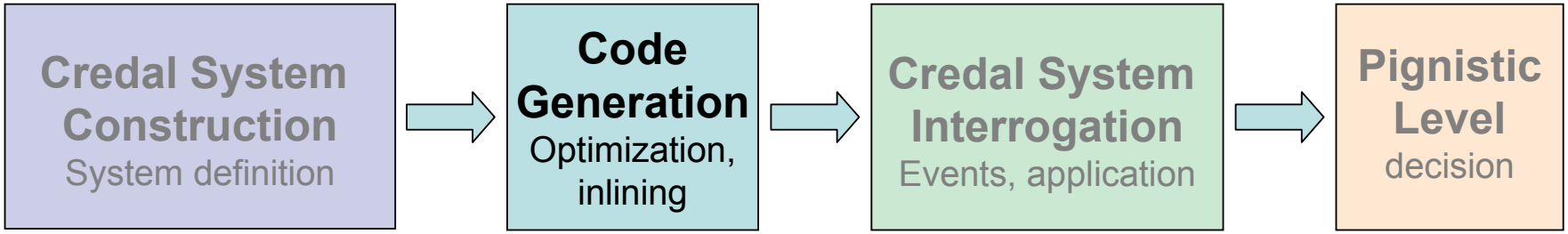
- Heuristics for **optimizing elimination sequences** (using **Join trees**)
- Decision algorithms, like **decision trees**
- **Memoizing** (set intersections and mass functions)

- **2 use cases:**
  - **Need of quick, unique application**:



  - **Approximations**

- **Need of multiple applications on different events**:

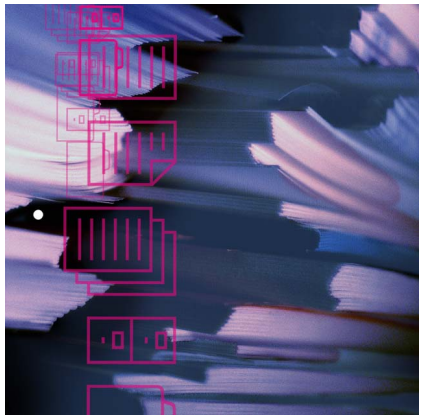


- **code generation**: generates an optimal system at the end of credal definition phase

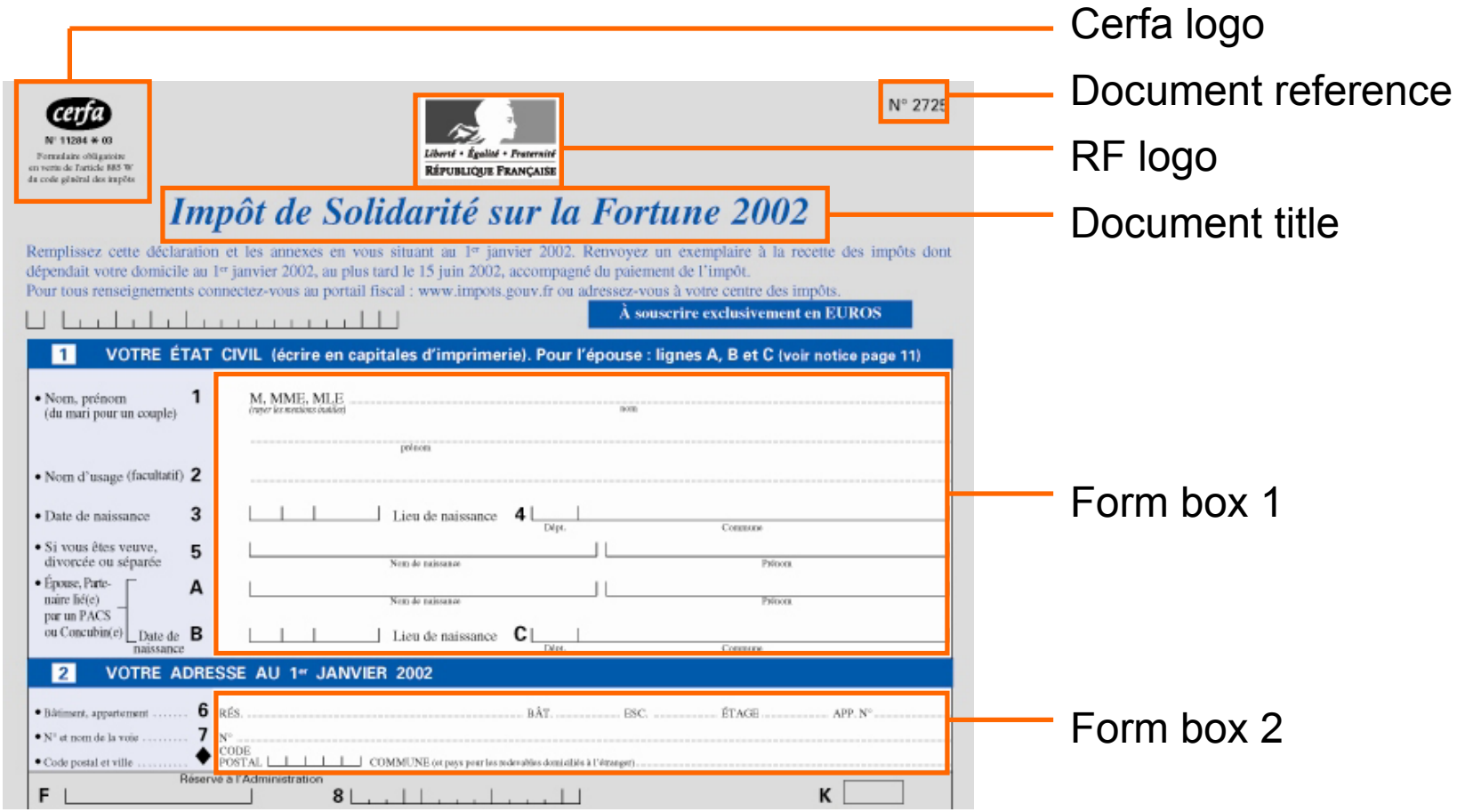
Applications

- Document classification
- Cytology (Adhoc^2)
- Satellite image classification

Document classification

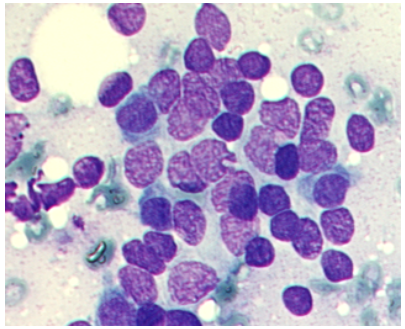


- Automatic classification and data extraction
  - ▷ to associate a model to a document
- A document contains **anchors**: logos, text boxes...
- Each anchor corresponds to one or several models

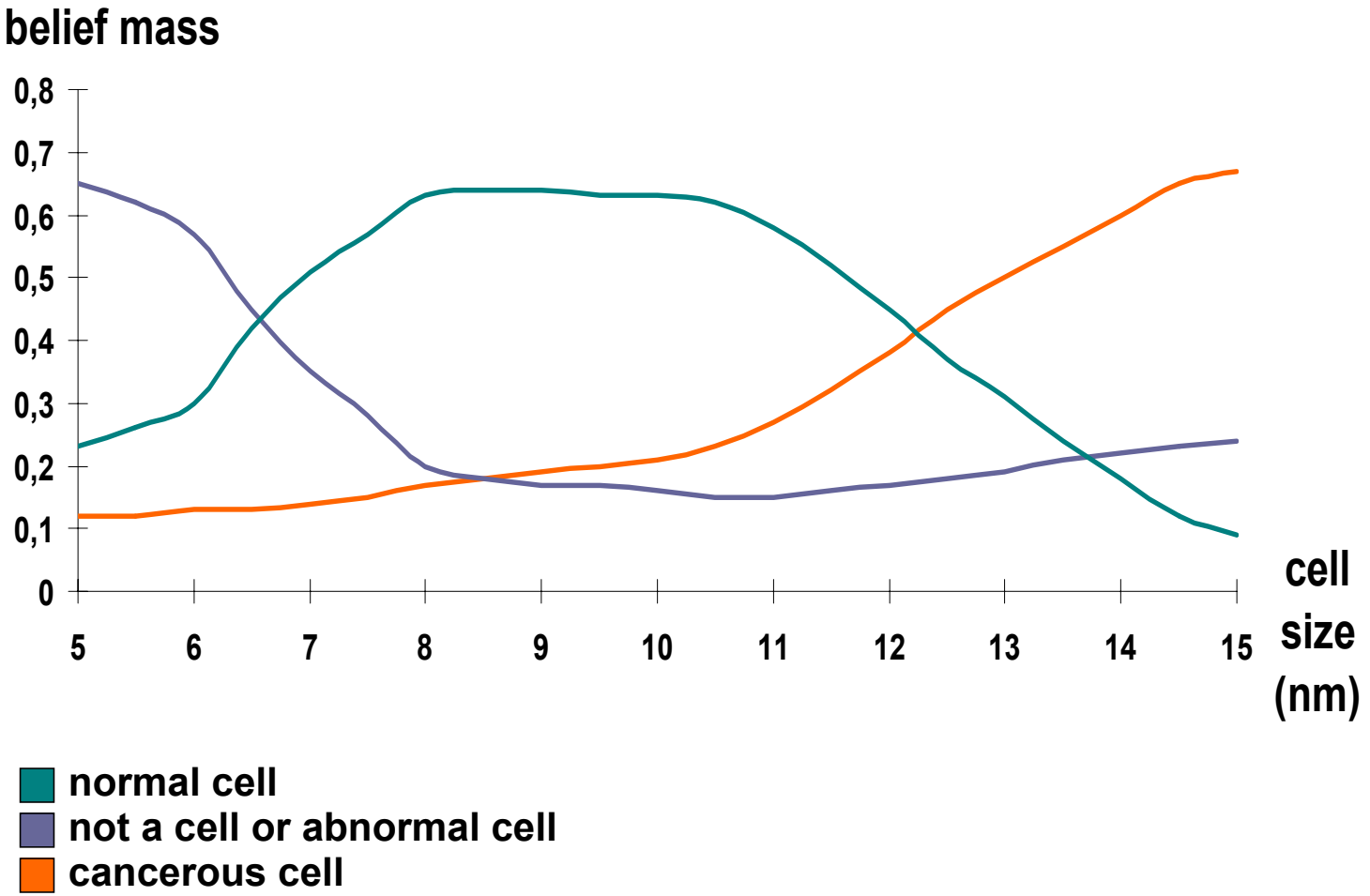

Cerfa logo
Document reference
RF logo
Document title
Form box 1
Form box 2
cerfa
N° 11284 * 03
Liberté • Égalité • Fraternité
RÉPUBLIQUE FRANÇAISE
N° 2725
Impôt de Solidarité sur la Fortune 2002
Remplissez cette déclaration et les annexes en vous situant au 1er janvier 2002. Renvoyez un exemplaire à la recette des impôts dont dépendait votre domicile au 1er janvier 2002, au plus tard le 15 juin 2002, accompagné du paiement de l'impôt.
Pour tous renseignements connectez-vous au portail fiscal : www.impots.gouv.fr ou adressez-vous à votre centre des impôts.
À souscrire exclusivement en EUROS
1 VOTRE ÉTAT CIVIL (écrire en capitales d'imprimerie). Pour l'épouse : lignes A, B et C (voir notice page 11)
• Nom, prénom (du mari pour un couple) 1
M, MME, MLE
• Nom d'usage (facultatif) 2
• Date de naissance 3
Lieu de naissance 4
• Si vous êtes veuve, divorcée ou séparée 5
• Épouse, Partenaire lié(e) par un PACS ou Concubin(e) A
Date de naissance B
Lieu de naissance C
2 VOTRE ADRESSE AU 1er JANVIER 2002
• Bâtiment, appartement 6
• N° et nom de la voie 7
• Code postal et ville
Réservé à l'Administration
F
8
K

- Modelisation:
  - ▷ 1 variable: $document\_model = \{model_1, model_2, model_3, ..., model_k\}$
  - ▷ potentials = anchors
- Anchor recognition ➔ belief masses for model sets
  - ▷ combining anchor masses ➔ global belief for each document model

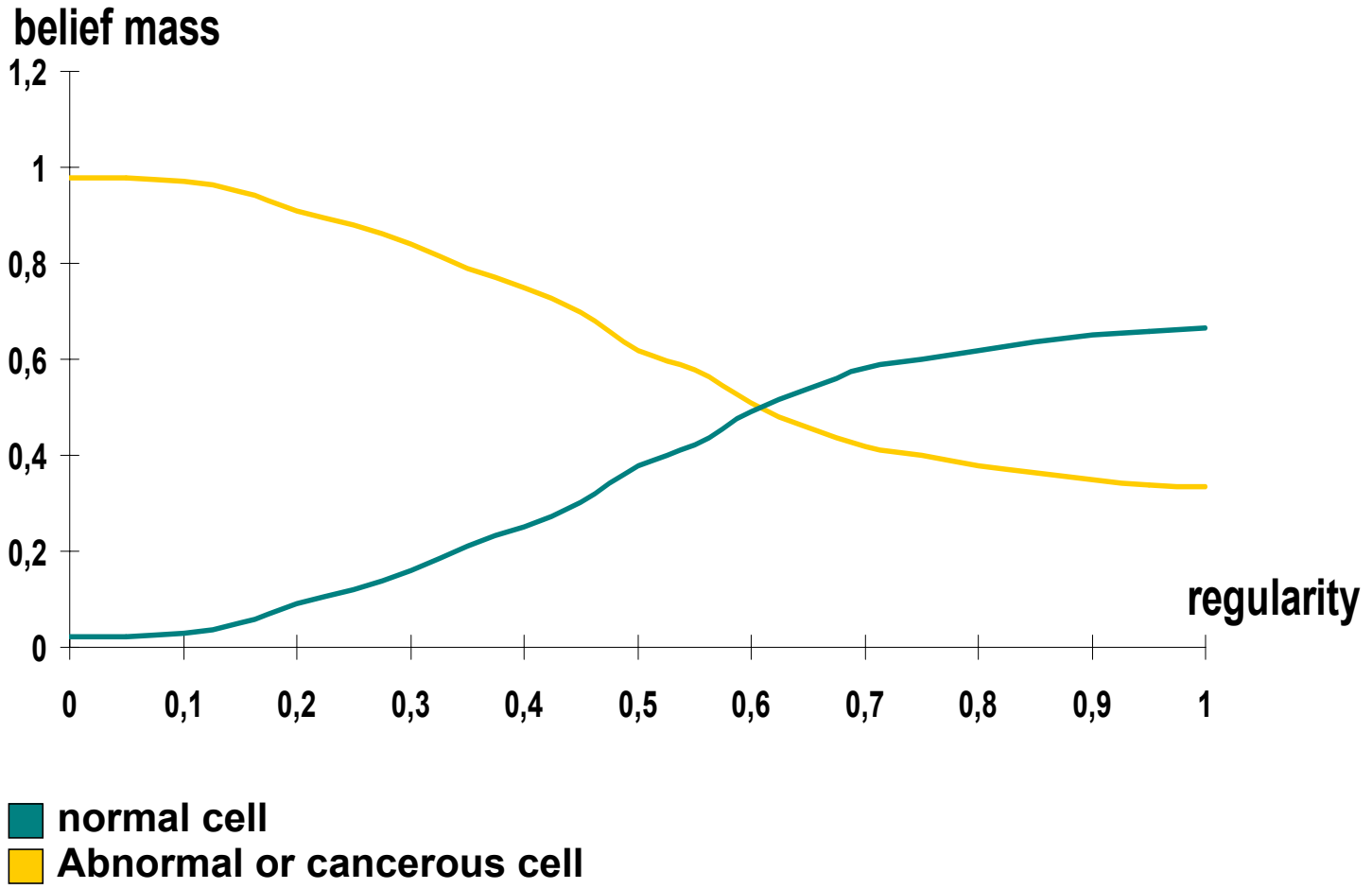
Cytology (Adhoc^2)



- Cancerous cells detection
  - ▷ diagnosis help
  - ▷ $\neq$ factors:
    - $*$ cell/nucleus size
    - $*$ cell/nucleus regularity
    - $*$ color...
- **Fuzzy**, **subjective** decision rules


belief mass
0,8
0,7
0,6
0,5
0,4
0,3
0,2
0,1
0
5
6
7
8
9
10
11
12
13
14
15
cell size (nm)
normal cell
not a cell or abnormal cell
cancerous cell

Size factor



Regularity factor

- Modelisation:
  - ▷ 1 variable: $cell\_type = \{normal, abnormal, not\_a\_cell, cancerous\}$
  - ▷ potentials = criters: size, regularity, color...

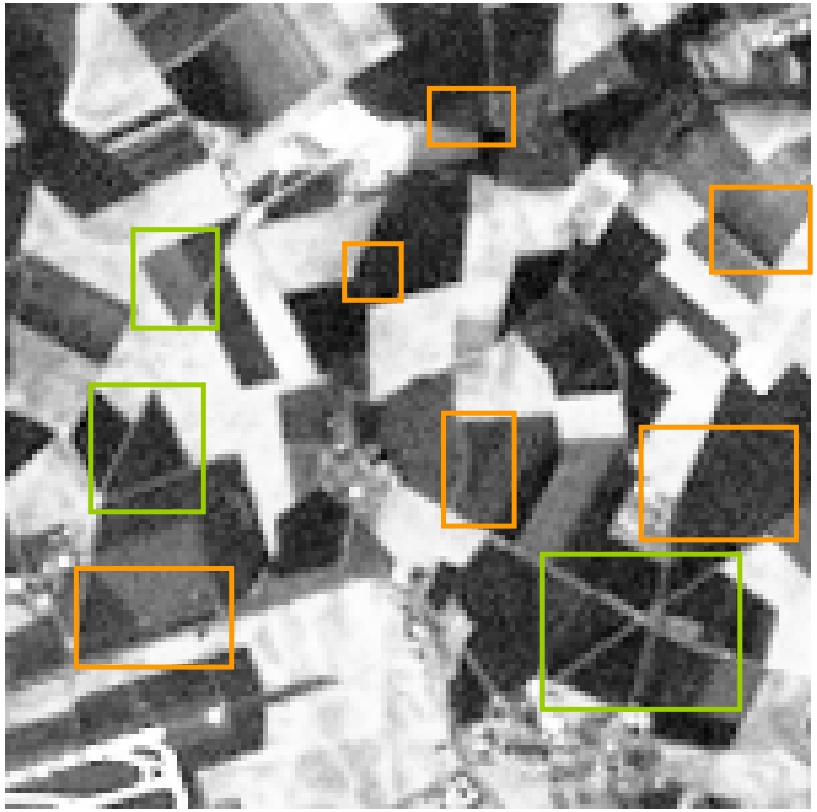
- Cell characteristics ➔ belief masses for cell type sets
  - ▷ combining criters ➔ global belief for each cell type
  - ▷ justified decision, quantified ignorance...

Satellite image classification

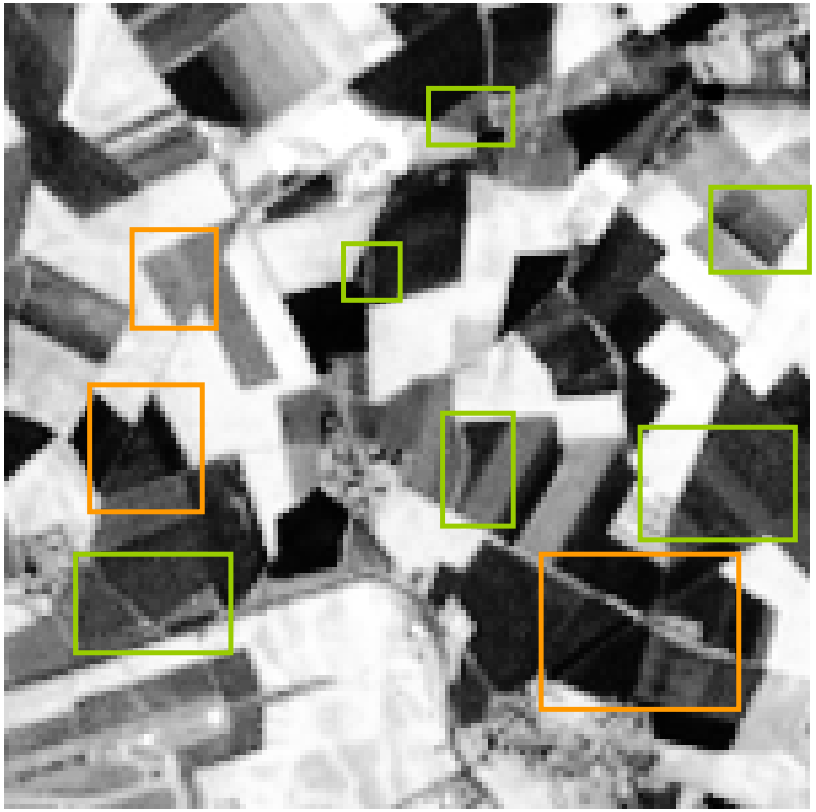
- Satellite images composed of several layers:
  - 7 layers according to wave length
  - Different visions of the same image ➔ different informations

Layer x

Layer y



Layer x



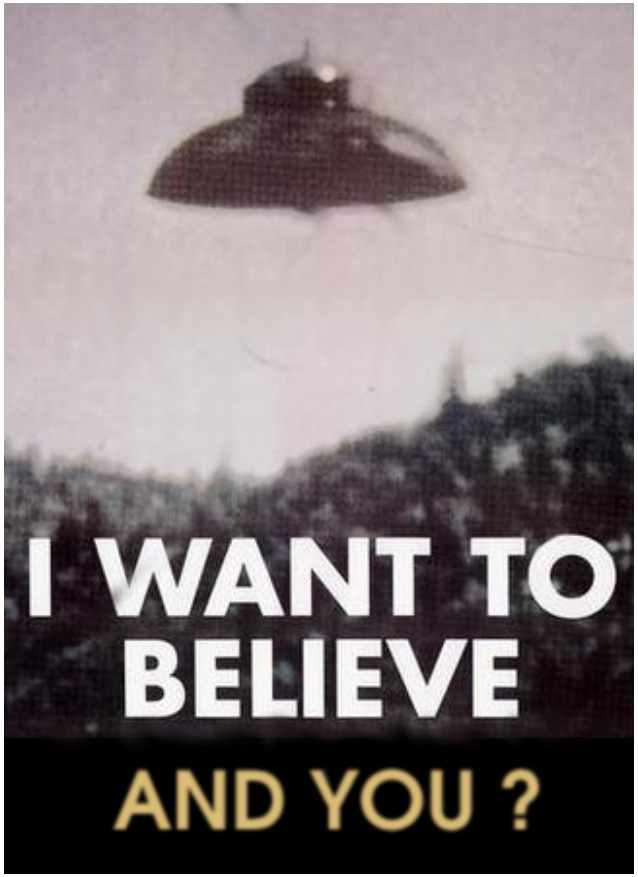
Layer y

- Modelisation:
  - ▷ 1 variable: $pixel\_type = \{forest, water, road, building, ...\}$
  - ▷ potentials = layers
- Pixel gray level (**context**) ➔ belief masses for pixel type sets
  - ▷ combining layers ➔ global belief for each pixel type

Conclusion

- Theory designed for dealing with uncertainty
  - ▷ Modelisation power
  - ▷ Computational complexity
- Evidenz engine
  - ▷ Code generation and algorithmic extension/improvement
- Future research applications:
  - ▷ Adhoc^2 cytology project
  - ▷ Satellite image classification

Questions?



References

Burrus, N. (2003). Evidence theory [part 1]. lrde seminar oral presentation.

Burrus, N. and Lesage, D. (2003). Evidence theory. Technical report, EPITA Research And Development Laboratory, Paris, France.

Haenni, R. and Lehmann, N. (2001). Implementing belief function computations. Technical Report 01-28, Department of Informatics, University of Fribourg.

Saffiotti, A. and Umkehrer, E. (1991). Pulcinella: a general tool for propagating uncertainty in valuation networks. pages 323–331.

Smets, P., Hsia, Y., Saffiotti, A., Kennes, R., Xu, H., and Umkehrer, E. (1992). The transferable belief model. pages 91–98.