The challenges of “intelligent” decision support: from preference learning to explaining recommendations

Wassila Ouerdane

MICS - CentraleSupélec
Université Paris-Saclay

Joint work, with

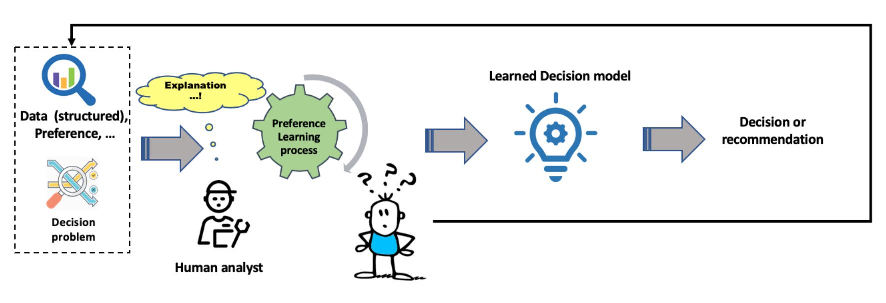
- Khaled Belahcène, Heudiasyc, UTC.
- Christophe Labreuche, Thales Research & Technology
- Nicolas Maudet, Lip6, Sorbonne Université
- Vincent Mousseau, MICS, CentraleSupélec

Outlines

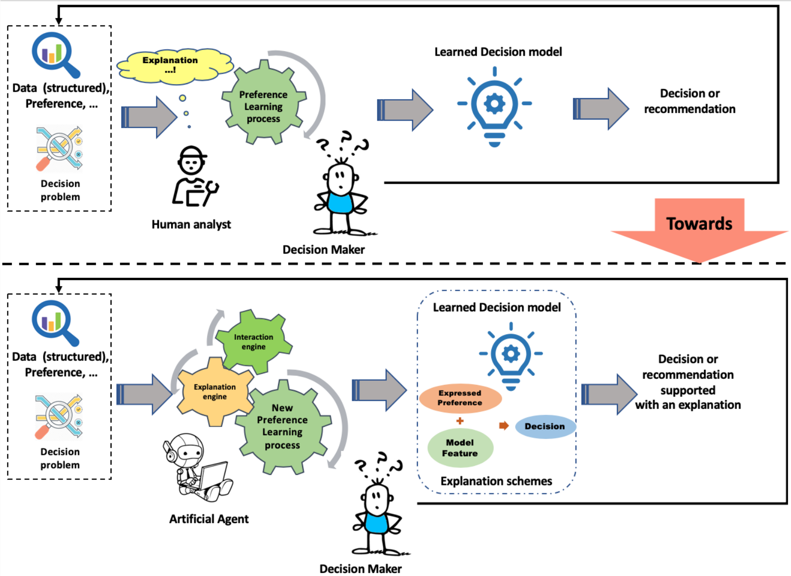
1. Positioning
2. Preference Elicitation and Learning
3. Explanation, argumentation,...
4. Towards Accountability!

Positioning

Decision Aiding Process



Decision Aiding Process ⇝ Artificial Agent?



Decision Aiding: examples

- **A decision maker needs to adjudicate a situation...**

  - A traveling scientist chooses a hotel.
  - A committee reviews candidates.

Decision Aiding: examples

- **A decision maker needs to adjudicate a situation...**
  - A traveling scientist chooses a hotel.
  - A committee reviews candidates.
- **There are several conflicting points of view**

Hotel	☆☆	🍴	🚆	$
h_A	$4^\star$	no	35 min	120 $
h_B	$4^\star$	yes	50 min	160 $
h_C	$2^\star$	yes	20 min	50 $
h_D	$2^\star$	no	30 min	40 $

1: $a \succ_1 b \succ_1 f \succ_1 e \succ_1 c \succ_1 d$

2: $e \succ_2 b \succ_2 c \succ_2 d \succ_2 a \succ_2 f$

3: $f \succ_3 a \succ_3 b \succ_3 d \succ_3 e \succ_3 c$

4: $d \succ_4 a \succ_4 c \succ_4 e \succ_4 f \succ_4 b$

5: $c \succ_5 e \succ_5 b \succ_5 f \succ_5 d \succ_5 a$

Decision Aiding: examples

- **A decision maker needs to adjudicate a situation...**
  - A traveling scientist chooses a hotel.
  - A committee reviews candidates.
- **There are several conflicting points of view**

Hotel	☆☆	🍴	🚆	$
h_A	$4^\star$	no	35 min	120 $
h_B	$4^\star$	yes	50 min	160 $
h_C	$2^\star$	yes	20 min	50 $
h_D	$2^\star$	no	30 min	40 $

1: $a \succ_1 b \succ_1 f \succ_1 e \succ_1 c \succ_1 d$

2: $e \succ_2 b \succ_2 c \succ_2 d \succ_2 a \succ_2 f$

3: $f \succ_3 a \succ_3 b \succ_3 d \succ_3 e \succ_3 c$

4: $d \succ_4 a \succ_4 c \succ_4 e \succ_4 f \succ_4 b$

5: $c \succ_5 e \succ_5 b \succ_5 f \succ_5 d \succ_5 a$

- **An analyst provides support**

Decision Aiding: examples

- **A decision maker needs to adjudicate a situation...**

  - A traveling scientist chooses a hotel.
  - A committee reviews candidates.

- **There are several conflicting points of view**

Hotel	☆☆	🍴	🚆	$
h_A	$4^\star$	no	35 min	120 $
h_B	$4^\star$	yes	50 min	160 $
h_C	$2^\star$	yes	20 min	50 $
h_D	$2^\star$	no	30 min	40 $

1: $a \succ_1 b \succ_1 f \succ_1 e \succ_1 c \succ_1 d$
2: $e \succ_2 b \succ_2 c \succ_2 d \succ_2 a \succ_2 f$
3: $f \succ_3 a \succ_3 b \succ_3 d \succ_3 e \succ_3 c$
4: $d \succ_4 a \succ_4 c \succ_4 e \succ_4 f \succ_4 b$
5: $c \succ_5 e \succ_5 b \succ_5 f \succ_5 d \succ_5 a$

- **An analyst provides support**
- **The process is subject to validation**

Principled decision aiding

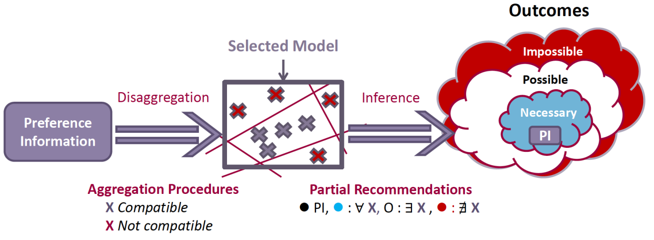
- not answering a query, but designing an **aggregation procedure** answering **any** query
- an aggregation **model** contains aggregation procedures satisfying common properties.
- a model is **selected** considering decision stance, expressiveness, tractability.
- The selected model is **elicited**, so as to determine a specific aggregation procedure

Preference Elicitation and Learning

Model-based aggregation of preferences

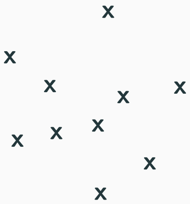
Approaches to model elicitation, based on collected Preference Information

Approach	Summary	Pros	Cons
Complete	Measuring via standard sequences of questions	Unequivocal	Demanding
Partial	Learning from the DM's statements	Efficient	Arbitrary
Robust	**Solving for every possible completion**	**Cautious**	**Indecisive**



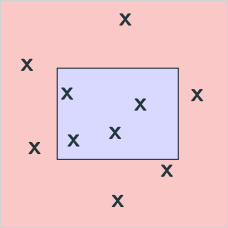
A toy example – Description

- Elicitation of a sorting model (= a classifier)
- Two categories : ✱/ 💣
- Alternatives (= data points) are points in the 2D plane
- Parameter = a cartesian products of intervals, i.e. a rectangle parallel to the axes
- Decision rule : points inside the rectangle are ✱, others are 💣



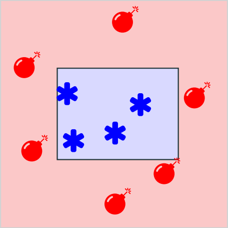
A toy example – Description

- Elicitation of a sorting model (= a classifier)
- Two categories : ✱/ 💣
- Alternatives (= data points) are points in the 2D plane
- Parameter = a cartesian products of intervals, i.e. a rectangle parallel to the axes
- Decision rule : points inside the rectangle are ✱, others are 💣



A toy example – Description

- Elicitation of a sorting model (= a classifier)
- Two categories : ✱/ 💣
- Alternatives (= data points) are points in the 2D plane
- Parameter = a cartesian products of intervals, i.e. a rectangle parallel to the axes
- Decision rule : points inside the rectangle are ✱, others are 💣

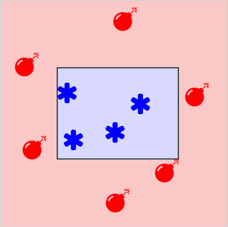


Approaches to elicitation

Approaches to elicitation

Learning

Computes 'fittest' model of the class

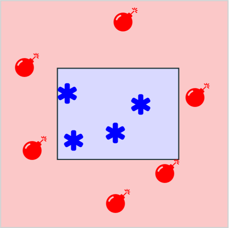


- Efficient: yields compiled knowledge
- Arbitrary wrt the incompleteness of information
- Opaque: lack of traceability

Approaches to elicitation

Learning

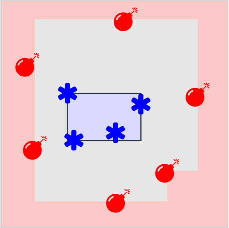
Computes 'fittest' model of the class



- Efficient: yields compiled knowledge
- Arbitrary wrt the incompleteness of information
- Opaque: lack of traceability

Robust Induction

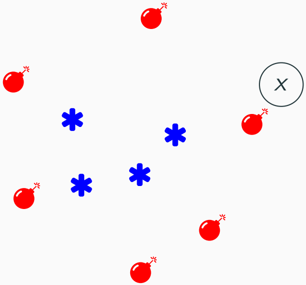
Solves for every possible completion



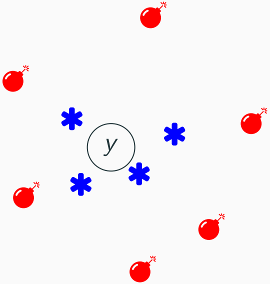
- Indecisive
- Inefficient: runtime depends on $|KB|$
- Cautious
- Traceable

Local recommendations supported

Argument for necessarily 💣

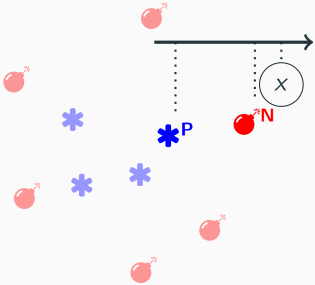


Argument for necessarily ✱



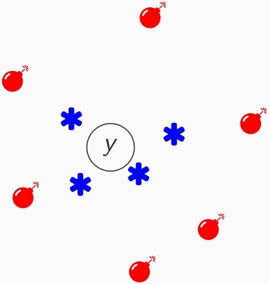
Local recommendations supported

Argument for necessarily 💣



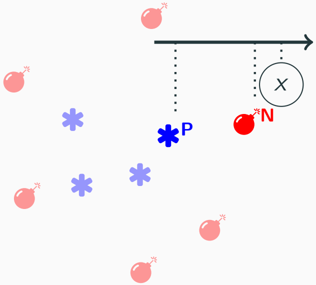
There is a positive example ✱^{P} and a negative example 💣^{N} and an axis i such that the values x_i and ✱$_i^N$ lie on both sides of 💣$_i^N$. Hence x is necessarily negative.

Argument for necessarily ✱



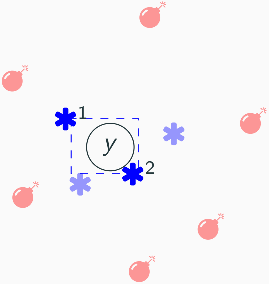
Local recommendations supported

Argument for necessarily 💣



There is a positive example ✱^{P} and a negative example 💣^{N} and an axis i such that the values x_i and ✱$_i^N$ lie on both sides of 💣$_i^N$. Hence x is necessarily negative.

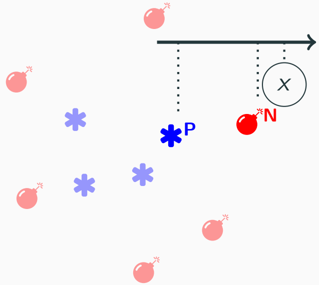
Argument for necessarily ✱



There are two positive examples ✱^{1} and ✱^{2} such that y lies in the rectangle with diagonal ✱^{1} and ✱^{2}. Hence, y is necessarily positive.

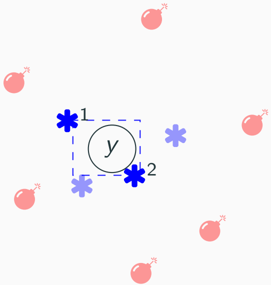
Local recommendations supported by argument schemes

Argument for necessarily 💣



There is a positive example ✱^{P} and a negative example 💣^{N} and an axis i such that the values x_i and ✱$_i^N$ lie on both sides of 💣$_i^N$. Hence x is necessarily negative.

Argument for necessarily ✱



There are two positive examples ✱^{1} and ✱^{2} such that y lies in the rectangle with diagonal ✱^{1} and ✱^{2}. Hence, y is necessarily positive.

Explanation, argumentation,...

Back to MCDA! Context Matters

Validation with the intention of delegation

- Supervised context: Human expert vs AI trainee
- Make sure AI takes good enough decision for good enough reasons

Elicitation, with the intention of mutual understanding

- Collaborative context: Human user ('DM') and AI analyst
- Make sure their respective representations align well enough

Accountability with the intention of justice

- Context: DM vs 3^{rd} party stakeholders of the decision
- Make sure their respective duties and rights have been duly accounted

Back to MCDA! Context Matters

Validation with the intention of delegation

- Supervised context: Human expert vs AI trainee
- Make sure AI takes good enough decision for good enough reasons

Elicitation, with the intention of mutual understanding

- Collaborative context: Human user ('DM') and AI analyst
- Make sure their respective representations align well enough

Accountability with the intention of justice

- Context: DM vs 3^{rd} party stakeholders of the decision
- Make sure their respective duties and rights have been duly accounted

⇝ **This necessary contextualization should be specified during the decision aiding process**

Explanation call for defeasible reasoning engine

When dealing with preferences, there is no no ground truth to be discovered

- paradoxes in Decision Theory
- impossibility results in Social Choice
- constructivist assumption
- right to call for a public deliberation

$\leadsto$ ties nicely with ‘provably beneficial AI’ assumptions

Explanations are called upon in case of conflicting views

- explainer may be right, wrong, or ...
- explainee mybe right, wrong, or ...

Explanation: Back to MCDA!

Our approach

- Explaining the reasoning itself, not its product
- A dialectical take to preference information
- Robust elicitation as deductive reasoning

Explanation: Back to MCDA!

Our approach

- Explaining the reasoning itself, not its product
- A dialectical take to preference information
- Robust elicitation as deductive reasoning

Purposes of an explanations

- allowing to scrutinize the reasoning, towards contestability
- highlight causes
- should be intelligible and sincere
- can be challenged

Using the argument schemes template for explanations

We propose to build explanations using the argument scheme template:

- computational model of a real-world argument [Walton, 1996]
- formally = operator tying premises to conclusions
- vehicle in a structured dialogue
- subject to critical questions: undercutting a premiss, rebutting a conclusion, warranting a rule

(Argumentation Theory)

Argumentation is a branch of the logic which is interested in non-monotonic logic *(Defeasible Reasoning)*. It formalizes this reasoning through the dialectical interaction between arguments and counter arguments.

Formal theories of argumentation have been extensively developed in the field of AI, in particular:

- by developing abstract models of aggregation of arguments [Dung, 1995];
- by using the structures of argumentation scheme as a tool for knowledge representation [Walton, 1996].

Application: Approval Sorting Procedure

Situation

- A comittee meets to decide upon the osrtiong of a number of candidates into two categories (to accept or not, projects to fund or not, etc.)
- It applies a decision process which is public, the outcomes are public as well, however the details of the votes are sensitive and should not be available.

Question?

To what extent can we make the decisions of a committee using approval sorting accountable while preserving as much as possible the details of the votes?

Approval sorting procedure

1. Preferences

Each juror has preferences over the candidates

1: $a \succ_1 b \succ_1 f \succ_1 e \succ_1 c \succ_1 d$

2: $e \succ_2 b \succ_2 c \succ_2 d \succ_2 a \succ_2 f$

3: $f \succ_3 a \succ_3 b \succ_3 d \succ_3 e \succ_3 c$

4: $d \succ_4 a \succ_4 c \succ_4 e \succ_4 f \succ_4 b$

5: $c \succ_5 e \succ_5 b \succ_5 f \succ_5 d \succ_5 a$

Approval sorting procedure

1. Preferences

Each juror has preferences over the candidates

1: $a \succ_1 b \succ_1 f \succ_1 e \succ_1 c \succ_1 d$

2: $e \succ_2 b \succ_2 c \succ_2 d \succ_2 a \succ_2 f$

3: $f \succ_3 a \succ_3 b \succ_3 d \succ_3 e \succ_3 c$

4: $d \succ_4 a \succ_4 c \succ_4 e \succ_4 f \succ_4 b$

5: $c \succ_5 e \succ_5 b \succ_5 f \succ_5 d \succ_5 a$

2. Approval

Each juror approves a subset of candidates

Individual rationality: approved are on the left

1: $\boxed{a \succ_1 b \succ_1 f} \succ_1 e \succ_1 c \succ_1 d$

2: $\boxed{e \succ_2 b \succ_2 c \succ_2 d} \succ_2 a \succ_2 f$

3: $\boxed{f \succ_3 a} \succ_3 b \succ_3 d \succ_3 e \succ_3 c$

4: $\boxed{d \succ_4 a \succ_4 c} \succ_4 e \succ_4 f \succ_4 b$

5: $\boxed{c} \succ_5 e \succ_5 b \succ_5 f \succ_5 d \succ_5 a$

Approval sorting procedure

1. Preferences

Each juror has preferences over the candidates

1: $a \succ_1 b \succ_1 f \succ_1 e \succ_1 c \succ_1 d$
2: $e \succ_2 b \succ_2 c \succ_2 d \succ_2 a \succ_2 f$
3: $f \succ_3 a \succ_3 b \succ_3 d \succ_3 e \succ_3 c$
4: $d \succ_4 a \succ_4 c \succ_4 e \succ_4 f \succ_4 b$
5: $c \succ_5 e \succ_5 b \succ_5 f \succ_5 d \succ_5 a$

2. Approval

Each juror approves a subset of candidates

Individual rationality: approved are on the left

1: $\boxed{a \succ_1 b \succ_1 f} \succ_1 e \succ_1 c \succ_1 d$
2: $\boxed{e \succ_2 b \succ_2 c \succ_2 d} \succ_2 a \succ_2 f$
3: $\boxed{f \succ_3 a} \succ_3 b \succ_3 d \succ_3 e \succ_3 c$
4: $\boxed{d \succ_4 a \succ_4 c} \succ_4 e \succ_4 f \succ_4 b$
5: $\boxed{c} \succ_5 e \succ_5 b \succ_5 f \succ_5 d \succ_5 a$

3. Tallying

Each candidate is approved by a coalition of jurors

a :	$\{\mathbf{1}, \mathbf{3}, \mathbf{4}\}$
b :	$\{\mathbf{1}, \mathbf{2}\}$
c :	$\{\mathbf{2}, \mathbf{4}, \mathbf{5}\}$
d :	$\{\mathbf{2}, \mathbf{4}\}$
e :	$\varnothing$
f :	$\{\mathbf{1}, \mathbf{3}\}$

Approval sorting procedure

1. Preferences

Each juror has preferences over the candidates

1: $a \succ_1 b \succ_1 f \succ_1 e \succ_1 c \succ_1 d$
2: $e \succ_2 b \succ_2 c \succ_2 d \succ_2 a \succ_2 f$
3: $f \succ_3 a \succ_3 b \succ_3 d \succ_3 e \succ_3 c$
4: $d \succ_4 a \succ_4 c \succ_4 e \succ_4 f \succ_4 b$
5: $c \succ_5 e \succ_5 b \succ_5 f \succ_5 d \succ_5 a$

2. Approval

Each juror approves a subset of candidates

Individual rationality: approved are on the left

1: $\boxed{a \succ_1 b \succ_1 f} \succ_1 e \succ_1 c \succ_1 d$
2: $\boxed{e \succ_2 b \succ_2 c \succ_2 d} \succ_2 a \succ_2 f$
3: $\boxed{f \succ_3 a} \succ_3 b \succ_3 d \succ_3 e \succ_3 c$
4: $\boxed{d \succ_4 a \succ_4 c} \succ_4 e \succ_4 f \succ_4 b$
5: $\boxed{c} \succ_5 e \succ_5 b \succ_5 f \succ_5 d \succ_5 a$

3. Tallying

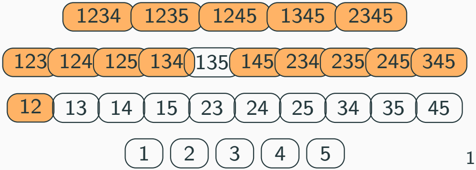
Each candidate is approved by a coalition of jurors

a :	$\{\mathbf{1}, \mathbf{3}, \mathbf{4}\}$
b :	$\{\mathbf{1}, \mathbf{2}\}$
c :	$\{\mathbf{2}, \mathbf{4}, \mathbf{5}\}$
d :	$\{\mathbf{2}, \mathbf{4}\}$
e :	$\varnothing$
f :	$\{\mathbf{1}, \mathbf{3}\}$

4. Aggregation

Sufficient coalitions of jurors are latent

Collective rationality: Sufficient coalitions are above insufficient ones



1234 1235 1245 1345 2345

123 124 125 134 135 145 234 235 245 345

12 13 14 15 23 24 25 34 35 45

1 2 3 4 5

Approval sorting procedure

1. Preferences

Each juror has preferences over the candidates

1: $a \succ_1 b \succ_1 f \succ_1 e \succ_1 c \succ_1 d$
2: $e \succ_2 b \succ_2 c \succ_2 d \succ_2 a \succ_2 f$
3: $f \succ_3 a \succ_3 b \succ_3 d \succ_3 e \succ_3 c$
4: $d \succ_4 a \succ_4 c \succ_4 e \succ_4 f \succ_4 b$
5: $c \succ_5 e \succ_5 b \succ_5 f \succ_5 d \succ_5 a$

2. Approval

Each juror approves a subset of candidates

Individual rationality: approved are on the left

1: $\boxed{a \succ_1 b \succ_1 f} \succ_1 e \succ_1 c \succ_1 d$
2: $\boxed{e \succ_2 b \succ_2 c \succ_2 d} \succ_2 a \succ_2 f$
3: $\boxed{f \succ_3 a} \succ_3 b \succ_3 d \succ_3 e \succ_3 c$
4: $\boxed{d \succ_4 a \succ_4 c} \succ_4 e \succ_4 f \succ_4 b$
5: $\boxed{c} \succ_5 e \succ_5 b \succ_5 f \succ_5 d \succ_5 a$

3. Tallying

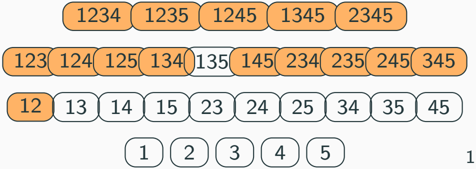
Each candidate is approved by a coalition of jurors

a :	$\{\mathbf{1},\mathbf{3},\mathbf{4}\}$
b :	$\{\mathbf{1},\mathbf{2}\}$
c :	$\{\mathbf{2},\mathbf{4},\mathbf{5}\}$
d :	$\{\mathbf{2},\mathbf{4}\}$
e :	$\varnothing$
f :	$\{\mathbf{1},\mathbf{3}\}$

4. Aggregation

Sufficient coalitions of jurors are latent

Collective rationality: Sufficient coalitions are above insufficient ones



5. Assignment

a ↦ ✔, b ↦ ✔, c ↦ ✔,
d ↦ ✘, e ↦ ✘, f ↦ ✘.

The Inverse NCS problem

Formulation

Given the jurors'preferences and a final assignment, can it be represented in the NCS model? I.e. is there a value of the parameter so that the final assignment has been obtained by applying NCS on the input preferences?

Resolution [Belahcene et al, Computers & OR 2018]

- the NCS model can be described in propositional logic
- feasibility of representation can be checked with a SAT solver
- a complete representation of the parameter space is exponential in #jurors

The Inverse NCS problem (contd.)

Pairwise separation

An assignment is pairwise separated if there is an individually rational tuple of approved set such that, for every pair of candidates (g accepted, b rejected), there is at least one juror approving g but not b.

Representation theorem

An assignment can be represented in the NCS model iff it is pairwise separated

Corollaries

- There is a short positive certificate for Inv-NCS
- Inv-NCS is NP-complete
- explanations for possibility based on pairwise separation are sound, complete and rather short

What about negative certificates?

No easy answer. As feasibility is decsribed by a CSP, the Minimal Unsatisfiable Subsets (MUSes) of clauses can be seen as an explanation of impossibility/necessity

Accountability ?

Important issue !
DARPA XAI program, [Doshi-Velez et al., 2017], [Wachter et al., 2017], . . .

- **Procedural regularity:** [Kroll et al., 2017]
  Has the committee complied with the publicly announced rule?
  ☞ checked by an audit agency
- **Contestability:**
  Could the decision for a single candidate have been reversed?
  ☞ refers to a number of reference case, e.g. jurisprudence
- **Sensitive information:**
  The details of the vote should be disclosed a minima

Auditing conformity: the design space

An independent audit agency has to check that the decision of the committee is a possible outcome of a NCS aggregation procedure ($\rightsquigarrow$ transparency).
Several rules can be envisioned:

1. The committee fully discloses the preferences of the jurors – the audit agency solves the NP-hard inverse problem with the SAT formulation
2. The committee also fully discloses the votes of the jurors – the audit agency solves the polytime inverse problem with fixed approved sets
3. The committee adopts an active stance and assumes the burden of proof. It leverages our Theorem (pairwise separation) to provide a certificate of feasibility

Auditing conformity: explanations of feasibility

- Public assignment:

  a : ✔, b : ✔, c : ✔, d : ✘, e : ✘, f : ✘.

- Private: jurors'approvals

  1: $\boxed{a \succ_1 b \succ_1 f} \succ_1 e \succ_1 c \succ_1 d$

  2: $\boxed{e \succ_2 b \succ_2 c \succ_2 d} \succ_2 a \succ_2 f$

  3: $\boxed{f \succ_3 a} \succ_3 b \succ_3 d \succ_3 e \succ_3 c$

  4: $\boxed{d \succ_4 a \succ_4 c} \succ_4 e \succ_4 f \succ_4 b$

  5: $\boxed{c} \succ_5 e \succ_5 b \succ_5 f \succ_5 d \succ_5 a$

- Public certificate:

1: $\boxed{a, b} \quad \succ_1 \quad e, d$

2: $\boxed{b} \quad \succ_2 \quad f$

4: $\boxed{a} \quad \succ_4 \quad f$

5: $\boxed{c} \quad \succ_5 \quad e, f, d$

		✘		
		d	e	f
	a	**1**	**1**	**4**
✔	b	**1**	**1**	**2**
	c	**5**	**5**	**5**

☞ Explanations are formalized into argument schemes – operators tying premisses to a conclusion [Walton, 1996]

Bad news: sometimes, explanations need to reference every juror

Auditing conformity: explanations of feasibility

$\mathbf{1}$: $\boxed{a, b} \quad \succ_1 \quad e, d$

according to $\mathbf{1}$, b is approved (and so is a which is better than b) whereas e is not (and neither is d which is worse than e), hence the procedure is able to discriminate a, b from d, e;

Definition (Argument Scheme (AS1))

We say a tuple $\langle (i_1, g_1, G_1, b_1, B_1), \ldots, (i_n, g_n, G_n, b_n, B_n) \rangle$ instantiates the argument scheme AS1 supporting the assignment α if: i) for all $k \in \{1 \ldots n\}$, $i_k \in \mathcal{N}$, $g_k \in G_k$, $\alpha(G_k) = \{\text{Good}\}$, $\forall g \in G_k, g \succsim_{i_k} g_k$, $b_k \in B_k$, $\alpha(B_k) = \{\text{Bad}\}$, $\forall b \in B_k, b_k \succsim_{i_k} b$ and $g_k \succ_{i_k} b_k$; and ii) $\bigcup_{k \in \{1 \ldots n\}} G_k \times B_k = \alpha^{-1}(\text{Good}) \times \alpha^{-1}(\text{Bad})$

Justifying individual decisions entailed by the jurisprudence

is unhappy ...

Justifying individual decisions entailed by the jurisprudence

is unhappy ...

The committee justifies its decision...

- Reference assignment (jurisprudence) $\alpha^\star$: a : ✔, b : ✔, c : ✔, d : ✘, e : ✘, f : ✘
- Position w.r.t. reference cases

1: $a \succ_1 b \succ_1 f \succ_1 e \succ_1 c \succ_1 d \succ_1$

2: $e \succ_2 b \succ_2 c \succ_2 d \succ_2 a \succ_2 f \succ_2$

3: $f \succ_3 a \succ_3 b \succ_3 d \succ_3$ $\succ_3 e \succ_3 c$

4: $d \succ_4 a \succ_4 c \succ_4 e \succ_4$ $\succ_4 f \succ_4 b$

5: $c \succ_5 e \succ_5 b \succ_5 f \succ_5$ $\succ_5 d \succ_5 a$

- It is not possible to represent $\alpha^\star \cup ($, ✔$)$ in the NCS model. Thus, is necessarily ✘.

Justifying individual decisions entailed by the jurisprudence

... by exhibiting some deadlock

- Assume is ✔, and consider the 3 pairs of candidates (✔, ✘):
  $\langle (c, e), ($ $, d), ($ $, f)\rangle$
- Each pair should be discriminated by at least one juror, but this is not possible simultaneously: **1**, **2**, **3** can not discriminate any pair, and **4**, **5** cannot discriminate more than one pair each, and there are 3 pairs to discriminate

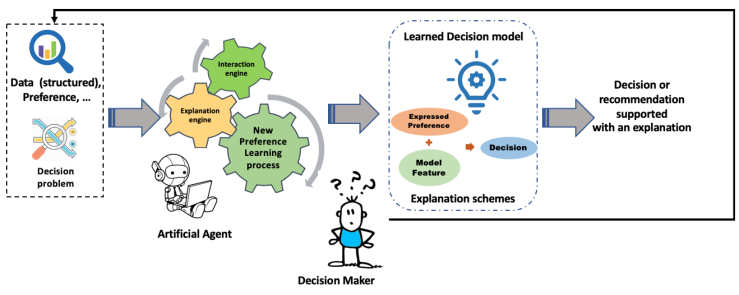
Justifying individual decisions entailed by the jurisprudence

… by exhibiting some deadlock

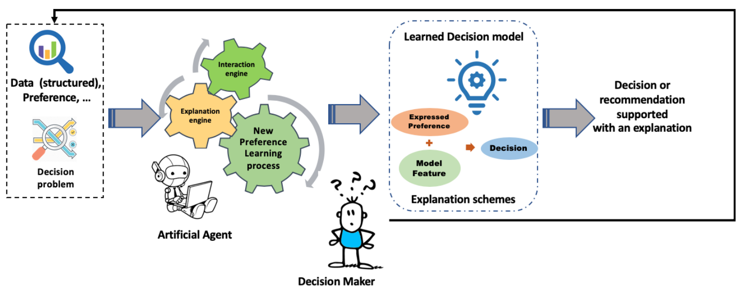
- Assume is ✔, and consider the 3 pairs of candidates (✔, ✘): $\langle (c, e),$ ($, d), ($ $, f)\rangle$
- Each pair should be discriminated by at least one juror, but this is not possible simultaneously: **1**, **2**, **3** can not discriminate any pair, and **4**, **5** cannot discriminate more than one pair each, and there are 3 pairs to discriminate

- this scheme is a sufficient condition … but necessary?
- sound, number of pairs ≡ measure of complexity
- complete ?

Towards Accountability!


Data (structured), Preference, ...
Decision problem
Interaction engine
Explanation engine
New Preference Learning process
Artificial Agent
Decision Maker
Learned Decision model
Expressed Preference
+
Model Feature
Decision
Explanation schemes
Decision or recommendation supported with an explanation

What about the dialectical aspect?



- The accountability in decision aiding has strong dialectical and adversarial components
- It could aptly be represented as a discussion between the decision maker and an agent discussing critically and in good faith various options.
- argumentation-based dialogue game [Labreuche et al, AAMAS 2015]

Ch. Labreuche, N. Maudet, W. Ouerdane, S. Parsons.. *A dialogue game for recommendation with adaptive preference models*. AAMAS'2015.

Towards Accounatbility!

- Decision aiding situation are pervasive in our daily life and in our society;
- We propose to build decision aiding systems that are *accountable* for their recommendations.
- Using formal tools from Decision theory and Artificial Intelligence aiming at
    - taking into account the decision maker's preferences and expertise
    - providing sound and complete explanation
    - handling the non-monotonic reasoning of a human decision maker