Designing and Computing Explanations for a Multicriteria Additive Value-based Model

Manuel Amoussou[1], Khaled Belahcene[2],
Nicolas Maudet[3], Vincent Mousseau[1], Wassila Ouerdane[1]

[1]MICS, CentraleSupélec, Université Paris-Saclay
[2] Université de Technologie de Compiègne, CNRS, UMR 7253 Heudiasyc
[3] Sorbonne Université, CNRS, Laboratoire d'Informatique de Paris 6

PhD thesis, May 2020

Interactive explanations in Multicriteria decision aiding.

- ❑ Elicitation procedures improvement
- ❑ Justification of decision models outcomes

Outline

Decision Context description

Decision Context description

Choosing the **best** 🥇 alternative among

a **finite** set $\mathbb{A}$ of alternatives

described over

m **bi-level scales** of criteria where

preferences are **additive**.

$\omega : \langle \omega_i \rangle_{i \in [m]}$ with $\omega_i : [m] \rightarrow \mathbb{R}^+$

Performance table

	a	b	c	d	e	f	g
W	✗	✓	✗	✗	✗	✓	✓
X	✗	✗	✓	✓	✓	✗	✗
Y	✓	✗	✗	✓	✗	✗	✗
Z	✗	✗	✓	✗	✓	✗	✓

W $\equiv$ **bfg**
X $\equiv$ **cde**
Y $\equiv$ **ad**
Z $\equiv$ **ceg**

Preference models representation

	a	b	c	d	e	f	g	🥇
ω^1	128	126	77	59	52	41	37	W
ω^2	128	77	126	59	52	41	37	X
ω^3	126	129	77	59	52	40	37	W

$$\omega^1(\text{W}) = 126 + 41 + 37 = 204$$

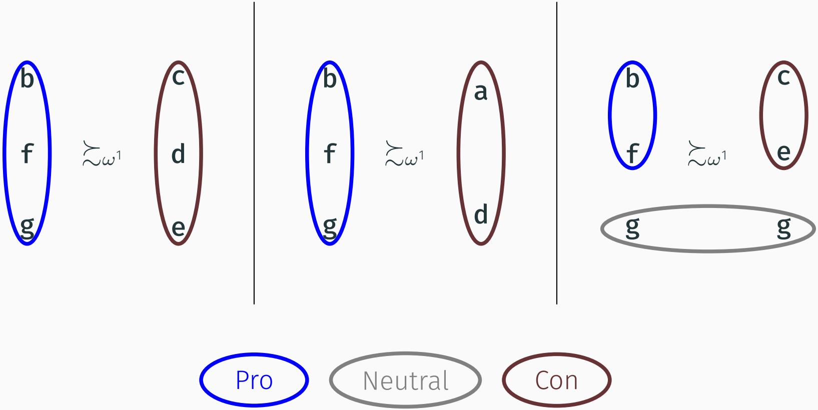
Explanation Design

Pro, Con and Neutral criteria

$$\omega^1 = \{\mathbf{a}: 128,\ \mathbf{b}: 126,\ \mathbf{c}: 77,\ \mathbf{d}: 59,\ \mathbf{e}: 52,\ \mathbf{f}: 41,\ \mathbf{g}: 37\}$$

$$W \equiv \mathbf{bfg}\ ;\ X \equiv \mathbf{cde}\ ;\ Y \equiv \mathbf{ad}\ ;\ Z \equiv \mathbf{ceg}$$

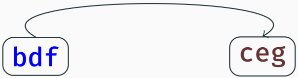
$$W \succsim_{\omega^1} X \mid W \succsim_{\omega^1} Y \mid W \succsim_{\omega^1} Z$$



Philosophical Foundations – What is Explanation ?[1]

"To explain an event is to provide some information about its causal history".

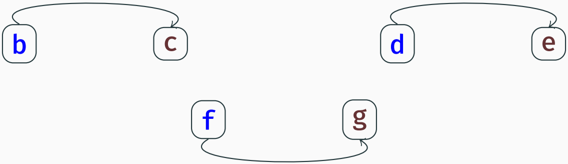
$\Longrightarrow$ ***an event:*** a pairwise comparison of alternatives given ω.



$\omega = \{\mathbf{a} : 128,\ \mathbf{b} : 126,\ \mathbf{c} : 77,\ \mathbf{d} : 59,\ \mathbf{e} : 52,\ \mathbf{f} : 41,\ \mathbf{g} : 37\}$

"Any event without a cause are root causes".

$\Longrightarrow$ ***Root causes:*** some ***atomic*** pairwise comparisons of bundles of criteria given ω.



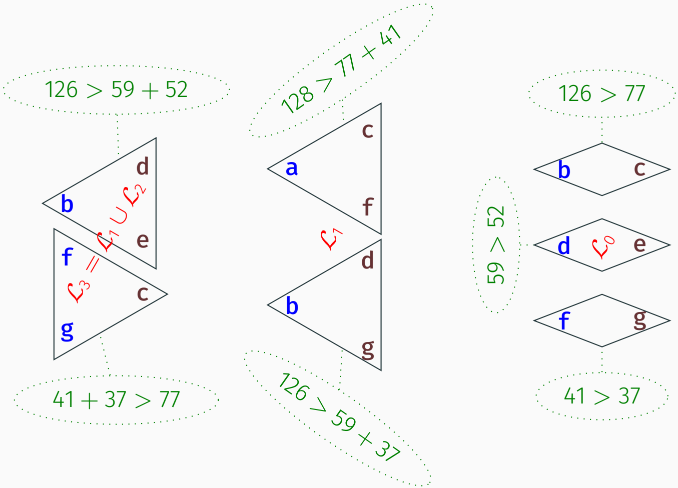
$\omega = \{\mathbf{a} : 128,\ \mathbf{b} : 126,\ \mathbf{c} : 77,\ \mathbf{d} : 59,\ \mathbf{e} : 52,\ \mathbf{f} : 41,\ \mathbf{g} : 37\}$

[1] Explanation in Artificial Intelligence: Insights from the Social Sciences, Tim Miller, 2019

(Covering) Explanation through decomposition languages

$\omega = \{\mathbf{a} : 128,\ \mathbf{b} : 126,\ \mathbf{c} : 77,\ \mathbf{d} : 59,\ \mathbf{e} : 52,\ \mathbf{f} : 41,\ \mathbf{g} : 37\}$

pro criteria vs. con criteria

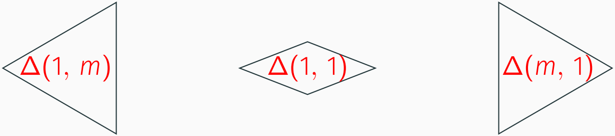


Decomposition Languages

- ❑ $\mathcal{L}_0 = \Delta(1, 1)$
- ❑ $\mathcal{L}_1 = \Delta(1, m)$
- ❑ $\mathcal{L}_2 = \Delta(m, 1)$
- ❑ $\mathcal{L}_3 = \mathcal{L}_1 \cup \mathcal{L}_2$

(Covering) Explanation through decomposition languages

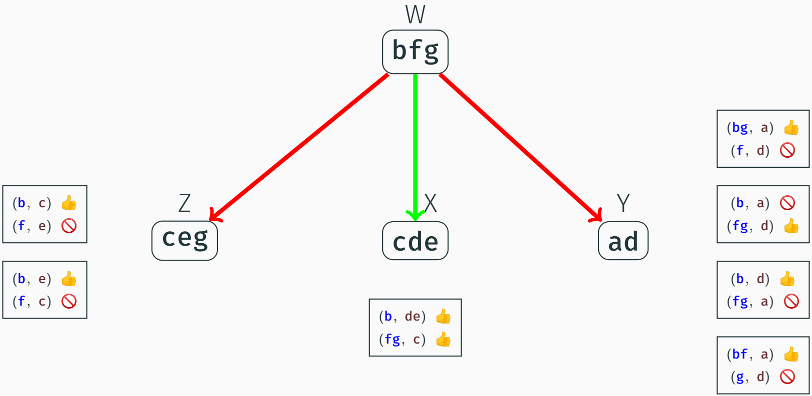
The chosen decomposition languages



- \+ Cognitively easy to grasp.
- \+ Easily scriptable in a natural language.
- \- Not complete.

Our languages are incomplete !!!

$\omega = \{\mathbf{a}: 128,\ \mathbf{b}: 126,\ \mathbf{c}: 77,\ \mathbf{d}: 59,\ \mathbf{e}: 52,\ \mathbf{f}: 41,\ \mathbf{g}: 37\}$



$\rightarrow$: not $\mathcal{L}_3$−covering explainable

$\rightarrow$: $\mathcal{L}_3$−covering explainable

Numerical Experiments (1/2)

Instance: $(\omega, \mathbb{A})$

- $|\mathbb{A}| = 10$
- $|\omega| = m \in [\![6; 15]\!]$

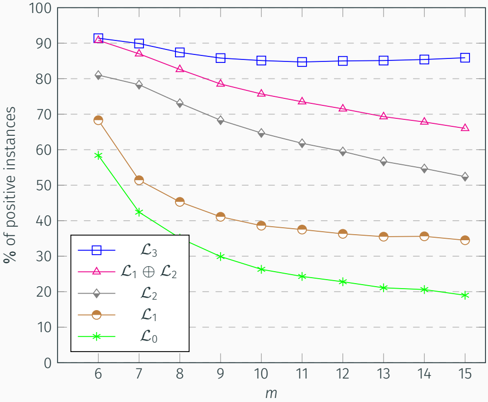
Sample: 500.000 instances

- 500 sets $\mathbb{A}$
- 1000 score functions ω

$\mathbb{A}$ set characteristics:

- no Pareto dominance
- significativity of all criteria

Percentage of positive instances



An instance is positive if and only if all "direct" statements (x^*, y) are $\mathcal{L}$−covering explainable.

m	6	7	8	9	10	11	12	13	14	15
$\mathcal{L}_3$	91.4%	89.9%	87.4%	85.8%	85.1%	84.7%	85.0%	85.1%	85.4%	85.9%
$\mathcal{L}_1 \oplus \mathcal{L}_2$	90.8%	87.0%	82.6%	78.5%	75.7%	73.5%	71.5%	69.3%	67.8%	66.0%
$\mathcal{L}_1$	81.0%	78.3%	73.1%	68.3%	64.7%	61.8%	59.1%	56.7%	54.7%	52.4%
$\mathcal{L}_2$	68.3%	51.4%	45.3%	41.1%	38.6%	37.5%	36.3%	35.5%	35.6%	34.5%
$\mathcal{L}_0$	58.4%	42.4%	35.1%	29.9%	26.3%	24.3%	22.8%	21.1%	20.6%	19.0%

Explanation Computation

Covering explanation: Computational complexity

Instance:

- A set $\mathbb{PRO}$ of pro criteria and a set $\mathbb{CON}$ of con criteria
- The preference model representation ω

$\mathcal{L}_0$

Question: Is there a bijective function f from $\mathbb{CON}$ to $\mathbb{PRO}$ such that for all $j \in \mathbb{CON}$, $\omega(f(j)) \geq \omega(j)$?

Complexity class: P

$\mathcal{L}_1$

Question: Is there a function f from $\mathbb{PRO}$ to $2^{\mathbb{CON}}$ such that :

1. $f(i) \cap f(i') = \emptyset$ if $i \neq i'$ with $i,\ i' \in \mathbb{PRO}$
2. $\bigcup_{i \in \mathbb{PRO}} f(i) = \mathbb{CON}$
3. $\omega(i) \geq \sum_{j \in f(i)} \omega(j)$ for all $i \in \mathbb{PRO}$

Complexity class: NP-hard

$\mathcal{L}_2$

Question: Is there a application g from $\mathbb{CON}$ to $2^{\mathbb{PRO}} \setminus \{\emptyset\}$ such that :

1. $g(j) \cap g(j') = \emptyset$ if $j \neq j'$ with $j,\ j' \in \mathbb{CON}$
2. $\sum_{i \in g(j)} \omega(i) \geq \omega(j)$ for all $j \in \mathbb{CON}$

Complexity class: NP-hard

$\mathcal{L}_3$

Question: Are there two disjoint subsets $\mathbb{PRO}^1$ and $\mathbb{PRO}^2$ of $\mathbb{PRO}$, two disjoint subsets $\mathbb{CON}^1$ and $\mathbb{CON}^2$ of $\mathbb{CON}$, a function f from $\mathbb{PRO}^1$ to $2^{\mathbb{CON}^1}$ and a application g from $\mathbb{CON}^2$ to $2^{\mathbb{PRO}^2} \setminus \{\emptyset\}$ such that:

1. $\mathbb{PRO}^1 \cup \mathbb{PRO}^2 = \mathbb{PRO}$ and $\mathbb{CON}^1 \cup \mathbb{CON}^2 = \mathbb{CON}$
2. $f(i) \cap f(i') = \emptyset$ if $i \neq i'$ with $i,\ i' \in \mathbb{PRO}^1$
3. $\bigcup_{i \in \mathbb{PRO}^1} f(i) = \mathbb{CON}^1$
4. $\omega(i) \geq \sum_{j \in f(i)} \omega(j)$ for all $i \in \mathbb{PRO}^1$
5. $g(j) \cap g(j') = \emptyset$ if $j \neq j'$ with $j,\ j' \in \mathbb{CON}^2$
6. $\sum_{i \in g(j)} \omega(i) \geq \omega(j)$ for all $j \in \mathbb{CON}^2$

Complexity class: NP-hard

$\mathcal{L}_3$—*Covering explanation*: Computation using ILP

Inputs

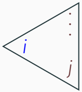
- The score function $\omega : \langle \omega_i \rangle_{i \in [m]}$.
- The pair (X, Y) to explain.

Notations

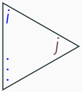
- *Pro* arguments set : $(X,\ Y)^+$
- *Con* arguments set : $(X,\ Y)^-$

(Binary) Variables

$$s^1_{ij} = \begin{cases} 1 & \text{if } j \in f(i). \\ 0 & \text{Otherwise.} \end{cases}$$



$$s^2_{ij} = \begin{cases} 1 & \text{if } i \in g(j). \\ 0 & \text{Otherwise.} \end{cases}$$



Constraints

- *Syntactic constraints*
  - For each *pro argument* i

$$s^1_{ij} + \sum_{j' \in (X,\ Y)^-} s^2_{ij'} \leq 1 \qquad \forall j \in (X,\ Y)^-$$

  - For each *con argument* j

$$\sum_{i' \in (X,\ Y)^+} s^1_{i'j} + s^2_{ij} \leq 1 \qquad \forall i \in (X,\ Y)^+$$

- *ω—compatibility constraints*
  - For each *pro argument* i

$$\omega_i \geq \sum_{j \in (X,\ Y)^-} \omega_j s^1_{ij}$$

  - For each *con argument* j

$$\sum_{i \in (X,\ Y)^+} (s^1_{ij} + s^2_{ij})\, \omega_i \geq \omega_j$$

Addressing incompleteness

$\omega = \{\mathbf{a} : 128,\ \mathbf{b} : 126,\ \mathbf{c} : 77,\ \mathbf{d} : 59,\ \mathbf{e} : 52,\ \mathbf{f} : 41,\ \mathbf{g} : 37\}$

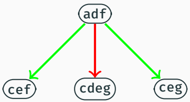
$\rightarrow$: not $\mathcal{L}_3$ — covering explainable

$\rightarrow$: $\mathcal{L}_3$ — covering explainable

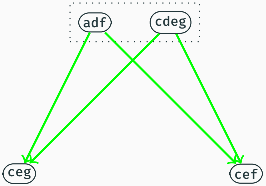
Addressing incompleteness

$\omega = \{\mathbf{a} : 128,\ \mathbf{b} : 126,\ \mathbf{c} : 77,\ \mathbf{d} : 59,\ \mathbf{e} : 52,\ \mathbf{f} : 41,\ \mathbf{g} : 37\}$

From 1—best



To 2—best recommendation



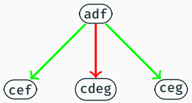
$\rightarrow$: not $\mathcal{L}_3$—covering explainable

$\rightarrow$: $\mathcal{L}_3$—covering explainable

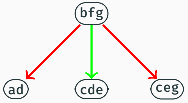
Addressing incompleteness

$\omega = \{\mathbf{a} : 128,\ \mathbf{b} : 126,\ \mathbf{c} : 77,\ \mathbf{d} : 59,\ \mathbf{e} : 52,\ \mathbf{f} : 41,\ \mathbf{g} : 37\}$

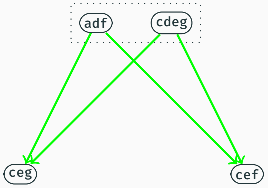
From 1—best



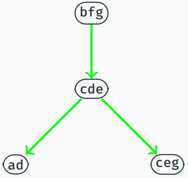
From *height* = 1



To 2—best recommendation



To *height* > 1

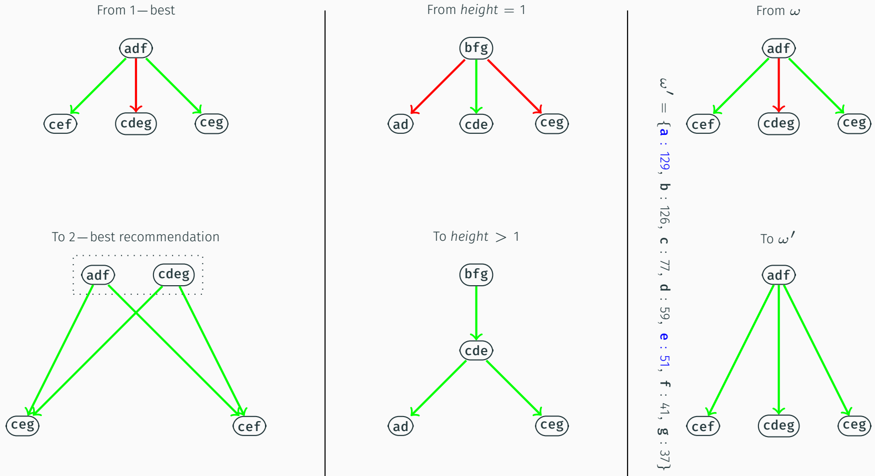


$\rightarrow$: not $\mathcal{L}_3$—covering explainable

$\rightarrow$: $\mathcal{L}_3$—covering explainable

Addressing incompleteness

$\omega = \{\mathbf{a} : 128,\ \mathbf{b} : 126,\ \mathbf{c} : 77,\ \mathbf{d} : 59,\ \mathbf{e} : 52,\ \mathbf{f} : 41,\ \mathbf{g} : 37\}$

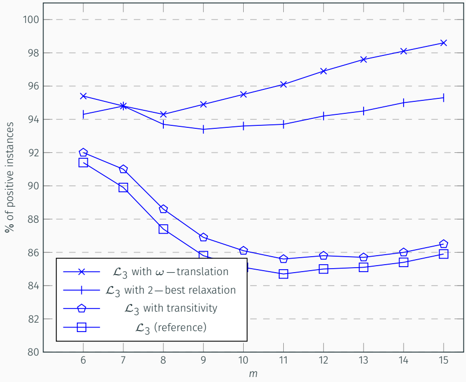


→ : not $\mathcal{L}_3$—covering explainable

→ : $\mathcal{L}_3$—covering explainable

Numerical experiments

Numerical Experiments (2/2)



Relaxation, Transitivity, Translation with $\mathcal{L}_3-$covering explanation

m	6	7	8	9	10	11	12	13	14	15
$\mathcal{L}_3$ with $\omega-$translation	95.4%	94.8%	94.3%	94.9%	95.5%	96.1%	96.9%	97.6%	98.1%	98.6%
$\mathcal{L}_3$ with 2−best relaxation	94.3%	94.8%	93.7%	93.4%	93.6%	93.7%	94.2%	94.5%	95.0%	95.3%
$\mathcal{L}_3$ with transitivity	92.0%	91.0%	88.6%	86.9%	86.1%	85.6%	85.8%	85.7%	86.0%	86.5%
$\mathcal{L}_3$ (reference)	91.4%	89.9%	87.4%	85.8%	85.1%	84.7%	85.0%	85.1%	85.4%	85.9%

Merci!