DA2PL 2020

Explaining Robust Additive Decision Models: Generation of Mixed Preference-Swaps chains Using MILP

Manuel Amoussou[1], Khaled Belahcene[2], Christophe Labreuche[3], Nicolas Maudet[4], Vincent Mousseau[1], Wassila Ouerdane[1]

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

Outline

1. Explanation in MultiCriteria Decision Aiding
2. Explanation of Robust Additive Decision Models
   - The decision situation
   - What do we explain ?
   - How do we explain?
3. Mixed-Explanation in binary core
   - Definition and Motivation
   - Computation using MILP
   - Extensions

Explanation in MultiCriteria Decision Aiding

Explanation in MCDA : MCDA in a nutshell!

- A **client** has a **problem** and asks for an **advice**.
- The **problem** involves **multi-attribute** alternatives.
- An **analyst** helps the **client** to find a **solution**.
- The **analyst** uses **formal models of rationality**.
- Such **model** inputs : data, **preference information** ...
- Such **model** output : a **recommendation**.

The Goals

- Interpretability ([Labreuche and Fossier, 2018])
- Accountability ([Belahcene et al., 2018])
- Trust and Acceptance from the user ([Labreuche et al., 2011])
- ...

Explanation in MCDA : a state of art

The Goals

- Interpretability ([Labreuche and Fossier, 2018])
- Accountability ([Belahcene et al., 2018])
- Trust and Acceptance from the user ([Labreuche et al., 2011])
- ...

The Means

- Argumentation ([Amgoud and Prade, 2009])
- Language ([Labreuche et al., 2011], [Labreuche et al., 2012])
- ...

Explanation of Robust Additive Decision Models

Contents

Explanation in MultiCriteria Decision Aiding

Explanation of Robust Additive Decision Models

- The decision situation
- What do we explain ?
- How do we explain?

Mixed-Explanation in binary core

- Definition and Motivation
- Computation using MILP
- Extensions

The decision situation

- **Two (02) characters** : a decision-maker ($\mathcal{DM}$) and an analyst ($\mathcal{DA}$).
- **A recommendation** : choosing, ranking or sorting.
- **A finite set of alternatives** : $\mathbb{A}$
    - n criteria
    - $\mathbb{X}_i = \{x_i^1, x_i^2, \ldots, x_i^{r_i-1}, x_i^{r_i}\}$ with $x_i^{r_i} \succsim_i x_i^{r_i-1} \succsim_i \cdots \succsim_i x_i^2 \succsim_i x_i^1$
    - $\mathbb{A} \subset \mathbb{X} = \prod_{i=1}^{n} \mathbb{X}_i$.
- **Preference information ($\mathbb{PI}$)** : holistic pairwise comparisons
- **Preference representation** : additive multi-attribute value model

$$x \succsim y \Leftrightarrow \mathcal{U}(x) = \sum_{i=1}^{n} u_i(x_i) \geq \sum_{i=1}^{n} u_i(y_i) = \mathcal{U}(y) \tag{1}$$

The decision situation : Selection of masks suppliers

- $\mathbb{A} = \{$ **c**,**o**,**r**,o,**n**,**a**,**v**,**i**,r,**u**,**s** $\}$; $|\mathbb{A}| = 9$.
- The recommendation : 4 best suppliers
- $n = 6$ criteria
  1. **customizable**: "yes" (+) or "no" (−)
  2. **washable**: "yes" (+) or "no" (−)
  3. **delivery time**: "1 - 14 days" (+) or "15 - 30 days" (−)
  4. **quality**: "high" (+) or "good" (−)
  5. **affordability**: "acceptable" (+) or "expensive" (−)
  6. **provider reputation**: "good" (+) or "fair" (−)
- $\forall i \in \{1, \ldots, n\},\ r_i = 2$
- Preference knowledge :
  $\mathbb{PI} = \{r \succsim i, v \succsim i, u \succsim s, r \succsim u\}$
  $\mathbb{PI} = \{(r,\ i), (v,\ i), (u,\ s), (r,\ u)\}$

	1	2	3	4	5	6
c	+	+	−	−	+	+
o	+	−	+	+	+	−
n	−	+	+	+	+	−
a	+	+	+	−	−	+
v	+	+	+	+	−	−
i	+	−	+	−	+	+
r	−	−	+	+	+	+
u	+	+	−	+	+	−
s	−	+	+	+	−	+

Performance table

Outline

Explanation in MultiCriteria Decision Aiding

Explanation of Robust Additive Decision Models

- The decision situation
- What do we explain ?
- How do we explain?

Mixed-Explanation in binary core

- Definition and Motivation
- Computation using MILP
- Extensions

Additive value function $\mathcal{U}$ is **additively compatible** with $\mathbb{PI}$, if for all $(a,\ b) \in \mathbb{PI}$,

$\mathcal{U}(a) = \sum_{i=1}^{n} u_i(a_i) \geq \sum_{i=1}^{n} u_i(b_i) = \mathcal{U}(b)$

Definition ([Greco et al., 2008]). Necessary preference relation

Given $\mathbb{PI}$, x is **necessarily preferred** to y – $(x, y) \in \mathcal{N}_{\mathbb{PI}}$ – if $\mathcal{U}(x) \geq \mathcal{U}(y)$ holds for every function $\mathcal{U} \in \mathbb{X} \to \mathbb{R}$ additively compatible with $\mathbb{PI}$.

Proposition ([Greco et al., 2008]).

Given $\mathbb{PI}$; $(x, y) \in \mathcal{N}_{\mathbb{PI}}$ iff, the following LP has **a non-negative solution**:

$$\mathrm{Min} \sum_{i=1}^{n} u_i(x_i) - \sum_{i=1}^{n} u_i(y_i)$$

$$s.t. \begin{cases} \sum_{i=1}^{n} u_i(a_i) \geq \sum_{i=1}^{n} u_i(b_i) & \forall (a, b) \in \mathbb{PI} \\ u_i(z_i^{r_i}) \geq u_i(z_i^{r_i - 1}) & \forall\, i \in \{1, \ldots, n\}, \forall\, r_i \geq 2 \end{cases}$$

The necessary preference relation

Proposition ([Belahcene et al., 2017]). Characterization of $\mathcal{N}_{\mathbb{PI}}$ using covectors

Given $\mathbb{PI}$, $(x, y) \in \mathcal{N}_{\mathbb{PI}}$ iff $(x, y)^\star$ can be written as a **linear combination with non-negative coefficients** of the covectors representing the preference information $\mathbb{PI}$ and the covectors of the dual base that represent Pareto dominance.

The necessary preference relation : Back to our example (1/2)

- $\mathbb{A} = \{$ **c**,**o**,**r**,o,**n**,**a**,**v**,**i**,r,**u**,**s** $\}$
- $|\mathbb{A}| = 9$.
- $\mathcal{B}$: the 4 best suppliers set.
- $x \in \mathbb{X}$ belongs to $\mathcal{B}$ iff :
  $|y \in \mathbb{A} : x \neq y \, and \, (x, z) \in \mathcal{N}_{\mathbb{PI}}| \geq 5$
- $\mathbb{PI} = \{(r, i), (v, i), (u, s), (r, u)\}$
- $r \in \mathcal{B}$ since
  $\{(r, i), (r, u), (r, s), (r, a), (r, c)\} \subset \mathcal{N}_{\mathbb{PI}}$

	1	2	3	4	5	6
c	+	+	−	−	+	+
o	+	−	+	+	+	−
n	−	+	+	+	+	−
a	+	+	+	−	−	+
v	+	+	+	+	−	−
i	+	−	+	−	+	+
r	−	−	+	+	+	+
u	+	+	−	+	+	−
s	−	+	+	+	−	+

Performance table

The necessary preference relation : Back to our example (2/2)

Why $\{(r, a), (r, c)\} \subset \mathcal{N}_{\mathbb{PI}}$?

$$(r,a)^\star = (r,i)^\star + (r,u)^\star + (u,s)^\star$$

$$\begin{array}{rrrrrrrl} & (-1 & 0 & 0 & 1 & 0 & 0) & = (r,i)^\star \\ + & (-1 & -1 & 1 & 0 & 0 & 1) & = (r,u)^\star \\ + & (1 & 0 & -1 & 0 & 1 & -1) & = (u,s)^\star \\ \hline & (-1 & -1 & 0 & 1 & 1 & 0) & = (r,a)^\star \end{array}$$

$$(r,c)^\star = 2 \times (r,u)^\star + (u,s)^\star + (v,i)^\star$$

$$\begin{array}{rrrrrrrl} & (-2 & -2 & 2 & 0 & 0 & 2) & = 2 \times (r,u)^\star \\ + & (1 & 0 & -1 & 0 & 1 & -1) & = (u,s)^\star \\ + & (0 & 1 & 0 & 1 & -1 & -1) & = (v,i)^\star \\ \hline & (-1 & -1 & 1 & 1 & 0 & 0) & = (r,c)^\star \end{array}$$

	1	2	3	4	5	6
c	+	+	−	−	+	+
o	+	−	+	+	+	−
n	−	+	+	+	+	−
a	+	+	+	−	−	+
v	+	+	+	+	−	−
i	+	−	+	−	+	+
r	−	−	+	+	+	+
u	+	+	−	+	+	−
s	−	+	+	+	−	+

Definition: Covector in the case of a binary core

The covector of (x, y) denoted by $(x, y)^\star$ is the $n-$ component vector $(x, y)^\star = (\lambda_i)_{1 \le i \le n}$ where :

$$\lambda_i = \begin{cases} -1 & \text{if } y_i \succ x_i \\ 0 & \text{if } x_i \sim y_i \\ 1 & \text{if } x_i \succ y_i \end{cases}$$

see [Belahcene et al., 2017] for the general definition.

Outline

Explanation in MultiCriteria Decision Aiding

Explanation of Robust Additive Decision Models

- The decision situation
- What do we explain ?
- How do we explain?

Mixed-Explanation in binary core

- Definition and Motivation
- Computation using MILP
- Extensions

How do we explain?

…Explaining via cancellation [Belahcene et al., 2019]

- Grounded on the *Cancellation axiom* of additive model.
- From *premises* ($\mathbb{PI}$ element) to a *conclusion* ($\mathcal{N}_{\mathbb{PI}}$ element).
- Canceling out terms across pairwise preference statements.

$$\mathbb{PI} = \{(r,\ i), (v,\ i), (u,\ s), (r,\ u)\}$$

$$\begin{array}{c cccccc c cccccc c}
 & 1 & 2 & 3 & 4 & 5 & 6 & & 1 & 2 & 3 & 4 & 5 & 6 & \\
r & \cancel{-} & \cancel{-} & \cancel{+} & \cancel{+} & \cancel{+} & \cancel{+} & \succsim & + & \cancel{-} & \cancel{+} & - & \cancel{+} & \cancel{+} & i \\
r & - & - & + & + & + & + & \succsim & \cancel{-} & + & + & \cancel{+} & - & + & s \\
\hline
r & - & - & + & + & + & + & \succsim & + & + & + & - & - & + & a
\end{array}$$

...Explaining via cancellation [Belahcene et al., 2019]

- Grounded on the *Cancellation axiom* of additive model.
- From *premises* ($\mathbb{PI}$ element) to a *conclusion* ($\mathcal{N}_{\mathbb{PI}}$ element).
- Canceling out terms across pairwise preference statements.

$$\mathbb{PI} = \{(r,\ i), (v,\ i), (u,\ s), (r,\ u)\}$$

$$\begin{array}{c|cccccc|c|cccccc|c}
 & 1 & 2 & 3 & 4 & 5 & 6 & & 1 & 2 & 3 & 4 & 5 & 6 & \\
r & \cancel{-} & \cancel{-} & \cancel{+} & \cancel{+} & \cancel{+} & \cancel{+} & \succsim & + & \cancel{-} & \cancel{+} & - & \cancel{+} & \cancel{+} & i \\
r & - & - & + & + & + & + & \succsim & \cancel{-} & + & + & \cancel{+} & - & + & s \\
\hline
r & - & - & + & + & + & + & \succsim & + & + & + & - & - & + & a
\end{array}$$

\+ Complete : all preference statement inferred are supported.

\- Require technical knowledge from the $\mathcal{DM}$.

\- Not suited for narration (parallel by nature).

...Explaining via sequence of preference swaps
[Belahcene et al., 2019]

"If every alternative for a given objective is rated equally ... you can ignore that objective in making your decision." ([Hammond et al., 1998])

Definition: Preference-swaps explanation

Given $\mathbb{PI}$, $(x, y) \in \mathcal{N}_{\mathbb{PI}}$, an **explanation** of (x, y) is **a sequence** :
$x := e^{(0)} \succsim e^{(1)} \succsim \cdots \succsim e^{(l)} := y$ of length l ($l > 1$), with $(e^{(j-1)}, e^{(j)}) \in \mathcal{N}_{\mathbb{PI}}$ $\forall j \in [\![1; l]\!]$ and differing on **at most 2** criteria (*the order*).

…Explaining via sequence of preference swaps [Belahcene et al., 2019]

"If every alternative for a given objective is rated equally … you can ignore that objective in making your decision." ([Hammond et al., 1998])

Definition: Preference-swaps explanation

Given $\mathbb{PI}$, $(x, y) \in \mathcal{N}_{\mathbb{PI}}$, an **explanation** of (x, y) is **a sequence** : $x := e^{(0)} \succsim e^{(1)} \succsim \cdots \succsim e^{(l)} := y$ of length l ($l > 1$), with $(e^{(j-1)}, e^{(j)}) \in \mathcal{N}_{\mathbb{PI}}$ $\forall j \in [\![1; l]\!]$ and differing on **at most 2** criteria (*the order*).

- \+ Easily scriptable in a natural language and well suited for narration.
- \+ Reduction to its minimum the cognitive effort required from the decision-maker.
- \- Not complete : some preference statements are not supported.
- \- In the general case, some shortest explanations could be arbitrarily long.

Explaining via sequence of preference swaps

...the case of a binary core

ie $\mathbb{PI}$ reference two levels according to each point of view .

+ Explanations can be kept short ($\leq \lfloor \frac{n}{2} \rfloor + 1$)

- Preference-swap of *order* 2:
  - **one *pro* argument** (i) vs. **one *con* argument** (j).
  - Double flip : $i : + \Rightarrow -$; $j : - \Rightarrow +$
- Preference-swap of *order* 1: **Monotonicity**.
- The explanation sequence :
  - counterbalance each *con* argument by one *pro*.
  - expression of Pareto Dominance.

Definition: Pros and Cons of a Statement

Given $\mathbb{PI}$ and $(x, y) \in \mathcal{N}_{\mathbb{PI}}$, we define :

$$(x, y)^{+} := \{i \in \{1, \ldots, n\} : (x, y)_i^{\star} = +1\}$$
$$(x, y)^{-} := \{i \in \{1, \ldots, n\} : (x, y)_i^{\star} = -1\}$$

In other words, $(x, y)^{+}$ is the subset of criteria i on which $x_i = +$ and $y_i = -$ and $(x, y)^{-}$, the subset of criteria i on which $x_i = -$ and $y_i = +$.

Explaining via sequence of preference swaps : Example

Can we explain why $\{(r,\ a), (r,\ c)\} \subset \mathcal{N}_{\mathbb{PI}}$?

$\mathbb{PI} = \{(r,\ i), (v,\ i), (u,\ s), (r,\ u)\}$

$(r,\ a) \in \mathcal{N}_{\mathbb{PI}}$

$(r,\ a)^+ = \{$quality, affordability$\}$; $(r,\ a)^- = \{$customizable, washable$\}$

$$r := - - + + + + \overset{(1)}{\succsim} - + + + - + \overset{(2)}{\succsim} + + + - - + := a$$

You prefer the supplier r over the supplier a because every thing else being equal:

(1) *you prefer an affordable mask to a washable one* and

(2) *you prefer a high quality mask to a customizable one.*

	1	2	3	4	5	6
r	$-$	$-$	$+$	$+$	$+$	$+$
a	$+$	$+$	$+$	$-$	$-$	$+$

Pros and Cons arguments of $(r,\ a) \in \mathcal{N}_{\mathbb{PI}}$

Explaining via sequence of preference swaps : Example

Can we explain why $\{(r, a), (r, c)\} \subset \mathcal{N}_{\mathbb{PI}}$?

$$\mathbb{PI} = \{(r, i), (v, i), (u, s), (r, u)\}$$

$(r, a) \in \mathcal{N}_{\mathbb{PI}}$

$(r, a)^+ = \{$quality, affordability$\}$; $(r, a)^- = \{$customizable, washable$\}$

$$r := --++++ \underset{(1)}{\succsim} -+++-+ \underset{(2)}{\succsim} +++--+ := a$$

You prefer the supplier r over the supplier a because every thing else being equal:

(1) *you prefer an affordable mask to a washable one* and

(2) *you prefer a high quality mask to a customizable one.*

	1	2	3	4	5	6
r	−	−	+	+	+	+
a	+	+	+	−	−	+

Pros and Cons arguments of $(r, a) \in \mathcal{N}_{\mathbb{PI}}$

$(r, c) \in \mathcal{N}_{\mathbb{PI}}$

$(r, c)^+ = \{$delivery time, quality$\}$; $(r, c)^- = \{$customizable, washable$\}$

- $r := --++++ \succsim -++-++ \succsim ++--++ := c$ X because $(-++-++, ++--++) \notin \mathcal{N}_{\mathbb{PI}}$
- $r := --++++ \succsim -+-+++ \succsim ++--++ := c$ X because $(--++++, -+-+++) \notin \mathcal{N}_{\mathbb{PI}}$

	1	2	3	4	5	6
r	−	−	+	+	+	+
c	+	+	−	−	+	+

Pros and Cons arguments of $(r, c) \in \mathcal{N}_{\mathbb{PI}}$

Mixed-Explanation in binary core

Outline

Explanation in MultiCriteria Decision Aiding

Explanation of Robust Additive Decision Models

- The decision situation
- What do we explain ?
- How do we explain?

Mixed-Explanation in binary core

- Definition and Motivation
- Computation using MILP
- Extensions

What is a Mixed explanation?

Definition: Mixed explanation

Given $\mathbb{PI}$, $(x,\ y) \in \mathcal{N}_{\mathbb{PI}}$. **A mixed explanation** corresponds to **a sequence**

$$x := e^{(0)} \succsim_{\mathcal{U}} e^{(1)} \succsim_{\mathcal{U}} \cdots \succsim_{\mathcal{U}} e^{(l)} := y \tag{2}$$

where :

- $\succsim_{\mathcal{U}}$ the binary relation induced by $\mathcal{U}$ on $\mathbb{X}$, $\mathcal{U}$ being $\mathbb{PI}$**-compatible**.
- $(e^{(j-1)}, e^{(j)})$ differing on **at most 2** criteria $\forall j \in [\![1; l]\!]$
- each *con* argument should be counterbalanced **exactly once**, i.e.

$$\forall j \in (x,y)^{-}, \left|\left\{k \in [\![1, l]\!] : (e^{(k-1)}, e^{(k)})^{-} = \{j\}\right\}\right| = 1; \tag{3}$$

- any *pro* argument can be used **at most once**, i.e.

$$\forall i \in (x,y)^{+}, \left|\left\{k \in [\![1, l]\!] : (e^{(k-1)}, e^{(k)})^{+} = \{i\}\right\}\right| \leq 1; \tag{4}$$

- there is a (potentially empty) set $\mathcal{M} \subset \{1, \ldots, l\}$ such as for all $m \in \mathcal{M}$, we have $(e^{(m-1)}, e^{(m)}) \in \mathcal{N}_{\mathbb{PI}}$

Motivation

Why going beyond the necessary explanation and embracing a mixed explanation ?

Motivation

Why going beyond the necessary explanation and embracing a mixed explanation ?

- *Necessary explanations* are less often computable than *Mixed explanations*.

Any decision maker would rather have an explanation than none at all.

Motivation

Why going beyond the necessary explanation and embracing a mixed explanation ?

- *Necessary explanations* are less often computable than *Mixed explanations*.

  Any decision maker would rather have an explanation than none at all.

- Suggest to decision-makers a way of reasoning they could appropriate.

  Some decision makers find the task of comparing alternatives difficult.

Motivation

Why going beyond the necessary explanation and embracing a mixed explanation ?

- *Necessary explanations* are less often computable than *Mixed explanations*.

  Any decision maker would rather have an explanation than none at all.

- Suggest to decision-makers a way of reasoning they could appropriate.

  Some decision makers find the task of comparing alternatives difficult.

- Mixed explanations help to collect additional preferential information.

  Snowball effect.

Motivation

Why going beyond the necessary explanation and embracing a mixed explanation ?

- *Necessary explanations* are less often computable than *Mixed explanations*.

  Any decision maker would rather have an explanation than none at all.

- Suggest to decision-makers a way of reasoning they could appropriate.

  Some decision makers find the task of comparing alternatives difficult.

- Mixed explanations help to collect additional preferential information.

  Snowball effect.

- Make use of the full capacity to explain via *preference-swaps* of *order* at most 2.

Computing Mixed Explanation : MILP formulation

Given $\mathbb{PI}$, $(x,\ y) \in \mathcal{N}_{\mathbb{PI}}$. Let $\mathbb{S} := (x,\ y)^{+} \times (x,\ y)^{-}$ be the *swaps space.* $|\mathbb{S}| \in \mathcal{O}(n^2)$.
Determine $\mathcal{S}(\subset \mathbb{S})$ the set of 2-order *preference-swaps* used in *mixed explanation*.

Computing Mixed Explanation : MILP formulation

Given $\mathbb{PI}$, $(x,\ y) \in \mathcal{N}_{\mathbb{PI}}$. Let $\mathbb{S} := (x,\ y)^+ \times (x,\ y)^-$ be the *swaps space*. $|\mathbb{S}| \in \mathcal{O}(n^2)$.
Determine $\mathcal{S}(\subset \mathbb{S})$ the set of 2-order *preference-swaps* used in *mixed explanation*.

Variables

- $u_i(+)$ and $u_i(-)$, $\forall i \in \{1 \ldots n\}$
- $$b_s = \begin{cases} 1 & \text{iff } s \in \mathcal{S} \\ 0 & \text{Otherwise.} \end{cases}$$

Computing Mixed Explanation : MILP formulation

Given $\mathbb{PI}$, $(x, y) \in \mathcal{N}_{\mathbb{PI}}$. Let $\mathbb{S} := (x, y)^+ \times (x, y)^-$ be the *swaps space*. $|\mathbb{S}| \in \mathcal{O}(n^2)$.
Determine $\mathcal{S}(\subset \mathbb{S})$ the set of 2-order *preference-swaps* used in *mixed explanation*.

Variables

- $u_i(+)$ and $u_i(-)$, $\forall i \in \{1 \ldots n\}$
-
$$b_s = \begin{cases} 1 & \text{iff } s \in \mathcal{S} \\ 0 & \text{Otherwise.} \end{cases}$$

Constraints

- Preference information :

$$\sum_{i=1}^{n} u_i(a_i) \geq \sum_{i=1}^{n} u_i(b_i) \; \forall \, (a, b) \in \mathbb{PI}$$

Computing Mixed Explanation : MILP formulation

Given $\mathbb{PI}$, $(x, y) \in \mathcal{N}_{\mathbb{PI}}$. Let $\mathbb{S} := (x, y)^+ \times (x, y)^-$ be the *swaps space*. $|\mathbb{S}| \in \mathcal{O}(n^2)$.
Determine $\mathcal{S}(\subset \mathbb{S})$ the set of 2-order *preference-swaps* used in *mixed explanation*.

Variables

- $u_i(+)$ and $u_i(-)$, $\forall i \in \{1 \ldots n\}$
- $$b_s = \begin{cases} 1 & \text{iff } s \in \mathcal{S} \\ 0 & \text{Otherwise.} \end{cases}$$

Constraints

- Preference information :

$$\sum_{i=1}^{n} u_i(a_i) \geq \sum_{i=1}^{n} u_i(b_i) \;\forall\, (a,\ b) \in \mathbb{PI}$$

- Normalization:

$$\begin{cases} u_i(-) = 0 \;\forall\, i \in \{1, \ldots, n\} \\ \sum_{i=1}^{n} u_i(+) = 1 \end{cases}$$

- Monotonicity :

$$u_i(+) - u_i(-) \geq 0 \quad \forall\, i \in \{1, \ldots, n\}.$$

Computing Mixed Explanation : MILP formulation

Given $\mathbb{PI}$, $(x,\ y) \in \mathcal{N}_{\mathbb{PI}}$. Let $\mathbb{S} := (x,\ y)^{+} \times (x,\ y)^{-}$ be the *swaps space*. $|\mathbb{S}| \in \mathcal{O}(n^2)$.
Determine $\mathcal{S}(\subset \mathbb{S})$ the set of 2-order *preference-swaps* used in *mixed explanation*.

Variables

- $u_i(+)$ and $u_i(-)$, $\forall i \in \{1 \dots n\}$
-
$$b_s = \begin{cases} 1 & \text{iff } s \in \mathcal{S} \\ 0 & \text{Otherwise.} \end{cases}$$

Constraints

- Preference information :

$$\sum_{i=1}^{n} u_i(a_i) \geq \sum_{i=1}^{n} u_i(b_i)\ \forall\ (a,\ b)\ \in\ \mathbb{PI}$$

- Normalization:

$$\begin{cases} u_i(-) = 0\ \forall\ i\ \in\ \{1, \dots, n\} \\ \sum_{i=1}^{n} u_i(+) = 1 \end{cases}$$

- Monotonicity :

$$u_i(+) - u_i(-) \geq 0\quad \forall\ i \in \{1, \dots, n\}.$$

- Each *con* conterbalanced exactly once :

$$\sum_{s=(i,\ j)\in\mathbb{S}} b_s = 1\quad \forall\ j\ \in\ (x,\ y)^{-}$$

Computing Mixed Explanation : MILP formulation

Given $\mathbb{PI}$, $(x, y) \in \mathcal{N}_{\mathbb{PI}}$. Let $\mathbb{S} := (x, y)^+ \times (x, y)^-$ be the *swaps space*. $|\mathbb{S}| \in \mathcal{O}(n^2)$.
Determine $\mathcal{S}(\subset \mathbb{S})$ the set of 2-order *preference-swaps* used in *mixed explanation*.

Variables

- $u_i(+)$ and $u_i(-)$, $\forall i \in \{1 \dots n\}$
- $$b_s = \begin{cases} 1 & \text{iff } s \in \mathcal{S} \\ 0 & \text{Otherwise.} \end{cases}$$

Constraints

- Preference information :

$$\sum_{i=1}^{n} u_i(a_i) \geq \sum_{i=1}^{n} u_i(b_i) \; \forall \, (a, b) \in \mathbb{PI}$$

- Normalization:

$$\begin{cases} u_i(-) = 0 \; \forall \, i \in \{1, \dots, n\} \\ \sum_{i=1}^{n} u_i(+) = 1 \end{cases}$$

- Monotonicity :

$$u_i(+) - u_i(-) \geq 0 \quad \forall \, i \in \{1, \dots, n\}.$$

- Each *con* conterbalanced exactly once :

$$\sum_{s=(i,j) \in \mathbb{S}} b_s = 1 \quad \forall \, j \in (x, y)^-$$

- Each *pro* is used at most once:

$$\sum_{s=(i,j) \in \mathbb{S}} b_s \leq 1 \quad \forall \, i \in (x, y)^+$$

Computing Mixed Explanation : MILP formulation

Given $\mathbb{PI}$, $(x,\ y) \in \mathcal{N}_{\mathbb{PI}}$. Let $\mathbb{S} := (x,\ y)^{+} \times (x,\ y)^{-}$ be the *swaps space*. $|\mathbb{S}| \in \mathcal{O}(n^2)$.
Determine $\mathcal{S}(\subset \mathbb{S})$ the set of 2-order *preference-swaps* used in *mixed explanation*.

Variables

- $u_i(+)$ and $u_i(-)$, $\forall i \in \{1 \ldots n\}$
- $$b_s = \begin{cases} 1 & \text{iff } s \in \mathcal{S} \\ 0 & \text{Otherwise.} \end{cases}$$

Constraints

- Preference information :

$$\sum_{i=1}^{n} u_i(a_i) \geq \sum_{i=1}^{n} u_i(b_i) \ \forall\ (a,\ b)\ \in \mathbb{PI}$$

- Normalization:

$$\begin{cases} u_i(-) = 0 \ \forall\ i \in \{1, \ldots, n\} \\ \sum_{i=1}^{n} u_i(+) = 1 \end{cases}$$

- Monotonicity :

$$u_i(+) - u_i(-) \geq 0 \quad \forall\ i \in \{1, \ldots, n\}.$$

- Each *con* conterbalanced exactly once :

$$\sum_{s=(i,\ j) \in \mathbb{S}} b_s = 1 \quad \forall\ j \in (x,\ y)^{-}$$

- Each *pro* is used at most once:

$$\sum_{s=(i,\ j) \in \mathbb{S}} b_s \leq 1 \quad \forall\ i \in (x,\ y)^{+}$$

- *Preference-swap* constraint :

$$\forall\ s = (i,\ j) \in \mathbb{S},\ [u_i(+) - u_i(-)] - [u_j(+) - u_j(-)] \geq b_s - 1$$

Computing Mixed Explanation : MILP formulation

Given $\mathbb{PI}$, $(x,\ y) \in \mathcal{N}_{\mathbb{PI}}$. Let $\mathbb{S} := (x,\ y)^+ \times (x,\ y)^-$ be the *swaps space*. $|\mathbb{S}| \in \mathcal{O}(n^2)$.
Determine $\mathcal{S}(\subset \mathbb{S})$ the set of 2-order *preference-swaps* used in *mixed explanation*.

Variables

- $u_i(+)$ and $u_i(-)$, $\forall i \in \{1 \dots n\}$
-

$$b_s = \begin{cases} 1 & \text{iff } s \in \mathcal{S} \\ 0 & \text{Otherwise.} \end{cases}$$

Constraints

- Preference information :

$$\sum_{i=1}^{n} u_i(a_i) \geq \sum_{i=1}^{n} u_i(b_i)\ \forall\ (a,\ b) \in \mathbb{PI}$$

- Normalization:

$$\begin{cases} u_i(-) = 0\ \forall\ i \in \{1, \dots, n\} \\ \sum_{i=1}^{n} u_i(+) = 1 \end{cases}$$

- Monotonicity :

$$u_i(+) - u_i(-) \geq 0\quad \forall\ i \in \{1, \dots, n\}.$$

- Each *con* conterbalanced exactly once :

$$\sum_{s=(i,\ j)\in\mathbb{S}} b_s = 1\quad \forall\ j \in (x,\ y)^-$$

- Each *pro* is used at most once:

$$\sum_{s=(i,\ j)\in\mathbb{S}} b_s \leq 1\quad \forall\ i \in (x,\ y)^+$$

- *Preference-swap* constraint :

$$\forall\ s = (i,\ j) \in \mathbb{S},\ [u_i(+) - u_i(-)] - [u_j(+) - u_j(-)] \geq b_s - 1$$

Objective

$\mathbb{S}^{\mathcal{N}_{\mathbb{PI}}}$: the subset of *necessary preference-swaps* in $\mathbb{S}$,
Maximize or **Minimize** :

$$\sum_{s\in\mathbb{S}^{\mathcal{N}_{\mathbb{PI}}}} b_s$$

Explanation in MultiCriteria Decision Aiding

Explanation of Robust Additive Decision Models

- The decision situation
- What do we explain ?
- How do we explain?

Mixed-Explanation in binary core

- Definition and Motivation
- Computation using MILP
- Extensions

Extensions (1/2)

- Use $\mathbb{PI}$ elements into the explanation sequence.

$\mathbb{PI} = \{(r,\ i), (v,\ i), (u,\ s), (r,\ u)\}$, $(r,\ c) \in \mathcal{N}_{\mathbb{PI}}$

- $r := --++++ \overset{①}{\succsim} \underbrace{++-++-}_{u} \overset{②}{\succsim} ++--++ := c$

versus

- $r := --++++ \overset{①}{\succsim} -+-+++ \overset{②}{\succsim} ++--++ := c$

You prefer the supplier r over the supplier c because :

1. *you tell me that you prefer supplier r to u and*
2. *every thing else being equal you prefer a high quality mask from a fair reputation supplier to a just good quality one supplied by a more well-known provider.*

versus

1. *you **might** prefer a quick delivery of non-washable masks to a late delivery of washable masks and*
2. *you prefer non-customizable masks of high quality to customizable ones and of good quality.*

Extensions (1/2)

- **Use $\mathbb{PI}$ elements into the explanation sequence.**

$\mathbb{PI} = \{(r, i), (v, i), (u, s), (r, u)\}$, $(r, c) \in \mathcal{N}_{\mathbb{PI}}$

*You prefer the supplier **r** over the supplier **c** because :*

- $r := --++++ \underset{1}{\succsim} \underbrace{++-++-}_{u} \underset{2}{\succsim} ++--++ := c$

(1) *you tell me that you prefer supplier **r** to **u** and*

(2) *every thing else being equal you prefer a high quality mask from a fair reputation supplier to a just good quality one supplied by a more well-known provider.*

versus

- $r := --++++ \underset{1}{\succsim} -+-+++ \underset{2}{\succsim} ++--++ := c$

(1) *you **might** prefer a quick delivery of non-washable masks to a late delivery of washable masks and*

(2) *you prefer non-customizable masks of high quality to customizable ones and of good quality.*

- **Between two (2) transitive explanations, which one to choose ?**
    - Ability to convince
    - Ability to contribute significantly to the elicitation process.

Extensions (2/2)

- How to efficiently build an explanation sequence optimizing a given measure ?
- Increase the order of *preference-swaps* to use.
    - 1 *pro* vs. 2 *cons*.
    - 1 *con* vs. 2 *pros*.
- About the finite set $\mathbb{A}$ and $\mathbb{PI}$ ...
    - What if $|\mathbb{X}_i| > 2$?
    - What if $\mathbb{X}_i$ is continuous ?
- What about the revision?
    - $\mathcal{DM}$ refutes a *necessary preference-swap*.

Conclusion

PhD thesis, May 2020

Interactive explanations in Multi-criteria decision aiding :
handling inconsistencies and levels of explanation.

INTERACTION - ELICITATION - EXPLANATION - REVISION

Amgoud, L. and Prade, H. (2009).
Using arguments for making and explaining decisions.
Artificial Intelligence, 173(3):413–436.

Belahcene, K., Chevaleyre, Y., Labreuche, C., Maudet, N., Mousseau, V., and Ouerdane, W. (2018).
Accountable approval sorting.
In *Proceedings IJCAI'18*, pages 70–76.

Belahcene, K., Labreuche, C., Maudet, N., Mousseau, V., and Ouerdane, W. (2017).
Explaining robust additive utility models by sequences of preference swaps.
Theory and Decision, 82(2):151–183.

Belahcene, K., Labreuche, C., Maudet, N., Mousseau, V., and Ouerdane, W. (2019).
Comparing options with argument schemes powered by cancellation.

In *Proc. of IJCAI'19*, pages 1537–1543.

Bouyssou, D., Marchant, T., Pirlot, M., Tsoukias, A., and Vincke, P. (2006).
***Evaluation and Decision models with multiple criteria: stepping stones for the analyst*, volume 86.**
Springer.

Greco, S., Mousseau, V., and Słowiński, R. (2008).
Ordinal regression revisited: Multiple criteria ranking using a set of additive value functions.
EJOR, 191(2):416–436.

Hammond, J., Keeney, R., and Raiffa, H. (1998).
Even swaps: A rational method for making trade-offs.
Harvard business review, 76:137–8, 143.

Labreuche, C. and Fossier, S. (2018).
Explaining multi-criteria decision aiding models with an extended shapley value.

In *Proc. of IJCAI'18*, pages 331–339.

Labreuche, C., Maudet, N., and Ouerdane, W. (2011).
Minimal and complete explanations for critical multi-attribute decisions.
In *Proc. of ADT*, pages 121–134, Piscataway, NJ, USA.

Labreuche, C., Maudet, N., and Ouerdane, W. (2012).
Justifying dominating options when preferences are incomplete.
In *Proceedings of ECAI*, pages 486–491, Montpellier, France.