Des explications transitives questionnables au service de l'élicitation de préférences additives

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

[1]MICS, CentraleSupélec, Université Paris-Saclay, France
[2]LIP6, Sorbonne Université, CNRS, France

{manuel.amoussou, khaled.belahcene}@centralesupelec.fr,
{vincent.mousseau, wassila.ouerdane}@centralesupelec.fr,
nicolas.maudet@lip6.fr

Résumé

On considère un modèle additif de décision multicritère dans lequel les critères sont évalués sur une échelle binaires. Le décideur fournit une information préférentielle ($\mathbb{PI}$) sur la base de laquelle une relation de préférence nécessaire est établie. Nous proposons d'expliquer une paire de cette préférence nécessaire par une chaîne transitive composée de swaps nécessaires, d'éléments de la $\mathbb{PI}$, et de swaps non-nécessaires. Une telle explication est appelée explication transitive questionnable. Nous proposons une méthode de calcul de telles explications et montrons comment de telles explications peuvent s'insérer naturellement dans une procédure d'élicitation de préférences.

Abstract

We consider an additive multi-criteria decision model in which the criteria are evaluated on a binary scale. The decider provides preferential information ($\mathbb{PI}$) on the basis of which a necessary preference relation is established. We propose to explain a pair of this necessary preference by a transitive chain composed of necessary swaps, elements of the $\mathbb{PI}$, and non-necessary swaps. Such an explanation is called questionable transitive explanation. We propose a method for the computation of such explanations and show how such explanations can fit naturally into a preference elicitation procedure.

1 Introduction

La question de l'explication en Aide MultiCritère à la Décision a fait l'objet, ces deux dernières décennies, de travaux scientifiques ([1], [2], [6], [7]) destinés à éclairer un décideur (le plus souvent peu connaisseur des modèles théoriques de décision) sur les éléments d'une recommandation produite par un système automatique piloté ou non par un analyste ; ces travaux visent à plus de transparence et ainsi à accroitre la confiance des utilisateurs de tels systèmes. La plupart de ces travaux se sont donc employés à produire des explications suffisamment simples pour être comprises d'un décideur humain tout en veillant à les garder le plus possible fidèles aux modèles de décision sous-jacents.

Dans cet article, nous nous intéressons à une seconde fonction de l'explication : celle de l'élicitation des préférences. En effet, on attend d'une explication intelligible proposée à un décideur qu'elle suscite chez ce dernier une réaction (acceptation, réfutation, ou clarification). La réaction du décideur devient ainsi, une occasion de collecte d'informations sur ses préférences, lesquelles peuvent être utilisées par l'analyste pour enrichir le modèle et fournir une recommandation plus adaptée.

L'article est organisé de la façon suivante : nous considérons des problèmes de décision dans lesquels les alternatives sont décrites sur un ensemble d'échelles binaires de critères et nous faisons l'hypothèse de la représentabilité des préférences du décideur par un modèle additif (voir Section 2). La Section 3 propose un algorithme de calcul d'explication des comparaisons par paire déduites du modèle basé sur la programmation mathématique. Nous proposons ensuite en Section 4 une ébauche de protocole interactif dans lequel l'explication pourrait remplir pleinement sa fonction d'élicitation ; ce que nous illustrons par un exemple. L'article s'achève par une conclusion et l'évocation de perspectives futures (Section 5).

2 Préliminaires

Dans cette section, nous précisons le contexte de décision considéré (Sous-section 2.1) ainsi que l'ensemble des comparaisons par paire déduites sujettes à explication (Sous-section 2.2) et détaillons notre proposition d'explication transitive questionnable (Sous-Section 2.3).

2.1 Contexte de décision

Le cadre que nous considérons est celui de l'Aide Multi-Critère à la Décision (AMCD) où sont en présence deux acteurs : un analyste et un décideur. Le décideur requiert « l'aide » de l'analyste pour opérer le choix d'un sous-ensemble $\mathbf{A}^*$ d'alternatives parmi un ensemble (plus grand) $\mathbf{A}$ d'alternatives décrites sur m critères binaires c'est-à-dire ayant exactement deux niveaux d'évaluation ("fort" et "faible"). On notera $\mathbf{X}$ le produit cartésien de m échelles binaires ordonnées $X_{\mathtt{i}} = \{\mathtt{0}, 1\}$ où $\mathtt{0}$ et 1 représentent respectivement les niveaux "faible" et "fort" de chaque critère :

$$\mathbf{X} = \prod_{\mathtt{i} \in [m]} X_{\mathtt{i}}$$

avec $[m]$ l'ensemble des critères. On a $\mathbf{A} \subset \mathbf{X}$.

Les préférences du décideur sont collectées sous la forme d'un ensemble noté $\mathbb{PI}$ (Preference Information) de comparaisons par paire d'alternatives (x, y) qui traduisent sa préférence de l'alternative x sur l'alternative y. Ces alternatives, sur lesquelles le décideur sait exprimer des préférences, forment l'ensemble $\mathbf{A}^R$ des alternatives dites de *référence* ; les ensembles $\mathbf{A}^R$ et $\mathbf{A}$ étant le plus souvent disjoints.

Nous faisons l'hypothèse que les préférences du décideur sont représentables par un modèle additif qui peut être décrit comme suit :

Définition 1 (Modèle additif avec critères binaires) *Une relation de préférence $\succsim$ sur $\mathbf{X}$ est représentable par un modèle additif si et seulement si il existe une fonction de score $\omega : \langle \omega_{\mathtt{i}} \rangle_{\mathtt{i} \in [m]}$ avec $\omega_{\mathtt{i}} : X_{\mathtt{i}} \to \mathbb{R}^+$ telle que :*

$$x \succsim y \iff \omega(x) = \sum_{\mathtt{i} \in [m]} \omega_{\mathtt{i}}(x_{\mathtt{i}}) \geq \sum_{\mathtt{j} \in [m]} \omega_{\mathtt{j}}(y_{\mathtt{j}}) = \omega(y)$$

où $x_{\mathtt{i}}$ (resp. $y_{\mathtt{j}}$) symbolise l'évaluation ($\mathtt{0}$ ou 1) de l'alternative x (resp. y) sur le $\mathtt{i}-$ème (resp. $\mathtt{j}-$ième) critère.

Sans perte de généralité, nous considérerons dans la suite que $\omega_{\mathtt{i}}(\mathtt{0}) = 0$ pour tout $\mathtt{i} \in [m]$; ce qui simplifiera entre autres l'écriture de la fonction de score ω.

Nous faisons également l'hypothèse que l'ensemble $\mathbb{PI}$ des préférences du décideur (collectées sous la forme de comparaisons par paire d'alternatives de référence) est *consistant* c'est-à-dire qu'il se « fond » dans au moins une relation de préférence $\succsim$ sur $\mathbf{X}$ représentable par un modèle additif : $\mathbb{PI} \subset \succsim$. L'ensemble des comparaisons composant $\mathbb{PI}$ peut donc être restitué par une ou plusieurs fonctions de score ω qui seront dites *compatibles* avec $\mathbb{PI}$.

Définition 2 (Fonction de score compatible avec $\mathbb{E}$) *Soit $\mathbb{E}$ un ensemble de comparaisons par paire (x, y). La fonction de score ω est dite compatible avec $\mathbb{E}$ si :*

$$\omega(x) \geq \omega(y) \quad \forall\ (x, y) \in \mathbb{E} \tag{1}$$

Lorsqu'il n'existe aucune fonction de score compatible avec $\mathbb{E}$ alors $\mathbb{E}$ est dit inconsistant. Dans le cas contraire, $\mathbb{E}$ est dit consistant.

L'existence d'une fonction de score compatible avec $\mathbb{E}$ peut être testée à l'aide d'un programme linéaire (sans fonction objectif) dont les contraintes sont définies par les inéquations (1).

Dans la suite, nous désignerons par $\mathbb{A}$ (resp. $\mathbb{X}$) l'ensemble des comparaisons par paire $\mathbf{A} \times \mathbf{A}$ (resp. $\mathbf{X} \times \mathbf{X}$).

Exemple 1 *Le présent exemple décrit une situation de décision dans laquelle un décideur sollicite l'aide d'un analyste pour le choix de quatre alternatives parmi les huit que compte l'ensemble $\mathbf{A}$ (voir Table 1). Chacune des alternatives est décrite sur $m = 6$ critères binaires symbolisés par des lettres ($[m] = \{\mathtt{a}, \mathtt{b}, \mathtt{c}, \mathtt{d}, \mathtt{e}, \mathtt{f}\}$).*

TABLE 1 – Description de l'ensemble $\mathbf{A}$.

	a	b	c	d	e	f
x^3	0	0	0	0	1	1
x^{13}	0	0	1	1	0	1
x^{21}	0	1	0	1	0	1
x^{24}	0	1	1	0	0	0
x^{34}	1	0	0	0	1	0
x^{37}	1	0	0	1	0	1
x^{40}	1	0	1	0	0	0
x^{49}	1	1	0	0	0	1

La Table 2 décrit l'ensemble des alternatives de référence (que le décideur « connaît bien ») sur lequel il fournit un ordre de préférence complet représenté par $\mathbb{PI}$:

TABLE 2 – Description de l'ensemble $\mathbf{A}^R$.

	a	b	c	d	e	f
r^7	0	0	0	1	1	1
r^{12}	0	0	1	1	0	0
r^{19}	0	1	0	0	1	1
r^{52}	1	1	0	1	0	0

$$\mathbb{PI} = \{(r^{52}, r^{19}), (r^{19}, r^7), (r^7, r^{12})\}$$

La fonction de score ω suivante est compatible avec $\mathbb{PI}$: $\omega = \{\boldsymbol{a} : 23,\ \boldsymbol{b} : 49,\ \boldsymbol{c} : 40,\ \boldsymbol{d} : 19,\ \boldsymbol{e} : 6,\ \boldsymbol{f} : 35\}$.

En effet, on a bien :

$\omega(r^{52}) = 23+49+19 = 91 \geq 90 = 49+6+35 = \omega(r^{19})$

$\omega(r^{19}) = 90 \geq 60 = \omega(r^{7})$ *et* $\omega(r^{7}) = 60 \geq 59 = \omega(r^{12})$

2.2 Relation nécessaire et recommandations

Nous nous proposons ici d'analyser l'impact de l'information préférentielle $\mathbb{PI}$ collectée auprès du décideur sur l'ensemble des comparaisons par paire d'alternatives (d'intérêt) $\mathbf{A}$ d'une part et sur l'ensemble des recommandations qui pourraient être faites au décideur d'autre part.

2.2.1 Relation nécessaire

En désignant par $\Omega_{\mathbb{PI}}$ l'ensemble des fonctions de score compatibles avec $\mathbb{PI}$ (Définition 2), il est évident que l'on a :

$$\omega(x) - \omega(y) \geq 0 \quad \forall\ (x,\ y) \in \mathbb{PI} \ \ \forall\ \omega \in \Omega_{\mathbb{PI}}.$$

Le concept de « relation nécessaire étant donné $\mathbb{PI}$ » va au-delà de la seule considération des comparaisons par paire appartenant à $\mathbb{PI}$ et étend la condition $\omega(x) - \omega(y) \geq 0$ à toutes les comparaisons par paire $(x,\ y) \in \mathbb{X}$:

Définition 3 (Relation nécessaire étant donné $\mathbb{PI}$) *Étant donnés un ensemble consistant de comparaisons par paire $\mathbb{PI}$ et deux alternatives $x,\ y \in \mathbf{X}$, on dit que x est nécessairement préférée à y si et seulement si on a : $\omega(x)-\omega(y) \geq 0$ pour toute fonction de score ω compatible avec $\mathbb{PI}$ ($\omega \in \Omega_{\mathbb{PI}}$).*

On notera $\mathbb{N}_{\mathbb{PI}}$ l'ensemble des comparaisons par paire $(x,\ y) \in \mathbb{X}$ telles que x est nécessairement préférée à y.

$\mathbb{N}_{\mathbb{PI}}$ est une relation binaire réflexive et transitive[3].

L'ensemble $\mathbb{P}$ des comparaisons par paire $(x,\ y)$ tel que : $y_{\mathtt{i}} = 1 \Rightarrow x_{\mathtt{i}} = 1$ traduisant la dominance de Pareto est (trivialement) inclus dans $\mathbb{N}_{\mathbb{PI}}$.

La Proposition 1 suggère une méthode de calcul des éléments de la relation nécessaire basée sur la résolution d'un programme linéaire. D'autres méthodes sont proposées dans la littérature, en particulier [3] qui est assez proche de la Proposition 1 et [2] basée sur l'utilisation des *cofacteurs*.

Proposition 1 *Étant donné un ensemble consistant de comparaisons par paire $\mathbb{PI}$, la comparaison par paire $(x,\ y) \in \mathbb{X}$ appartient à $\mathbb{N}_{\mathbb{PI}}$ si l'ensemble $\mathbb{PI} \cup \{(y,\ x)\}$ est inconsistant (voir Définition 2).*

Comme $\mathbf{A}$ est l'ensemble des alternatives sur lequel va porter la recommandation de l'analyste, l'ensemble $\mathbb{N}_{\mathbb{PI}} \cap \mathbf{A}$ que nous noterons par la suite $\mathbb{N}^{\mathbf{A}}_{\mathbb{PI}}$ revêt une importance particulière : il s'agit de comparaisons par paire déduites évidentes dont on peut dire que le décideur est « convaincu » (étant donnée l'hypothèse de représentabilité de ses préférences par un modèle additif) et par rapport auxquelles il est potentiellement à même d'exprimer les raisons de ces préférences. Les expliquer de façon « simple et intelligible » pourrait permettre à l'analyste de vérifier l'alignement du raisonnement véhiculé par les explications produites avec les convictions du décideur et donc de capturer le cas échéant de nouvelles informations préférentielles qui serviraient à réduire l'incertitude autour de l'ensemble $\mathbf{A}^*$ des alternatives à recommander.

2.2.2 Recommandations admissibles

Nous avons vu dans le paragraphe précédent, que l'information préférentielle $\mathbb{PI}$ pourrait permettre de conclure qu'une alternative $x \in \mathbf{A}$ est nécessairement préférée à une autre alternative $y \in \mathbf{A}$: $(x,\ y) \in \mathbb{N}^{\mathbf{A}}_{\mathbb{PI}}$. Cette déduction pourrait également se traduire en la réduction du nombre des ensembles $\mathbf{A}_{cand}$ « candidats » à recommander au décideur. En effet, la déduction $(x,\ y) \in \mathbb{N}^{\mathbf{A}}_{\mathbb{PI}}$ pourrait par exemple disqualifier une ou plusieurs recommandations candidates $\mathbf{A}_{cand}$ telle que $y \in \mathbf{A}_{cand}$.

Plus généralement, en considérant pour chaque recommandation candidate $\mathbf{A}_{cand}$ l'ensemble caractéristique de ses comparaisons par paire $\{(x,\ y),\ x \in \mathbf{A}_{cand}$ et $y \notin \mathbf{A}_{cand}\}$, il est possible de déterminer si $\mathbf{A}_{cand}$ peut être recommandé ou non.

Proposition 2 (Recommandation admissible étant donné $\mathbb{PI}$) *Un sous-ensemble de candidats $\mathbf{A}_{cand} \subset \mathbf{A}$ tel que $|\mathbf{A}_{cand}| = |\mathbf{A}^*|$ est admissible (ou recommandable) si l'ensemble des comparaisons par paire $\mathbb{PI} \cup \{(x,\ y),\ x \in \mathbf{A}_{cand}$ et $y \notin \mathbf{A}_{cand}\}$ est consistant (voir Définition 2).*

Exemple 2 (Suite de l'Exemple 1) *Étant donné $\mathbb{PI}$, les comparaisons par paires de $\mathbb{A}$ déduites sont les suivantes :*

$$\mathbb{N}^{\mathbf{A}}_{\mathbb{PI}} = \{(x^{49},\ x^{37}),\ (x^{37},\ x^{3}),\ (x^{49},\ x^{3})\}$$

et le nombre de recommandations admissibles (de 4 alternatives parmi les 8) est de 20.

L'objet de notre contribution est de montrer comment est-ce que l'explication des éléments de la relation nécessaire pourrait permettre la collecte d'information préférentielle supplémentaire avec comme corollaire la diminution du nombre de recommandations admissibles.

2.3 Explication de la relation nécessaire

Dans [8] on peut lire : « Expliquer un évènement, c'est fournir des informations sur ses causes [1] ». Dans le domaine

1. Traduction de *"To explain an event is to provide some information about its causal history"[8].*

de l'Aide MultiCritère à la Décision (AMCD), un certain nombre de travaux ont développé, à travers des propositions diverses, une logique similaire dans le cas où l'« évènement » à expliquer est un élément de la relation nécessaire.

Ici, nous nous appuierons sur deux propositions de la littérature ([5] et [2]) pour définir une nouvelle structure d'explication qui combine les approches qui y sont développées.

2.3.1 Explication sous forme de *preferential reduct* [5]

Ce que les auteurs entendent par "*preferential reduct*", c'est tout simplement l'ensemble minimal des éléments de la $\mathbb{PI}$ qui justifie l'appartenance de la comparaison par paire déduite à l'ensemble $\mathbb{N}_{\mathbb{PI}}$. Cette façon d'expliquer pointe donc directement les comparaisons par paire fournies par le décideur qui sont la cause de la déduction faite. On remarquera qu'elle est complète dans le sens où l'ensemble explicatif d'éléments de la $\mathbb{PI}$ ne sera jamais vide. Elle ne révèle cependant pas les mécanismes (applications des propriétés du modèle additif) qui sont à l'oeuvre dans cette déduction; ce qui, dans le cadre de la délivrance d'une explication à un décideur humain, peut s'avérer insuffisant.

Exemple 3 (Suite des Exemples 1 et 2) *Le sous-ensemble* $\{(r^{52}, r^{19}), (r^{19}, r^7)\}$ *de* $\mathbb{PI}$ *est le preferential reduct de la déduction* $(x^{49}, x^3) \in \mathbb{N}^{\mathbf{A}}_{\mathbb{PI}}$.

Les mécanismes qui sont à l'oeuvre dans cette déduction sont :

- *(i) l'application de la propriété de transitivité qui permet d'obtenir* (r^{52}, r^7) *à partir de* (r^{52}, r^{19}) *et* (r^{19}, r^7).
- *(ii) l'application de la propriété de cancellation d'ordre 1 qui autorise un raisonnement « toute chose égale par ailleurs » :*
  $(r^{52}, r^7) \equiv (\mathtt{110101}, \mathtt{000111})$ *qui est équivalent à* $(x^{49}, x^3) \equiv (\mathtt{110001}, \mathtt{000011})$.
  Les critères colorés étant ceux communs aux deux alternatives et le critère $\mathtt{d}$ *en bleu étant celui sur lequel s'est appliqué le changement conjoint d'évaluation.*

2.3.2 Explication à l'aide de *preference swaps* [2]

En guise de résumé très sommaire de la contribution de cet article ([2]), nous pouvons dire que l'explication de la comparaison par paire $(x, y) \in \mathbb{N}_{\mathbb{PI}}$ y est conçue comme une chaîne transitive reliant x à y et transitant par des alternatives réelles ($\in \mathbf{A} \cup \mathbf{A}^R$) ou fictives ($\notin \mathbf{A} \cup \mathbf{A}^R$) telle que les comparaisons par paire consécutives le long de cette chaîne appartiennent (aussi) à $\mathbb{N}_{\mathbb{PI}}$ et composées de (2) alternatives qui ne divergent que d'au plus deux critères (d'où le terme *swaps* dans *preference swaps*). On utilisera les crochets pour représenter le swap $[\mathtt{i}, \mathtt{j}]$ que l'on distinguera plus aisément de l'écriture d'une éventuelle comparaison par paire $(\mathtt{i}, \mathtt{j})$ qui utilise les parenthèses.

Dans ce type d'explication, on note la matérialisation du souci de révéler les mécanismes sous-jacents à la déduction (en particulier la propriété de transitivité) doublé du désir de ramener chacune des comparaisons par paire d'alternatives consécutives (grâce à la propriété de cancellation d'orde 1) soit à l'expression d'une préférence de type dominance de Pareto soit à une confrontation entre exactement 2 critères. Ce faisant, on comprend aisément que cette façon d'expliquer offre un maximum de garantie d'intelligibilité pour un décideur humain car la "complexité" des atomes qui composent la chaîne explicative est réduite à son minimum [2].

Sans surprise, l'inconvénient majeur de ce type d'explication est son incomplétude : certaines comparaisons par paire déduites n'admettent pas d'explication comme on peut le constater dans l'exemple suivant.

Exemple 4 (Suite des Exemples 1 et 2) *La comparaison par paire* $(x^{49}, x^3) \in \mathbb{N}^{\mathbf{A}}_{\mathbb{PI}}$ *n'est pas explicable à l'aide de preference swaps. Cette comparaison, par application de la propriété de cancellation d'ordre 1 est équivalente à la confrontation* $(\{\mathtt{a}, \mathtt{b}\}, \{\mathtt{e}\})$ *de sous-ensembles de critères. Pour que* (x^{49}, x^3) *soit explicable à l'aide de preference swaps, il aurait fallu avoir soit* $(x^{\{\mathtt{a}\}}, x^{\{\mathtt{e}\}}) \in \mathbb{N}_{\mathbb{PI}}$ *soit* $(x^{\{\mathtt{b}\}}, x^{\{\mathtt{e}\}}) \in \mathbb{N}_{\mathbb{PI}}$ *où l'alternative* $x^{\{\mathtt{i}\}}$ *vaut* $\mathtt{0}$ *sur tous les critères sauf sur le critère* $\mathtt{i}$ *; or on n'a ni* $(x^{\{\mathtt{a}\}}, x^{\{\mathtt{e}\}}) \in \mathbb{N}_{\mathbb{PI}}$ *ni* $(x^{\{\mathtt{b}\}}, x^{\{\mathtt{e}\}}) \in \mathbb{N}_{\mathbb{PI}}$.

Par contre, la comparaison par paire déduite : $(x^{49}, x^{37}) \equiv (\mathtt{110001}, \mathtt{100101}) \equiv (x^{\{\mathtt{b}\}}, x^{\{\mathtt{d}\}})$ *en tant que preference swap est (trivialement) explicable.*

2.3.3 Explication transitive questionable

Comme indiqué en conclusion de la sous-section 2.2 et en introduction de la sous-section 2.3, notre proposition d'explication a pour ambition de combiner les avantages des approches d'explication résumées aux paragraphes 2.3.1 et 2.3.2 et de permettre la capture d'information préférentielle supplémentaire susceptible de réduire le nombre de recommandations admissibles.

En effet, les explications à construire (voir Section 3) seront :

- transitives et composées de *preference swap(s)* et d'élément(s) de la $\mathbb{PI}$
- et questionnables c'est-à-dire avec au moins un *swap* $[\mathtt{i}, \mathtt{j}]$ dont la projection (voir Définition 5) dans $\mathbb{X}$ n'appartenant pas à $\mathbb{N}_{\mathbb{PI}}$.

Une explication transitive questionnable conserve donc les garanties d'intelligibilité pour un décideur humain du fait de sa structure et du caractère atomique (les *swaps*) des éléments qui la composent. À défaut de constituer une preuve de la déduction de la comparaison à expliquer étant donné $\mathbb{PI}$, elle fournit une **autre** preuve potentielle de cette

2. Le lecteur intéressé pourra consulter [4] qui détaille une méthode « rationnelle » destinée à faciliter la prise de décision et basée sur les *even-swaps*.

déduction dont la validation ou la contestation des swaps contribue à acquérir de l'information préférentielle supplémentaire.

Exemple 5 (Suite des Exemples 1 et 2)

$$[x^{49}]\ \underbrace{110001 \gtrsim 110100}_{[\boldsymbol{f},\, \boldsymbol{d}]}\overbrace{[r^{52}] \gtrsim [r^{19}]}^{\in \mathbb{PI}}\underbrace{010011 \gtrsim 100011}_{[\boldsymbol{b},\, \boldsymbol{a}]}$$

$$:= \underbrace{100011 \gtrsim 100101}_{[\boldsymbol{e},\, \boldsymbol{d}]}[x^{37}]$$

La séquence ci-dessus est un exemple d'explication transitive questionnable de la déduction (x^{49}, x^{37}). *Elle est composée d'une séquence de comparaisons par paires qui « lient »* x^{49} *à* x^{37}. *Sont mis en évidence les comparaisons par paire qui constituent des swaps et les éléments empruntés à la* $\mathbb{PI}$. *Le swap* $[\mathbf{e}, \mathbf{d}]$ *de projection dans* $\mathbb{X}$, *la comparaison par paire* $(100011, 100101)$ *est un « chaînon » dont la validation ou la contestation permettra d'enrichir la* $\mathbb{PI}$ *et potentiellement de réduire le nombre de recommandations admissibles.*

3 Calcul d'explications transitives questionnables

Dans cette section, nous allons montrer comment est-ce que le problème du calcul d'une explication transitive questionnable peut être modélisé sous la forme d'un programme mathématique mixte (Mixed Integer Program en anglais). Nous tenterons dans un premier temps une formalisation du problème en le ramenant à la recherche d'un chemin dans un graphe à définir (Sous-section 3.2). Nous donnerons ensuite les détails du programme mathématique dont la résolution fournit, lorsqu'elle existe, l'explication (la plus courte) de la déduction $(x, y) \in \mathbb{N}^{\mathbf{A}}_{\mathbb{PI}}$ (Sous-section 3.3).

Mais avant, nous introduisons un certain nombre de définitions de concepts empruntés à [2] et qui nous seront utiles par la suite.

3.1 Quelques définitions

Étant donnée une comparaison par paire $(u, v) \in \mathbb{X}$, nous rappelons ici les définitions des ensembles de critères $\mathtt{pro}(u, v)$, $\mathtt{con}(u, v)$ et $\mathtt{neutr}(u, v)$ (Définition 4) qui nous permettront de mieux appréhender la notion de swap : expression de la préférence d'un critère sur un autre [3] qui peut se concevoir à travers différentes comparaisons par paire de $\mathbb{X}$ (Définition 5). Ces définitions seront complétées par l'introduction de la notion de *décomposition à base de swap(s)* (Définition 6) et de celle de longueur de la comparaison par paire $(u, v) \in \mathbb{X}$ en termes de swap(s) (Définition 7).

Définition 4 (Ensembles $\mathtt{pro}$, $\mathtt{con}$ et $\mathtt{neutr}$) *Soient* u *et* v *deux alternatives de* $\mathbf{X}$ *décrites sur l'ensemble* $[m]$ *de critères.*

Les ensembles $\mathtt{pro}(u, v)$, $\mathtt{con}(u, v)$ *et* $\mathtt{neutr}(u, v)$ *de la comparairon par paire* (u, v) *sont définis comme suit :*

$$\mathtt{pro}(u, v) = \{\mathtt{i} \in [m],\ u_{\mathtt{i}} = 1 \text{ et } v_{\mathtt{i}} = 0\}$$

$$\mathtt{con}(u, v) = \{\mathtt{j} \in [m],\ u_{\mathtt{j}} = 0 \text{ et } v_{\mathtt{j}} = 1\}$$

$$\mathtt{neutr}(u, v) = \{\mathtt{i} \in [m],\ u_{\mathtt{i}} = v_{\mathtt{i}}\}$$

On remarquera que les trois ensembles $\mathtt{pro}(u, v)$, $\mathtt{con}(u, v)$ *et* $\mathtt{neutr}(u, v)$ *partitionnent l'ensemble des critères* $[m]$ *et représentent respectivement l'ensemble de critères en faveur* ($\mathtt{pro}$), *en défaveur* ($\mathtt{con}$) *et neutre(s)* ($\mathtt{neutr}$) *de la préférence traduite par la comparaison par paire* (u, v).

Définition 5 (Projection d'un swap dans $\mathbb{X}$) *Lorsque* $\mathtt{pro}(u, v) = \{\mathtt{i}\}$ *et* $\mathtt{con}(u, v) = \{\mathtt{j}\}$, *on dit que* (u, v) *est une projection du swap* $[\mathtt{i}, \mathtt{j}]$ *dans* $\mathbb{X}$.

Il existe exactement 2^{m-2} *projections différentes du swap* $[\mathtt{i}, \mathtt{j}]$ *dans* $\mathbb{X}$.

Définition 6 (Décomposition à base de swap(s)) *Soit* $(u, v) \in \mathbb{X}$ *une comparaison par paire. Une décomposition à base de swap(s) de* (u, v) *est une application injective* ψ *de* $\mathtt{con}(u, v)$ *vers* $\mathtt{pro}(u, v)$. *L'ensemble des swaps :* $\{[\psi^{-1}(\mathtt{j}), \mathtt{j}],\ \mathtt{j} \in \mathtt{con}(u, v)\}$ *caractérisent la décomposition lorsqu'elle existe. En effet, les comparaisons par paire* (u, v) *telles que :* $|\mathtt{con}(u, v)| > |\mathtt{pro}(u, v)|$ *n'admettent (par définition) pas de décomposition à base de swap(s).*

Eu égard à la définition précédente, toutes les *décompositions à base de swap(s)* de (u, v) sont caractérisées chacune par des ensembles de swaps $\{[\psi^{-1}(\mathtt{j}), \mathtt{j}],\ \mathtt{j} \in \mathtt{con}(u, v)\}$ qui sont tous de taille égale à $|\mathtt{con}(u, v)|$. Ce constat permet donc d'introduire la définition suivante :

Définition 7 (Longueur de (u, v) en termes de swap(s)) *Soit* $(u, v) \in \mathbb{X}$. *On définit comme suit la longueur* $\mathrm{L}(u, v)$ *en termes de swap(s) de la comparaison par paire* (u, v) *:*

$$\mathrm{L}(u, v) = \begin{cases} +\infty & \textit{si } (u, v) \textit{ n'admet pas de décomposition à base de swaps} \\ |\mathtt{con}(u, v)| + \mathrm{I}_{(u, v)} & \textit{sinon} \end{cases}$$

avec $\mathrm{I}_{(u, v)} = 1$ *si* $|\mathtt{pro}(u, v)| > |\mathtt{con}(u, v)|$ *et* 0 *sinon.*

Exemple 6 (Suite des Exemples 1 et 5) *Considérons les 3 comparaisons par paires suivantes :* (x^{40}, x^{21}), (x^{49}, x^{3}) *et* (r^{19}, x^{37}). *La Table 3 en fournit les ensembles* $\mathtt{pro}$ *et* $\mathtt{con}$, *une décomposition à base de swap(s) (lorsqu'elle existe) et sa longueur en termes de swap(s) :*

3. Plus rigoureusement, le swap $[\mathtt{i}, \mathtt{j}]$ traduit en termes de préférence la supériorité de l'intensité de la différence entre les niveaux "fort" et "faible" du critère $\mathtt{i}$ sur celle du critère $\mathtt{j}$.

Table 3 – Décomposition à base de swap(s)

(u, v)	$\mathtt{con}(u, v)$	$\mathtt{pro}(u, v)$	ψ	$\mathrm{L}(u, v)$
(x^{40}, x^{21})	$\{\mathbf{b}, \mathbf{d}, \mathbf{f}\}$	$\{\mathbf{a}, \mathbf{c}\}$	n'existe pas	-
(x^{49}, x^{3})	$\{\mathbf{e}\}$	$\{\mathbf{a}, \mathbf{b}\}$	$\psi_1(\mathbf{e}) = \mathbf{a}$	2
(r^{19}, x^{37})	$\{\mathbf{a}, \mathbf{d}\}$	$\{\mathbf{b}, \mathbf{e}\}$	$\psi_2(\mathbf{a}) = \mathbf{b}$ et $\psi_2(\mathbf{d}) = \mathbf{e}$	2

L'ensemble (singleton) $\{[\mathbf{a}, \mathbf{e}]\}$ caractérise ψ_1 tandis que l'ensemble $\{[\mathbf{b}, \mathbf{a}], [\mathbf{e}, \mathbf{d}]\}$ caractérisent ψ_2. Remarquons, dans le cas de ψ_2, que ce sont justement les swaps $[\mathbf{b}, \mathbf{a}]$ et $[\mathbf{e}, \mathbf{d}]$ de projections respectives dans $\mathbb{X}$ ($\mathtt{010011}$, $\mathtt{100011}$) et ($\mathtt{100011}$, $\mathtt{100101}$) qui composent la (sous-)chaîne reliant r^{19} à x^{37} dans l'explication transitive questionnable de (x^{49}, x^{37}).

3.2 Formalisation du problème

Étant donnés un ensemble (consistant) $\mathbb{PI} = \{(u^1, v^1), \dots, (u^k, v^k)\}$ de k comparaisons par paires et deux alternatives x, y telles que $(x, y) \in \mathbb{N}_{\mathbb{PI}}$, on définit le graphe orienté pondéré $\mathcal{G}^{\mathbb{PI}}_{(x,y)} = (V, E)$ appelé *graphe d'explication de (x, y) étant donné* $\mathbb{PI}$ comme suit :

— $V = \{x, y\} \cup \{u^1, \dots, u^k\} \cup \{v^1, \dots, v^k\}$
Les sommets V de $\mathcal{G}^{\mathbb{PI}}_{(x,y)}$ représentent donc l'ensemble des alternatives impliquées dans $\mathbb{PI}$ complétées de x et y.

— $E = E_1 \cup E_2 \cup E_3 \cup E_4 \cup E_5$ avec
  — $E_1 = \{(x, u^1), \dots, (x, u^k)\}$
  — $E_2 = \{(v^1, y), \dots, (v^k, y)\}$
  — $E_3 = \{(v^t, u^{t'}), \forall t, t' \in \{1, \dots, k\}$ avec $t \neq t'$
  — $E_4 = \{(u^1, v^1), \dots, (u^k, v^k)\} = \mathbb{PI}$
  — $E_5 = \{(x, y)\}$

Les arcs E_4 représentent exactement les comparaisons par paire de la $\mathbb{PI}$. Les arcs E_1 symbolisent les liens directs entre le sommet-alternative x et les sommets entrants de E_4 tandis que les arcs E_2 représentent les liens directs entre les sommets sortants de E_4 ($\mathbb{PI}$) et le sommet-alternative y. Quant aux arcs E_3, ils représentent les connections directes entre toute paire d'arcs distincts de E_4. Enfin, l'ensemble E_5 contient uniquement l'arc direct (x, y).

— La fonction g de pondération des arcs est définie comme suit :

$$g((u, v)) = \begin{cases} 1 & \text{si } (u, v) \in E_4\ (\mathbb{PI}) \\ \mathrm{L}(u, v) & \text{sinon} \end{cases} \tag{2}$$

Une schématisation graphique d'un graphe d'explication est illustrée à la Figure 1. Sur ce graphique, on peut remarquer que les arcs de chaque catégorie ($E_1, \dots, E_5$) ont une couleur caractéristique. La $\mathbb{PI}$ est composée de deux comparaisons par paires qui suffisent à induire $(x, y) \in \mathbb{N}_{\mathbb{PI}}$. Ces comparaisons par paire sont représentées par des arcs en trait plein pour indiquer qu'elles apparaîtraient telles quelles dans une explication transitive questionnable. Les autres arcs par contre sont représentés en pointillés car les comparaisons par paire correspondantes devront faire l'objet d'une décomposition à base de swap(s) s'ils étaient empruntés dans la chaîne explicative de la déduction (x, y).

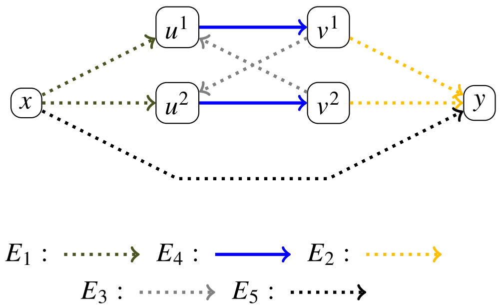

Figure 1 – Graphe d'explication $\mathcal{G}^{\mathbb{PI}}_{(x,y)}$

De façon formelle, le problème de l'existence d'une explication transitive questionnable peut être décrit comme suit :

Entrées :
— Un ensemble de comparaisons par paire $\mathbb{PI}$
— Une comparaison par paire $(x, y) \in \mathbb{N}_{\mathbb{PI}}$
— Un entier naturel L_{max}

Question :
Existe-t-il dans le graphe d'explication $\mathcal{G}^{\mathbb{PI}}_{(x,y)}$ un chemin $C = [z^0, \dots, z^{t-1}, z^t, \dots, z^l]$ de longueur l tel que :
— $l \leq L_{max}$
— reliant x à y c'est-à-dire : $z^0 = x$ et $z^l = y$
— il existe $t \in [[1; l]]$ tel que $(z^{t-1}, z^t) \notin \mathbb{PI}$ et (z^{t-1}, z^t) admet une décomposition ψ à base de swap(s) dont au moins un des swaps caractéristiques a pour projection dans $\mathbb{X}$ (voir Définition 5) une comparaison par paire qui n'appartient pas à la relation nécessaire (voir Définition 3).

Nous avons fait le choix dans la sous-section 3.3 de modéliser le problème décrit ci-dessus dans sa version optimisation : celle conduisant à retenir l'explication la plus courte. En effet, l'explication est destinée à un décideur (le plus souvent humain) dont il convient de tenir compte des limites de la capacité à traiter plusieurs informations à la fois ([7]).

3.3 Modélisation

La modélisation du problème du calcul de la plus courte explication transitive questionnable va s'appuyer sur l'ensemble noté $\widetilde{S}_{\mathbb{PI}}$ de swaps dont les projections dans $\mathbb{X}$ (Définition 5) n'appartiennent pas à la relation nécessaire (Définition 3) afin de garantir le caractère questionnable des explications à produire.

3.3.1 Les variables

On distinguera trois catégories de variables :

1. Variables continues $\mathtt{w_i}$
   Il s'agit des m variables $\mathtt{w_i}$ continues (strictement) positives dont les valeurs correspondent aux évaluations des niveaux "fort" des critères binaires. On rappelle (Définition 1) que nous avons fait l'hypothèse d'une évaluation nulle des niveaux "faible" des m critères.
2. Variables binaires $\mathsf{arc}_{(u,\,v)}$
   Pour chaque arc $(u,\,v)$ du graphe d'explication $\mathcal{G}^{\mathbb{PI}}_{(x,\,y)}$, on définit une variable binaire $\mathsf{arc}_{(u,\,v)}$ valant 1 si et seulement si $(u,\,v)$ appartient au chemin solution du problème.
3. Variables binaires $\mathsf{s}^{[\mathtt{i},\,\mathtt{j}]}_{(u,\,v)}$
   Pour chaque arc $(u,\,v) \notin E_4$, on définit $|\mathtt{pro}(u,\,v)| \times |\mathtt{con}(u,\,v)|$ variables binaires qui permettront de déterminer (lorsqu'elle existe) la décomposition à base de swap(s) (Définition 6) de la comparaison par paire $(u,\,v)$ utilisée dans l'explication transitive de la déduction $(x,\,y)$. $\mathsf{s}^{[\mathtt{i},\,\mathtt{j}]}_{(u,\,v)}$ vaut 1 si et seulement si le swap $[\mathtt{i},\,\mathtt{j}]$ appartient aux swaps caractéristiques de ladite décomposition.

3.3.2 La fonction objectif

Étant donnée la fonction de pondération des arcs (Équation 2), la fonction objectif s'écrit (linéairement) comme suit :

$$\sum_{(u,\,v)\in E} arc_{(u,\,v)} \times g((u,\,v))$$

où E représente l'ensemble des arcs du graphe $\mathcal{G}^{\mathbb{PI}}_{(x,\,y)}$.

3.3.3 Les contraintes

On distinguera six grandes catégories de contraintes :

1. Normalisation de la fonction de score et (stricte) positivité de ses composantes
$$\sum_{\mathtt{i}\in[m]} \mathtt{w_i} = 1 \tag{3}$$
$$\mathtt{w_i} \geq \epsilon \quad \forall\, \mathtt{i} \in [m] \tag{4}$$
   avec ϵ une réel positif arbitrairement petit.
2. Prise en compte de la $\mathbb{PI}$
$$\sum_{\mathtt{i}\,\in\,\mathtt{pro}(u,\,v)} \mathtt{w_i} \geq \sum_{\mathtt{j}\,\in\,\mathtt{con}(u,\,v)} \mathtt{w_j} \quad \forall\,(u,\,v) \in \mathbb{PI} \tag{5}$$
3. Contraintes de « chemin »
$$\sum_{(x,\,u)\in E} \mathsf{arc}_{(x,\,u)} = 1 \tag{6}$$
$$\sum_{(u,\,z)\in E} \mathsf{arc}_{(u,\,z)} = \sum_{(z,\,v)\in E} \mathsf{arc}_{(z,\,v)} \quad \forall\, z \in V\backslash\{x,\,y\} \tag{7}$$
   avec E (resp. V) l'ensemble des arcs (resp. sommets) du graphe $\mathcal{G}^{\mathbb{PI}}_{(x,\,y)}$.
4. Contraintes liant les variables $\mathsf{s}^{[\mathtt{i},\,\mathtt{j}]}_{(u,\,v)}$ et $\mathsf{arc}_{(u,\,v)}$
   Pour chaque arc $(u,\,v) \notin \mathbb{PI}$,
$$\sum_{i\,\in\,\mathtt{pro}(u,\,v)} \mathsf{s}^{[\mathtt{i},\,\mathtt{j}]}_{(u,\,v)} = \mathsf{arc}_{(u,\,v)} \quad \forall\, \mathtt{j} \in \mathtt{con}(u,\,v) \tag{8}$$
$$\sum_{i\,\in\,\mathtt{con}(u,\,v)} \mathsf{s}^{[\mathtt{i},\,\mathtt{j}]}_{(u,\,v)} \leq \mathsf{arc}_{(u,\,v)} \quad \forall\, \mathtt{i} \in \mathtt{pro}(u,\,v) \tag{9}$$
5. « Au moins un swap de $\widetilde{S}_{\mathbb{PI}}$ utilisé »
$$\sum_{(u,\,v)\,\in\,E\backslash E_4} \;\sum_{[i,\,j]\,\in\,\widetilde{S}_{\mathbb{PI}}} \mathsf{s}^{[\mathtt{i},\,\mathtt{j}]}_{(u,\,v)} \geq 1 \tag{10}$$
   avec E l'ensemble des arcs du graphe $\mathcal{G}^{\mathbb{PI}}_{(x,\,y)}$ et E_4 ceux représentant les éléments de la $\mathbb{PI}$ (Sous-section 3.2).
6. Contraintes liant les variables $\mathsf{s}^{[\mathtt{i},\,\mathtt{j}]}_{(u,\,v)}$ à $\mathtt{w_i}$ et $\mathtt{w_j}$
$$\mathtt{w_i} - \mathtt{w_j} \geq (1+\epsilon) \times \mathsf{s}^{[\mathtt{i},\,\mathtt{j}]}_{(u,\,v)} - 1 \tag{11}$$
   avec ϵ une réel positif arbitrairement petit.

3.3.4 Récupération de la solution

Le programme mathématique décrit ci-dessus peut admettre ou non une solution. Si son exécution échoue, cela signifie qu'il n'existe pas d'explication transitive questionnable de la déduction $(x,\,y) \in \mathbb{N}_{\mathbb{PI}}$. Par contre, si elle réussit, la valeur optimale fournit la longueur de l'explication. Les valeurs des variables $\mathsf{arc}_{(u,\,v)}$ indiquent les arcs empruntés et celles des variables $\mathsf{s}^{[\mathtt{i},\,\mathtt{j}]}_{(u,\,v)}$ renseignent sur leur décomposition à base de swap(s).

4 Mobilisation au sein d'un dispositif interactif

L'objet de cette section est de décrire une ébauche de protocole interactif à travers lequel les explications transitives questionnables rempliront une fonction d'élicitation et d'en fournir une instanciation illustrée. Nous rappelons que les préférences du décideur sont additives. À cette hypothèse, nous ajoutons le fait que l'information préférentielle qu'il fournit est « sûre » c'est-à-dire que :

(i) les comparaisons par paires composant $\mathbb{PI}$ ne seront donc pas remises en cause lors du processus interactif

(ii) le décideur ne se trompe pas dans l'appréciation des swaps qui lui sont fournis.

Sur la base de ces hypothèses, on remarquera que les seules informations préférentielles additionnelles capturées sont des comparaisons par paire résultant de la validation ✔ ou de la contestation ✖ des swaps utilisés dans les explications transitives questionnables. Aussi, remarquerons-nous que le processus « converge » : le nombre d'éléments de la relation nécessaire sur $\mathbf{A}$ ($|\mathbb{N}^{\mathbf{A}}_{\mathbb{PI}}|$) croît tandis que le nombre de recommandations admissibles (Paragraphe 2.2.2) décroît. L'augmentation du nombre d'éléments de la relation nécessaire sur $\mathbf{A}$ offre de nouvelles occasions d'explication qui permettront de capturer de l'information préférentielle additionnelle orchestrant ainsi une sorte d'« effet boule de neige » qui pourrait aboutir à l'identification exacte de l'ensemble $\mathbf{A}^*$ à recommander (comme c'est le cas dans l'illustration de la Sous-section 4.2).

4.1 Ébauche de protocole interactif

Dans ce qui suit, nous identifions différents instants t consécutifs à la « collecte des réactions du décideur » sur la t−ième explication transitive questionnable qui lui est soumise. Ainsi, l'instant $t = 0$ correspond au tout début de l'interaction et $\mathbb{PI}^t$ à l'ensemble des comparaisons par paire à l'instant t. On note : $\mathbb{S}^t$ l'ensemble des comparaisons par paire projections dans $\mathbb{X}$ des swaps (validés ou contestés) de la t−ième explication transitive questionnable.

Squelette d'un protocole interactif

Entrées. $\mathbf{A}$, $\mathbf{A}^R$ et $\mathbb{PI}^0$.

Objectif. Éliciter les préférences du décideur à l'aide d'explications transitives questionnables.

Le protocole.

$t = 0$

Tant que $|\mathbb{N}^{\mathbf{A}}_{\mathbb{PI}^t}| \neq 0$ **et** `ConditionArret` non vérifiée

  `Choisir` $(x, y) \in \mathbb{N}^{\mathbf{A}}_{\mathbb{PI}^t}$

  Tenter d'expliquer (x, y) (Section 3)

  Si l'explication existe :

    Collecter $\mathbb{S}^t$ (Les réactions du décideur)

    Mises à jour :

     ☛ $\mathbb{PI}^{t+1} = \mathbb{PI}^t \cup \mathbb{S}^t$

     ☛ $t \leftarrow t + 1$

Dans le squelette décrit ci-dessus, on voit bien (**Tant que** $|\mathbb{N}^{\mathbf{A}}_{\mathbb{PI}^t}| \neq 0$) qu'un facteur limitant de la collecte d'information préférentielle additionnelle est l'absence de déductions de comparaisons par paire d'alternatives de l'ensemble $\mathbf{A}$. Lorsque cette condition est vérifiée, le choix (Procédure `Choisir`) de la paire (x, y) à expliquer peut être aléatoire, conforme à une stratégie bien définie ou encore explicitement désignée par le décideur. Quant à la condition d'arrêt `ConditionArret`, elle pourrait traduire une contrainte sur la valeur de t (avoir t faible pour « pas trop fatiguer » le décideur) ou une condition sur l'évolution du nombre de recommandations admissibles (seuil ou convergence).

4.2 Illustration

Nous proposons ici, une instanciation du protocole détaillé précédemment. Les données du problème sont données par les Tables 4 et 5. On remarquera que les ensembles d'alternatives d'intérêt $\mathbf{A}$ et de référence $\mathbf{A}^R$ ne sont pas disjoints ($\mathbf{A} \cap \mathbf{A}^R = \{x^{50}, x^7\}$). Le type de recommandation considéré est un choix ($|\mathbf{A}^*| = 1$) et l'on dispose à l'étape $t = 0$ de l'information préférentielle ci-dessous :

$$\mathbb{PI}^0 = \{(r^{43}, r^{24}), (r^{24}, x^{50}), (x^{50}, x^7), (x^7, r^{17})\}$$

L'ensemble $\mathbb{N}^{\mathbf{A}}_{\mathbb{PI}^0}$ des déductions par calcul de la relation nécessaire sur l'ensemble des comparaisons par paires $\mathbb{A}$ se présente comme suit :

$$\mathbb{N}^{\mathbf{A}}_{\mathbb{PI}^0} = \{(x^{56}, x^{50}), (x^{56}, x^{37}), (x^{56}, x^7), (x^{26}, x^{50}), (x^{26}, x^7), (x^{25}, x^{50}), (x^{25}, x^7), (x^{50}, x^7)\}$$

L'ensemble des (5) recommandations admissibles étant donné $\mathbb{PI}^0$ est :

$$\{x^{56}\}, \{x^{26}\}, \{x^{25}\}, \{x^{42}\}, \{x^{11}\}$$

En effet, la présence des comparaisons par paire (x^{56}, x^{50}), (x^{56}, x^{37}), (x^{56}, x^7) (mises en évidence par la couleur violet dans $\mathbb{N}^{\mathbf{A}}_{\mathbb{PI}^0}$) suffit à montrer que les alternatives x^{50}, x^{37}, et x^7 ne sauraient être recommandées.

Table 4 – Description de l'ensemble $\mathbf{A}$.

	a	b	c	d	e	f
x^{56}	1	1	1	0	0	0
x^{26}	0	1	1	0	1	0
x^{25}	0	1	1	0	0	1
x^{42}	1	0	1	0	1	0
x^{11}	0	0	1	0	1	1
x^{50}	1	1	0	0	1	0
x^{37}	1	0	0	1	0	1
x^7	0	0	0	1	1	1

Le déroulé complet de l'illustration s'étend sur 4 étapes (décrites de façon uniforme) : l'explication transitive questionnable de la paire déduite $(x, y) \in \mathbb{N}^{\mathbf{A}}_{\mathbb{PI}^t}$ est fournie par la mise en évidence d'une chaîne dans un sous-graphe du *graphe d'explication* $\mathcal{G}^{\mathbb{PI}^t}_{(x,y)}$ et la donnée des swaps explicatifs. Les réactions du décideur par rapport à ces derniers

TABLE 5 – Description de l'ensemble $\mathbf{A}^R$.

	a	b	c	d	e	f
r^{43}	1	0	1	0	1	1
r^{24}	0	1	1	0	0	0
x^{50}	1	1	0	0	1	0
x^{7}	0	0	0	1	1	1
r^{17}	0	1	0	0	0	1

sont symbolisées par ✔ pour exprimer une validation et ✖ pour signifier une contestation. La déduction d'une paire supplémentaire suffisante pour réduire le nombre de recommandations admissibles est indiquée le cas échéant. La mise à jour de la $\mathbb{PI}$ ne s'effectue qu'avec les swaps de l'ensemble $\widetilde{S}_{\mathbb{PI}^t}$ (swaps dont les projections dans $\mathbb{X}$ n'appartiennent pas à la relation nécessaire (voir Sous-section 3.3)) : réaliser la mise à jour avec des swaps n'appartenant pas à $\widetilde{S}_{\mathbb{PI}^t}$ ne rajoute que de l'information redondante.

Étape $t = 1$: Explication de (x^{50}, x^7) de longueur 2

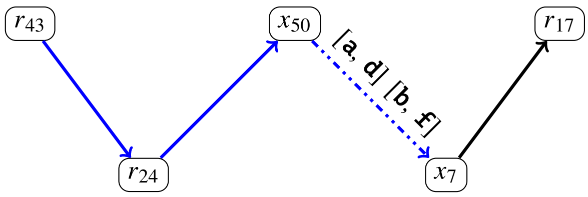

FIGURE 2 – Étape $t = 1$

Réactions du décideur : [a, d] : ✔ [b, f] : ✔

Mise à jour : $\mathbb{PI}^1 = \mathbb{PI}^0 \cup \{$[a, d], [b, f]$\}$

On a : $(x^{26}, x^{11}) \in \mathbb{N}^{\mathbf{A}}_{\mathbb{PI}^1}$, ce qui suffit à disqualifier la recommandation admissible $\{x^{11}\}$ à l'Étape $t = 0$.

L'ensemble des (4) recommandations admissibles est donc désormais :

$$\{x^{56}\}, \{x^{26}\}, \{x^{25}\}, \{x^{42}\}$$

Étape $t = 2$: Explication de (x^{56}, x^7) de longueur 3

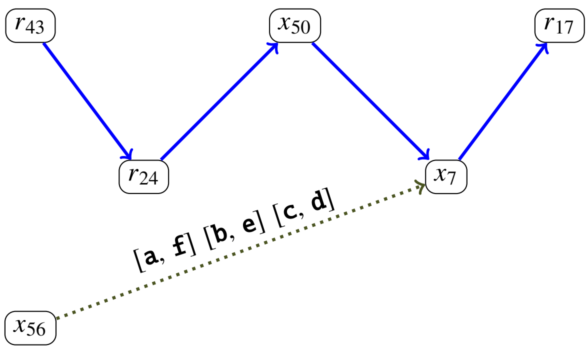

FIGURE 3 – Étape $t = 2$

Réactions du décideur : [a, f] : ✔ [b, e] : ✔ [c, d] : ✔

Mise à jour : $\mathbb{PI}^2 = \mathbb{PI}^1 \cup \{$[a, f], [b, e]$\}$ ([c, d] $\notin \widetilde{S}_{\mathbb{PI}^1}$)

On a : $(x^{56}, x^{25}) \in \mathbb{N}^{\mathbf{A}}_{\mathbb{PI}^2}$ et $(x^{56}, x^{42}) \in \mathbb{N}^{\mathbf{A}}_{\mathbb{PI}^2}$ et l'ensemble des (2) recommandations admissibles devient :

$$\{x^{56}\}, \{x^{26}\}$$

Étape $t = 3$: Explication de (x^{42}, x^7) de longueur 3

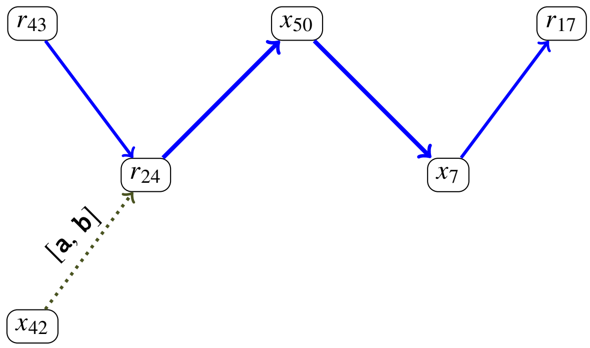

FIGURE 4 – Étape $t = 3$

Réactions du décideur : [a, b] : ✖

Mise à jour : $\mathbb{PI}^3 = \mathbb{PI}^2 \cup \{$[b, a]$\}$)

On n'a ni : $(x^{56}, x^{26}) \in \mathbb{N}^{\mathbf{A}}_{\mathbb{PI}^3}$ ni $(x^{26}, x^{56}) \in \mathbb{N}^{\mathbf{A}}_{\mathbb{PI}^3}$ et donc l'ensemble des recommandations admissibles reste :

$$\{x^{56}\}, \{x^{26}\}$$

Étape $t = 4$: Explication de (x^{56}, x^{37}) de longueur 3

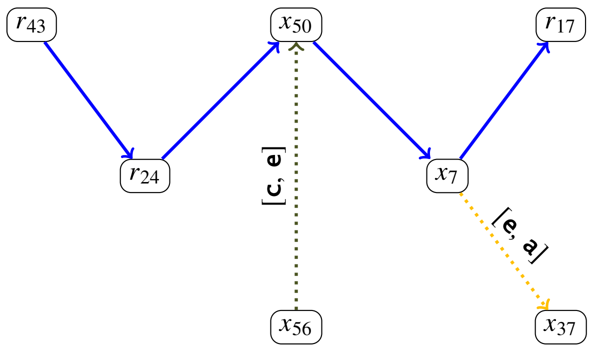

FIGURE 5 – Étape $t = 4$

Réactions du décideur : [c, e] : ✔ [e, a] : ✖

Mise à jour : $\mathbb{PI}^4 = \mathbb{PI}^3 \cup \{$[c, e], [a, e]$\}$)

On a : $(x^{56}, x^{26}) \in \mathbb{N}^{\mathbf{A}}_{\mathbb{PI}^4}$ et donc l'alternative à recommander est :

$$\{x^{56}\}$$

Pour convaincre le lecteur de la représentabilité par un modèle additif des préférences exprimées dans cette illustration, nous fournissons la fonction de score suivante :

$$\omega = \{\mathtt{a} : 29, \mathtt{b} : 32, \mathtt{c} : 66, \mathtt{d} : 24, \mathtt{e} : 20, \mathtt{f} : 18\}$$

5 Conclusion et Perspectives

Dans cet article, nous avons essayé de montrer que l'explication peut, en Aide MultiCritère à la Décision (AMCD), remplir une fonction d'élicitation de préférences. En faisant l'hypothèse de préférences représentables par un modèle additif sur domaine de critères binaire, nous avons proposé un programme mathématique qui calcule la plus courte explication transitive questionnable d'un élément (x, y) de la relation nécessaire. Pour rappel, la relation nécessaire sur un ensemble $\mathbf{X}$ d'alternatives est l'ensemble des comparaisons par paire (x, y) soutenues par l'ensemble des modèles compatibles avec l'information préférentielle $\mathbb{PI}$ fournie par le décideur.

La présente contribution complète la proposition [1] qui décrit une palette de schémas d'arguments utilisés pour justifier la comparaison par paire (x, y) soutenue par un modèle précis ω. En effet, elle décrit un nouvel opérateur qui combine suivant des règles précises une ensemble donné de connaissances (ici des préférences) pour en construire de nouvelles (définition du schéma d'arguments) ; ce nouvel opérateur s'inspirant des schémas transitif et de couverture. Notre contribution se distingue cependant desdites propositions de [1] d'une part, en ce qu'elle s'inscrit dans un cadre où les préférences du décideur sont exprimées sous la forme de l'ensemble des comparaisons par paire $\mathbb{PI}$ qui peuvent être soutenues par plusieurs modèles précis différents ω et d'autre part en ce ce que l'explication produite, du fait de son caractère questionnable, contribuera à « aider » l'analyste à déterminer de façon précise l'alternative à recommander.

Les explications transitives questionnables ont donc vocation à être utilisées dans un protocole interactif (entre un décideur et un analyste) devant aboutir à la formulation d'une recommandation. Les grandes lignes d'un tel protocole ont été esquissées en Sous-section 4.1 et une illustration en a été donnée (Sous-section 4.2). Dans cette dernière, on a pu remarquer que la collecte d'information préférentielle à l'aide des explications transitives questionnables a permis de déterminer avec précision la recommandation à faire au décideur. Il convient donc, dans des travaux futurs, d'essayer de quantifier cet apport en réalisant des simulations numériques portant sur une plus large variété de problèmes de décision (nombre de critères supérieur à 6, calibrage de la quantité d'information préférentielle déduite . . .). Les résultats de telles expérimentations seront édifiantes et permettront une instanciation plus précise de protocoles d'interaction (stratégie de choix de la paire à expliquer ou de l'explication à proposer lorsqu'il en existe plusieurs . . .) afin que cet apport soit maximal.

Références

[1] Amoussou, M., Kh. Belahcène, N. Maudet, V. Mousseau et W. Ouerdane: *Step-wise Explanations for the Additive Model*. Working Paper (hal-03964933), 2021.

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

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

[4] Hammond, J., R. Keeney et H. Raiffa: *Even Swaps : A Rational Method for Making Trade-offs*. Harvard business review, 76 :137–8, 143, 1998.

[5] Kadzinski, Milosz, Salvatore Corrente, Salvatore Greco et Roman Słowiński: *Preferential reducts and constructs in robust multiple criteria ranking and sorting*. Operations Research-Spektrum, 36 :1021–1053, octobre 2014.

[6] Labreuche, Christophe: *A general framework for explaining the results of a multi-attribute preference model*. Artificial Intelligence, 175(7) :1410–1448, 2011, ISSN 0004-3702. `https://www.sciencedirect.com/science/article/pii/S0004370210001979`, Representing, Processing, and Learning Preferences : Theoretical and Practical Challenges.

[7] Miller, George A.: *The Magical Number Seven, Plus or Minus Two : Some Limits on Our Capacity for Processing Information*. The Psychological Review, 63(2) :81–97, March 1956. `http://www.musanim.com/miller1956/`.

[8] Miller, Tim: *Explanation in Artificial Intelligence : Insights from the Social Sciences*. ArXiv, abs/1706.07269, 2019.