Explications de recommandations fondées sur des principes d'équité à l'aide de transferts

Hénoïk Willot, Khaled Belahcene, Sébastien Destercke

Explications de recommandations fondées sur des principes d'équité à l'aide de transferts

Hénoïk Willot Khaled Belahcene Sébastien Destercke

Heudiasyc, Université de Technologie de Compiègne, France

{henoik.willot, khaled.belahcene, sebastien.destercke}@hds.utc.fr

Résumé

Nous explorons la génération d'explication pour des préférences basées sur un modèle de somme pondérée ordonnée (*OWA*) favorisant une répartition équilibrée des performances relatives à différents points de vue. Nous proposons des explications, correctes vis-à-vis du modèle, fondées sur la composition par transitivité d'arguments élémentaires fondés sur la dominance de Pareto, des transferts de Pigou-Dalton, et l'information préférentielle fournie par le décideur. Nous proposons plusieurs approches heuristiques permettant de calculer ces explications, validées par une campagne expérimentale montrant que les explications obtenues sont souvent de longueur optimale.

1 Introduction

Our aim is to propose explanation tools for recommendations based on preference aggregation. The need for decision-theoretic recommender systems–tools helping decision makers to formalize and support their judgment in a principled manner–has in turn given rise to a need for tools allowing people–be they actors of the decision process, or third parties impacted by it–to understand, scrutinize, validate or contradict the functioning of such recommender systems.

Preference aggregation is the process of merging comparative judgments expressed from various points of view into a single ranking. Points of view can represent various aspects of a situation (such as in multiple criteria decision aiding–MCDA), be expressed by various agents, or several possible worlds when modelling uncertainty. In turn, the aggregated judgment can be used as a basis for decision, supporting tasks such as choosing the best alternative, comparing them, or sorting them into ordered categories [3].

The MCDA literature usually distinguishes three approaches to aggregation : "aggregate then compare", where judgments are normatively described as complete preorders and described numerically with a score–the higher, the better–and the aggregator is a multi-attribute utility function [11] ; "compare then aggregate", where the aggregated judgment is represented by an outranking relation constructed from the preference profiles [16] ; and models based on logic. Following the numeric approach, it is customary to decide on several high-level features of the aggregator–either technical, such as possessing an additive form [12] , or decision-theoretic, such as being compatible to Pareto-dominance, satisfying anonymity or idempotence, etc. Usually, these requirements are chosen so as to define, either directly or indirectly via a representation theorem, a parametric family of aggregators. When the decision task requires to be able to compare any two alternatives, the usual approach, called preference elicitation, is to select a specific, precise value of the preference parameter. It is common to use indirect elicitation techniques, where the aggregator is fitted to preference information (PI) given by the decision maker in the form of comparative statements about alternatives, as opposed to statements concerning the parameters [10]. Full elicitation is not mandatory, though : skeptical recommendations can be derived considering the whole set of aggregators of the family that are compatible to the PI.

In the context of MCDA, Belahcene et al. have recently shown that, for the class of additive aggregators, it is possible to provide structured explanations where elementary arguments are organized according to a specific scheme. In [1], an explanation of a comparative statement is a decomposition into elementary swaps linked together by transitivity. In [2], an explanation is a decomposition into preference statements committed by the decision maker, assembled together by a high-order cancellation property.

When points of view are assessed on the same scale, the Choquet integral is a convenient class of aggregators,

offering a good mixture of expressiveness, interpretabilty and computational tractability [7]. We focus on the subclass of *anonymous* aggregators, where the respective identities of the points of view play no role into their aggregation. Evaluations can be permuted, and the importance of a given score is related to its rank in the ordering of scores. These aggregators are thus named *ordered weighted average* (OWAs). Introduced in MCDA by Yager [18], they form a family of function parameterized by a tuple of weights, one per criteria, similarly to the weighted sum, and encompass the minimum, maximum, median and mean operators as particular cases. Moreover, by imposing the weights to be non-increasing w.r.t. the rank, it is possible to favour balanced scores over imbalanced ones, thus representing a sense of fairness. The explanatory engine described in [1] relies on swaps between criteria and is inspired by the notion of *even swaps* [9]. In the field of welfare economy, many methods and criteria are used to rank sets of incomes depending on the distribution of wealth among agents. The Pigou-Dalton principle [17] provides a similar notion of acceptable transfers in this context : the inequality between agents is reduced when a rich agent gives a small portion ϵ of its wealth to a poorer agent.

Our contribution is the definition of structured explanations for recommendations based on fairness-oriented OWAs. We propose to arrange comparative statements based on Pareto dominance, Pigou-Dalton transfers, and PI into a transitive structure. We begin by introducing the OWA operator, the Pigou-Dalton Principle and other definitions we will use in Section 2. In Section 3, we deal with the case where preference and explanations can be constructed without relying on preference information, and propose a heuristic to compute short explanations. In section 4, we address the case where preference is inferred from PI, and propose to find an additive decomposition of a comparative statement as an intermediate step towards finding a transitive explanation. Finally we will run an example combining both in Section 5 and give some insight about the performances of the method in Section 6.

2 Preliminaries

2.1 Ordered Weighted Averages

In this section we will introduce the decision problems and the formulation of the OWA operator. The MCDA problem we consider is ranking alternatives over a set of n criteria, defined on the same domain $\mathcal{X}$, which can be $[0, 1]$ or $\mathbb{R}$. The model of preference should then create an order over the candidates represented by their vectors $x \in \mathcal{X}^n$.

Definition 1 (Reordering function)**.** *We define the reordering function $^\uparrow$ as the permutation function over $\mathcal{X}^n$ s.t. $x \mapsto x^\uparrow$ with $x_1^\uparrow \leq x_2^\uparrow \leq \cdots \leq x_n^\uparrow$.*
We denote by $\mathcal{X}^{n\uparrow}$ the domain of such vectors $x^\uparrow$.

Definition 2 (Ordered Weighted Average [18])**.** *The OWA operator is a function $\mathcal{X}^n \to \mathbb{R}^+$ defined by a vector of weights $w \in \mathcal{W}$ s.t. $\forall\ i\ w_i \in [0, 1]$ and $\sum_{i=1}^n w_i = 1$:*

$$OWA_w(x) = \sum_{i=1}^{n} w_i x_i^\uparrow$$

Even if it is presented as a weighted sum, thanks to the reordering function $^\uparrow$ the weights are assigned to the rank of the criteria, allowing to represent non linear preference operators, such as the min operator $w_{min} = (1, 0, \ldots, 0)$, the max operator $w_{max} = (0, \ldots, 0, 1)$, and in general any quantile as well as the arithmetic mean operator $w_{mean} = (\frac{1}{n}, \ldots, \frac{1}{n})$.

Definition 3 (Fairness-oriented OWA (FOWA))**.** *An OWA operator is fairness-oriented if its weight vector w also satisfy $w_1 \geq w_2 \geq \cdots \geq w_n$. We note by $\mathcal{W}^\searrow \subset \mathcal{W}$ the domain of such weights.*

A FOWA represents preferences oriented toward equity because, with a higher emphasis put on the smaller values, it gives higher scores to vectors with balanced values than to vectors where modalities are concentrated on a small subset of criteria. We can note that an increase in a small variable has a bigger impact on the aggregated value than the same increase in a big variable, *i.e.* $\forall\ i < j\ OWA_w(x^\uparrow + k \times e_i) \geq OWA_w(x^\uparrow + k \times e_j)$, with e_l the vector that is one for its l^{th} element and zero everywhere else.

Definition 4 (Ranking relation)**.** *We define the ranking relation $\succeq_w$ induced by the operator OWA_w by :*

$$a \succeq_w b \iff OWA_w(a) \geq OWA_w(b)$$

Example 1. *The Decision Maker is asked to rank students over their results (scaled between 0 and 1) in 3 main courses {"Science", "Literature", "Language"}. She prefers students who are balanced between the 3 courses, and with an analyst the OWA with the following vector of weights has been designed : $w = (0.6, 0.3, 0.1)$.*

If we consider three students a, b and c such that $a = (0.7, 1, 0.5)$, $b = (0.7, 0.7, 0.7)$ and $c = (1, 0.7, 0.8)$. Their scores defined by the OWA w will be computed on their reordered vectors $a^\uparrow = (0.5, 0.7, 1)$, $b^\uparrow = (0.7, 0.7, 0.7)$ and $c^\uparrow = (0.7, 0.8, 1)$, and are :

$$OWA_w(a) = 0.5 \times 0.6 + 0.7 \times 0.3 + 1 \times 0.1 = 0.61$$
$$OWA_w(b) = 0.7 \times 0.6 + 0.7 \times 0.3 + 0.7 \times 0.1 = 0.7$$
$$OWA_w(c) = 0.7 \times 0.6 + 0.8 \times 0.3 + 1 \times 0.1 = 0.76$$

Therefore we obtain the preferences $c \succeq_w b \succeq_w a$.

In this example, we do not exactly know why the weight vector was $w = (0.6, 0.3, 0.1)$. The issue with determining

a specific set of weights is that it produces a total preorder (with possible ties) and may produce knowledge that the DM is not aware of and could potentially disagree with. Furthermore, obtaining precise values is cognitively demanding and require strong efforts. To circumvent this problem of finding the right set of weights, we can robustify our model using a set of models [14]. The set of OWA is defined as the set which respects the information obtained from the DM, her preferential information (PI), through an interactive process. In our case of study, the information collected is of the shape of m preference statements $a^j \succeq_{PI} b^j$, $j \in \{1, \dots, m\}$, with a^j, b^j alternatives.

Definition 5 (Robust OWA)**.** *We define a robust OWA operator the set $W_{PI} \subseteq \mathcal{W}$ of OWA weights :*

$$W_{PI} = \{w \in \mathcal{W} \ : \ \forall j \in \{1, \dots, m\} \ a^j \succeq_w b^j\}$$

And we note $W_{PI}^{\searrow} = \mathcal{W}^{\searrow} \cap W_{PI}$

As we now have a set of models instead of a single vector of weights, we have to adapt our process for producing preferences. We can define two relations of preferences from the set W_{PI}, a necessary and a possible preference relations [5]. In this paper we only focus on the necessary preference.

Definition 6 (Necessary preference)**.** *We define the necessary preference $\mathcal{N}_{PI}^{OWA^{\searrow}}$ of a robust FOWA with respect to preference information PI as :*

$$a \mathcal{N}_{PI}^{OWA^{\searrow}} b \Leftrightarrow \forall w \in W_{PI}^{\searrow}, \ a \succeq_w b$$

In order to compute the set W_{PI} and to reason with the necessary preference relation $\mathcal{N}_{PI}^{OWA^{\searrow}}$, we can adapt the GRIP method [4] that allows to decide whether a pair of alternatives belongs to the necessary preference relation, given some PI, for the additive value model, by solving a linear program [1]. In fact, to represent OWA operators, we only need to feed the method with vectors already reordered by $^{\uparrow}$, and for the representation of FOWA operators we need to add $n-1$ linear constraints $w_i \geq w_{i+1}$ [2], $i \in \{1, \dots, n-1\}$ to constraint the weights of the additive value function in the linear program to be decreasing.

2.2 Pigou-Dalton Principle and Dominance relations

In this section we will first connect the OWA aggregators to the Pigou-Dalton principle. To do so, we will introduce the dominance relations and the Pigou-Dalton principle that we will use in our explanation engine and its potential use cases.

1. Such a LP formulation could already be found in [8], but we opt to use the more streamlined formalism of GRIP.

2. to follow the GRIP methods notations, the constraints are $u_i(\beta_i) - u_{i+1}(\beta_{i+1}) \geq 0$

The Pigou-Dalton Principle was first introduced in welfare economic problem, where the components of the vector to compare are the incomes of economical agents, ranked from bottom to top [17]. In this context, the Pigou-Dalton Principle defines a relation $\succeq_{PDP}$ between two distributions over n agents.

Definition 7 (Pigou-Dalton Principle)**.** *Let x be the income vector of n agents such that $x = (x_1, \dots, x_n)$, $x_1 \leq x_2 \leq \dots \leq x_n$.*
The vector x' is favoured to x by the Pigou-Dalton Principle, noted $x \preceq_{PDP} x'$ if there exists points of view $i, j \in \{1, \dots, n\}$, $i < j$ and quantity $\epsilon > 0$ s.t. :

$$\begin{cases} \forall k \in \{1, \dots, n\}, \ k \neq i, \ k \neq j, \ x'_k = x_k; \\ x'_i = x_i + \epsilon \ \leq x_{i+1}; \ \textit{and} \\ x'_j = x_j - \epsilon \geq x_{j-1}. \end{cases}$$

When these conditions are met, we also use the notation $x \underset{PDP}{\overset{j \to i}{\leq}} x'$ to account for the witnesses i and j.

The Pigou-Dalton Principle therefore favours vectors of incomes where a "rich" agent j gives a positive portion ϵ of its wealth to a "poorer" agent i in order to reduce the inequality. We also add explicitly another constraint on ϵ which is not always clear in the literature : the order in the distribution is preserved, $x_j - \epsilon \geq x_{j-1}$ and $x_i + \epsilon \leq x_{i+1}$. This principle of equity is respected by FOWA operators as we discussed earlier after definition 3.

We will now introduce two preorders relations, compatible with FOWA operators, which will be used by our explanation engine : the Pareto and Lorenz dominance, denoted respectively as $\succeq_P$ and $\succeq_L$.

Definition 8 (Pareto-dominance)**.**

$$\forall a, b \in \mathcal{X}^n, \ a \succeq_P b \iff \forall i \in \{1, \dots, n\} \ a_i \geq b_i$$

Pareto dominance embodies the desirable property of monotonicity of a preference aggregator : if an alternative is better on every aspect than another, then it should be preferred.

Definition 9 (Lorenz vector)**.** *We call the Lorenz vector of a candidate a the cumulative vector $L(a)$ of $\mathbb{R}^n$ whose components are defined by :*

$$L(a)_i = \sum_{j=1}^{i} a_j^{\uparrow}$$

Definition 10 (Lorenz-dominance)**.** *We define the Lorenz-dominance $\succeq_L$ by :*

$$a \succeq_L b \iff \forall i \in \{1, \dots, n\} \ L(a)_i \geq L(b)_i$$

Example 2. *(Example 1 continued) The Lorenz vector of the three candidates* $\{a, b, c\}$ *are* $L(a) = (0.5, 1.2, 2.2)$*,* $L(b) = (0.7, 1.4, 2.1)$ *and* $L(c) = (0.7, 1.4, 2.1)$*.*
By comparing the Lorenz vectors we obtain the following Lorenz-dominance relation statements : $c \succeq_L a$ *and* $c \succeq_L b$*. We can note that the Lorenz-dominance is a partial preorder, as neither a or b Lorenz-dominates the other.*

FOWA operators are highly linked to Lorenz Dominance as shown by Golden and Perny [6].

Proposition 1 (Reformulation from Lemma 2 in [6])**.**

$$a \succeq_L b \Leftrightarrow \forall w \in \mathcal{W}^{\searrow}\ a \succeq_w b \Leftrightarrow a \mathcal{N}_{\emptyset}^{OWA^{\searrow}} b$$

Therefore, with proposition 1 we have that the Lorenz-dominance is compatible to any FOWA operator, meaning that these results will also appear in robust FOWA but are not depending on the DM preferential information. This first set of results makes up the core of our explanatory engine described in Section 3.

3 Transitive explanations for Lorenz dominance

In this section we will present an algorithm which computes efficiently an explanation for a Lorenz dominance in the form of a transitive chain of transfers using the Pigou-Dalton Principle. The former is only a small part of the results a robust FOWA can produce and the rest will be addressed in section 4.

We saw from proposition 1 that some results, those corresponding to the Lorenz dominance $\succeq_L$, are compatible with every FOWA operator. It naturally follows that these results, will appear in the necessary preferences of any robust FOWA operator. It has also be known since the 1960s that the Lorenz dominance and the Pigou-Dalton Principle are closely related.

Definition 11 (Transitive explanation (TE))**.** *Given a set of binary relations over alternatives Y, we call transitive explanation of* $a \succeq b$ *using Y, a tuple* $(x^0, \ldots, x^{k+1}) \in (\mathcal{X}^n)^{k+2}$ *such that*

$$a = x^0,\ b = x^{k+1}, \forall i \in \{1, \ldots, k\} x^i\ \mathcal{R}_i\ x^{i+1},\ \textit{with}\ \mathcal{R}_i \in Y$$

Proposition 2 ([15], reformulation from Proposition 3.1)**.** $a \succeq_L b$ *iff there exists a transitive explanation* $(x^0, \ldots, x^{k+1}) \in \mathcal{X}^{n\uparrow}$ *using* $Y = \{\succeq_{PDP}, \succeq_P\}$

In [15], Lorenz dominance is only considered over alternatives which have the same last value in their Lorenz vectors, i.e. for which the sum of all values is the same. As we want the scope of our explanation engine to be as broad as possible, we imbue it with the capability of inserting Pareto dominance statements in the explanation sequence to overcome this limitation and deal with the potential surplus.

From proposition 2, we can build a transitive sequence of preferences, a transitive explanation, combining only progressive Pigou-Dalton transfers and Pareto dominance to explain every pair (a, b) such that $a \succeq_L b$. Note that, because of the equivalence in proposition 1, the Lorenz-dominated alternatives are exactly the ones for which explanations can solely be based on Pareto and Pigou-Dalton transfers and we will need other explanation mechanisms to explain necessary preferences of a robust FOWA operator when alternatives are not Lorenz-dominated. As Pigou-Dalton transfers only redistribute a portion ϵ among candidates without breaking the order of criteria, and as Pareto dominance is only here to remove the surplus that can remain between the last intermediate candidate and b, every intermediate candidate from $(x^0, \ldots, x^{k+1})$ is within $\mathcal{X}^{n\uparrow}$.

Hence this sequence can be presented to the DM as an explanation because :

— it is of finite length ;
— the mechanisms are plausible, given the explainee adheres to the principles of monotonicity and fairness they embody ;
— the mechanisms used are of small cognitive load (Pareto dominance does not require any trade-off, while a Pigou-Dalton transfer can be described as occurring between two points of view, ignoring the rest) ; and
— the intermediate candidates used are plausible, even though they are not present in the set of candidates to rank.

The question of finding an algorithm to build a (not necessarily unique) sequence of Pigou-Dalton transfers has been answered in a close but not identical domain, on a problem called Majorization [13].
It is defined as a preorder over vectors using their values reordered in a decreasing way, so if we take similar notations as in definition 1 we would be dealing with vectors $x^{\downarrow}$. Majorization $a \succeq_M b$ occurs when, for each criterion k, we have $\sum_{i=1}^{k} a_i^{\downarrow} \geq \sum_{i=1}^{k} b_i^{\downarrow}$. The link with our problem is therefore clear as $a \succeq_M b \Leftrightarrow b \succeq_L a$.

In this context, the majorization is explained using a sequence of "Robin Hood transfers", which are Pigou-Dalton transfers, produced with a polynomial time algorithm. Without going into the mathematical details, we can give simply the idea of their (Lorenz-revisited) algorithm. We have $a \succeq_L b$, which means that a is more balanced than b. Especially, we can find some index j where $b_j^{\uparrow} > a_j^{\uparrow}$, and

some index $k < j$ where $a_k^{\uparrow} > b_k^{\uparrow}$. The idea is to perform an exchange between these two points of view of a quantity wich is as large as possible, i.e. $\epsilon = min(a_j^{\uparrow} - b_j^{\uparrow}, b_k^{\uparrow} - a_k^{\uparrow})$. Their idea for choosing suitable values for j and k is left vague; it is usually the smallest j possible and for this j the biggest k possible.

We can pinpoint three possible drawbacks :

1. the value of ϵ does not guarantee the candidate built by the transfer to be ordered;
2. it is limited to the case where $\sum_{i=1}^{n} a_i = \sum_{i=1}^{n} b_i$; and
3. the algorithm will find a sequence but does not aim at making it short.

Point #2 can easily be solved by allowing the explainer to use arguments based on Pareto dominance. However, this increase in flexibility makes point #3 even more prominent. Indeed, we have more flexibility in finding the criteria k so that $\sum_{i=1}^{k} a_i^{\downarrow} \geq \sum_{i=1}^{k} b_i^{\downarrow}$ as we have "surplus" (non zero Pareto-dominance implies $\sum_{i=1}^{n} a_i > \sum_{i=1}^{n} b_i$). We now present our (heuristic) Algorithm 1, which is similar to the idea given above but tries to solve the three points mentioned. A natural idea consists in minimizing the length of the explanation, but this problem seems computationally hard, even though we were not able to assess its theoretical difficulty. Consequently, we propose a heuristic method to compute short transitive explanations for a Lorenz dominance statement. In Section 6, we compare this heuristic to a A* algorithm computing an explanation of proven minimal length. Experiments tend to show we achieve nearly minimal length in a fraction of the time required to perform the exact search.

By focusing on Pigou-Dalton transfers occurring between variables which can receive (*Step 1*) or give (*Step 3 (I)*) the complete difference with the loser we ensure that the candidate obtained after the transfer is ordered. It also allows us to have a maximum length of explanation of n. Indeed at each step we remove at least one criteria from the set of criteria on which there is a non zero difference with the looser, bounding the explanation by the cardinal of this set, itself bounded by the number n of criteria.

Once we cannot find a rank j for which the attribute value x_j is bigger than b_j, we remove every surplus that could exists with a single Pareto dominance statement. Unfortunately, our heuristic algorithm does not always leads to the smallest explanation length as shown in example 3.

Example 3. *In the same context as example 1, we want to rank students, this time over their grades in 5 courses. The two students at hand are* $d = (0.6\ 0.7\ 0.5\ 0.7\ 0.8)$ *and* $e = (0.8\ 1\ 0.6\ 0.4\ 0.4)$. *It is easy to compute that* $d \succeq_L e$, *therefore we can apply our algorithm 1 and a A* search to have two explanations.*

Algorithm 1: Algorithm explaining Lorenz-dominance with Pareto dominance and Pigou-Dalton transfers

Input: $a, b \in \mathcal{X}^{n\uparrow}$ s.t. $a \succeq_L b$
Output: C
$x = a$; C = a
Compute $\mathcal{J}$, the set of indices j s.t. $x_j < b_j$ and s.t. we can perform a trade $\epsilon_j = b_j - x_j$
If $\mathcal{J} == \emptyset$ go to *Steps 5*
For each $j \in \mathcal{J}$:
  (I) Compute $\mathcal{K}$, the set of indices $k < j$ s.t. $x_k > b_k$ and s.t. we can perform a trade $\epsilon_k = x_k - b_k$
  (II) Find the index $k' \in \mathcal{K}$ allowing to perform the biggest trade $\epsilon = \max_{k' \in \mathcal{K}} min(\epsilon_j, \epsilon_{k'})$ (if draws take the largest index)
  (III) $X = (x_1, \ldots, x_{k'} - \epsilon, \ldots, x_j + \epsilon, \ldots, x_n)$
  (IV) If we don't have $X \succeq_L b$, go back to (II) to find another index in $\mathcal{K} \setminus \{k'\}$
  (V) $x = X$; C = C $\underset{PDP}{\overset{j \to k'}{\geq}}$ x
Go back to *Step 2*
Compute $\mathcal{K}$, the set of indices k s.t. $x_k > b_k$
If $\mathcal{K} \neq \emptyset$: C = C $\succeq_P b$

We reverse the explanation returned by the algorithm as it is easier to read and understand when the Pigou-Dalton transfers are performed in the reading direction. We note with $\bar{i}$ *the criterion receiving and* $\underline{i}$ *the criterion giving. We obtain for our algorithm the explanation :*

$e^{\uparrow} = (0.4\ 0.4\ 0.6\ 0.8\ 1) \preceq_P (0.4\ \overline{0.5}\ 0.6\ 0.8\ 1) \underset{PDP}{\overset{5 \to 2}{\preceq}}$
$(0.4\ \overline{0.6}\ 0.6\ 0.8\ \underline{0.9}) \underset{PDP}{\overset{4 \to 1}{\preceq}} (\overline{0.5}\ 0.6\ 0.6\ \underline{0.7}\ 0.9) \underset{PDP}{\overset{5 \to 3}{\preceq}}$
$(0.5\ 0.6\ \overline{0.7}\ 0.7\ \underline{0.8}) = d^{\uparrow}$ *of length 4.*
With the A search we obtain a different explanation :*
$e^{\uparrow} = (0.4\ 0.4\ 0.6\ 0.8\ 1) \preceq_P (0.4\ 0.4\ \overline{0.7}\ 0.8\ 1) \underset{PDP}{\overset{5 \to 2}{\preceq}}$
$(0.4\ \overline{0.6}\ 0.7\ 0.8\ \underline{0.8}) \underset{PDP}{\overset{4 \to 1}{\preceq}} (\overline{0.5}\ 0.6\ 0.7\ \underline{0.7}\ 0.8 = d^{\uparrow})$ *of length 3.*

In conclusion, our algorithm 1 computes in polynomial time a chain of Pigou-Dalton transfers and Pareto dominance statement to explain any Lorenz dominance statement with a bounded length of n statements. The length of the explanation is unfortunately not minimal, but the true minimal length can be computed for example with a A* algorithm over a graph exponential in size in the number of criteria.

As we saw previously, Lorenz dominance statements form a subset of the preference yielded by a robust FOWA aggregator, missing the part entailed by the specific PI obtained from the decision maker. Thus the idea we will deve-

lop in Section 4 is to complete Pigou-Dalton transfers and Pareto dominance with a combination of statements deduced from the PI to produce a sequence of preferences, or at least a decomposition of preferences, to explain every necessary preference statements obtained by a robust FOWA operator.

4 Additive and transitive decompositions of necessary preference statements

In this section we introduce and justify a linear programming model producing decompositions for every necessary preference statement compatible with the robust FOWA $W_{PI}^{\searrow} \subseteq \mathcal{W}^{\searrow}$ constrained by the PI. We start by introducing the notion of decomposition, which is weaker than the one of transitive explanation, we then try to assemble them into a transitive sequence of preferences. We will first introduce notations for the problem, then we present the model and finally we explain every variable and constraint and discuss it.

The problem we want to solve is finding an explanation for the necessary preference statement $c \; \mathcal{N}_{PI}^{OWA^{\searrow}} \; d$, $c, d \in \mathcal{X}^n$. We have 3 mechanisms at our disposal : PI statements provided by the DM, Pigou-Dalton transfers and Pareto dominance. Trying to find directly a valid transitive explanation $(x^0, \ldots, x^{k+1}) \in \mathcal{X}^{n\uparrow}$ using $\{\preceq_P, \preceq_{PDP}, \preceq_{PI}\}$ is a difficult planning problem, so we begin by building an additive decomposition of this statement.

Definition 12 (Decomposition)**.** *We call decomposition of $a \succeq b$ by Y, Y a set of explanation mechanisms, the "proto-explanation" defined by :*

$$\forall i \in \{1, \ldots, n\} \; a_i - b_i = \sum_{y \in Y} \gamma_{yi}$$

γ_y is the contribution vector of explanation mechanism y to the preference $a \succeq b$

Remark 1. *If we take a transitive explanation $\{x^0, \ldots, x^{k+1}\}$ of $a \succeq b$ using Y, we have $a = x^0, b = x^{k+1}$ and $\forall j \in \{0, \ldots, k\} \; x^j \; \mathcal{R}_j \; x^{j+1}$, $\mathcal{R}_k \in Y$.*
We can rewrite the latter as $x^j - x^{j+1} = \gamma_{\mathcal{R}_j}$. By summation we obtain $a - b = \sum_{j=0}^{k} x^j - x^{j+1} = \sum_{j=0}^{k} \gamma_{\mathcal{R}_j}$. Therefore an only decomposition based "proto-explanation" is weaker than a transitive explanation in the sense that any transitive explanation can be rewritten as a decomposition.

As we have seen, invoking the anonymity of the model, we rewrite the statement $a^j \succeq_{PI} b^j$ by $a^{j\uparrow} - b^{j\uparrow}$. We can then build a $m \times n$ matrix PI containing these m statements. Abusing notations, we will use interchangeably the set of statement PI and the matrix PI.
Previously, when we wanted to compute necessary and possible preference statements (with the GRIP method), we used the PI as a constraint to find the necessary relation, so we used $(a^{j\uparrow} - b^{j\uparrow}) \times w \geq \vec{0}$ as a constraint of the model. Here however we will interpret the statement $a^{j\uparrow} - b^{j\uparrow} = \mathrm{PI}_j$ as a trade-off between ranks of attributes, for which the DM agreed it increases the quality of a candidate. This positive trade-off could be interpreted *ceteris paribus* as an argument in support of having $x^{\uparrow} + \mathrm{PI}_j$ preferred to $x^{\uparrow}$, for any alternative $x \in \mathcal{X}^n$, provided $x^{\uparrow} + \mathrm{PI}_j$ also defines an alternative belonging to $\mathcal{X}^{n\uparrow}$. Moreover, we augment this argument by allowing it to invoke the *positive homogeneity* of the model.

Definition 13 (PI dominance)**.**

$$b^{\uparrow} \underset{PI}{\overset{\lambda_j \times \mathrm{PI}_j}{\leq}} a^{\uparrow} \Leftrightarrow a^{\uparrow} = b^{\uparrow} + \lambda_j \times \mathrm{PI}_j$$

We note that this dominance is a refinement of a $\mathcal{N}_{PI_j \setminus \emptyset}^{OWA^{\searrow}} \; b$

We can then compute the contribution of the PI as $\gamma_{PI} = \mathrm{PI}^T \times \lambda$ with $\lambda \in \mathbb{R}_m^+$.

Small reminder, a Pigou-Dalton transfer occurs between two variables i, j with $i < j$ and we remove a quantity ϵ from criterion j to give it to criterion i. We obtain a contribution $\gamma_{PDP} = (e_k - e_j) \times \epsilon$. As several Pigou-Dalton transfer can occur in the decomposition, the total contribution of the transfers can be written for each criterion i as $\gamma_{PDP,i} = \tau$ with $\tau \in \mathbb{R}_n$. To be able to write linear constraints over this type of transfers, we will divide this vector τ in two vectors τ^+ and τ^-, with $\tau_i^+ \in \mathbb{R}^-$ (resp. $\tau_i^- \in \mathbb{R}^+$) the quantity criterion i have to give (resp. receive).

Pareto-dominance between a and b can be simply written as a vector $\mu \in \mathbb{R}_n^+$, so the contribution of Pareto-dominance is $\gamma_P = \mu$.

We now have the contribution of our 3 explanation mechanisms and we can sum them up for every criteria $i \in \{1, \ldots, n\}$ and obtain :

$$c_i^{\uparrow} - d_i^{\uparrow} = \tau_i^+ + \tau_i^- + \mu_i + \sum_{j=1}^{m} \lambda_j \times (a_i^{j\uparrow} - b_i^{j\uparrow}) \qquad (1)$$

The set of equation we build from (1) and the domains form a linear program, but unfortunately we have to introduce some binary variables to constrain the Pigou-Dalton variables τ^+ and τ^- left unconstrained. We will have to go back to the definition of Pigou-Dalton transfers to build the constraints.

First a Pigou-Dalton transfer is balanced, we only redistribute ϵ between criteria. This can be written as a linear constraint :

$$\left\|\tau^+\right\|_1 - \left\|\tau^-\right\|_1 = 0 \qquad (2)$$

Then, we know that the transfers take from a criterion j a positive quantity to give to a criterion i such that $i < j$.

We can then write $\forall k \in \{1, \ldots, n\} \sum_{i=k}^{n} \tau_i^+ + \tau_i^- <= 0$ which can be synthesized with the matrix constraint :

$$\mathrm{U} \times (\tau^- + \tau^+) \leq \vec{0} \tag{3}$$

with U the upper triangular matrix of size $n \times n$ with ones in the upper triangle.

Finally we need to introduce some auxiliary binary variables t^+ and t^- used to avoid useless Pigou-Dalton transfers. Their goal is to ensure that a criterion k is used only as a receiver or a giver in the transfers (otherwise we will perform useless trades using k as an unnecessary intermediate). We obtain the mixed integer linear equations using M a big constant we will discuss after :

$$\mathrm{M} \times t^+ + \tau^+ \geq \vec{0} \tag{4}$$

$$\mathrm{M} \times t^- - \tau^- \geq \vec{0} \tag{5}$$

$$t^+ + t^- \leq \vec{1} \tag{6}$$

Our big M has to be scaled as an upper bound of the maximum width that can be used in a Pigou-Dalton transfer (as receiver or giver). Its computation is fairly easy, in the normalised and general case. In the normalized case we can use $\mathrm{M} = 1$ as the maximum value we can add or retrieve to a criteria is 1. In the general case, we can only redistribute values through criteria, so if we take $\mathrm{M} = \|c\|_1 + \|d\|_1$, we are sure to cover every case.

Example 4. *We take a small illustrative example. We suppose that the DM has expressed her preferences over two candidates a^1 and b^1, resulting into the statement* $\mathrm{PI} = (-0.05\ 0.1\ -0.05\ 0)$.
The problem is then of ranking two candidates $a^\uparrow = (0.5\ 0.5\ 0.7\ 0.9)$ *and* $b^\uparrow = (0.3\ 0.7\ 0.8\ 0.8)$. *As* $L(a) = (0.5\ 1\ 1.7\ 2.6)$ *and* $L(b) = (0.3\ 1\ 1.8\ 2.6)$ *there is no Lorenz dominance between them but the GRIP method finds that* $a\ \mathcal{N}_{PI}^{OWA^\searrow}\ b$.
For the statement $a^\uparrow - b^\uparrow = (-0.05\ 0.1\ 0\ -0.05)$ *a decomposition can be :*

- $PI = 1 \times (-0.05\ 0.1\ -0.05\ 0)$
- $\tau^+ = (0\ 0\ 0\ -0.05)$
- $\tau^- = (0\ 0\ 0.05\ 0)$
- $\mu = \vec{0}$

Theorem 1 (Characterization of the necessary preference with decomposition)**.**

$$a \mathcal{N}_{PI}^{OWA^\searrow} b \Leftrightarrow \exists \mu, \tau^+, \tau^-, \lambda \ s.t. \ (a^\uparrow - b^\uparrow) = \mathrm{PI}^T \times \lambda + \mu + \tau^+ + \tau^-$$

Démonstration. Proving that if we have such a decomposition for $(a^\uparrow - b^\uparrow)$ then we have $a\ \mathcal{N}_{PI}^{OWA^\searrow}\ b$ is obvious.
Proving the converse will require the use of Farkas' lemma. First, we suppose that the polytope obtained in $\mathcal{W}^\searrow$ by the intersection of PI statements, *i.e.* the intersection of the hyperplanes $(a^{j\uparrow} - b^{j\uparrow}) \times w = 0$, is consistent, meaning that the robust FOWA $W_{PI}^\searrow$ is non-empty.

Then, we have $a \mathcal{N}_{PI}^{OWA^\searrow} b$, meaning that it is not possible to find a set of weights $w \in W_{PI}^\searrow$ such that $b \succ_w a$, i.e. $(a^\uparrow - b^\uparrow) \times w < 0$.
Therefore by using Farkas' lemma, we can write $(a^\uparrow - b^\uparrow) \times w$ as a linear combination of our constraints, *i.e.* PI statements, Pigou-Dalton transfers and Pareto dominance. □

To complete our linear program we have to define our objective. We want to provide an explanation, as easy and as short as possible. Our goal is then first to minimize the contribution of our explanation mechanisms in the decomposition. But our explanation mechanisms are not on the same scale of complexity for the user. Pareto dominance is natural and easy to understand. Pigou-Dalton transfers are a little harder to interpret but as it is the reason we use fairness-oriented OWA and are binary exchanges, it is still easy to understand, with a small cognitive cost. However, using a positive linear combination of PI statements is hard to understand, as it is performed with the anonymized variables and not the base variables, especially if the computed coefficients are complicated and the statements include multiple variables, so it should be limited as much as possible.

We represent this complex situation with multiple objectives :

- minimize the number of PI statements involved
- reduce the part of Pigou-Dalton in the decomposition, *i.e.* maximize Pareto dominance
- find the nicest coefficients values for the PI statements

The idea of finding the nicest coefficients is that when we use the PI dominance $\overset{\lambda_j \times \mathrm{PI}_j}{\underset{PI}{\leq}}$, we want the λ_j displayed to be at least rational [3], and preferably integer. To do so we will introduce in equation 1 an integer coefficient α, corresponding to the denominator of our λ_j which becomes also an integer. We obtain :

$$\alpha \times (c_i^\uparrow - d_i^\uparrow) = \mathrm{PI}^T \times \lambda + \mu + \tau^+ + \tau^- \tag{7}$$

with $\lambda \in \mathbb{N}^m$ and $\alpha \in \mathbb{N}^*$.

Subsequently, we used the (single-)objective function $\mathcal{F} = \|\lambda\|_1 + \alpha + \frac{1}{\mathcal{M}} \|t^-\|_1$ to minimize, but other approaches can be considered, such as minimising a norm L_0 or canceling balancing effects between λ and α.

To summarize our mixed integer program is represented in figure 1.

3. This is always possible, because the constraints are expressed using integers.

$$
\begin{aligned}
&F.obj : Min\ \mathcal{F}\\
&\quad \text{subject to}\\
&\alpha \times (c_i^{\uparrow} - d_i^{\uparrow}) = \mathrm{PI}^T \times \lambda + \mu + \tau^+ + \tau^-\\
&\mathrm{M} \times t^+ + \tau^+ \geq \vec{0}\\
&\mathrm{M} \times t^- - \tau^- \geq \vec{0}\\
&t^+ + t^- \leq \vec{1}\\
&\left\|\tau^+\right\|_1 - \left\|\tau^-\right\|_1 = 0\\
&\mathrm{U} \times (\tau^- + \tau^+) \leq \vec{0}\\
&\lambda \in \mathbb{N}^m \quad \mu \in \mathbb{R}_+^n, \quad \tau^+ \in \mathbb{R}_-^n, \quad \tau^- \in \mathbb{R}_+^n\\
&\quad \alpha \in \mathbb{N}^*, \quad t^+ \in \{0,1\}^n, \quad t^- \in \{0,1\}^n.
\end{aligned}
$$

FIGURE 1 – A linear program allowing to find an additive decomposition of a comparative statement.

5 An illustrative example

In this section we will run an example, close to a real-case study, to use our algorithm and linear program to explain inferred preferences. The decision maker is the person in charge of the recruitment of students in a large french University. Her University has a large variety of formations but the policy is to form students toward general knowledge with specialisation in the last year. Therefore her preferences are oriented toward students with balanced overall good results instead of students specialised in some field only.

She is asked by the University her preferences over a set of 7 students, represented by their grades (on a scale from 0 to 20) in 4 important subjects : {"Mathematics", "Philosophy", "Biology", "English" }. This small subset of students corresponds to students from "*classes préparatoires*" and is fundamentally smaller that the set of students she will receive later in the year from students which just obtained their baccalaureate exam. Therefore she wants to automatize the process of building her preferences, even if it returns incomplete rankings, as long as she can have non technical explanations of the results, both for her personal use and for answering questions coming e.g. from rejected students or from a supervising regulatory institution. For all these reasons we presented her (without explicitly tell her the name) our robust FOWA operator.

To avoid any biases based on the names of candidates, we randomized them and assigned them a capital letter in $\{a, b, c, d, e, f, g\}$. The results of the candidates (reordered) to rank are :

Student	#1	#2	#3	#4
$a^{\uparrow}$	5	13	14	18
$b^{\uparrow}$	5	15	15	16
$c^{\uparrow}$	6	13	16	16
$d^{\uparrow}$	7	10	17	18
$e^{\uparrow}$	8	9	16	20
$f^{\uparrow}$	6	11	17	17
$g^{\uparrow}$	7	11	16	17

To start our model with preferences, we asked the decision maker to rank two pairs of candidates and she replied that $b \succeq_{PI} c$ and $d \succeq_{PI} e$.

We then compute the robust FOWA operator $W_{PI}^{\searrow}$ with a linear program inspired from the GRIP method [4] and obtain the necessary preference order which is : $b\ \mathcal{N}_{PI}^{OWA^{\searrow}}\ c$, $b\ \mathcal{N}_{PI}^{OWA^{\searrow}}\ g$, $b\ \mathcal{N}_{PI}^{OWA^{\searrow}}\ d$, $g\ \mathcal{N}_{PI}^{OWA^{\searrow}}\ a$, $d\ \mathcal{N}_{PI}^{OWA^{\searrow}}\ e$ and $d\ \mathcal{N}_{PI}^{OWA^{\searrow}}\ f$, and all preferences deduced by transitivity. We can also compute the Lorenz dominance for every pair of candidates, obtaining preference statements $b \succeq_L a$, $g \succeq_L a$, $c \succeq_L a$, $c \succeq_L f$, $d \succeq_L f$.

By comparing the two preference sets, we can see that adding the preferential information $c \overset{1\times \mathrm{PI}_1}{\underset{PI}{\preceq}} b$ and $e \overset{1\times \mathrm{PI}_2}{\underset{PI}{\preceq}} d$ creates new preferences among candidates such as $c\ \mathcal{N}_{PI}^{OWA^{\searrow}}\ g$ and $c\ \mathcal{N}_{PI}^{OWA^{\searrow}}\ d$ and transitive ones such as $b\ \mathcal{N}_{PI}^{OWA^{\searrow}}\ g$, $b\ \mathcal{N}_{PI}^{OWA^{\searrow}}\ d$, $b\ \mathcal{N}_{PI}^{OWA^{\searrow}}\ e$ and $b\ \mathcal{N}_{PI}^{OWA^{\searrow}}\ f$. To sum up, our robust FOWA produces 14 preference relations, including 2 PI from the user and 12 statements to explain, 5 with Lorenz-dominance explanations and 7 with our decomposition program.

We will not detail all of these 12 statements, only one based on Lorenz-dominance such as $g \succeq_L a$ and two from our linear program such as $b\ \mathcal{N}_{PI}^{OWA^{\searrow}}\ e$ and $c\ \mathcal{N}_{PI}^{OWA^{\searrow}}\ g$. To present thing shortly we will refer as $PI_1 = (-1\ 2\ -1\ 0)$ and $PI_2 = (-1\ 1\ 1\ -1)$ the vectors corresponding respectively to the PI statements $b \succeq_{PI} c$ and $d \succeq_{PI} e$.

With algorithm 1, we obtain as explanation for $g \succeq_L a$: $a^{\uparrow} = (5\ 13\ 14\ 18) \overset{2\to 1}{\underset{PDP}{\preceq}} (\overline{7}\ \underline{11}\ 14\ 18) \overset{4\to 3}{\underset{PDP}{\preceq}} (7\ 11\ \overline{16}\ \underline{16}) = g^{\uparrow}$.

With linear program (1) we find two decompositions for $b\ \mathcal{N}_{PI}^{OWA^{\searrow}}\ e$ and $c\ \mathcal{N}_{PI}^{OWA^{\searrow}}\ g$:

— $b^{\uparrow} - e^{\uparrow} = (-3\ 6\ -1\ -4) = PI_1 + 2 \times PI_2 + \nu^+ + \nu^-$
with $\nu^- = (0\ 2\ 0\ 0)$ and $\nu^+ = (0\ 0\ -2\ 0)$

— $c^{\uparrow} - g^{\uparrow} = (-2\ 4\ -1\ -1) = PI_1 + \nu^+ + \nu^-$
with $\nu^- = (0\ 1\ 0\ 0)$ and $\nu^+ = (0\ 0\ 0\ -1)$

We were also able to find through a brute force algorithm (testing all permutation) a sequence of preferences using the decomposition to produce a transitive explanation sequence, but it is not a general result.

For $b \succ_{PI}^{OWA^{\searrow}} e$ we found : $e^{\uparrow} = (8\ 9\ 16\ 20) \overset{3\to 2}{\underset{PDP}{\preceq}} (8\ \overline{11}\ \underline{14}\ 20) \overset{2\times PI_2}{\underset{PI}{\preceq}} (\underline{6}\ \overline{13}\ \overline{16}\ \underline{16}) \overset{1\times PI_1}{\underset{PI}{\preceq}} (\underline{5}\ \overline{15}\ \underline{15}\ 16) = b^{\uparrow}$

For $c \succ_{PI}^{OWA^{\searrow}} g$ we found : $g^{\uparrow} = (7\ 11\ 16\ 17) \overset{1\times PI_1}{\underset{PI}{\preceq}} (\underline{6}\ \overline{13}\ \underline{15}\ 17) \overset{4\to 3}{\underset{PDP}{\preceq}} (6\ 13\ \overline{16}\ \underline{16}) = c^{\uparrow}$

We can note that if we tried to apply the decomposition in another order, it would build intermediate candidates which are not ordered.

Length	#Values	% identical values	Max difference
1	82 334	100 %	-
2	160 431	96.97 %	4
3	162 967	89.83 %	5
4	135 623	80.61 %	4
5	91 777	71.46 %	3
6	45 148	66.09 %	2
7	13 120	68.53 %	1
8	1539	100 %	-

TABLE 1 – Length difference between algorithm 1 and A* algorithm.

6 Experimental results

In this section we will present some results obtained from experiments. We have generated randomly 1000 samples of 50 candidates (non Pareto-dominated) over 8 criteria with 5 PI statements. To compute these PI statements we draw from a Dirichlet distribution a set of ground-truth FOOWA weigth $w^{\searrow}$, and drew randomly 5 pairs (a^i, b^i) of non Lorenz-dominated candidates and added to the robust FOOWA the preference induced by $w^{\searrow}$. For the rest of the experiment the ground-truth will be hidden. From these sample we studied two aspects of our method. In the first place, we computed for Lorenz-dominated pairs of candidates the results of our algorithm 1 against A* algorithm and we compared the obtained length to the true minimal length and the computation times. In the second place, we tried through several means to compute a transitive explanation from the decompositions obtained by our MILP formulation detailed in figure 1.

6.1 Algorithm 1 against A* algorithm

Our goal is to compare the explanation length and the computation time our polynomial heuristic algorithm 1 with a method guaranteeing to find the "true" minimum length. As introduced in Section 3 we do so by the means of a A* search. Indeed, the idea underlying the A* algorithm is to find the shortest path between the winning candidate to the loosing candidate using Pigou-Dalton transfers and Pareto-dominance. The heuristic we use to estimate the distance to the loosing candidate is the number of positions which need to receive from a Pigou-Dalton transfer, plus 1 if there are more variables in a position to give than to receive. This estimate is indeed a lower bound of the number of remaining trades.

We represent in table 1 the difference between the length found by the A* algorithm and our algorithm 1. Each row i of the table corresponds to a true length of i Pigou-Dalton transfers and Pareto-dominance. We can see that overall the results obtained by our algorithm 1 do not differ from the A* algorithm, except in some cases where up to 34% (in line 6) of the results are worst. But still, our algorithm is of interest, especially when comparing computation times from table 2 and keeping in mind that during the process of finding a transitive explanation we can call for Pigou-Dalton transfers several times.

Length	#Values	Whisker	Q1	Median	Q3
1	82 334	1.37%	5.26%	6.54%	7.85%
2	160 431	11.74%	12.53%	20.10%	23.5%
3	162 967	-3.92%	20.6%	28.5%	33.61%
4	135 623	39.94%	26.54%	33.79%	39.21%
5	91 777	1.09%	38.00%	46.57%	62.05%
6	45 148	53.12%	79.18%	86.16%	91.23%
7	13 120	7.55%	93.47%	96.59%	97.96%
8	1 539	58.20%	98.10%	99.26%	99.6%

TABLE 2 – Percentage of reduction of compute time between algorithm 1 and A* algorithm.

We have represented in table 2 the percentage of reduction in the computation time by using algorithm 1 instead of the A* algorithm. We can see that only very little data (in line 3) are corresponding to the A* algorithm performing better, and with high number of criteria we see a strict dominance of algorithm 1. Therefore, for the rest of the study combining the resolution of finding a concise explanation for Lorenz-dominance and PI dominance, we will be using algorithm 1 for its benefits in computational cost.

6.2 Feasibility of transitive explanation for decomposition computed by the MILP formulation

Our goal is to study the availability of an explanation from the decomposition found by our MILP formulation in figure 1 for a necessary preference statement. We recall that a valid explanation is a sequence of progressive transfers from the winner to the looser, using our 3 explanation mechanisms $\{\preceq_P, \preceq_{PDP}, \preceq_{PI}\}$, while verifying that each intermediate candidate used in the explanation is valid, *i.e.* with values of criteria ordered and belonging to the domain.

To compute such a sequence from a set of values $(\frac{1}{\alpha} \times \lambda, \mu, \nu^+, \nu^-)$, we search for a permutation verifying our validity constraint. We allow to permute each individual PI statement with the Pareto dominance and the total Pigou-Dalton transfers. This algorithm does not try to divide any of this statements. We obtain some encouraging results on the length of explanation, displayed in table 3, unfortunately too many preferences fail to have a valid permutation : we found 192 436 cases where we cannot build the transitive explanation, meaning that 73% of our explanations cannot be interpreted into valid sequences.

PI involved	#Explanations	Q1	Median	Q3
1	64 450	4	6	8
2	5 941	7	8	9
3	491	8	9	10
4	16	9.75	11	12
Not found	192 436	-	-	-

TABLE 3 – Distribution of explanation length by quantity of PI involved.

A first hypothesis to explain these results outside our valid domain could be that, because the PI itself is not linear and does not guarantee to stay inside the domain (where Pigou-Dalton transfers and Pareto dominance do), maybe some statements are required to be performed but cannot be taken separately. They should be performed in sequence but without looking into the intermediary result. Therefore we build a unique "whole PI" statement to recompute permutation with. New sequences are found but their number, 914, is not sufficient to conclude for the significance of the hypothesis.

Another cause for the unfeasibility of scheduling the decomposition into a valid transitive explanation could be the proximity of candidates to the boundaries of the domain $\mathcal{X}^n$, making the use of the PI statements hard without fragmentation. We therefore performed the experience again with candidates sampled over $\mathcal{X}^n = [0.1; 0.9]^n$ instead of $[0; 1]^n$, leaving this space for the explanations. We obtain the results in table 4.

PI involved	#Explanations	Q1	Median	Q3
1	84 347	4	6	8
2	8 315	7	8	9
3	596	8	9	10
4	22	10	11	11
Not found	161 797	-	-	-

TABLE 4 – Distribution of explanation length by quantity of PI involved for candidates sampled over $[0.1; 0.9]^n$.

We obtain better results, even adding 1 229 more with the "one PI statement" augmentation, but we still have a humongous quantity (161 797 so around 63%) of non valid sequences. Finally, it could be worthwhile to consider a finer-grained scheduling of PI statements, where a statement $\lambda_k \times PI_k$ could be split into $\delta_k \lambda_k \times PI_k$ and $(1-\delta_k) \times \lambda_k \times PI_k$. We could use these statements in different locations in the explanation to build a sequence which could seem like $c^{\uparrow} \underset{PI}{\geq} c^{\uparrow} - \delta \times \lambda \times PI \succeq_L d^{\uparrow} + (\vec{1} - \delta) \times \lambda \times PI \underset{PI}{\geq} d^{\uparrow}$.

7 Conclusion and future works

In this paper we looked into the problem of producing explanations for the robust fairness-oriented OWA. The explanations are designed to be the least technical possible to be shared with external users which are possibly not involved in the conception of the system and have very little knowledge about it. To understand our explanations, it is only required to know why the model is used, preferring candidates with a balanced profile and therefore agrees with the Pigou-Dalton principle of transfers on the ordered profile of the candidate.

We presented an algorithm which builds a transitive chain for Lorenz-dominance as an explication. The existence of the chain is known from a long time [15] as a result in welfare economy and also the link with the OWA operator [6]. This algorithm works in a polynomial time in the number n of criteria, with a length of maximum n transfers. It can be used to explain Lorenz dominance in general and not only in the fairness-oriented OWA setting, where it corresponds to the results obtained by the necessary preference relation $\mathcal{N}_{\emptyset}^{OWA^{\searrow}}$.

We then presented a mixed integer linear program to compute an additive decomposition of a preference in $\mathcal{N}_{PI}^{OWA^{\searrow}}$, composed by a positive linear combination of the preference statements given by the DM, Pigou-Dalton transfers and Pareto-Dominance. This decomposition however is not a transitive explanation, and the existence of the latter for a given decomposition is still unknown yet, even though we can always build a decomposition (Theorem 1). We plan to adopt planning tools to further investigate the problem of computing transitive explanations, and the question whether guiding this search by the pre-computation of an additive decomposition is efficient or not.

Références

[1] Belahcene, Khaled, Christophe Labreuche, Nicolas Maudet, Vincent Mousseau et Wassila Ouerdane:

Explaining robust additive utility models by sequences of preference swaps. Theory and Decision, 82(2) :151–183, 2017. Number : 2 Publisher : Springer.

[2] Belahcène, Khaled, Christophe Labreuche, Nicolas Maudet, Vincent Mousseau et Wassila Ouerdane: *Comparing Options with Argument Schemes Powered by Cancellation*. Dans Kraus, Sarit (rédacteur) : *Proceedings of the Twenty-Eighth International Joint Conference on Artificial Intelligence, IJCAI 2019, Macao, China, August 10-16, 2019*, pages 1537–1543. ijcai.org, 2019.

[3] Bouyssou, Denis, Thierry Marchant, Marc Pirlot, Alexis Tsoukiàs et Philippe Vincke: *Evaluation and decision models with multiple criteria : Stepping stones for the analyst*. International Series in Operations Research and Management Science, Volume 86. Boston, 1st édition, 2006, ISBN 0-387-31098-3.

[4] Figueira, José Rui, Salvatore Greco et Roman Słowiński: *Building a set of additive value functions representing a reference preorder and intensities of preference : GRIP method*. European Journal of Operational Research, 195(2) :460–486, 2009. Number : 2 Publisher : Elsevier.

[5] Giarlotta, Alfio et Salvatore Greco: *Necessary and possible preference structures*. Journal of Mathematical Economics, 49(2) :163–172, 2013. Number : 2.

[6] Golden, Boris et Patrice Perny: *Infinite order Lorenz dominance for fair multiagent optimization*. Dans *Proceedings of the 9th International Conference on Autonomous Agents and Multiagent Systems : volume 1*, pages 383–390, 2010.

[7] Grabisch, Michel et Christophe Labreuche: *A decade of application of the Choquet and Sugeno integrals in multi-criteria decision aid*. Annals of Operations Research, 175(1) :247–286, 2010. Number : 1 Publisher : Springer.

[8] Greco, Salvatore, Vincent Mousseau et Roman Slowinski: *Ordinal regression revisited : Multiple criteria ranking using a set of additive value functions*. Eur. J. Oper. Res., 191(2) :416–436, 2008.

[9] Hammond, John S., Ralph L. Keeney et Howard Raiffa: *The Even- Swap Method for Multiple Objective Decisions*. Dans Haimes, Yacov Y. et Ralph E. Steuer (rédacteurs) : *Research and Practice in Multiple Criteria Decision Making*, pages 1–14, Berlin, Heidelberg, 2000. Springer Berlin Heidelberg, ISBN 978-3-642-57311-8.

[10] Jacquet-Lagreze, Eric et Jean Siskos: *Assessing a set of additive utility functions for multicriteria decision-making, the UTA method*. European journal of operational research, 10(2) :151–164, 1982. Number : 2 Publisher : Elsevier.

[11] Keeney, Ralph et Howard Raiffa: *Decisions with multiple consequences : preferences and value tradeoffs*, 1976.

[12] Krantz, D.H., R.D. Luce, P. Suppes et A. Tversky: *Foundations of measurement*, tome 1 : Additive and polynomial representations. Academic Press, New York, 1971.

[13] Marshall, A.W., I. Olkin et B.C. Arnold: *Inequalities : Theory of Majorization and Its Applications*. Springer Series in Statistics. Springer New York, 2010.

[14] Mitchell, Tom M: *Generalization as search*. Artificial intelligence, 18(2) :203–226, 1982. Number : 2 Publisher : Elsevier.

[15] Moyes, Patrick *et al.*: *Gini or Lorenz : Does it make a difference for inequality measurement ?* Dans *21st Annual European Economic Association Congress-24-28 août*, 2006.

[16] Roy, B.: *The outranking approach and the foundations of ELECTRE methods*. Theory and Decision, 31 :49–73, 1991.

[17] Shorrocks, Anthony F.: *Ranking Income Distributions*. Economica, 50(197) :3–17, 1983.

[18] Yager, Ronald R.: *On Ordered Weighted Averaging Aggregation Operators in Multicriteria Decisionmaking*. Dans Dubois, Didier, Henri Prade et Ronald R. Yager (rédacteurs) : *Readings in Fuzzy Sets for Intelligent Systems*, pages 80–87. Morgan Kaufmann, 1993.