A regret-based preference elicitation approach for sorting with multicriteria reference profiles

Nawal Benabbou, Patrice Perny, Paolo Viappiani

A regret-based preference elicitation approach for sorting with multicriteria reference profiles

Nawal Benabbou and **Patrice Perny** and **Paolo Viappiani** [1]

Abstract. In this paper we present an incremental elicitation method to determine the importance of the coalitions of criteria in a multicriteria sorting method. The method is designed to assign alternatives to predefined categories by comparing their performance vector to reference profiles. These comparisons lead to binary preference indices that are aggregated to determine the membership of the alternatives to predefined categories. We present an active learning process to determine the weighting coefficients modeling the importance of criteria in the aggregation process. Learning examples are generated one by one and presented to the Decision Maker to efficiently reduce the uncertainty attached to criteria weights. The process is stopped when all alternatives can be assigned to a category with the desired guarantee. We present the formal elicitation method as well as numerical tests showing its practical efficiency.

Keywords: Multicriteria sorting, capacity, Choquet integral, fuzzy preference relations, ordered categories.

1 Introduction

The evaluation of alternatives is a critical task in all decision support methods. When the alternatives must be evaluated with respect to multiple criteria, several aggregation procedures have been proposed to assess the overall value of the alternatives, either to rank them by decreasing order of preference, or to assign them to predefined ordered categories on the basis of their intrinsic value. In this paper, we focus on the latter objective and we consider multicriteria evaluation methods assessing the intrinsic value of alternatives by comparing their performances to predefined reference levels.

In this family of methods, we can distinguish two approaches. The '*aggregate then compare*' approach consists in a two stage process that first aggregates the performances of every alternative to produce an overall rating (e.g. using a weighted sum) and then compares the resulting rating to reference levels to assign the alternative in a predefined category (e.g., good, medium, bad). For example, alternatives are labelled as 'good' when they rate beyond 10, as 'bad' when they rate below 5, and as 'medium' otherwise. The main drawback of this first approach is that it is often difficult to derive a significant rating from criterion values expressed on different scales.

The second approach overcomes this problem by swapping the aggregation and comparison steps; for this reason, it is named the '*compare then aggregate*' approach. This approach requires an ordered sequence of reference profiles to be defined, representing different levels of requirements in the space of criteria; these profiles are totally ordered using Pareto dominance. A category is then implicitly defined by all the alternatives that beat a given reference profile but no other reference profile ranked higher in the dominance order. Thus, reference profiles act as upper and lower bounds of categories. This way of assigning alternatives to ordered categories while using multicriteria evaluations has been initally introduced by Roy in the Electre TRI method [11, 15, 16, 4]. Then multiple variants of this method have been proposed, based on a similar scheme [12, 14, 21, 8] and known as '*multicriteria sorting methods*'. They have been widely used in various applications (see the above references) and also investigated axiomatically [2, 3].

The '*compare then aggregate*' approach requires a preference relation to be defined in the space of criteria in order to compare performance vectors of alternatives to reference profiles. This preference relation is usually defined by aggregating the preference relations derived from each criterion considered separately. The aim of this paper is to learn from examples the proper aggregation method to be used to produce the aggregated preference relation. The most common aggregation method used for aggregating preference relations derived from criteria is a weighted majority rule. In this case, the weights of criteria must be learned from examples. This problem has been studied in [12, 13, 10, 17].

More recently, an extension of this approach has been proposed in [18] based on a generalization of weighted majority using a set function (namely a *capacity*) weighting any subset of criteria. This generalization enhances the descriptive power of the weighted majority model by allowing positive or negatives synergies among criteria due to the use of a non-necessarily additive definition of criterion weights. In this paper, we generalize this approach in a number of directions. First, we consider a generalized aggregation method to assign alternatives to categories. This method, introduced in [14], uses fuzzy preference relations to compare alternatives to references profiles and defines a degree of membership of any alternative to any categories (this method will be formally specified in Section 2). Secondly, the aggregation of fuzzy preference relations is performed using a possibly non linear weighted aggregator, for example the Choquet integral that combines fuzzy preferences with a non-additive measure of the importance of criteria. The learning of classifiers based on a Choquet integral has been investigated in [9]. Our approach is different because we use here the Choquet integral to aggregate fuzzy preferences rather than criterion values. Thirdly, instead of proposing a passive learning approach aiming to assess weighting parameters from a given database of examples, we propose here an active learning of the *capacity* that progressively asks examples of assignment to the Decision Maker (DM), selected one by one to efficiently reduce the set of possible capacities until a complete assignment of the alternatives can be defined with confidence.

The paper is organized as follows: in Section 2, we recall the

[1] Sorbonne Universités, UPMC Univ Paris 06 and CNRS, LIP6, UMR 7606, Paris, France, email: name.surname@lip6.fr

main features of the assignment method introduced in [14]. Then, the incremental approach proposed to elicit the aggregation function is presented in Section 3. Finally, numerical tests showing the efficiency of the approach are presented in Section 4.

2 Sorting using reference profiles

Let K_ℓ, $\ell \in \{1,\dots,q\}$, be a set of q ordered categories, K_1 being the best category and K_q being the worst one. Let X be the set of alternatives that must be assigned into one of these categories. Let us assume that every alternative is represented by a performance vector $x = (x_1,\dots,x_n) \in \mathbb{R}^n$ where x_i denotes the performance of x with respect to criterion i, $i \in N = \{1,\dots,n\}$; we assume here that criteria are to be maximized. Let us denote $\{r^0, r^1, \dots, r^q\}$ the set of reference profiles used to defined the bounds of the categories. Each element $r^\ell \in \mathbb{R}^n$ defines the lower boundary of category K_ℓ, $\ell \in \{1,\dots,q\}$, while $r^0 \in \mathbb{R}^n$ represents a top performance profile (not feasible) chosen to bound above all possible criterion values. These reference profiles are defined in such a way that $r_i^\ell > r_i^{\ell+1}$ for all $i \in N$ and $\ell \in \{0,\dots,q-1\}$.

These reference profiles being defined, the principle of a preference-based assignment method is to assign x to category K_ℓ when x is preferred to r^ℓ and x is not preferred to $r^{\ell-1}$. In the method introduced in [14], this principle is implemented by comparing x to all profiles r^ℓ using a preference index defined as follows:

$$P(x, r^\ell, \theta) = f_\theta(P_1(x, r^\ell), \dots, P_n(x, r^\ell)) \tag{1}$$

where f_θ is an aggregation function (compatible with Pareto dominance) parameterized by θ and $P_i(x, r^\ell)$ is a monocriterion preference index defined by:

$$P_i(x, r^\ell) = \begin{cases} 1 & \text{if} \quad x_i - r_i^\ell > \gamma_i^+ \\ \frac{x_i - r_i^\ell - \gamma_i^-}{\gamma_i^+ - \gamma_i^-} & \text{if} \quad \gamma_i^- < x_i - r_i^\ell \le \gamma_i^+ \\ 0 & \text{if} \quad x_i - r_i^\ell \le \gamma_i^- \end{cases}$$

The variation of index $P_i(x, r^\ell)$ as the difference $x_i - r_i^\ell$ increases is represented here below:

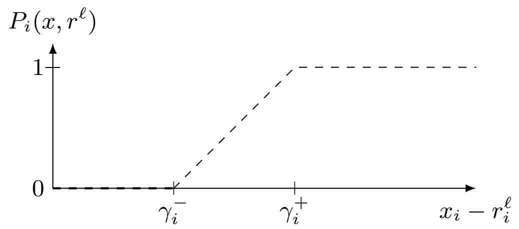

Figure 1. Preference index $P_i(x, r^\ell)$ as a function of $x_i - r_i^\ell$

This index represents the credibility of the statement "x is better than r^ℓ w.r.t criterion i". The index is maximal (i.e. equals 1) when the score difference $x_i - r_i^\ell$ exceeds the preference threshold γ_i^+. It is minimal when the difference $x_i - r_i^\ell$ falls below the indifference threshold γ_i^-. These thresholds are part of the definition of the criterion scale and are defined by the DM in such a way that $\gamma_i^+ > \gamma_i^-$. The preference threshold γ_i^+ defines the minimal difference compatible with a strict preference, whereas the indifference threshold γ_i^- represents the maximal score difference compatible with an indifference (absence of preference). Between these two thresholds, there is an area where we may hesitate between strict preference and indifference; in this area, the preference index grows linearly with $x_i - r_i^\ell$ (see Figure 1).

A standard choice for f_θ is a compromise operator, i.e. that verifies $\min_{i\in N}\{x_i\} \le f_\theta(x) \le \max_{i\in N}\{x_i\}$ for all $x = (x_1,\dots,x_n) \in \mathbb{R}^n$. When using such an operator, the overall preference index $P(x, r^\ell, \theta)$ defined by Equation (1) necessary belongs to $[0,1]$. Value 1 is achieved when the criteria are unanimously in favor of strict preference, whereas value 0 is achieved when criteria are unanimously against preference. Between these two extreme cases, the overall index measures the strength of arguments supporting the statement "x is better than r^ℓ". A common choice for f_θ is the weighted sum $f_\theta(p_1,\dots,p_n) = \sum_{i=1}^n \theta_i p_i$ where θ_i is the weight attached to criterion i. A more interesting choice is to define f_θ as a Choquet integral [5, 6] which provides a more general and more flexible aggregator. In this case, θ is the set function defining the weight of any coalition of criteria. We will come back to this case in Section 3.2.

After the aggregation step, the index measuring on the $[0,1]$ scale the membership of any alternative x to any category K_ℓ, $\ell \in \{1,\dots,q\}$, is defined by:

$$m_\ell(x,\theta) = \min\left\{P(x,r^\ell,\theta),\ 1 - P(x,r^{\ell-1},\theta)\right\} \tag{2}$$

and finally, alternative x is assigned to category K_{ℓ^*} where ℓ^* is the smallest index such that:

$$m_{\ell^*}(x,\theta) = \max_{k\in\{1,\dots,q\}} m_k(x,\theta)$$

This procedure is the general preference-based filtering method introduced in [14]. Equation (2) ensures that, for any fixed x and θ, $m_\ell(x,\theta)$ (seen as a function of ℓ) is unimodal, with a maximum at least equal to 0.5 (for more details see [14]).

This method is typical of the '*compare then aggregate*' approach. We first compare any alternative x to all reference profiles by considering each criterion separately, which leads to indices $P_i(x, r^\ell), i \in N$. These indices are then aggregated using function f_θ to define the membership of x to any category K_ℓ. The advantage of this approach, compared to the more classical '*aggregate then compare*' approach lies in the use of reference vectors r^ℓ instead of scalar thresholds on aggregated values. This provides a finer control in the definition of category boundaries in the multiobjective space. The decision of assigning an alternative is based on the scoring vector and not on an aggregated value: two alternatives having the same "average" value but different profiles may enter into different categories, as shown in the following example:

Example 1. *Let us consider two alternatives x, y and a sequence of 3 reference profiles r^0, r^1, r^2 with the following grades on three criteria:*

	1	2	3
x	6	15	15
y	30	3	3
r^0	50	50	50
r^1	12	12	12
r^2	0	0	0

The three reference profiles allow the definition of two categories K_1 and K_2. We use here the same valuation scale with a preference threshold $\gamma_i^+ = 1$ and indifference threshold $\gamma_i^- = 0$, for all $i \in N = \{1,2,3\}$. Let us assume that f_θ is the weighted sum with weights $\theta = (1/3, 1/3, 1/3)$, i.e. $f_\theta(z_1,z_2,z_3) = (z_1+z_2+z_3)/3$ for all $z \in \mathbb{R}^n$. We obtain $P(x,r^0,\theta) = P(y,r^0,\theta) = 0$ and $P(x,r^2,\theta) =$

$P(y, r^2, \theta) = 1$. *The only difference between x and y are due to profil r^1. We indeed have: $P(x, r^1, \theta) = 2/3$ and $P(y, r^1, \theta) = 1/3$. Hence, using Equation (2), we get the following membership values:*

	m_1	m_2
x	$2/3$	$1/3$
y	$1/3$	$2/3$

As a consequence, x is here assigned to category K_1 whereas y is assigned to category K_2.

In this example, x obtains a better position than y (since K_1 is better than K_2). Note that alternatives x and y would be undiscernible with the '*aggregate then compare*' approach since they have the same average: $f_\theta(x) = f_\theta(y)$. Moreover, if the grade of y increases from 30 to 50 on criterion 1, it can easily be checked that y remains in K_2. There is no improvement despite the fact that the average score of y defined by f_θ increases. This is due to the fact that here, there is no point in improving its performance on criterion 1 since it already exceeds the value of reference profile r^1 while the weaknesses of y on criteria 2 and 3 remain. This example illustrates the non-compensatory nature of this sorting procedure, where difference of grades does not play any role beyond a given threshold; this is a clear difference with procedures based on the direct aggregation of criterion values.

3 An incremental approach for sorting alternatives using reference profiles

The procedure introduced in the previous section involves at some step an aggregation operation f_θ in which parameter θ controls the importance of criteria and coalitions of criteria in the overall assessment of alternatives. Our aim is to propose, in the framework of the approach presented above, an incremental procedure to assess parameter θ. We assume that this parameter is initially unknown and we want to use an active learning process to progressively reduce the uncertainty attached to θ. Examples will be selected one by one and presented to the DM that will be asked to classify them; these new classifications will induce constraints restricting the space of admissible θ, and the process will be repeated until being able to classify all the alternatives in X.

Our procedure relies on the notion of minimax regret, allowing to make robust decisions in face of uncertainty, and to ask informative queries to further reduce the uncertainty on the space of admissible θ. This can be seen as an adaptation to sorting problems of incremental elicitation mechanisms designed for choice problems (e.g., [20, 1]).

Whenever θ is precisely known, we wish to assign x to the category K_ℓ such that $m_\ell(x, \theta) \geq m_k(x, \theta)$ for all $k \in \{1, \ldots, q\}$. Therefore, it is natural to define the *loss* or *regret* associated to assigning x to K_ℓ rather than assigning it to K_k as:

$$\mathrm{R}(x, K_\ell, K_k, \theta) = m_k(x, \theta) - m_\ell(x, \theta).$$

When we only know that θ belong to an uncertainty set Θ, we may be interested in computing the following regrets:

Definition 1. *For any alternative $x \in X$, the pairwise max regret (PMR) of assigning x to the category K_ℓ instead of assigning it to the category K_k is defined by:*

$$\begin{aligned}\mathrm{PMR}(x, K_\ell, K_k, \Theta) &= \max_{\theta \in \Theta} \mathrm{R}(x, K_\ell, K_k, \theta) \\ &= \max_{\theta \in \Theta} \Big\{ m_k(x, \theta) - m_\ell(x, \theta) \Big\}.\end{aligned}$$

$\mathrm{PMR}(x, K_\ell, K_k, \Theta)$ is the maximum feasible gap between the membership indices of alternative x with respect to categories K_k and K_ℓ. It represents the worst-case loss that we may incur by assigning x to category K_ℓ instead of category K_k when parameter θ belongs to Θ.

Definition 2. *The max regret (MR) of assigning $x \in X$ to category K_ℓ is defined by:*

$$\mathrm{MR}(x, K_\ell, \Theta) = \max_{k \in \{1,\ldots,q\}} \mathrm{PMR}(x, K_\ell, K_k, \Theta)$$

$\mathrm{MR}(x, K_\ell, \Theta)$ is the worst-case loss that we may incur by assigning alternative x to category K_ℓ instead of any other categories. We now define the notion of *minimax regret* and the *regret-optimal category* associated to alternative $x \in X$:

Definition 3. *The minimax regret (mMR) of alternative $x \in X$ is:*

$$\mathrm{mMR}(x, \Theta) = \min_{\ell \in \{1,\ldots,q\}} \mathrm{MR}(x, K_\ell, \Theta)$$

Considering the uncertainty set Θ, a cautious decision rule would consist in assigning alternative x to the category minimizing $\mathrm{MR}(x, K_\ell, \Theta)$, named the regret-optimal category of x here below. In order to assess the maximal error on the complete assignment when using this rule, we introduce now the notion of *maximum minimax regret*:

Definition 4. *The maximum minimax regret (MmMR) is:*

$$\mathrm{MmMR}(X, \Theta) = \max_{x \in X} \mathrm{mMR}(x, \Theta)$$

In sorting problems, MmMR plays the role of measuring the current decision quality; in particular, if $\mathrm{MmMR} = 0$, then we know that the complete assignment is valid. However, it might be the case that the aggregate value MmMR is too large according to the DM. In that case, it is natural to consider incremental elicitation procedures to make this value decrease until a given tolerance threshold $\delta \geq 0$. This is actually possible due to the fact that the inequality $\mathrm{MmMR}(X, \Theta') \leq \mathrm{MmMR}(X, \Theta)$ is true for all $\Theta' \subseteq \Theta$, meaning that the value MmMR cannot increase when including new preference information obtained from the DM; actually, in the subsection devoted to numerical tests, we will see that, in practice, it strictly decreases when queries are chosen in a reasoned way.

As in a choice problems, comparison queries between alternatives might be asked; however, in our context, it is less straightforward to identify a pair of alternatives forming an informative preference query (i.e. which is likely to induce a regret reduction). Instead, we can ask the DM to choose, for a given alternative, the category that fits best (either among a pair of categories, among a given subset of categories, or among all possible categories). For this type of queries, we will denote $K_\ell \succsim_x K_k$ the preference for category K_ℓ over category K_k concerning the assignement of alternative $x \in X$. For choosing which query to ask next, we propose to focus on an alternative x that is associated with the largest value mMR in the current regret-optimal assignment; by asking a query involving such an alternative, we are indeed likely to reduce the value MmMR since $\mathrm{MmMR}(X, \Theta) = \mathrm{mMR}(x, \Theta)$ holds. There are at least the two following possibilities:

- we can ask the DM to assign x to the most relevant category or
- we can ask the DM to compare the regret-optimal category K_{ℓ^*} of alternative x with its *regret-maximizing adversarial category* (the one that maximizes $\mathrm{PMR}(x, K_{\ell^*}, K_k, \Theta_{\mathcal{P}})$) and state which one is the most relevant among the two.

The latter approach has the advantage of requiring less effort from the DM while focusing on the pair of categories inducing the current value MmMR.

3.1 Determination of the optimal assignment using mixed integer linear programming

Let $\mathcal{P}$ be the set gathering all preference statements of type $K_\ell \succsim_x K_k$ collected so far, and let $\Theta_\mathcal{P}$ be the set containing all parameters θ consistent with information $\mathcal{P}$, i.e. such that $m_\ell(x,\theta) \geq m_k(x,\theta)$ for all examples $K_\ell \succsim_x K_k \in \mathcal{P}$. We assume here that operator f_θ is linear in θ (e.g., a weighted sum or a Choquet integral). In order to determine the regret-optimal assignment, we need a method that efficiently computes $\mathrm{PMR}(x, K_\ell, K_k, \Theta_\mathcal{P})$ for any $x \in X$ and any $\ell, k \in \{1, \ldots, q\}$. This computation is challenging because it requires a maximization of the difference of two minima. Nevertheless, we will show that this optimization problem can be decomposed in two mixed integer linear programs. In order to do that, we need first to show that the answers to preference queries induce linear constraints over the set of parameters $\Theta_\mathcal{P}$.

Proposition 1. *Let $x \in X$ be any alternative and K_ℓ, K_k two categories such that $\ell \neq k$. If $\ell > k$, then $\Theta_{\mathcal{P}\cup K_\ell \succsim_x K_k}$ is:*

$$\Big\{\theta \in \Theta_\mathcal{P}, 1-P(x,r^{\ell-1},\theta) \geq \min\{P(x,r^k,\theta), 1-P(x,r^{k-1},\theta)\}\Big\}$$

otherwise $\Theta_{\mathcal{P}\cup K_\ell \succsim_x K_k}$ is:

$$\Big\{\theta \in \Theta_\mathcal{P}, P(x,r^\ell,\theta) \geq \min\{P(x,r^k,\theta), 1-P(x,r^{k-1},\theta)\}\Big\}$$

Proof : If we observe that category K_ℓ is preferred to category K_k for the assignment of alternative $x \in X$, then we want to restrict $\Theta_\mathcal{P}$ to all parameters θ such that $m_\ell(x,\theta) \geq m_k(x,\theta)$, which can be rewritten $\min\{P(x,r^\ell,\theta), 1-P(x,r^{\ell-1},\theta)\} \geq \min\{P(x,r^k,\theta), 1-P(x,r^{k-1},\theta)\}$. This is actually equivalent to imposing the two following constraints:

$$P(x,r^\ell,\theta) \geq \min\{P(x,r^k,\theta),\ 1-P(x,r^{k-1},\theta)\} \quad (3)$$
$$1-P(x,r^{\ell-1},\theta) \geq \min\{P(x,r^k,\theta),\ 1-P(x,r^{k-1},\theta)\} \quad (4)$$

First, assume that K_ℓ, K_k are such that $\ell > k$. In that case, we know that we have $P_i(x,r^\ell) \geq P_i(x,r^k)$ for all $i \in N$. Therefore, since f_θ is compatible with Pareto dominance, then we have $P(x,r^\ell,\theta) \geq P(x,r^k,\theta)$ for all $\theta \in \Theta_\mathcal{P}$. As a consequence, Equation (3) holds for all $\theta \in \Theta_\mathcal{P}$ and so the associated constraint is not needed for updating the set of feasible parameters $\Theta_\mathcal{P}$ according to the observed preference $(x, K_\ell \precsim K_k)$. Hence, only Equation (4) remains in that case. Now, assuming that $\ell < k$, we have $P_i(x,r^{\ell-1}) \leq P_i(x,r^{k-1})$ for all $i \in N$. Therefore, since f_θ is compatible with Pareto dominance, then $P(x,r^{\ell-1},\theta) \leq P(x,r^{k-1},\theta)$ for all $\theta \in \Theta_\mathcal{P}$, i.e. $1-P(x,r^{\ell-1},\theta) \geq 1-P(x,r^{k-1},\theta)$. Therefore, Equation (4) holds for all $\theta \in \Theta_\mathcal{P}$ and so the associated constraint is not needed for updating the set of feasible parameters $\Theta_\mathcal{P}$ according to $K_\ell \succsim_x K_k$. Hence, only Equation (3) remains. ■

Thus, if the DM states that, for a given alternative $x \in X$, category K_ℓ is better suited than category K_k, then it is sufficient to impose the following constraint over the set of feasible parameters:

- $1-P(x,r^{\ell-1},\theta) \geq \min\{P(x,r^k,\theta), 1-P(x,r^{k-1},\theta)\}$ if $\ell > k$,
- $P(x,r^\ell,\theta) \geq \min\{P(x,r^k,\theta), 1-P(x,r^{k-1},\theta)\}$ otherwise.

These constraints can be linearized using standard linearization of the min aggregator. In particular, this is done in the following way:

If $\ell > k$, we impose:

$$\begin{cases} Mb + 1 - P(x,r^{\ell-1},\theta) \geq P(x,r^k,\theta) \\ M(1-b) + 1 - P(x,r^{\ell-1},\theta) \geq 1 - P(x,r^{k-1},\theta) \end{cases}$$

If $\ell < k$, we impose:

$$\begin{cases} Mb + P(x,r^\ell,\theta) \geq P(x,r^k,\theta) \\ M(1-b) + P(x,r^\ell,\theta) \geq 1 - P(x,r^{k-1},\theta) \end{cases}$$

where b is a boolean variable and M is a numerical scalar value greater than one.

Hence, we proved that $\Theta_\mathcal{P}$ can be described with linear constraints when $\mathcal{P}$ is composed of preferences of type (x, K_ℓ, K_k). Now, the following proposition proves that PMR-optimizations can be performed using mixed integer linear programming.

Proposition 2. *For any $x \in X$ and any $\ell, k \in \{1, \ldots, q\}$, we have:*

$$\mathrm{PMR}(x, K_\ell, K_k, \Theta_\mathcal{P}) = \max\{\beta_1, \beta_2\}$$

where β_1 and β_2 are respectively the optimum values of the following mixed integer linear programs:

$$\max_{\substack{\theta\in\Theta_\mathcal{P}\\ t\in\mathbb{R}}} \{t - P(x,r^\ell,\theta)\} \quad \text{s.t. } t \leq P(x,r^k,\theta),\ t \leq 1-P(x,r^{k-1},\theta)$$

$$\max_{\substack{\theta\in\Theta_\mathcal{P}\\ t\in\mathbb{R}}} \{t + P(x,r^{\ell-1},\theta) - 1\} \quad \text{s.t. } t \leq P(x,r^k,\theta),\ t \leq 1-P(x,r^{k-1},\theta)$$

Proof : For any alternative $x \in X$ and any two categories K_ℓ, K_k:

$$\begin{aligned}\mathrm{PMR}(x, K_\ell, K_k, \Theta_\mathcal{P}) &= \max_{\theta\in\Theta_\mathcal{P}} \mathrm{R}(x, K_\ell, K_k, \theta) \\ &= \max_{\theta\in\Theta_\mathcal{P}} \Big\{m_k(x,\theta) - m_\ell(x,\theta)\Big\}\end{aligned}$$

Since $m_k(x,\theta) = \min\{P(x,r^k,\theta), 1-P(x,r^{k-1},\theta)\}$, we can compute $\mathrm{PMR}(x, K_\ell, K_k, \Theta_\mathcal{P})$ by solving the following program:

$$\max_{\substack{\theta\in\Theta_\mathcal{P}\\ t\in\mathbb{R}}} \Big\{t - m_\ell(x,\theta)\Big\} \quad \text{s.t. } t \leq P(x,r^k,\theta),\ t \leq 1-P(x,r^{k-1},\theta)$$

This program is obtained by using standard linearization of the min aggregator. Then, we have:

$$\begin{aligned} t - m_\ell(x,\theta) &= t - \min\{P(x,r^\ell,\theta),\ 1-P(x,r^{\ell-1},\theta)\} \\ &= t + \max\{-P(x,r^\ell,\theta),\ P(x,r^{\ell-1},\theta) - 1\} \\ &= \max\{t - P(x,r^\ell,\theta),\ t + P(x,r^{\ell-1},\theta) - 1\}\end{aligned}$$

Therefore, $\mathrm{PMR}(x, K_\ell, K_k, \Theta_\mathcal{P})$ can be computed by solving the following optimization problem:

$$\max_{\substack{\theta\in\Theta_\mathcal{P}\\ t\in\mathbb{R}}} \Big\{\max\{t - P(x,r^\ell,\theta),\ t + P(x,r^{\ell-1},\theta) - 1\}\Big\} \quad \text{s.t. } t \leq P(x,r^k,\theta),\ t \leq 1-P(x,r^{k-1},\theta)$$

The result is finally obtained by interchanging the max operators. ■

Therefore, computing $\mathrm{PMR}(x, K_\ell, K_k, \Theta_\mathcal{P})$ can be easily performed by solving two mixed integer linear programs and then selecting the greatest optima.

3.2 Application to Choquet integrals

In this subsection, we will focus on a particular instance obtained by interpreting f_θ as a Choquet integral in Equation (1). This allows positive or negative synergies among arguments when aggregating preference indices $P_i(x, r^\ell)$ into an overall index $P(x, r^\ell, \theta)$. We now recall the definition of Choquet capacities and (discrete) Choquet integrals[2]. A normalized Choquet capacity v is a real-valued set-function defined on 2^N such that $v(\emptyset) = 0$, $v(N) = 1$ and $v(A) \leq v(B)$ for all $A \subseteq B \subseteq N$; value $v(A)$ is the weight attached to coalition A, for any $A \subseteq N$. The Choquet integral is then defined by:

$$C_v(x) = \sum_{i=1}^{n} \left[x_{(i)} - x_{(i-1)}\right] v(X_{(i)}) \text{ with } x_{(0)} = 0$$

where $(.)$ is a permutation of $\{1, \ldots, n\}$ which sorts the components of x by increasing order (i.e. $x_{(i)} \leq x_{(i+1)}$ for $i \in \{1, \ldots, n-1\}$) and $X_{(i)} = \{(i), \ldots, (n)\}$. In the following, the uncertain capacity function v takes the role of θ and the set of feasible parameters $\Theta_\mathcal{P}$ is the set of all normalized capacities compatible with $\mathcal{P}$. Alternative $x \in X$ is now compared to any profile r^ℓ using the *preference index*:

$$P(x, r^\ell, v) = C_v(P_1(x, r^\ell), \ldots, P_n(x, r^\ell))$$

The membership of solution x to category K_ℓ is now defined by:

$$m_\ell(x, v) = \min\{P(x, r^\ell, v),\ 1 - P(x, r^{\ell-1}, v)\}$$

In the particular case where $\Theta_\mathcal{P}$ is a set of strictly monotonic capacities (i.e. $v(A) < v(B)$ for all $A \subset B \subseteq N$), we can linearize the constraints induced by $\mathcal{P}$ in a simpler way, so that we can avoid the use of boolean variables.

Proposition 3. *For any $x \in X$ and any $\ell, k \in \{1, \ldots, q\}$:*
If $\ell > k$ and $P_i(x, r^{\ell-1}) \neq P_i(x, r^{k-1})$ for some $i \in N$, then

$$\Theta_{\mathcal{P} \cup K_\ell \succsim_x K_k} = \{v \in \Theta_\mathcal{P},\ 1 - P(x, r^{\ell-1}, v) \geq P(x, r^k, v)\}$$

If $\ell < k$ and $P_i(x, r^\ell) \neq P_i(x, r^k)$ for some $i \in N$, then

$$\Theta_{\mathcal{P} \cup K_\ell \succsim_x K_k} = \{v \in \Theta_\mathcal{P},\ P(x, r^\ell, v) \geq 1 - P(x, r^{k-1}, v)\}$$

Otherwise, $\Theta_{\mathcal{P} \cup K_\ell \succsim_x K_k} = \Theta_\mathcal{P}$.

Proof : Assume that $\ell > k$. According to Proposition 1, capacities $v \in \Theta_\mathcal{P}$ that are compatible with $K_\ell \succsim_x K_k$ are those verifying:

$$1 - P(x, r^{\ell-1}, v) \geq \min\{P(x, r^k, v), 1 - P(x, r^{k-1}, v)\} \quad (5)$$

Since $\ell > k$, we know that $P_i(x, r^{\ell-1}) \geq P_i(x, r^{k-1})$ for all $i \in N$. In the case where $P_i(x, r^{\ell-1}) = P_i(x, r^{k-1})$ for all $i \in N$, we have $P(x, r^{\ell-1}, v) = P(x, r^{k-1}, v)$, i.e. $1 - P(x, r^{\ell-1}, v) = 1 - P(x, r^{k-1}, v)$; therefore, for all $v \in \Theta_\mathcal{P}$, Equation (5) is satisfied and so $\Theta_{\mathcal{P} \cup K_\ell \succsim_x K_k} = \Theta_\mathcal{P}$. Now, consider the case where $P_i(x, r^{\ell-1}) \neq P_i(x, r^{k-1})$ for some $i \in N$. Since $v \in \Theta_\mathcal{P}$ is strictly monotonic, then C_v is strictly increasing with Pareto dominance, and so we have $P(x, r^{\ell-1}, v) > P(x, r^{k-1}, v)$, i.e. $1 - P(x, r^{\ell-1}, v) < 1 - P(x, r^{k-1}, v)$. As a consequence, Equation (5) is satisfied if and only if we have $1 - P(x, r^{\ell-1}, v) \geq P(x, r^k, v)$. Hence, $\Theta_{\mathcal{P} \cup K_\ell \succsim_x K_k} = \{v \in \Theta_\mathcal{P},\ 1 - P(x, r^{\ell-1}, v) \geq P(x, r^k, v)\}$.

Now, assume that $\ell < k$. In that case, according to Proposition 1, capacities $v \in \Theta_\mathcal{P}$ that satisfy $K_\ell \succsim_x K_k$ are those verifying:

$$P(x, r^\ell, v) \geq \min\{P(x, r^k, \theta),\ 1 - P(x, r^{k-1}, v)\} \quad (6)$$

[2] Refer, for instance, to [5, 6] for a much more detailed description.

Since $\ell < k$, we have $P_i(x, r^\ell) \leq P_i(x, r^k)$ for all $i \in N$. First, assume that $P_i(x, r^\ell) = P_i(x, r^k)$ for all $i \in N$. In that case, we have $P(x, r^\ell, v) = P(x, r^k, v)$ and so Equation (6) is verified by all capacities $v \in \Theta_\mathcal{P}$; hence $\Theta_{\mathcal{P} \cup K_\ell \succsim_x K_k} = \Theta_\mathcal{P}$. Now, assume that $P_i(x, r^\ell) \neq P_i(x, r^k)$ for some $i \in N$. Since $v \in \Theta_\mathcal{P}$ is strictly monotonic, then C_v is strictly increasing with Pareto dominance, and so we necessarily have $P(x, r^\ell, v) < P(x, r^k, v)$. Therefore, Equation (6) is satisfied if and only if $P(x, r^\ell, v) \geq 1 - P(x, r^{k-1}, v)$ and so $\Theta_{\mathcal{P} \cup K_\ell \succsim_x K_k} = \{v \in \Theta_\mathcal{P},\ P(x, r^\ell, v) \geq 1 - P(x, r^{k-1}, v)\}$.■

Thus, according to Proposition 3, observing that category K_ℓ is preferred to category K_k for alternative x amounts to imposing:

- the linear constraint $1 - P(x, r^{\ell-1}, v) \geq P(x, r^k, v)$ if $\ell > k$ and $P_i(x, r^{\ell-1}) \neq P_i(x, r^{k-1})$ for some $i \in N$,
- the linear constraint $P(x, r^\ell, v) \geq 1 - P(x, r^{k-1}, v)$ if $\ell < k$ and $P_i(x, r^\ell) \neq P_i(x, r^k)$ for some $i \in N$ and
- no additional constraints otherwise.

As a consequence, when assuming that the DM's preferences can be modeled by a Choquet integral with a strictly monotonic capacity, the PMR-optimization consists in solving two linear programs (instead of mixed integer linear programs in the general case) and then selecting the maximum between the two optima (see Proposition 2).

4 Numerical tests

In this section, we consider datasets of alternatives uniformly drawn within $[0, 1]^n$ and simulated DMs answer to preference queries according to a randomly generated Choquet integral (defined with a strictly monotonic capactity). First, we want to compare the following query selection strategies in terms of MmMR reduction:

- **S0**: this strategy asks the DM to state which among all the categories suits most a randomly chosen alternative in the dataset.
- **S1**: this strategy asks to compare the regret-optimal category of x with its regret-maximizing adversarial category, where x is the alternative associated with the largest mMR (see Section 3).
- **S2**: this strategy asks which among all the categories suits most the alternative associated with the largest mMR value.

Linear optimizations are performed by the Gurobi solver called from a program written in Java. The reference profiles under consideration are constant utility profiles dividing the utility scale $[0, 1]$ into intervals of same size on every criterion. In Figure 2, we report the MmMR at each iteration step of the incremental procedures; results are obtained by averaging over 100runs.

First, we can see that the MmMR value reduces significantly faster with S1 and S2 than with S0; for instance, after at most 20 queries on average, the MmMR value is around 40 percent of the maximum regret in the dataset with S1 and S2, while still remaining above 80 percent with S0. Moreover, as expected, S2 turns out to be more informative than S1 (since the MmMR reduces more quickly). However, we also note that S1 needs just approximately 3 additional queries on average to achieve the same level of regret as S2. This empirical evidence suggests to use S1 instead of S2 since the former has lower cognitive cost than the latter while still being very effective (asking the DM to state which among all the categories suits most a given alternative indeed requires much more cognitive effort from the DM than the comparison of two categories).

The next numerical tests aim to evaluate the impact of the following parameters on computation times of MmMR calculations: q the number of categories, n the number of criteria, p the number of observed

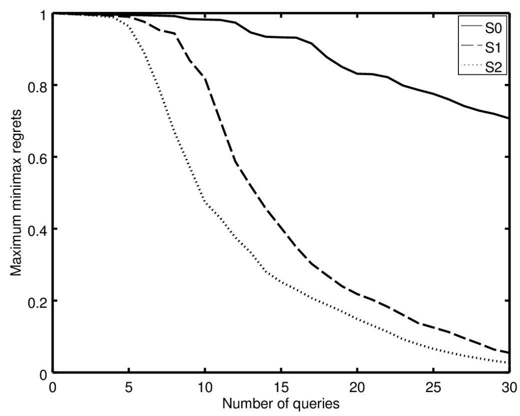

Figure 2. Maximum minimax regret reduction for strategies S0, S1 and S2 ($n = 5$, 150 alternatives, 5 categories).

preferences and the number of alternatives. In Table 1, computation times are obtained by averaging over 100 runs.

In particular, computation times drastically increase with q, the number of categories (due to the quadratic number of PMR-computations). Moreover, computation times are significantly impacted by the number of criteria. This is due to the fact that the number of variables and constraints of the linear programs grows exponentially with the number of criteria (in order to ensure monotonicity of the Choquet capacity); note however that computation times can be further reduced when considering some particular subclasses of capacities (2-additive capacities, belief functions).

Table 1. Computation times (in seconds) of MmMR calculations; results for instances with $|X| = 50, 100, 200$ are respectively given in lines 1,2,3.

$q=5$				$q=10$			
$n=5$		$n=7$		$n=5$		$n=7$	
$p=0$	$p=5$	$p=0$	$p=5$	$p=0$	$p=5$	$p=0$	$p=5$
0.8	1.4	1.8	2.3	3.5	5.6	6.7	8.6
1.1	1.8	2.1	3.3	4.6	6.8	9.1	10.8
2.2	3.5	6.9	7.1	7.6	10.7	18.2	21.1

5 Conclusion

The main advantage of our approach is that assignment queries are selected using the minimax regret criterion. The constraints derived from these examples indeed allow an efficient reduction of uncertainty where it is decisive, thus facilitating the assignment of remaining alternatives to a category. This applies to a wide family of weighted aggregation functions, including weighted sums, ordered weigthed averaging operators and Choquet integrals.

For Choquet integrals, the proposed approach is practically feasible provided that the number of criteria is not too large (about 10). For problems involving a larger number of criteria, the linear programs to be solved for computing regrets is computationally demanding, due to monotonicity constraints. In this case, a first solution consists in using capacity admitting compact representations, e.g. k-additive capacities [7] for a bounded k; in this particular case, the elicitation of the capacity remains tractable (see e.g. [19] for $k = 2$). Another interesting option would be to use fictitious alternatives with very simple profiles as learning example. This allows to drastically simplify the management of the monotonicity constraint, even with large numbers of criteria, as shown for choice problems in [1].

REFERENCES

[1] Nawal Benabbou, Patrice Perny, and Paolo Viappiani, 'Incremental Elicitation of Choquet Capacities for Multicriteria Decision Making', in *Proceedings of ECAI'14*, pp. 87–92, (2014).

[2] D. Bouyssou and Th. Marchant, 'An axiomatic approach to noncompensatory sorting methods in MCDM, I: The case of two categories', *European Journal of Operational Research*, **178**(1), 217 – 245, (2007).

[3] D. Bouyssou and Th. Marchant, 'An axiomatic approach to noncompensatory sorting methods in MCDM, II: More than two categories', *European Journal of Operational Research*, **178**(1), 246 – 276, (2007).

[4] J. Almeida Dias, J. Rui Figueira, and B. Roy, 'Electre tri-c : A multiple criteria sorting method based on characteristic reference actions', *European Journal of Operational Research*, **204**(3), (2010).

[5] M. Grabisch, 'The application of fuzzy integrals in multicriteria decision making', *European Journal of Operational Research*, **89**(3), 445–456, (1996).

[6] M. Grabisch, J.-L. Marichal, R. Mesiar, and E. Pap, *Aggregation Functions (Encyclopedia of Mathematics and Its Applications)*, Cambridge University Press, New York, NY, USA, 2009.

[7] Michel Grabisch, 'The application of fuzzy integrals in multicriteria decision making', *European journal of operational research*, **89**(3), 445–456, (1996).

[8] S. Greco, V. Mousseau, and R. Slowinski, 'Multiple criteria sorting with a set of additive value functions', *European Journal of Operational Research*, **207**(3), 1455–1470, (2010).

[9] E. Hüllermeier and A. Fallah Tehrani, 'Efficient learning of classifiers based on the 2-additive Choquet integral', in *Computational Intelligence in Intelligent Data Analysis Studies in Computational Intelligence Volume*, volume 445, pp. 17–29, (2013).

[10] A. Leroy, V. Mousseau, and M. Pirlot, 'Learning the paraeters of a multiple criteria sorting method based on a majority rule', in *proceedings of ADT'11*, volume 6992 of *LNCS*, pp. 219–233. Springer, (2011).

[11] J. Moscarola and B. Roy, 'Procdure automatique d'examen de dossiers fonde sur une segmentation trichotomique en prsence de critres multiples', *RAIRO*, **11**(2), 145–173, (1977).

[12] V. Mousseau and R. Slowinski, 'Inferring an ELECTRE TRI model from assignment examples', *Journal of Global Optimization*, **12**, 157–174, (1998).

[13] V. Mousseau, R. Slowinski, and P. Zielniewicz, 'A useroriented implementation of the ELECTRE-TRI method integrating preference elicitation support', *Computers and Operations Research*, **27**, 757–777, (2000).

[14] P. Perny, 'Multicriteria filtering methods based on concordance and nondiscordance principles', *Annals of operations Research*, **80**, 137–165, (1998).

[15] B. Roy, 'A multicriteria analysis for trichotomic segmentation problems', in *Multiple Criteria Analysis: Operational Methods*, eds., Peter Nijkamp and Jaap Spronk, pp. 245–257. Gower Press, (1981).

[16] B. Roy, 'The outranking approach and the foundations of electre methods', *Theory and Decision*, **31**, 49–73, (1991).

[17] O. Sobrie, V. Mousseau, and M. Pirlot, 'Learning a majority rule model from large sets of assignment examples', in *Proceedings of ADT'13*, pp. 336–350, (2013).

[18] O. Sobrie, V. Mousseau, and M. Pirlot, 'Learning the parameters of a non compensatory sorting model', in *Proceedings of ADT'15*, pp. 153–170, (2015).

[19] Ali Fallah Tehrani, Weiwei Cheng, Krzysztof Dembczynski, and Eyke Hüllermeier, 'Learning monotone nonlinear models using the Choquet integral', *Machine Learning*, **89**(1-2), 183–211, (2012).

[20] T. Wang and C. Boutilier, 'Incremental Utility Elicitation with the Minimax Regret Decision Criterion', in *Proceedings of IJCAI'03*, pp. 309–316, (2003).

[21] C. Zopounidis and M. Doumpos, 'Multicriteria classification and sorting methods: A literature review', *European Journal of Operational Research*, **138**(2), 229–246, (2002).