Multiple Criteria Sorting models and methods. Part II: Theoretical results and general issues

Khaled Belahcène*, Vincent Mousseau†, Wassila Ouerdane†, Marc Pirlot‡
Olivier Sobrie‡

December 2, 2022

Abstract

Multiple criteria sorting methods assign objects into ordered categories while objects are characterized by a vector of n attributes values. Categories are ordered, and the assignment of the object is monotonic w.r.t. to some underlying order on the attributes scales (criteria). We drew a landscape of these methods in Part I "Survey of the literature" (published in a previous issue of the present journal) and we aim to provide a theoretical view of the field in this second part. We describe a general framework for MCS models and position some existing models in the picture. Issues related to imperfect or insufficient information are then discussed. We also address questions that arise in the final phase of a decision aiding process as, e.g., explaining a recommendation or suggesting efficient ways of improving an object assignment.

Keywords: multiple criteria decision making, multiple criteria sorting, monotone classification, preference learning.

1 Introduction

Part I: "Survey of the literature" of this paper (previously published) was devoted to the presentation of a panorama of the literature related to Multiple Criteria Sorting (MCS) models and methods. The emphasis was put on two families of models, one based on additive value functions (AVF), running on from the pioneering UTADIS method [Jacquet-Lagrèze and Siskos, 1982], the other based on outranking relations, starting from Electre Tri [Wei, 1992, Roy and Bouyssou, 1993]. From the methodological point of view, we gave an account of direct and indirect methods for eliciting the model parameters. In the indirect elicitation approach, the model's parameters are induced from a set of assignment examples. Typically in MCS applications, the scarcity of available assignment examples leads to indeterminacy of the model's parameters. We analyzed different ways of dealing with parameters indeterminacy (selection of central models, robust and stochastic approaches). We also discussed the relationships between MCS methods and other approaches such as monotone classification, machine/preference learning, and rules induction techniques (DRSA).

*Heudiasyc, Université de Technologie de Compiègne, France
†MICS, CentraleSupélec, Université Paris-Saclay, Gif-Sur-Yvette, France
‡MATHRO, Faculté Polytechnique, Université de Mons, Belgium

The rest of the paper (Part II) is organized as follows.

Section 2 gives an overview of theoretical results characterizing several models. Such results allow us to understand the relationships (inclusion, equivalence) between some models and how general these models are.

Section 3 discusses issues raised by imperfect information: How to deal with sets of assignment examples that are not fully compatible with a model? How to deal with insufficient, imprecise or uncertain information ?

Section 4 addresses issues related to the final phase of a decision-aiding process, such as: How to explain the recommendations and how to improve an object's assignment?

Section 5 closes the survey with some conclusions and research perspectives.

A list of abbreviations is available in Appendix A.

2 Understanding models

Are there characteristics of a model that make it legitimate for use in a particular decision-aiding process? As far as the DM is concerned, any model or method that helps her to understand her problem is fine. The expert in MCDA methods who helps the DM has other requirements regarding models and methods. These should be logically correct and use data in a way compatible with their meaning (e.g., they should not make sums of ordinal data). The expert in MCDA methods is expected to have a deep understanding of the underlying models. In particular, she knows how to interpret the model's parameters to question the DM appropriately. The expert should also have a global picture of the models in mind and be aware of their interrelations and logical dependencies.

In order to understand models and dispose of a global picture, characterizing the models helps a lot. Basically, in the context of MCS, a characterization tells which ordered partitions can be represented in a given model by adequately tuning its parameters. In the sequel, we outline what is known regarding the characterization of sorting models. We also position some of the main models in the picture.

2.1 The Monotone Sorting Model

It is easy to characterize the set of all ordered partitions. They constitute the most general output of MCS methods, without any peculiarity except for monotonicity. Let us recall some notation. The result of applying an MCS method is an ordered partition $\mathcal{C} = (C^1, \dots, C^h, \dots, C^p)$ of the set $X = \Pi_{i=1}^n X_i$. The latter is the Cartesian product of the scales X_i of the criteria. An object is described by an n-dimensional vector $(x_1, \dots x_n)$ of evaluations $x_i \in X_i$, for $i = 1, \dots, n$. Therefore, one can consider that X represents the set of all objects. We assume that an MCS method is able to assign any object from X to one of the categories $C^h, h = 1, \dots, p$.

Note that, for the moment, we do not assume that there is a unidimensional preference order on each set X_i. We assume here that all sets X_i are finite[1]. This restricted setting is sufficient in most MCDA contexts. The sets X_i need not be sets of numbers. They may be sets of labels or, if they are sets of numbers, the preference may neither be "the larger the better" nor the "smaller the better". This setting encompasses, in particular, the models mentioned in Part I, Section 4.6, item 4 ("Non-monotone criteria or attributes").

[1]See Bouyssou and Marchant [2007a,b] for a general presentation including infinite criteria scales.

Monotone Ordered Partitions (MOP). An ordered partition $\mathcal{C}$ of X is said to be *monotone* if there are (preference) orderings $\succsim_i$ on $X_i, i = 1, \ldots, n$, such that, whenever one improves an object x on some criterion i (w.r.t. the preference order $\succsim_i$), the resulting object is not assigned into a worse category of the partition than x. More formally, if $x \in X$, we denote by (y_i, x_{-i}), the object that has value y_i on criterion i and the same evaluations as x on all criteria but criterion i (we thus have $x_{-i} \in X_{-i} = \prod_{j \neq i} X_j$). The monotonicity condition amounts to imposing the following: $y_i \succsim_i x_i$ and $x \in C^h$, then (y_i, x_{-i}) is assigned into $C^{\geq h} = \bigcup_{l \geq h} C^l$, i.e., into a category at least as good as C^h. A partition satisfying this property is called a *Monotone Ordered Partition* (MOP).

In case we do not know of preference orderings on the attributes scales X_i, how can we tell that an ordered partition is monotone ? Actually, MOPs are exactly the ordered partitions that satisfy a property called *Linearity* by Bouyssou and Marchant [2007a,b]. This property reads as follows: for all $i \in N$, $h, h' \in \{1, \ldots, p\}$, $x_i, y_i \in X_i$, $a_{-i}, b_{-i} \in X_{-i}$,

$$\left\{ \begin{array}{l} (x_i, a_{-i}) \in C^h \\ \text{and} \\ (y_i, b_{-i}) \in C^{h'} \end{array} \right. \quad \Rightarrow \quad \left\{ \begin{array}{l} (y_i, a_{-i}) \in C^{\geq h} \\ \text{or} \\ (x_i, b_{-i}) \in C^{\geq h'}. \end{array} \right. \tag{1}$$

Intuitively, this property means that for all pairs $x_i, y_i \in X_i$, either substituting the evaluation x_i of an object by y_i is beneficial for a better assignment or it is the opposite. In other words, if $\mathcal{C} = (C^1, \ldots, C^h, \ldots, C^p)$ is a MOP, one cannot find $x_i, y_i \in X_i$, $a_{-i}, b_{-i} \in X_{-i}$, $h, h' \in \{1, \ldots, p\}$ such that :

$$\left\{ \begin{array}{l} (x_i, a_{-i}) \in C^h, \quad (y_i, a_{-i}) \notin C^{\geq h} \\ \text{and} \\ (y_i, b_{-i}) \in C^{h'}, \quad (x_i, b_{-i}) \notin C^{\geq h'}. \end{array} \right. \tag{2}$$

The presence of such a configuration would be the sign that the ordered partition is *not* monotone.

Preference orderings induced by a MOP. Any ordered partition $\mathcal{C}$ induces a relation $\succsim_i$ on X_i as follows: for all $x_i, y_i \in X_i$,

$$x_i \succsim_i y_i \;\text{ if }\; \forall a_{-i} \in X_{-i}, \forall h, \;\; [(y_i, a_i) \in C^h \;\Rightarrow\; (x_i, a_{-i}) \in C^{\geq h}]. \tag{3}$$

This relation is reflexive and transitive by definition. It is complete if and only if the Linearity property (1) holds. Thus, if $\mathcal{C}$ is a MOP, it induces a complete preorder relation $\succsim_i$ on each criterion scale X_i. Such relations are interpreted as induced preference orderings.

As shown by (3), a MOP is thus monotone w.r.t. each induced preference ordering $\succsim_i$. Let us define the *induced dominance relation* $\succsim$ as follows: for all $x, y \in X$,

$$x \succsim y \text{ if } \; x_i \succsim y_i, \forall i = 1, \ldots, n. \tag{4}$$

Obviously, a MOP is monotone w.r.t. the induced dominance relation, which is a partial order.

Monotonicity w.r.t. natural orders. In case the scale X_i of criterion i is a set of numbers or whenever it is a qualitative ordered scale, there is a natural ordering $\geq_i$ on X_i, which may or may not be *compatible* with the preference ordering $\succsim_i$. "Compatible" means that, for all $x_i, y_i \in X_i$, we have that $x_i \geq_i y_i$ implies $x_i \succsim_i y_i$. In this case, the induced relation $\succsim_i$ is in general coarser than the natural one $\geq_i$. Some values x_i, y_i with $x_i >_i y_i$ may not be distinguished by the induced relation, i.e.,, we may have $x_i \sim_i y_i$ (but never $y_i \succ_i x_i$, if the relations are compatible). If the relations $\geq_i$ and $\succsim_i$ are compatible, the ordered partition is monotone w.r.t. the natural order on X_i.

In many applications, there is a natural order $\geq_i$ on the scales of all criteria and it is intuitively clear that the ordered partition should be compatible with these natural orders (either in the sense "the larger the better" or the opposite sense). In such a case, the orders $\succsim_i$ induced by the ordered partition are compatible with the natural orders $\geq_i$ or the inverse natural order $\leq_i$ depending on the sense of the preference. Assuming that the sense of preference is "the larger the better" on all criteria, the ordered partition is then also compatible with the *natural dominance relation* $\geq$ (defined in the obvious manner by $x \geq y$ if $x_i \geq_i y_i$) for all i).

The decomposable threshold model [Goldstein, 1991]. The following simple result characterizes MOPs. It was first established by Goldstein [1991] in the case of two and three categories and generalized by Greco et al. [2001], Słowínski et al. [2002], Bouyssou and Marchant [2007a,b].

Theorem 1
The ordered partition $\mathcal{C} = (C^1, \ldots, C^h, \ldots, C^p)$ of the finite set $X = \Pi_{i=1}^n X_i$ is a monotone ordered partition (MOP) iff there are

- real valued functions $u_i : X_i \to \mathbb{R}, i = 1, \ldots, n$,
- real valued thresholds $\lambda_h, h = 1, \ldots, p+1$,
- a nondecreasing function $F : \mathbb{R}^n \to \mathbb{R}$,

such that, for all $x = (x_1, \ldots, x_n) \in X$, for all $h \in \{1, \ldots, p\}$,

$$x \in C^h \quad \text{iff} \quad \lambda_h \leq F(u_1(x_1), \ldots, u_n(x_n)) < \lambda_{h+1}. \tag{5}$$

In words, in case $\mathcal{C}$ is a MOP, it is always possible to represent the assignment rule of an object x to a category C^h by means of a value function F and thresholds. If the value of F associated to x falls between the thresholds delimiting a category C^h, then x is assigned to C^h. It can be assumed, without loss of generality, that the functions u_i are compatible with the induced preference order $\succsim_i$ on X_i, i.e., if $y_i \succ_i x_i$, then $u_i(y_i) > u_i(x_i)$.

Examples. It is clear that the AVF model satisfies (5), with F, an additive value function:

$$F(u_1(x_1), \ldots, u_n(x_n)) = \sum_{i=1,\ldots,n} u_i(x_i).$$

Note however that not all MOPs can be represented in the AVF model [see supplementary material in Bouyssou et al., 2022, Section E].

It is also clear that assignment rules based on any score that is monotone w.r.t. the preference order on criteria scales leads to a MOP. This is the case, in particular, for the methods discussed in Part I, Section 4.1.1, including those relying on the distance to an ideal point and/or an anti-ideal point.

It is less straightforward to exhibit an appropriate function F in the case of Electre Tri. Of course, such a representation does exist since Electre Tri is a monotone sorting method [Bouyssou et al., 2022, Prop. 1].

Non-monotonicity w.r.t. natural orders on criteria scales. If there is a natural order $\geq_i$ on the scale X_i and the ordered partition is not monotone with this order, it may be monotone w.r.t. a generally unknown preference order $\succsim_i$. In that case, the scale X_i has to be transformed. It is the role of function u_i to re-code the values $x_i \in X_i$ in order that the ordered partition is monotone w.r.t. the natural order on $u_i(X_i) \subseteq \mathbb{R}$. A typical example is the blood glucose level (glycemia) criterion in medical diagnosis. Glycemia values (measured while fasting) between 70 and 100 mg/dl are considered normal (i.e., most desirable). Above 100 mg/dl, are different grades of hyperglycemia, possibly linked with diabetes. Blood glucose level below 70 mg/dl is hypoglycemia. Re-coding the values of blood glucose level in increasing order of preference typically requires an unimodal (single-peaked) function u_i. The papers by Guo et al. [2019], Minoungou et al. [2022] incorporate the possibility of such re-codings of non-monotone criteria in indirect elicitation of the parameters of an AVF model and an MR-Sort model, respectively.

2.2 Decision rules

There is a trivial way of reformulating Theorem 1 in terms of decision rules. Let $\mathcal{C} = (C^1, \ldots, C^p)$ be a MOP of X that can thus be represented as in equation (5). $\mathcal{C}$ is monotone w.r.t. the induced dominance relation $\succsim$ (defined by equation (4)).

"At least" decision rules. For all $a \in X$, we may define a decision rule d_a, which applies to all $x \in X$ that dominate a w.r.t. the induced dominance relation, i.e., such that $x \succsim a$. We define, for $a \in C^h$,

$$d_a(x) = \begin{cases} C^h & \text{if } x = a \\ C^{\geq h} & \text{if } x \succsim a. \end{cases} \tag{6}$$

Such decision rules are referred to as "at least" decision rules because they say that x is assigned to C^h or better in case it dominates an object in C^h. Using "at least" decision rules, we may recover the partition $\mathcal{C}$ by assigning x to C^h whenever there is $a \in C^h$ such that $d_a(x) = C^{\geq h}$ and there is no $a' \in C^{h+1}$ such that $d_{a'}$ applies to x.

This essentially proves the following result.

Theorem 2
The ordered partition $\mathcal{C} = (C^1, \ldots, C^h, \ldots, C^p)$ of the finite set $X = \Pi_{i=1}^n X_i$ is a monotone ordered partition (MOP) iff there are

- complete preorders $\succsim_i$ on $X_i, i = 1, \ldots, n$,
- a subset of objects $A \subseteq X$, with $A \cap C^h = A^h \neq \varnothing$, for $h = 1, \ldots, p$,
- a set of decision rules d_a indexed by the objects $a \in A$,

such that, for all $x \in X$, for all $h = 1, \ldots, p$,

$$x \in C^h \quad \text{iff} \quad x \text{ satisfies a decision rule } d_a \text{ for some } a \in A^h \text{ and no decision rule } d_{a'} \text{ for } a' \in A^{h+1}.$$

As explained above, a trivial choice for the set A in the theorem is $A = X$. Do we need to consider as many rules as there are objects in X? Certainly not. The dominance relation $\succsim$ on X determined by the orderings $\succsim_i$ on X_i is a partial order on X, which also partially orders each category C^h. Since we have assumed that X is finite, there is a set of minimal objects $\min(C^h)$ in C^h. Any object that dominates an object in $\min(C^h)$ is in C^h or a better category. If it doesn't dominate any object in $\min(C^{h+1})$, it is in C^h. A minimal set of "at least" rules is thus obtained by associating a rule d_a to each $a \in A = \bigcup_{h=1}^{p} \min(C^h)$. We thus have that x is assigned to C^h iff there is a in $\min(C^h)$ such that x satisfies d_a and there is no a' in $\min(C^{h+1})$ such that x satisfies $d_{a'}$.

Example. Let us illustrate the above by an example. Let $X_i = \{0, 1, 2\}$, for $i = 1, 2, 3$. Consider the partition $\mathcal{C}$ of $X = \Pi_{i=1}^{3} X_i$ in three ordered categories (C^1, C^2, C^3). Let C^2 (resp. C^3) be the set of objects $x \in X$ such that $x_1 + x_2 + x_3 \geq 2$ (resp. $x_1 + x_2 + x_3 \geq 5$). This partition is monotone w.r.t. the natural order on X_i, hence the induced preference ordering $\succsim_i$ is compatible with the natural order on X_i, for all i. We have $\min(C^1) = \{(0, 0, 0)\}$ and $\min(C^2) = \{(1, 1, 0), (1, 0, 1), (0, 1, 1), (2, 0, 0), (0, 2, 0), (0, 0, 2)\}$ and $\min(C^3) = \{(2, 2, 1), (2, 1, 2), (1, 2, 2)\}$. In order to assign $x = (2, 0, 1)$, we observe that $x \succsim (1, 0, 1) \in \min(C^2)$ and x does not dominate any element in $\min(C^3)$. Therefore, x is assigned to C^2. In a similar way, $x = (1, 0, 0)$ is assigned to C^1 because x dominates $(0, 0, 0) \in \min(C^1)$ and does not dominate any element in $\min(C^2)$.

"At most" decision rules. Symmetrically to "at least" decision rules, we may define *"at most" decision rules*. For $b \in X$, the decision rule d'_b assigns b to C^h in case $b \in C^h$ and assigns x to $C^{<h}$ if $b \succsim x$. It is easy to give a version of Theorem 2 for "at most" decision rules. If we take for A the union of maximal elements (w.r.t. $\succsim$) in the different categories $\bigcup_{h=1}^{p} \max(C^h)$, we assign x to C^h if there is $b \in \max(C^h)$ with $d'_b(x) = C^{\leq h}$ and there is no $b' \in \max(C^{h-1})$ such that $d'_{b'}$ applies to x.

Using the above example, we have $\max(C^1) = \{(1, 0, 0), (0, 1, 0), 0, 0, 1\}$, $\max(C^2) = \{(2, 1, 1), (1, 2, 1), (1, 1, 2), (2, 2, 0), (2, 0, 2), (0, 2, 2)\}$. For instance, object $x = (2, 0, 1)$ is assigned to C^2 because x is dominated by $(2, 1, 1) \in \max(C^2)$, hence it satisfies the rule $d'_{(2,1,1)}$ and x does not satisfy any rule associated with the elements in $\max(C^1)$.

We may also use both "at least" and "at most" decision rules in a third version of Theorem 2. The assignment rule is thus expressed as follows: x is assigned to C^h iff there is a in $\min(C^h)$ such that x satisfies d_a and there is b in $\max(C^h)$ such that x satisfies d'_b.

For instance, using the above example, object $x = (2, 0, 1)$ is assigned to C^2 because x satisfies the "at most" rule $d'_{(2,1,1)}$ with $(2, 1, 1) \in \max(C^2)$ and x satisfies the "at least" rule $d_{(1,0,1)}$ with $(1, 0, 1) \in \min(C^2)$.

Reshaped decision rules. Theorem 2 and the two characterizations that can be obtained by using "at most" decision rules and combinations of "at least" and "at most" decision rules are quite similar to Theorems 4.3, 4.4 and 4.5 in Słowínski et al. [2002]. The decision rules used in the latter look different from those described above but are logically equivalent. Słowínski et al. [2002] consider "at least" decision rules δ parameterized by

- a subset $B = \{i_1, \ldots, i_\beta\}$ of the set of criteria $\{1, \ldots, n\}$, with β denoting the number of criteria in B;
- a β-tuple of real numbers $r_B = (r_{i_1}, \ldots, r_{i_\beta})$;
- one of the categories, C^h.

It is assumed that a function u_i from X_i into the reals is defined for all i and respects the preference ordering $\succsim_i$, i.e., $x_i \succ_i y_i$ implies $u_i(x_i) > u_i(y_i)$. An object x satisfies the decision rule $\delta(B, r_B, C^h)$

if $u_i(x) \geq r_i$ for all $i \in B$. In such a case, the rule assigns x to category C^h or better. It is easy to see that any $\delta(B, r_B, C^h)$ rule is satisfied by exactly the same objects as some d_a decision rule, and conversely. The object a corresponding to the δ rule is as follows :

- for $i \notin B$, a_i is a value in X_i for which u_i is minimum;
- for $i \in B$, a_i is a value in X_i such that $u_i(a_i) = \min\{u_i(x_i) : u_i(x_i) \geq r_i\}$.

"At most" decision rules $\delta'(B, r_B, C^h)$ are defined symmetrically and are equivalent to decision rules d'_b associated to a maximal element in C^h.

Using again the above example, we have that the "at least" decision rule d_a with $a = (1, 0, 1)$ corresponds to the decision rule δ with $B = \{1, 3\}$, $r_B = \{1, 1\}$ and $C^h = C^2$. Note that, in this example, we may take for u_i the identity function, i.e., $u_i(x_i) = x_i$, since the ordered partition is monotone w.r.t. the natural order on X_i.

DRSA. In the Dominance based Rough Sets Approach (see Part I, Section 4.5.1), "at least" and/or "at most" decision rules, (respectively in the shape of the δ and/or δ' rules just described) are induced from a set of known assignments. In the limit case in which the assignment of *all* objects in X is known and these assignments correspond to a MOP of X, the certain rules induced by a DRSA approach would perfectly restate the MOP. This is established by Theorem 2 and its variants or by Theorems 4.3, 4.4 and 4.5 in Słowínski et al. [2002]. In practice, only the assignments of a subset of objects $Y \subseteq X$ are known. The induced rules, in general, imprecisely describe the set of known assignments. Even with certain rules, some objects in Y may not be assigned to the precise category they belong to; some objects in Y may fail to match the condition part of all induced rules. The goal of the DRSA rule induction process is to predict the assignment of objects that do not belong to Y. In order to predict as many unknown assignments as possible, it is beneficial to have rules whose condition is the least possible specific, i.e., δ or δ' rules in which the set of criteria B is as reduced as possible (*reducts* in rough sets terminology). Therefore, DRSA selects minimally specific rules that describe as well as possible the set of known assignments. In general, the rules induced in DRSA are compatible with several MOPs on X. The set of known assignment examples is said to be *inconsistent* when no MOP on X is compatible with all of them. This is for instance the case when there are dominance violations within the known assignments.

In the sequel we call *Decision Rule model* the description of a MOP on a product set by means of a set of rules that determine it as in Theorem 2 or in Słowínski et al. [2002, Theorems 4.3, 4.4, and 4.5]. The Decision Rule model is the model underlying DRSA, but the term DRSA refers to the techniques developed for inducing rules from a (sub)set of known assignments.

2.3 The Non-Compensatory Sorting model

Bouyssou and Marchant [2007a,b] characterize an idealized version of the (pessimistic or pseudo-conjunctive) Electre Tri model that they call the Non-Compensatory Sorting (NCS) model (already introduced in Part I, Section 4.3). As a reminder, this model assigns an object to category C^h iff the object is at least as good as the category lower limit profile on a sufficient subset of criteria while it doesn't meet this condition w.r.t. category C^{h+1}. Bouyssou and Marchant position the NCS model within the general monotone sorting model. For each criterion scale X_i, the NCS model makes no difference between evaluations that lie between the values of successive profiles. Therefore, X_i is partitioned in at most $p + 1$ equivalence classes of values delimited by the values of the categories lower limit profiles and ordered in the same order as the category limit profiles. These ordered equivalence classes define a complete preorder $\succsim_i$, which is compatible with the "natural" preference order $\geq_i$ on X_i. The former is generally coarser than the latter since it does not distinguish values between consecutive limiting profiles. Basically, Bouyssou and Marchant's characterization amounts to restricting the number of equivalence classes of values on each criterion scale to, at most, the number of categories plus one. Note that Bouyssou and Marchant [2007a,b] also characterize a version of the NCS model with a veto.

Relationships. The pseudo-conjunctive (pessimistic) Electre Tri model contains MR-Sort, i.e., the former is more general than the latter. The NCS model also contains MR-Sort (see Part I, Section 4.3). NCS is not contained in the pseudo-conjunctive Electre Tri model because the concordance index in the latter is computed using additive weights, while there are sets of sufficient subsets of criteria in an NCS model that cannot be represented using weights and a threshold.

The pseudo-conjunctive Electre Tri-nB [Fernández et al., 2017] of course contains the pseudo-conjunctive Electre Tri. Actually, in case the number of limiting profiles of each category is not upper-bounded, it is easy to prove that Electre Tri-nB is equivalent to the monotone sorting model [Bouyssou et al., 2022], i.e., any MOP can be represented in the Electre Tri-nB model, using an appropriate number of limiting profiles. Hence, Electre Tri-nB is also equivalent to the Decision Rule Model underlying DRSA.

2.4 A picture of models

Figure 1 represents inclusion relationships between models, including those that have been axiomatically characterized. The box at the top is the most general monotone sorting model. Any model for sorting in ordered categories while respecting monotonicity w.r.t. some preference ordering of the criteria scales (which is not necessarily the natural ordering) is contained in this box. Any such model can be described in several different ways, including by a set of rules (as in the Decision Rule model) or an Electre Tri-nB model with an adequate number of limiting profiles separating the categories. Below, in the left part of the figure, are some models belonging to the Electre Tri family. Of course, the family of Electre Tri-nB model with fixed maximal number $n \geq 2$ of limiting profiles per category contains Electre Tri (which is Electre Tri-nB, with $n = 1$). The family of Electre Tri-nC models could have been represented in the graph, as included in the general model, but without any other known inclusion. On the righthand side, we find, in particular, the family of models based on an additive value function and the more general case of a value function (VF) in which criteria may interact[2]. Other methods based on monotone scores (see Part I, Section 4.1.1) could have been represented there; they are all included in the general monotone model while other inclusions have not been established.

2.5 Further results

Electre Tri. For further results and discussion on the relationships between sorting models in the Electre Tri family, the interested reader may want to see Bouyssou et al. [2022, 2021].

[2]The main model of this type is the Choquet integral model [see, e.g., Grabisch, 2016, Chapter 6]. This model has been characterized by Wakker [1989, Theorem VI.5.1] in a ranking context. However, the latter result applies only in the case all criteria are evaluated on a common scale. This is a convenient setting for decision under uncertainty, but raises difficulties related to criteria scales commensurateness in a MCDM/A context [Grabisch, 2016, p. 343].

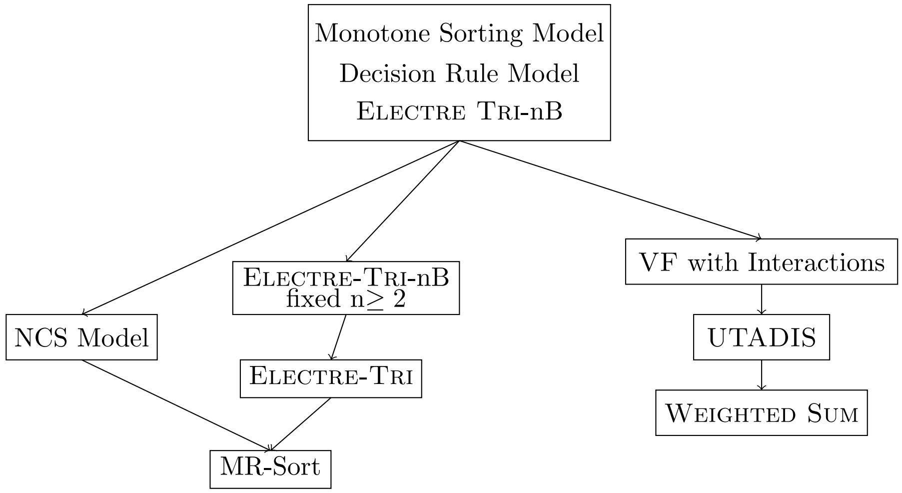

Figure 1: Inclusion relationships between models

In another paper, Bouyssou and Marchant [2015] investigate the relationship between the Electre Tri models based on limiting profiles (Electre Tri-B) and those based on central or characteristic profiles (Electre Tri-C). They show in particular that these models are not equivalent, i.e., not all ordered partitions representable in one type of model can be represented in the other. They also analyze the relationship between the pseudo-conjunctive (pessimistic) and the pseudo-disjunctive (optimistic) version of Electre Tri-B, which do not correspond through the application of a natural operation of *transposition* (or duality). They propose a variant of Electre Tri in which the two versions correspond through transposition.

Sugeno integral. The *Sugeno integral* is a (monotone) score computed by using only max and min operations and a capacity which weighs the subsets of criteria [see, e.g., Dubois et al., 2001]. Therefore it is suitable for aggregating ordinal data. Ordinal aggregation operators such as $\max, \min$, the median or all order statistics are particular cases of the Sugeno integral. As the Choquet integral, the Sugeno integral can be used, together with thresholds, to assign objects to ordered categories, provided the criteria scales are commensurate or made commensurate. Słowínski et al. [2002] gave a characterization of monotone ordered partitions that can be obtained by using a Sugeno integral and thresholds. Actually, since the criteria scales X_i are different in general and may even be non-numeric, the Sugeno integral for an object $x = (x_1, \ldots, x_n)$ is computed after the evaluations x_i have been re-coded into $u_i(x_i)$, where u_i is a function mapping the scale X_i into the nonnegative reals. The role of the u_i's is to make the scales X_i commensurate [see Bouyssou et al., 2009, Section 6.3]. Bouyssou and Marchant [2007b] establish that the Sugeno integral-based sorting model is equivalent to the non-compensatory sorting (NCS) model. Any partition obtained by the NCS model can thus also be described by score and thresholds, the score being a Sugeno integral.

Additive value function. Bouyssou and Marchant [2010] have characterized an additive value function model for a particular type of ordered partitions. These partitions have three categories: the attractive objects on top, the unattractive, on the bottom, and the neutral ones, in between. The neutral category is special. It is "thin" in the sense that any improvement (resp. deterioration) of the performance of a neutral object on any criterion leads to its assignment into the attractive

(resp. unattractive) category. The neutral category can be seen as a frontier. An additive value function of such a partition assigns a positive (resp. null, negative) value to the objects in the attractive (resp. neutral, unattractive) objects. The authors sketch an elicitation procedure of the model's parameters (i.e., the marginal value functions). This procedure presents well-chosen objects to the DM and asks her to which category they should be assigned. All questions are of that type.
The authors have extended their characterization to more categories, actually, to the case of p "slightly" overlapping ordered categories. Successive categories slightly overlap in the sense that their intersection is a "thin frontier" between them, as is the neutral category in the initial model, which can be seen as a thin frontier between attractive and unattractive objects. In the additive value function representing the categories, the objects in the "thin" frontier are all assigned the same value.

2.6 Remarks

What are models characterizations good for?

- Having a characterization of a model allows to give a clear interpretation to the model's parameters. In a direct elicitation perspective (Part I, Section 2.4), this helps in determining the parameters values and supports the explanation of the recommendations. For instance, in the NCS and the MR-Sort models, the limiting profiles determine classes of values on the criteria scales. Within a class, value differences make no difference in terms of assignment. More efficient questioning strategies can be elaborated on the basis of these precise interpretations of the parameters.

- Results regarding models inclusion relationships (such as illustrated in Figure 1) allow to assess the degree of generality of the models used. Since information scarcity is typical of multiple criteria decision processes, preferably using simple models reduces the indeterminacy in the model's parameters. Using very general models in the presence of little information, is contrary to the *parsimony principle*, which is well-known in statistical modelling and could be recommended in MCDM/A too. Taking all statements made by the DM into account, hence using more general models, could be counter-productive. For instance, the DM's claiming that some criteria interact positively or negatively should not necessarily imply shifting to a model with criteria interactions. The need for modelling criteria interactions should be challenged through questions raised to the DM, in terms of appropriate objects assignments. A tradeoff should be made between model complexity, or generality, and the respect of all the available information. In cases, one may prefer a simpler model that doesn't fit perfectly the available information over a more complex model.

3 Dealing with imperfect information

In this section, we discuss issues related to the quality of the information contained, e.g., in a set of assignment examples. The first subsection addresses the case in which a set of known assignment examples can not be fully represented in the chosen model. In the second subsection, we review how insufficient or imprecise information can be dealt with.

3.1 Inconsistent sets of assignment examples

In the above, we hardly mentioned the issues raised by "inconsistent data". Stating that a set of assignment examples is inconsistent is relative to some presupposed MCS model. In our context, it means that some assignment or set of assignments is incompatible with a sorting model supposedly suitable in the current decision context. Actually, a single assignment cannot be inconsistent by itself. It can be such when added to a set of previously known assignments. The simplest inconsistent set of assignment examples is composed of two objects, one dominating the other (w.r.t. known preference orderings of the criteria scales), and the dominated one assigned to a better category than the dominant. In such a case, no model instance will restore these two assignments, whatever the monotone sorting model considered. Assignments violating the dominance relation could be called an *absolutely inconsistent* case since no monotone sorting model can accommodate such assignments. In case no preference ordering is presumed on the criteria scales, inconsistency means violating the Linearity property; such a violation is revealed by a quadruplet of objects as in equation (2). In such a case no monotone sorting model can accommodate the assignment examples. In other cases, less radical inconsistencies can be dealt with, for instance, by considering a more general MCS model.

Whenever a set of assignments proves incompatible with a sorting model postulated as suitable, there are several ways of dealing with the problem, such as:

- detecting the inconsistencies and "solve" them;
- moving to a more general model or another type of model (with a different logic) that is able to represent the currently inconsistent assignments;
- splitting the set of assignment examples in parts and fitting a model to each subset;
- using a model and/or an indirect elicitation procedure that is tolerant to inconsistent assignments.

We briefly address each of these cases and point to some work that has been done in each direction. The handling of inconsistencies interferes with other types of "data imperfections" such as hesitation, uncertainty, and imprecision (see Sections 3.1.4 and 3.2).

3.1.1 Detecting and handling inconsistencies

Indirect elicitation of an AVF model or an Electre Tri model when only the criteria weights and the thresholds are unknown goes through a linear programming formulation. Assignments induce linear inequalities involving the model's parameters. Mousseau et al. [2003] have designed two algorithms to identify all minimal (in cardinality) subsets of constraints to be deleted in order to restore consistency. Building on this, Mousseau et al. [2006] propose alternative approaches to resolve inconsistencies. Instead of deleting assignment examples, one may relax them, i.e., consider an interval of possible assignments. In case the DM is able to associate a confidence degree to each assignment example, the authors elaborate algorithms that present the sets of constraints to be deleted or relaxed, starting with the least confident constraints.

We are not aware of other work in this vein.

3.1.2 Moving to another model

In case a model proves unable to restore all known assignments, one may consider using a more general, hence more expressive, model. For instance, when working with UTADIS and using piecewise linear marginal value functions, one option is to increase the number of pieces in the marginals. Another option consists of considering as variables all marginal values associated with evaluations of the objects in the set of assignment examples w.r.t. all criteria (as done in Greco et al. [2010]). Inconsistencies w.r.t. the latter formulation mean that the assignment examples are not compatible with a general AVF model. Still more general, criteria interactions may be added as done by Greco et al. [2014]. This example illustrates a possible modeling methodology which implements a parsimony principle using a family of increasingly general embedded models. This strategy would start with the simplest model in the family. If it restores all assignments examples, we keep the fitted model. Otherwise, we move to a more complex model in the family, and so on, until a model is possibly found that restates all assignment examples.

What are the families of embedded models that can be used in such an approach? We just point out the family of AVF models, possibly extended by VF models involving criteria interactions. The former does not allow to capture all MOPs (as mentioned in Section 2.1), while it is unknown whether the latter do. A family that indeed allows representing all MOPs is Electre Tri-nB, using an unbounded number of limiting profiles. Unfortunately, indirect elicitation algorithms for such models rely on MIP or SAT solvers, which cannot handle large numbers of limiting profiles.

3.1.3 Splitting the learning set

In group decision-making, group members may assign the same object to different categories. If this is the case, no single model (at least no model that assigns objects to a single category) is able to restate the assignments of all group members. Group decision methods for aiding to sort often rely on assignment examples provided by each group member (see Part I, Appendix D.4.3). These assignments may be contradictory. A frequent strategy is to consider a sorting model representing the preferences of each group member and then build a consensus model or, more modestly, consensual assignments.

In group decision-making, each assignment example is labelled by the group member who provided it; hence there is a straightforward way of splitting the set of assignment examples. This is not the case in general. In situations where a single model cannot represent all assignment examples, an option is to find an adequate way of splitting them and develop a sorting model for each subset. Kadziński et al. [2020] do that in the context of the AVF model. They assume that the model to be used for sorting may depend on the performance profile of the objects. They motivate and illustrate their approach by the example of the evaluation of research units in different fields of science. Their method involves the learning of a decision tree that splits the objects on the basis of their performance w.r.t. specific criteria.

3.1.4 Keeping pace with inconsistencies

A frequent way of dealing with possible inconsistencies is by tolerating them and finding a model that minimizes their number or degree of inconsistency. In the early times of development of the UTA and UTADIS methods, linear programming formulations have been devised, which find an AVF (and thresholds) if there is one restoring all known assignments; otherwise, they produce one that restores the assignments as well as possible[3].

In (supervised) machine learning, model-based approaches to preference learning [see, e.g., Fürnkranz and Hüllermeier, 2010], proceed from specific assumptions about the type of preference structure to be learned and about the data generation mechanism, usually a stochastic process. Data (i.e., assignments, in our case) are thus noisy. Stochastic modeling allows using induction principles such as *maximum likelihood*. The model identified is the *most likely* given the hypotheses and the data. In this setting, "inconsistencies" result from noise superimposed to a "ground truth". An example of such an approach is the choquistic regression proposed by Tehrani and Hüllermeier [2013], Tehrani et al. [2012] in which a latent value function (which is a Choquet integral of the object evaluations) is estimated.

Within DRSA, the presence of inconsistencies in the set of known assignments results in inducing the so-called *ambiguous* rules (see Part I, Section 4.5.1). For instance, consider the following example involving two criteria, both evaluated on the scale $X_i = \{1, 2, 3\}$, $i = 1, 2$. The preference order on X_i is the natural order (the larger, the better). Let there be four categories C^1 to C^4, labelled in increasing order of preference. Assume we have, among the known assignments, object $(1, 2)$ assigned to C^3 and object $(3, 2)$ assigned to C^2, which violates dominance. Rules potentially induced by these data are:

δ : If $x_1 \geq 1$ and $x_2 \geq 2$, then $x = (x_1, x_2)$ belongs to C^2 or better;
δ' : If $x_1 \leq 3$ and $x_2 \leq 2$, then $x = (x_1, x_2)$ belongs to C^3 or worse.

These two assignment examples that violate dominance induce rules that do not allow to unambiguously tell into which category objects $(x_1, 2)$, with $1 \leq x_1 \leq 3$ are assigned. The rules imply that they are assigned either to C^2 or C^3. In the case of DRSA, inconsistencies in the assignment data result in imprecision or hesitation in assignments provided by the induced rules.

3.2 Insufficient, imprecise and uncertain information

Various forms of imperfection affect the information used to determine MCS models. First of all, in the case of indirect elicitation of a model's parameters from assignment examples, it is exceptional that the available examples completely determine a model instance. In other words, for the chosen model type, several sets of model parameters allow us to restore the assignment examples (in case there is no inconsistency in the data) or restore them equally well. The case is not so much different when assignment examples are imprecise (examples are not assigned to a single category but to a category interval or a set of categories) and/or when model parameter values are imprecisely known. Such imprecise input information tends to increase the set of model instances compatible with the available data. The same methods as those used for dealing with model indeterminacy generally allow to handle the additional imprecision.

The ROR approach (mainly applied to the AVF model, see Part I, Section 3.4.1) considers all instances of the AVF model and thresholds that restore the assignment examples. The resulting output regarding an object is the set of its possible assignments. This approach is also able to

[3]Different implementations of an "as well as possible" requirement can be found in the literature. Usually, slack variables are introduced in the linear programming formulation. There are several ways of doing so [see, e.g., Jacquet-Lagrèze, 1982, Doumpos and Zopounidis, 2002]. The objective function of the LP minimizes the sum of the slack variables, or the maximal value of slack variables, or a linear combination of such objectives. A positive slack variable means that a constraint is violated. Minimizing the *number* of wrong assignments requires a MIP formulation.

deal with imprecise assignments of examples in the learning set (i.e., some objects in the learning set may be assigned to a category interval instead of a single category, which reveals a form of uncertainty about the "right" assignment from the DM's part). The ROR approach does not take the leap: the models considered are *univocal*, i.e., they assign each object to a single category. Simply, when a single model cannot be identified, one considers all models compatible with the assignment examples, all univocal models. The recommendation regarding an object is the set of all categories to which a compatible univocal model assigns the object.

SMAA (see Part I, Section 3.4.2) proposes another way of dealing with imprecision or indeterminacy. If the model's parameters cannot be completely determined, one may sample the parameters space of the compatible models (e.g., using a uniform distribution). To each sampled set of parameters corresponds a model, a univocal one. Using the CAI (class acceptability index), one obtains a distribution on categories for each object to be assigned. This technique also keeps the notion of the univocal model as central. The distribution on compatible models (via a "prior" distribution on the corresponding part of the parameters space) yields a distribution on categories for each object.

INTERCLASS [Fernández et al., 2019, see Part I, Section 4.2] revises Electre Tri-B in the case where all information obtained in a direct elicitation process (no assignment example) may be affected by imprecision. All parameters may be given as value intervals. Curiously enough, the revised assignment rule remains univocal. The rule assigns to a specific category, not to a category interval.

DRSA is a method for inducing rules from assignment examples (in case of sorting problems). The underlying model is the Decision Rule Model alias, the Monotone Sorting Model, which is the most general model respecting the dominance relation (see Sections 5.1.1 and 5.1.2). The DRSA method yields such a model only when information is complete and consistent. In this case, every object is assigned to a precise category with certainty. Otherwise, DRSA yields imprecise assignments in the form of sets of categories, which are furthermore qualified as necessary or possible.

Moves towards models assigning to sets of categories. In some papers, the authors move, rather timidly, from univocal models to *models assigning to category intervals*. This can be done by considering jointly two related univocal assignment models. With Electre Tri-C [Almeida-Dias et al., 2010], the assignment rule uses both the assignments provided by the descending and the ascending rule. This leads to the possibility of category interval assignments. This idea of combining the two assignment rules related to the Electre Tri model and its various extensions, especially those using central profiles, is also present in Fernández et al. [2020]. This paper extends INTERCLASS to adapt it to Electre Tri-nB and Electre Tri-nC. In contrast with INTERCLASS, the authors now advocate using the two rules conjointly.

In a similar spirit, Janssen and Nemery [2013] extend $\mathcal{F}$low$\mathcal{S}$ort (see Part I, Section 4.1.2) to deal with interval-valued evaluations and parameters. The output is, in general, a category interval.

Such a departure from working with instances of univocal sorting models compatible with input information occurs in a different manner with Rocha and Dias [2008], Köksalan et al. [2009], Kadziński and Ciomek [2016] (see Part I, Section 4.4). The authors do not use any underlying *sorting model*. They do, however, postulate an outranking model for the DM's preferences on pairs of objects. They impose as constraints that assignments should not violate the outranking relations compatible with the available information, i.e., an object that outranks another may not be assigned to a worse category (thus extending the *respect of dominance* principle into the *respect of outranking*). From a robust learning perspective, they output all assignments that respect the information provided by the DM and do not violate the outranking relations compatible with this information.

Forms of imprecision in input data and output. In addition to the above, we list a number of different ways of modelling imperfection in input data and the corresponding output, which have been considered in the MCS or monotone classification literature.

Probabilistic assignment. Logistic regression and choquistic regression (see Part I, Section 4.5.3) associate a probability distribution on the set of categories to each object. The probabilistic assignment model is learned from a set of assignment examples. The latter are viewed as realizations of random variables depending on the object's evaluations. An object assignment is predicted on the basis of the probability distribution associated to the object by the learned model. For instance, the predictor assigning to the category having maximal probability is a predictor that minimizes the 0/1 risk.

Assignment with credibility degree. Liu et al. [2020] deal with valued assignment examples, i.e., examples can be assigned to multiple classes with a credibility degree associated with each of them. They work in the context of the AVF sorting model. They compute a vector of credibility degrees of assignment to each category for the objects out of the set of assignment examples.

Missing or imprecise evaluations. Meyer and Olteanu [2019] deal with imprecise (intervals) or missing evaluations in an extension of the MR-Sort model. The resulting assignments are category intervals in general.

Fuzzy sets. Pereira et al. [2019a] reinterpret the single criterion concordance indices of Electre Tri-C in terms of membership functions. Determining the $p-1$ profiles values and the preference and indifference thresholds for each criterion is reformulated as determining p trapezoidal fuzzy numbers associated with the categories. No discordance index is considered. The descending and ascending rules of Electre Tri-C are adapted, but each of them yields a univocal assignment. Each object thus receives at most two different assignments. Pereira et al. [2019b] extend the previous approach using *hesitant fuzzy numbers*; with these, the membership degree is imprecisely known, which corresponds to imprecise knowledge of preference and indifference thresholds. The authors adapt the descending and the ascending rules to this setting and apply these rules using three different aggregations of the single criterion concordance indices. This may result in several different assignments.

Other papers use fuzzy sets or numbers (in particular hesitant fuzzy numbers) in relation to ELECTRE methods, but none other specifically for sorting. See Pereira et al. [2019b, for references].

The reader may also want to see the review on techniques for modelling uncertain input data in MCDM/A [Pelissari et al., 2021]. However, this review does not specifically address MCS methods, but rather all MCDM/A methods.

Summing up. Most methods for dealing with imperfect input data keep relying upon an univocal sorting model. Imprecision in input data (parameters or examples) and indeterminacy often result in using a set of model instances for assigning objects to a category or a subset of categories.

Assuming a probability distribution on the instances or the model's parameter values may lead to a probability distribution on an object's possible assignments. Other methods adapt classical univocal models sorting rules to imprecise evaluations, parameter values or assignment examples. This may result in univocal assignment rules or assigning objects to category intervals or subsets. One type of model (logistic or choquistic regression), coming from the machine learning field, postulates that assignment examples are realizations of random variables. Consequently, the learned model produces a probability distribution of categories as a probabilistic assignment for each object. This approach, however, seems better suited to model the variability of the behavior of many agents rather than that of a single DM.

4 Beyond recommendation

At the end of the process in which a MCS model has been selected, its parameters have been elicited and recommendations have emerged, further issues may arise. These pertain to a phase of the decision aiding process that could be called *post-treatment*. "Benchmarking" and explanation of the recommendations are among the questions that have attracted attention in recent years.

4.1 Benchmarking

How to improve the evaluations of an object which is assigned to a certain category in such a way that it will be assigned to a better category? This problem is known under various denominations: *benchmarking* [Petrovic et al., 2014] , *post factum analysis* [Kadziński et al., 2016], *inverse multicriteria sorting problem* [Mousseau et al., 2018][4].

Benchmarking is originally a business management tool that compares a company's performances to that of market leaders and finds improvement paths inspired by more successful competitors (*benchmarks*). This idea was previously implemented in DEA. Ghahraman and Prior [2016] search for optimal paths proceeding by stepwise improvements towards the efficient frontier. This concept of stepwise benchmarking is adapted by Petrovic et al. [2014] to a multiple criteria *ranking* context in which the Electre I method is used for ranking. Paths with equal improvements on each criterion at each step are searched for.

In the framework of robust ordinal regression (ROR), thus working with all AVF and thresholds compatible with assignment examples, Kadziński et al. [2016] answer questions such as: "What are the minimal performance improvements that lead an object to be necessarily (or possibly) assigned to a better category ? ".

Kadziński et al. [2022] formulate the search for improvement paths as optimization problems. The steps in a path consist of real or fictitious objects. Their method is very general. It applies when the parameters of the sorting model are precisely specified and it can also deal with all models compatible with a set of assignment examples in the framework of the robust approach. It can be adapted to UTADIS, Electre Tri or DRSA. Constraints defining feasible improvements can be imposed.

[4] Note that the latter term is classical in this sense in the field of classification [Aggarwal et al., 2010] but may be ambiguous because some authors use it designate the learning of a sorting model from assignment examples.

4.2 Model interpretation and results explainability

In a decision-aiding process, explaining to the DM why an object is assigned to a specific category is important to build trust in the approach and the recommendations. The simplest way of doing so is to work preferably with simple models based on clear principles. This rejoins the methodological parsimony principle alluded to in previous Section 3.1.4, i.e., start with the simplest possible model in a family of models of incremental complexity and move to a more complex model only in case of necessity. In this way, the DM will gradually understand the basic model and acknowledge the necessity of a more complex model. So, starting with a weighted sum in an approach based on value functions or with an MR-Sort model in an outranking-based approach seems advisable.

Things may be different if the DM's involvement is limited or in a pure preference learning context. Assume that a model, a family of compatible models, or a set of decision rules have been obtained. The validity of such models for the DM or the client depends on the plausibility of assignments of objects other than the assignment examples. Again, if the obtained model is simple, the DM or the client could be able to understand it and elaborate an interpretation of the sorting model as a whole.

In all cases, an alternative approach consists of developing an explanatory system, usually relying on the same principles as the models used in the indirect elicitation or the learning phase. This kind of approach is recent and relies on the notion of *accountability*, i.e., the ability of a human DM to own a recommendation made by a system [Doshi-Velez et al., 2017, Wachter et al., 2017]. Such concern is closely related to the emergence of similar approaches regarding Artificial Intelligence-based systems under the name of "eXplainable AI (XAI)", see [Gunning and Aha, 2019].

The generation of explanations in a multicriteria setting is not a straightforward task because different criteria are at stake, and the DM is not necessarily able to fully assess their importance or to understand how they interact. Moreover, once the DM is confronted with the result and the explanation, she may realize that it is not exactly what she expected. Therefore, she may make changes or provide new information that will have some effects on the elicitation phase (choosing the next questions, adapting the parameters of the model, etc.). Thus, beyond the acceptance facility, presenting an explanation may have an impact on the representation of the user's mode of reasoning, which is the basis of building the recommendation.

Various works were devoted to the question of designing explanations of the outcomes of multiple criteria decision models. Several works were focused on additive value models, see for instance [Belahcène et al., 2017, Belahcène et al., 2017, Labreuche, 2011, Labreuche and Fossier, 2018, Labreuche et al., 2011, 2012, Zhong et al., 2019, Bazin et al., 2020]. Overall, the main idea is to break down the decision into simple statements presented to the DM. The whole sequence of statements should formally support the decision. For instance, Belahcène et al. [2017] explain why an object x is ranked before an object y by exhibiting a sequence of *preference swaps* (that is, pairwise comparisons of alternatives varying on two criteria only) that starts from y and finishes in an object that is dominated by x. All objects in the sequence are preferred to y (because of preference swaps), and the final object is dominated by x. This establishes, by transitivity, that x is preferred to y.

For sorting methods, things are still in progress regarding this question. Works has been initiated in [Belahcène et al., 2017, Belahcène et al., 2018], to tackle the accountability problem of decisions issued from a Non-Compensatory Sorting (NCS) model. For instance, [Belahcène et al., 2018] discussed a context where a committee (criteria) meets to decide upon sorting several candidates (alternatives) into two categories (e.g., Bad and Good). The committee applies a public decision process; the outcomes are also public. However, the details of the votes are sensitive and should not be made available. Thus, to what extent can the committee be accountable for its decisions? The authors formalize accountability using a feasibility problem expressed as a Boolean satisfiability formulation. Two situations were considered. In the first one, the committee needs to justify that its decision is *a possible* NCS assignment. A characterization result helps to turn the existence of such an assignment into finding separations of the pairs of Good and Bad candidates over at least one point of view, which can be formulated as a SAT problem. This allows generating a single *argument scheme* that can explain all possible NCS assignments. The second situation arises when the assignment of a new candidate is *necessarily* derived from jurisprudence. Thanks to the characterization result, one can also construct an argument scheme representing deadlock situations. An argument scheme is a tool borrowed from Artificial Intelligence. They are operators tying a tuple of premises – pieces of information satisfying some conditions – to a conclusion [Walton, 1996], allowing for capturing prototypical reasoning patterns.

5 Conclusion

The topic of MCS is torn between two domains, MCDM/A and Machine/Preference learning. Originally, MCDM/A postulates interactions with a DM. Very soon, MCS methods focused on indirect elicitation of the model's parameters on the basis of assignment examples. This brought the field close to ordinal or monotone classification and machine learning. The specificity of MCS is that model's parameters have to be elicited using few assignment examples. Therefore, the indeterminacy of the model is more of an issue than the incompatibility of the assignments with a model. Of course, nothing precludes using MCS models to compete with machine learning methods for monotone classification. However, few indirect elicitation methods developed for MCS models have been tested on large sets of assignment examples. Some indirect elicitation methods for "simple" MCS models, such as MR-Sort and UTADIS scale up reasonably to deal with large sets of assignments examples and have been compared to monotone classification algorithms on benchmarks. Such methods do not fully fit, however, into the Machine/Preference Learning paradigm, which postulates a ground truth, a stochastic error model and determines a model's parameters by optimizing an objective function that is meaningful in this setting as, e.g., maximizing the likelihood of the observed assignments. In contrast, DRSA is based on a very general model, is suited to work with large sets of assignments and is positioned from the outset as a knowledge discovery tool.

Current state of development

The major part of the published literature on MCS methods is devoted to proposing (i) new sorting models, (ii) variants of the two landmark models UTADIS and Electre Tri, and (iii) indirect methods for eliciting the parameters of MCS models.

Regarding indirect elicitation, we observe the following.

- Methods for eliciting the parameters of an MCS model based on a *value function* are richly developed. They handle all sorts of variants of the model (criteria interactions, hierarchy of criteria, imprecise assignments, etc.). They deal with model indeterminacy in various ways (central or representative model, possible and necessary assignments, probability of assignment, etc.).

- Indirect elicitation of the parameters of MCS models based on an outranking relation (such as Electre Tri) is technically more difficult because limiting or central profiles are to be determined, which requires solving a MIP, a SAT or a nonlinear optimization problem. Therefore, the means for handling model indeterminacy are also poorer.

In contrast, direct elicitation is a poorly developed issue. Few questioning strategies based on rigorous principles (i.e., on primitives that are related to a theoretical characterization of a model) have been elaborated. This is also due to the small number of methods that have been characterized. An exception is Bouyssou and Marchant [2010] for a particular case of the AVF model.

Research topics deserving further investigation

Given the current state of development of the field, we suggest pushing forward three research directions that are not yet much developed or deserve further attention.

Active learning. In a decision-aiding process, the DM's availability is usually limited. Therefore, it is important to ask the DM informative questions. In the standard situation where the DM is asked to assign some objects, these objects should be chosen so that their assignment efficiently reduces the set of compatible model's parameters. In some settings, a "budget of questions" is available. They should be chosen adequately, either in sequence or all from the start. Appropriate criteria for selecting questions have to be studied. Optimal questions in terms of reduction of the compatible parameters set are likely to be cognitively difficult questions for the DM, since the best objects are intuitively close to the border of the categories. Some sort of compromise between questioning efficiency and cognitive difficulty has to be studied.

Results explainability. In case the recommendations are obtained through learning a model based on assignment examples; it is important to convince the DM that these recommendations are reliable. If the model considered is simple and interpretable, the recommendations may be justified by invoking the model itself. This is especially true in case the DM actively participated in the decision-aiding process. In case the model is complicated and its parameters have been learned on the basis of assignment examples, it is helpful to dispose of methods able to generate arguments justifying the recommendations.

Models axiomatization and direct elicitation. Having at disposal axiomatized models has several virtues. It provides a clear picture of models inclusion, allows to the design of direct elicitation procedures that only rely on the primitives of the model, and gives clues for explaining the recommendations. While it is hopeless to try to characterize complicated models that have been proposed, further efforts could identify the essential features of some methods and characterize simplified versions of these methods. Direct elicitation methods can also benefit from research on the first topic, i.e., active learning.

Acknowledgments

We are grateful to Denis Bouyssou for reading a previous version of the manuscript and making a number of relevant comments. We also thank the Editors for inviting us to write this survey and for their observations on the final draft. Of course, the responsibility for errors and omissions in this paper as well as the opinions that are expressed remains entirely with the authors.

References

C.C. Aggarwal, C. Chen, and J. Han. The inverse classification problem. *Journal of Computer Science and Technology*, 25(3):458–468, 2010.

J. Almeida-Dias, J. R. Figueira, and B. Roy. ELECTRE TRI-C: A multiple criteria sorting method based on characteristic reference actions. *European Journal of Operational Research*, 204(3):565–580, 2010.

A. Bazin, M. Couceiro, M.-D. Devignes, and A. Napoli. Explaining Multicriteria Decision Making with Formal Concept Analysis. In *Concept Lattices and Applications 2020*, volume 2668 of *CEUR Workshop Proceedings*, pages 119–130, 2020.

K. Belahcène, C. Labreuche, N. Maudet, V. Mousseau, and W. Ouerdane. A model for accountable ordinal sorting. In Carles Sierra, editor, *Proceedings of the Twenty-Sixth International Joint Conference on Artificial Intelligence, IJCAI 2017, Melbourne, Australia, August 19-25, 2017*, pages 814–820, 2017.

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

K. Belahcène, C. Labreuche, N. Maudet, V. Mousseau, and W. Ouerdane. Accountable approval sorting. In *Proceedings of the 27th International Joint Conference on Artificial Intelligence (IJCAI 2018)*, 2018.

D. Bouyssou and T. 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, 2007a.

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

D. Bouyssou and T. Marchant. Additive conjoint measurement with ordered categories. *European Journal of Operational Research*, 203(1):195–204, 2010.

D. Bouyssou and T. Marchant. On the relations between ELECTRE TRI-B and ELECTRE TRI-C and on a new variant of ELECTRE TRI-B. *European Journal of Operational Research*, 242 (1):201–211, 2015.

D. Bouyssou, T. Marchant, and M. Pirlot. A conjoint measurement approach to the discrete Sugeno integral. In S. Brams, W. V. Gehrlein, and F. S. Roberts, editors, *The Mathematics of Preference, Choice and Order. Essays in Honor of Peter C. Fishburn*, pages 85–109. Springer, 2009.

D. Bouyssou, T. Marchant, and M. Pirlot. A note on ELECTRE TRI-nB with few limiting profiles. *4OR*, June 2021.

D. Bouyssou, T. Marchant, and M. Pirlot. A theoretical look at Electre Tri-nB and related sorting models. *4OR*, February 2022.

F. Doshi-Velez, M. Kortz, R. Budish, Ch. Bavitz, S. Gershman, D. O'Brien, S. Schieber, J. Waldo, D. Weinberger, and A. Wood. Accountability of AI under the law: The role of explanation. *CoRR*, abs/1711.01134, 2017.

M. Doumpos and C. Zopounidis. *Multicriteria decision aid classification methods.* Kluwer Academic Publishers, Dordrecht Boston, 2002. ISBN 1402008058.

D. Dubois, J.-L. Marichal, H. Prade, M. Roubens, and R. Sabbadin. The use of the discrete Sugeno integral in decision making: A survey. *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems*, 9:539–561, 2001.

E. Fernández, J. R. Figueira, J. Navarro, and B. Roy. ELECTRE TRI-nB: A new multiple criteria ordinal classification method. *European Journal of Operational Research*, 263(1):214–224, 2017.

E. Fernández, J. R. Figueira, and J. Navarro. An interval extension of the outranking approach and its application to multiple-criteria ordinal classification. *Omega*, 84:189–198, 2019.

E. Fernández, J. R. Figueira, and J. Navarro. Interval-based extensions of two outranking methods for multi-criteria ordinal classification. *Omega*, 95:102065, 2020.

J. Fürnkranz and E. Hüllermeier. Preference learning: An introduction. In J. Fürnkranz and E. Hüllermeier, editors, *Preference Learning*, pages 1–17. Springer, 2010.

A. Ghahraman and D. Prior. A learning ladder toward efficiency: Proposing network-based stepwise benchmark selection. *Omega*, 63:83–93, 2016.

W. M. Goldstein. Decomposable threshold models. *Journal of Mathematical Psychology*, 35(1): 64–79, 1991.

M. Grabisch. Remarkable polyhedra related to set functions, games and capacities. *TOP*, 24: 301–326, 2016.

S. Greco, B. Matarazzo, and R. Słowiński. Conjoint measurement and rough set approach for multicriteria sorting problems in presence of ordinal criteria. In A. Colorni, M. Paruccini, and B. Roy, editors, *A-MCD-A, Aide Multicritère à la Décision / Multiple Criteria Decision Aid*, pages 117–144. European Commission, Joint Research Centre, Luxembourg, 2001.

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

S. Greco, V. Mousseau, and R. Słowiński. Robust ordinal regression for value functions handling interacting criteria. *European Journal of Operational Research*, 239(3):711–730, 2014.

D. Gunning and D. Aha. DARPA's Explainable Artificial Intelligence (XAI) Program. *AI Magazine*, 40(2):44–58, 2019.

M. Guo, X. Liao, and J. Liu. A progressive sorting approach for multiple criteria decision aiding in the presence of non-monotonic preferences. *Expert Systems and Applications*, 123:1–17, 2019.

E. Jacquet-Lagrèze. Binary preference indices: A new look on multicriteria aggregation procedures. *European Journal of Operational Research*, 10(1):25–32, 1982.

E. Jacquet-Lagrèze and Y. Siskos. Assessing a set of additive utility functions for multicriteria decision making: the UTA method. *European Journal of Operational Research*, 10:151–164, 1982.

P. Janssen and P. Nemery. An extension of the flowsort sorting method to deal with imprecision. *4OR*, 11(2):171–193, 2013.

M. Kadziński and K. Ciomek. Integrated framework for preference modeling and robustness analysis for outranking-based multiple criteria sorting with ELECTRE and PROMETHEE. *Information Sciences*, 352-353:167–187, 2016.

M. Kadziński, K. Ciomek, P. Rychly, and R. Słowiński. Post factum analysis for robust multiple criteria ranking and sorting. *Journal of Global Optimization*, 65(3):531–562, 2016.

M. Kadziński, M. Ghaderi, and M. Dąbrowski. Contingent preference disaggregation model for multiple criteria sorting problem. *European Journal of Operational Research*, 281(2):369–387, 2020.

M. Kadziński, M. Stamenković, and M. Uniejewski. Stepwise benchmarking for multiple criteria sorting. *Omega*, 108:102579, 2022.

M. Köksalan, V. Mousseau, Ö. Özpeynirci, and S. B. Özpeynirci. A new outranking-based approach for assigning alternatives to ordered classes. *Naval Research Logistics*, pages 74–85, 2009.

C. Labreuche. A general framework for explaining the results of a multi-attribute preference model. *Artificial Intelligence*, 175(7):1410–1448, 2011.

C. Labreuche and S. Fossier. Explaining multi-criteria decision aiding models with an extended shapley value. In *IJCAI International Joint Conference on Artificial Intelligence*, page 331 – 339, 2018.

C. Labreuche, N. Maudet, and W. Ouerdane. Minimal and complete explanations for critical multi-attribute decisions. In *Algorithmic Decision Theory*, pages 121–134, 2011.

C. Labreuche, N. Maudet, and W. Ouerdane. Justifying dominating options when preferential information is incomplete. In *Proceedings of the 20th European Conference on Artificial Intelligence*, ECAI'12, page 486–491. IOS Press, 2012.

J. Liu, M. Kadziński, X. Liao, X. Mao, and Y. Wang. A preference learning framework for multiple criteria sorting with diverse additive value models and valued assignment examples. *European Journal of Operational Research*, 286(3):963–985, 2020.

P. Meyer and A.-L. Olteanu. Handling imprecise and missing evaluations in multi-criteria majority-rule sorting. *Computers & Operations Research*, 110:135–147, 2019.

P. Minoungou, V. Mousseau, W. Ouerdane, and P. Scotton. A MIP-based approach to learn MR-Sort models with single-peaked preferences. *Annals of Operations Research*, 2022.

V. Mousseau, J. R. Figueira, L. C. Dias, C. Gomes da Silva, and J. Clímaco. Resolving inconsistencies among constraints on the parameters of an MCDA model. *European Journal of Operational Research*, 147(1):72–93, 2003.

V. Mousseau, L.C. Dias, and J. Figueira. Dealing with inconsistent judgments in multiple criteria sorting models. *4OR*, 4(3):145–158, 2006.

V. Mousseau, Ö. Özpeynirci, and S. Özpeynirci. Inverse multiple criteria sorting problem. *Annals of Operations Research*, 267(1-2):379–412, 2018.

R. Pelissari, M. C. Oliveira, A. J. Abackerli, S. Ben Amor, and M. R. Pontes Assumpção. Techniques to model uncertain input data of multi-criteria decision-making problems: a literature review. *International Transactions in Operational Research*, 28(2):523–559, 2021.

J. Pereira, E. C. B. de Oliveira, L. F. A. M. Gomes, and R. M. Araujo. Sorting retail locations in a large urban city by using ELECTRE TRI-C and trapezoidal fuzzy numbers. *Soft Computing*, 23(12):4193–4206, 2019a.

J. Pereira, E. C. B. de Oliveira, D. C. Morais, A. P. C. S. Costa, and P. A. López. ELECTRE TRI-C with hesitant outranking functions: Application to supplier development. *Journal of Intelligent Fuzzy Systems*, 37(6):7923–7933, 2019b.

M. Petrovic, N. Bojkovic, I. Anic, M. Stamenkovic, and S. Pejcic-Tarle. An ELECTRE-based decision aid tool for stepwise benchmarking: An application over EU Digital Agenda targets. *Decision Support Systems*, 59:230–241, 2014.

C. Rocha and L. C. Dias. An algorithm for ordinal sorting based on ELECTRE with categories defined by examples. *Journal of Global Optimization*, 42(2):255–277, 2008.

B. Roy and D. Bouyssou. *Aide multicritère à la décision: méthodes et cas.* Economica Paris, 1993.

R. Słowínski, S. Greco, and B Matarazzo. Axiomatization of utility, outranking and decision-rule preference models for multiple-criteria classification problems under partial inconsistency with the dominance principle. *Control and Cybernetics*, 31(4):1005–1035, 2002.

A. F. Tehrani and E. Hüllermeier. Ordinal Choquistic regression. In *Proceedings of the 8th conference of the European Society for Fuzzy Logic and Technology.* Atlantis Press, 2013.

A. F. Tehrani, W. Cheng, K. Dembczyński, and E. Hüllermeier. Learning monotone nonlinear models using the Choquet integral. *Machine Learning*, 89(1–2):183–211, 2012.

S. Wachter, B. Mittelstadt, and L. Floridi. Why a Right to Explanation of Automated Decision-Making Does Not Exist in the General Data Protection Regulation. *International Data Privacy Law*, 7(2):76–99, 2017.

P. P. Wakker. *Additive representations of preferences: A new foundation of decision analysis.* Kluwer, Dordrecht, 1989.

D. Walton. *Argumentation schemes for presumptive reasoning.* Mahwah, N. J.,Erlbaum, 1996.

Y. Wei. *Aide multicritère à la décision dans le cadre de la problématique du tri : concepts, méthodes et applications.* Thèse de doctorat, Université Paris Dauphine, Paris, France, 1992. (in French).

Q. Zhong, X. Fan, X. Luo, and F. Toni. An explainable multi-attribute decision model based on argumentation. *Expert Systems with Application*, 117:42–61, 2019.

Appendix A List of abbreviations

For the reader's convenience, we list below, in alphabetic order, the acronyms used in the text, except for acronyms of sorting methods.

- AVF: Additive Value Function
- CAI: Class Acceptability Index
- DM: Decision Maker
- DRSA: Dominance based Rough Sets Approach
- LP: Linear Program
- MCDM/A: Multiple Criteria Decision Making / Aiding
- MCS: Multiple Criteria Sorting
- MILP: Mixed Integer Linear Program
- ML: Machine Learning
- MOP: Monotone Ordered Partition
- PL: Preference learning
- ROR: Robust Ordinal Regression
- SMAA: Stochatic Multicriteria Acceptability Analysis
- VF: Value Function