How to Handle Missing Values in Multi-Criteria Decision Aiding?

Christophe Labreuche, Sébastien Destercke

How to Handle Missing Values in Multi-Criteria Decision Aiding?

Christophe Labreuche[1] , **Sébastien Destercke**[2]
[1]Thales Research & Technology, Palaiseau, France
[2]UMR CNRS 7253 Heudiasyc, Sorbonne universités,
Université de Technologie de Compiègne, Compiègne, France
christophe.labreuche@thalesgroup.com, sebastien.destercke@hds.utc.fr

Abstract

It is often the case in the applications of Multi-Criteria Decision Making that the values of alternatives are unknown on some attributes. An interesting situation arises when the attributes having missing values are actually not relevant and shall thus be removed from the model. Given a model that has been elicited on the complete set of attributes, we are looking thus for a way – called restriction operator – to automatically remove the missing attributes from this model. Axiomatic characterizations are proposed for three classes of models. For general quantitative models, the restriction operator is characterized by linearity, recursivity and decomposition on variables. The second class is the set of monotone quantitative models satisfying normalization conditions. The linearity axiom is changed to fit with these conditions. Adding recursivity and symmetry, the restriction operator takes the form of a normalized average. For the last class of models – namely the Choquet integral, we obtain a simpler expression. Finally, a very intuitive interpretation is provided for this last model.

1 Introduction

Many decision problems involve multiple and conflicting criteria. The aim of Multi-Criteria Decision Aiding (MCDA) is to formalize such decision situations by constructing a model representing the preferences of a decision maker and then exploiting this model on a set of alternatives in order to make some recommendation (e.g. find the most preferred option).

In applications, alternatives are often imperfect in the sense that their values on some decision attributes might be imprecise or missing. When these values are represented by probability distributions, one can derive a probability distribution over the possible recommendations [Durbach and Stewart, 2012]. Another possibility is to combine MCDA and Decision Under Uncertainty techniques [Dominiak, 2006; Gaspars-Wieloch, 2015]. We focus in this work in the second type of imperfectness where the values of alternatives are missing for some attributes. In the context of MCDA, there are basically two possible interpretations: In the first one, a missing value is a relevant criterion that is important to make the decision. Hence the missing value is seen as an extremely imprecise evaluation. Such uncertainty is propagated in the MCDA model and one can analyze whether it is possible to discriminate among the alternatives despite this uncertainty [Lahdelma *et al.*, 1998]. In the second interpretation, the missing value is considered as an irrelevant criterion, i.e. the criterion becomes irrelevant for this alternative when its value is missing. The following example illustrates this situation.

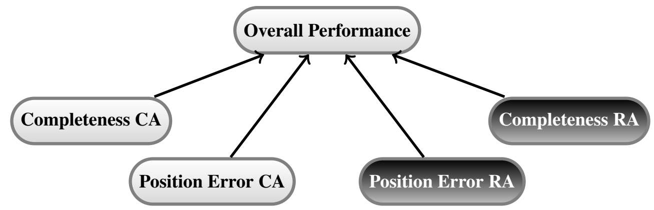

Figure 1: Description of the MCDA model. The overall performance aggregates the two performance criteria for each of the two classes. There is one value of position error and one value of completeness gathering all the trajectories in a given class. For instance, “Position Error RA” is the position RMS error among all trajectories of RAs.

1.1 Motivating Example

Example 1 *Consider the problem of designing an airspace supervision system for Air Traffic Management (ATM). One needs to select the “best” system among several candidate solutions, based on the assessment of their quality [ESASSP, 2012]. In order to make this choice, the decision maker simulates each candidate solution on several scenarios representative of real ATM traffic. The aim of the supervision system is to estimate the trajectories of aircrafts thanks to tracking algorithms. Two elementary attributes are considered to measure the quality of the supervision system: “position error” (Root Mean Square [RMS] error between the estimated trajectories and the real ones), and “completeness” (percentage of the trajectories for which aircrafts are detected and tracked). Attributes position error and completeness are computed separately for the existing classes of aircrafts: CA (Commercial Airplanes) and RA (Recreational airplanes). An MCDA model is then obtained, as shown in Figure 1.*

Each scenario depicts a different situation that might occur, and involves a predefined set of trajectories. There are, in particular, scenarios in which there is no aircraft in a given *category – e.g. no RA. In this case, this means that there is no evaluation of the two attributes "Position Error RA" and "Completeness RA".* ■

In the previous example, two values of the attributes are missing. In order to assess the candidate solutions on the scenario without RA, one needs a new MCDA model having as attributes only the ones that are evaluated. The corresponding tree appears in Figure 1 in light gray (the nodes in dark gray are removed). It is not realistic to ask the decision maker to calibrate this new model, because of time constraints (the decision maker already spent some significant time to calibrate the full model; one does not wish to spent more time calibrating sub-models), of feasibility (the set of possible missing values might be combinatorial, yielding a combinatorial set of potential sub-models), and of consistency with the original model (there might be unwished important gaps between the original model and the learned restricted one).

1.2 Related Works

In the context of data fusion, missing values can be handled in different ways [Saar-Tsechansky and Provost, 2007; Garca-Laencina *et al.*, 2010]: (1) they can be merely ignored in the fusion process; (2) the missing value can be replaced by a default value, which can be for instance the average value of this variable over a dataset; (3) the fusion mechanism can implicitly take into account missing elements.

In statistics, there are basically three types of missing data: the probability of having a missing value for a variable can be completely random (Missing Completely At Random – MCAR), can be independent of the variable value but dependent on the value of the other variables (Missing At Random – MAR), or can be dependent on the values of all variables (Missing not at Random – NMAR) [Little and Rubin, 2002].

Handling missing data is an important challenge in machine learning [Goodfellow *et al.*, 2016]. The simplest approach consists in completing the missing data by a single value (e.g. the average value over the dataset) [Goodfellow *et al.*, 2016; Lundberg and Lee, 2017]. Other methods fill the gaps with (estimated) probability density functions [Smieja *et al.*, 2018], variational autoencoders [McCoy *et al.*, 2018] or modified generative adversarial network [Yeh *et al.*, 2017].

In Social Choice, when the preferences of the voters are incomplete, a safe approach is to determine whether a candidate is the winner for every possible completion of the preferences – called the necessary winner [Konczak and Lang, 2005]. This idea of robust recommendation is also used in multi-criteria decision making [Boutilier *et al.*, 2004; Wilson, 2014]. If this entailment is empty, one can then compute the proportion of completions which yields each possible outcome [Lahdelma *et al.*, 1998].

The existing works aim at filling the missing values in one way or another. To our knowledge, there is no work in which the missing data is considered as non relevant and shall be removed from the model.

1.3 Contributions

We assume we are given a preference model that has been elicited on the complete set of attributes. Our aim is to automatically construct a sub-model restricted to a subset of attributes, by removing the missing attributes from the original model. Such process is called "restriction operator".

This situation arises in the design of complex system such as in Ex. 1, in which attributes are computed by simulation on operational scenarios. Another application is online monitoring of an industrial system through several Key Performance Indicators (KPIs) measuring the instantaneous performance of the system or its performance over a small interval of time. KPIs are often based on events that occur or not during the time window. A criterion becomes irrelevant if the KPI is a performance attached to events that did not occur.

Section 2 gives the basic definitions. We restrict ourselves to utility models. Three classes of preference models are considered: general quantitative models, monotone quantitative models satisfying normalization conditions, and lastly a versatile class of monotone quantitative models – namely the Choquet integral. This model has many applications in AI, for instance in combinatorial optimization [Galand *et al.*, 2013], incremental elicitation [Benabbou *et al.*, 2017], preference learning [Tehrani *et al.*, 2012], machine leaning with SVM [Li *et al.*, 2015].

The concept of restriction operator is introduced in Section 3. Its expression shall take the form of some kind of combination of the values of the full model for some special alternatives. It is not easy to provide a satisfactory expression of the restriction operator. Hence we start by defining properties that are wished and then we derive the expression of this operator by an axiomatic approach. Axiomatic characterizations are proposed for the three classes of models – see Section 4, 5 and 6 respectively. Different expressions are obtained for these three classes. This first one is linear in the original model and is an average of the model over the missing values. The other two take the form of a normalized average due to the normalization conditions. Moreover the expression of the restriction operator for the last two models differ with regard to the values over which the average is computed.

2 Notation

We consider an MCDA problem with a set $N = \{1, \ldots, n\}$ of attributes. We assume a finite set of values Y_i on each attribute $i \in N$. For $B \subseteq N$, we set $Y_B := \prod_{i \in B} Y_i$. We assume that the preference model is represented by a quantitative utility over Y_B. This covers many widely-used models. For model U defined on B, $x, y \in Y_B$ and $A \subseteq B$, we use the notation $U(x_A, y_{-A})$ to denote the value of U on the compound alternative taking value x on A and value y on $B \setminus A$.

The rest of this section is devoted to describing the three classes of models going from the more general one to the less general one.

2.1 General Class of Models

The first class is the family of quantitative preference models. For $B \subseteq N$, let $\mathcal{U}_{\mathrm{G}}(B)$ (where G stands for General) be the set of functions $U : Y_B \to \mathbb{R}$.

We define, for $i \in N$ and $y_i \in Y_i$, the delta function by $\delta_{y_i}(x_i) = 1$ if $x_i = y_i$, and $\delta_{y_i}(x_i) = 0$ otherwise.

The Generalized Additive Independence (GAI) model is a very general class of models in class $\mathcal{U}_{\mathrm{G}}(B)$ that is only supposed to fulfill additivity across subsets of B [Fishburn, 1967; Bacchus and Grove, 1995]. It takes the form

$$U(x) = \sum_{S \in \mathcal{S}} u_S(x_S), \tag{1}$$

where $\mathcal{S}$ is a collection of subsets of B and $u_S : Y_S \to \mathbb{R}$. Note that the additive utility model is a particular case where $\mathcal{S}$ contains only singletons [Fishburn, 1970].

Due to the presence of sums in (1), the utility functions shall correspond to scales of difference [Krantz *et al.*, 1971], that is they are given up to an affine transformation. We say that two functions f and g on B are two "equivalent scales" (of difference) – denoted by $f \propto_B g$ – if there exists coefficients c_1, c_0 such that $f(x_B) = c_1\, g(x_B) + c_0$.

The following example (taken from [Boutilier *et al.*, 2004]) illustrates the GAI model.

Example 2 *The problem is to choose a menu composed of a main dish (attribute 1 with two values {Fish,Meat}) and some wine (attribute 2 with two values {White,Red}). Consider the following GAI model* $U(x_1, x_2) = u_1(x_1) + u_{1,2}(x_1, x_2)$ *with* $u_1(\text{Fish}) = 2$, $u_1(\text{Meat}) = 0$, $u_{1,2}(\text{Fish,White}) = 1$, $u_{1,2}(\text{Fish,Red}) = 0$, $u_{1,2}(\text{Meat,White}) = 0$, $u_{1,2}(\text{Meat,Red}) = 1$. *According to this model, Fish is preferred to Meat, White wine is preferred to Red wine for Fish, and Red wine is preferred to White wine for Meat.*

2.2 Monotone Models

In Ex. 2, the preferences over the elements of attribute "wine" are conditional on the value taken on attribute "main dish". Yet we do not encounter such conditional preferences in most of MCDA problems. In this case, there exists a weak order $\succsim_i$ on each Y_i, where $x_i \succsim_i y_i$ means that $x_i \in Y_i$ is at least as good as $y_i \in Y_i$ regardless of the values on the other attributes. In Ex. 1, we have such unconditional preferences over the attributes: the smaller the position error the better; the larger the completeness the better. The least preferred and most preferred elements in Y_i according to $\succsim_i$ are denoted by $\bot_i$ and $\top_i$ respectively. We also set $\overline{Y}_i = \{\bot_i, \top_i\}$, and $\succ_i$ is the asymmetric part of $\succsim_i$. Utility U is said to be "monotone" if $U(x) \geq U(y)$ whenever $x_i \succsim_i y_i$ for all $i \in N$.

The process of removing some criteria A from U makes sense only if the remaining criteria $B \setminus A$ have some influence on U. An attribute $i \in B$ is said to be "influential" if there exists $x_i, y_i \in Y_i$ and $z_{-i} \in Y_{B\setminus\{i\}}$ such that $U(x_i, z_{-i}) \neq U(y_i, z_{-i})$. As U is monotone in its coordinates, we obtain:

$$\forall i \in N\ \exists z_{-i} \in Y_{B\setminus\{i\}}, \quad U(\top_i, z_{-i}) > U(\bot_i, z_{-i}). \tag{2}$$

We can restrict the previous relation to $z_{-i} \in \overline{Y}_{B\setminus\{i\}}$:

$$\forall i \in N\ \exists z_{-i} \in \overline{Y}_{B\setminus\{i\}}, \quad U(\top_i, z_{-i}) > U(\bot_i, z_{-i}). \tag{3}$$

In general, (2) does not imply (3). We will only assume (3) in this paper.

We can assume w.l.o.g. that U takes values in $[0, 1]$. An alternative taking value $\bot_i$ (resp. $\top_i$) on all attributes is naturally assigned to the worst (resp. best) possible evaluation 0 (resp. 1). These are two normalization conditions on U.

Definition 1 $\mathcal{U}_{\mathrm{M}}(B)$ *is the set of monotone* $U \in \mathcal{U}_{\mathrm{G}}(B)$ *fulfilling (3),* $U(\bot_B) = 0$ *and* $U(\top_B) = 1$.

Example 3 *Consider two attributes,* $Y_1 = \{a, b, c\}$ *(with* $\top_1 = c \succ_1 b \succ_1 a = \bot_1$*) and* $Y_2 = \{a, b, c, d\}$ *(with* $\top_2 = d \succ_2 c \succ_2 b \succ_2 a = \bot_2$*) and the normalized monotone utility function* $U : Y_1 \times Y_2 \to \mathbb{R}$ *defined by*

U	a	b	c	d
a	0	0	0	$\frac{1}{2}$
b	0	$\frac{1}{2}$	$\frac{1}{2}$	$\frac{1}{2}$
c	0	$\frac{1}{2}$	1	1

2.3 Idempotent Models

We consider the special case, where all attributes are evaluated on the same scale, i.e. $Y_1 = \cdots = Y_n =: Y$. The values of Y can be interpreted as reference satisfaction levels that have the same interpretation across the criteria. For instance, if value $a \in Y$ is interpreted as a bad evaluation, then alternative $(a, a, \ldots, a)$ is bad on all criteria and thus deserves the bad evaluation a as overall evaluation. More formally, aggregation function $U \in \mathcal{U}_{\mathrm{G}}(B)$ is said to be "idempotent" if

$$U(a, \ldots, a) = a \qquad \forall a \in Y. \tag{4}$$

As U takes values in $[0, 1]$, we shall have $Y \subseteq [0, 1]$. Not so many aggregation models satisfy this property. The most well-known and versatile class of idempotent aggregation functions is the Choquet integral [Choquet, 1953]. It is a numerical function that applies to numerical values. We restrict ourselves here to the Choquet integral. We assume that Y takes numerical values inside interval $[0, 1]$: $Y = \{a_0, a_1, \ldots, a_p\}$ with $0 = a_0 < a_1 < \cdots < a_p = 1$, with $\succ_i = >$. The extreme points of Y are $\overline{Y} = \{0, 1\}$.

The Choquet integral is a versatile model able to represent interaction among criteria [Grabisch *et al.*, 2000; Grabisch and Labreuche, 2010]. It is characterized by a "capacity" [Choquet, 1953] which is a set function $\mu : 2^B \to [0, 1]$ such that $\mu(\emptyset) = 0$, $\mu(B) = 1$ (Normalization conditions) and $\mu(A) \leq \mu(A')$ for all $A \subseteq A' \subseteq B$ (Monotonicity conditions). Criterion $i \in B$ is said to be "non-degenerate" in μ if there exists $A \subseteq B \setminus \{i\}$ such that $\mu(A \cup \{i\}) > \mu(A)$.

The Choquet integral of $x \in [0, 1]^B$ w.r.t. a capacity μ has the following expression [Choquet, 1953] :

$$C_\mu(x) = \sum_{j=1}^{b} x_{\tau(j)} [\mu(S_\tau^B(j)) - \mu(S_\tau^B(j+1))], \tag{5}$$

where $\tau : \{1, \ldots, b\} \to B$ is a permutation s.t. $x_{\tau(1)} \leq \cdots \leq x_{\tau(b)}$, $b = |B|$, $S_\tau^B(j) = \{\tau(j), \ldots, \tau(b)\}$, $S_\tau^B(b+1) = \emptyset$.

Definition 2 $\mathcal{U}_{\mathrm{C}}(B)$ *is the set of Choquet integrals on* B *w.r.t. capacities that are non-degenerate in all criteria.*

The Choquet integral satisfies (4). One can readily see that for the Choquet integral, relations (2) and (3) are equivalent and hold iff all criteria are non-degenerate. Hence, due to the monotonicity conditions on capacities, $\mathcal{U}_{\mathrm{C}}(B) \subseteq \mathcal{U}_{\mathrm{M}}(B)$.

Model (5) contains many interesting sub-models as particular cases. One of them is the Choquet integral w.r.t. a 2-additive capacity. It takes the following expression:

$$C_\mu^2(x) = \sum_{i \in B} v_i\, x_i - \sum_{\{i,j\} \subseteq B} I_{i,j} \frac{|x_i - x_j|}{2}. \tag{6}$$

The first part is a weighted sum, where v_i represents the mean importance of criterion i. Interaction between criteria comes from the second part. Coefficient $I_{i,j}$ is the intensity of interaction between criteria i and j, which belongs to $[-1, 1]$. If $I_{i,j} > 0$ (resp. $I_{i,j} < 0$), the interaction term penalizes (resp. increases) the overall assessment C^2_μ, proportionally to the difference of score between criteria i and j.

Example 4 (Ex. 1 cont.) *The elicited model is:*

$$U(x) = \frac{1}{6}\Big(2x^{\mathrm{CA}}_{\mathrm{C}} + 2x^{\mathrm{RA}}_{\mathrm{C}} + x^{\mathrm{CA}}_{\mathrm{PE}} + x^{\mathrm{RA}}_{\mathrm{PE}} - \frac{|x^{\mathrm{CA}}_{\mathrm{PE}} - x^{\mathrm{RA}}_{\mathrm{PE}}|}{2} - \frac{|x^{\mathrm{CA}}_{\mathrm{C}} - x^{\mathrm{RA}}_{\mathrm{C}}|}{2} - \frac{|x^{\mathrm{CA}}_{\mathrm{PE}} - x^{\mathrm{CA}}_{\mathrm{C}}|}{2} - \frac{|x^{\mathrm{RA}}_{\mathrm{PA}} - x^{\mathrm{RA}}_{\mathrm{C}}|}{2}\Big),$$

where x^j_i, $i \in \{\mathrm{PE}, \mathrm{C}\}$ (for Position Error and Completeness resp.) and $j \in \{\mathrm{CA}, \mathrm{RA}\}$ denote the values of the attributes. Completeness is more important than Position Error as one needs first to track the aircrafts as much as possible, before caring about their position accuracy. Moreover there are positive interactions giving bonus when Completeness and Position Error are simultaneously satisfied, or when the system is good at both CA and RA.

3 Problem Statement

Consider model $U \in \mathcal{U}_{\mathrm{T}}(N)$ where $\mathrm{T} \in \{\mathrm{G}, \mathrm{M}, \mathrm{C}\}$, and alternative x which values are unknown on the attributes $A \subset N$. These values are called "missing". Several approaches can be used to assess the alternative given only x_{-A}. One can model uncertainty on the missing information by the uniform probability distribution over the possible completions of x_{-A}, using the maximum entropy principle. The evaluation of x_{-A} is then obtained by the following expected utility:

$$\frac{1}{\prod_{i\in A}|Y_i|}\sum_{y_A\in Y_A} U(y_A, x_{-A}). \tag{7}$$

We explore here another venue. Given a model elicited on the full set of attributes, we define a restriction operator returning a preference model defined on a subset of attributes. As the process of removing some attributes can be applied several times, the support of the initial model can be a strict subset B of N, and the restricted function shall belong to the same class of models as the initial function. The restriction operator has thus a closed form (with $A \subset B \subseteq N$)

$$[\cdot]^{\mathrm{T},B}_{-A} : \mathcal{U}_{\mathrm{T}}(B) \to \mathcal{U}_{\mathrm{T}}(B \setminus A). \tag{8}$$

The motivation of using a restriction operator rather than directly a formula such as (7) is twofold. Firstly, it explicitly exhibits a model on the set of attributes with known values. This model can be shown and explained to the user, which is helpful in practice. Secondly, we are interested in the case where all attributes are intrinsically relevant in model U, but an attribute becomes "irrelevant" when its value is unknown (missing). Hence it makes sense to remove the missing attributes and restrict the elicited model.

The rest of this paper is devoted to defining $[U]^{\mathrm{T},B}_{-A}(x_{B\setminus A})$ for $x_{B\setminus A} \in Y_{B\setminus A}$. We adopt an axiomatic approach: instead of starting from an expression (such as (7)), we propose axioms on $[\cdot]^{\mathrm{T},B}_{-A}$ that are important for an end-user, and then we show that there is a unique restriction operator fulfilling these properties. This approach is applied for the three classes of models previously described.

4 Non-Monotone Preference Models

We consider a general model $U \in \mathcal{U}_{\mathrm{G}}(B)$. We first introduce the axioms on $[\cdot]^{\mathrm{G},B}_{-A}$ and then give the axiomatic result.

4.1 Axioms and Axiomatic Result

Linearity

Let us restrict ourselves to the case where A is a singleton $A = \{i\}$. The models we consider satisfy some kind of additivity, like for the GAI model (1). If U is a linear combination of two simpler models U' and U'', then $[U]^{\mathrm{G},B}_{-i}$ shall be a linear combination of $[U']^{\mathrm{G},B}_{-i}$ and $[U'']^{\mathrm{G},B}_{-i}$. Standard linearity of $[\cdot]^{\mathrm{G},B}_{-i}$ means that $[\beta'\, U' + \beta''\, U'']^{\mathrm{G},B}_{-i} = \beta'\, [U']^{\mathrm{G},B}_{-i} + \beta''\, [U'']^{\mathrm{G},B}_{-i}$ where $\beta', \beta'' \in \mathrm{I\!R}$ are constant. Here in order to put β', β'' outside operator $[\cdot]^{\mathrm{G},B}_{-i}$, it is sufficient that they are constant in i. Hence β', β'' can depend on attributes in $B \setminus \{i\}$.

Axiom Linearity (L): *Let $B \subseteq N$, $i \in B$ with $B \neq \{i\}$, $\beta', \beta'' \in \mathcal{U}_{\mathrm{G}}(B \setminus \{i\})$, and $U', U'' \in \mathcal{U}_{\mathrm{G}}(B)$. Then*

$$[\beta'\, U' + \beta''\, U'']^{\mathrm{G},B}_{-i} = \beta'\, [U']^{\mathrm{G},B}_{-i} + \beta''\, [U'']^{\mathrm{G},B}_{-i}.$$

Most of MCDA models fulfill some linearity. This ensures a kind of understandability by humans. If the restriction operator were completely non-linear we would obtain a complex and not easy to interpret restricted function.

Proposition 1 *Under* **L***, there exists coefficients $\{\gamma^B_i(y_i)\}_{y_i\in Y_i}$ such that for all $x_{-i} \in Y_{B\setminus\{i\}}$*

$$[U]^{\mathrm{G},B}_{-i}(x_{-i}) = \sum_{y_i\in Y_i} \gamma^B_i(y_i)\, U(y_i, x_{-i}). \tag{9}$$

Proof : By **L**, as δ is a base of U, $[U]^{\mathrm{G},B}_{-i}(x_{-i}) = \sum_{y_i\in Y_i} \gamma^B_i(y_i)\, U(y_i, x_{-i})$ where $\gamma^B_i(y_i) := [\delta_{y_i}]^{\mathrm{G},B}_{-i}$. ■

Constant

If U is constant, then removing attribute i does not change the function.

Axiom Constant (C): *Let $B \subseteq N$ and $i \in B$ with $B \neq \{i\}$. If $U(x) = C$ for all $x \in Y_B$, with $C \in \mathrm{I\!R}$, then $[U]^{\mathrm{G},B}_{-i}(x_{-i}) = C$.*

Proposition 2 *Under* **L** *and* **C***, there exists coefficients $\{\gamma^B_i(y_i)\}_{y_i\in Y_i}$ such that (9) holds, with*

$$\sum_{y_i\in Y_i} \gamma^B_i(y_i) = 1. \tag{10}$$

For space limitation, the proofs of this result and others are omitted.

When only one variable needs to be removed, the expression of the restriction operator (see Prop. 2) is similar to a general probabilistic approach computing the expected utility over the unknown values with a non-uniform distribution.

Single Value Contribution

Assume that $U \in \mathcal{U}_{\mathrm{G}}(B)$ takes the form (for $i \in B$ and $y_i \in Y_i$)

$$\forall x \in Y_B \qquad U(x) = \delta_{y_i}(x_i) \tag{11}$$

Then the result of removing attribute i from U shall not depend on value y_i, as stated by the following axiom.

Axiom Single Value Contribution (SVC): *Let $B \subseteq N$, $i \in B$ with $B \neq \{i\}$. If U satisfies (11), then $[U]_{-i}^{\mathrm{G},B}$ does not depend on y_i.*

Proposition 3 *Under* **L**, **C** *and* **SVC**, *then*

$$[U]_{-i}^{\mathrm{G},B}(x_{-i}) = \frac{1}{|Y_i|} \sum_{y_i \in Y_i} U(y_i, x_{-i}).$$

The three axioms uniquely specify the restriction w.r.t. one variable, and $[U]_{-i}^{\mathrm{G},B}$ is an average of U over the missing values. We now consider the restriction w.r.t. more variables.

Recursivity

When we need to remove more than one attribute, we can proceed in several ways: remove all of them at the same time, or remove one at a time (with all possible ordering of the variables). There is no reason why these different ways shall provide different outcomes. This idea is expressed in the following axiom defined for $\mathrm{T} \in \{\mathrm{G}, \mathrm{M}, \mathrm{C}\}$.

Axiom Recursivity (R[T]): *For all $A \subset B$, $i \in B \setminus A$ with $A \cup \{i\} \neq B$, and all $U \in \mathcal{U}_{\mathrm{T}}(B)$, we have*

$$[U]_{-A\cup\{i\}}^{\mathrm{G},B}(x_{B\setminus(A\cup\{i\})}) = \Big[[U]_{-A}^{\mathrm{G},B}\Big]_{-i}^{\mathrm{G},B\setminus A}(x_{B\setminus(A\cup\{i\})}).$$

Note that this axiom will be used for other assumptions on U, and is thus indexed by $\mathrm{T} \in \{\mathrm{G}, \mathrm{M}, \mathrm{C}\}$.

Theorem 1 $[\cdot]_{-\cdot}^{\mathrm{G},\cdot}$ *satisfies* **L**, **C**, **SVC** *and* **R**[G], *if and only if for all $B \subseteq N$, $A \subset B$ and all $x_{-A} \in Y_{B\setminus A}$,*

$$[U]_{-A}^{\mathrm{G},B}(x_{-A}) = \frac{1}{\prod_{i\in A}|Y_i|} \sum_{y_A \in Y_A} U(y_A, x_{-A}). \tag{12}$$

This result shows that there is a unique restriction operator fulfilling the four axioms. Interestingly, we recover exactly expression (7) in which the uncertainty on the missing information is modeled by the uniform probability distribution over the possible completions of x_{-A}.

Sketch of the proof: One can easily check that (12) satisfies to all axioms.

Conversely consider $[\cdot]_{-\cdot}^{\mathrm{G},\cdot}$ satisfying all axioms. One can easily show relation (12) by induction on $|A|$, thanks to Proposition 3 and **R**[G]. ■

Example 5 (Ex. 2 cont.) *We obtain* $[U]_{-1}^{\mathrm{G},N}(\mathrm{White}) = \frac{1}{2}\big[U(\mathrm{Fish},\mathrm{White}) + U(\mathrm{Meat},\mathrm{White})\big] = \frac{3}{2}$, $[U]_{-1}^{\mathrm{G},N}(\mathrm{Red}) = \frac{1}{2}\big[U(\mathrm{Fish},\mathrm{Red}) + U(\mathrm{Meat},\mathrm{Red})\big] = \frac{3}{2}$, $[U]_{-2}^{\mathrm{G},N}(\mathrm{Fish}) = \frac{1}{2}\big[U(\mathrm{Fish},\mathrm{White}) + U(\mathrm{Fish},\mathrm{Red})\big] = \frac{5}{2}$, $[U]_{-2}^{\mathrm{G},N}(\mathrm{Meat}) = \frac{1}{2}\big[U(\mathrm{Meat},\mathrm{White}) + U(\mathrm{Meat},\mathrm{Red})\big] = \frac{1}{2}$. *So there is no intrinsic preference of a particular type of wine, which is consistent with $u_{1,2}$. Moreover, "fish" is intrinsically preferred to "meat", as expected given u_1.*

4.2 Justification of the Axioms

In order to see the importance of the axioms, we provide examples of restriction operators violating one axiom and fulfilling the other ones. Note that the existence of such examples also show the independence of the axioms.

Example 6 *Let us define*

$$[U]_{-A}^{\mathrm{G},B}(x_{-A}) = \min_{y_A \in Y_A} U(y_A, x_{-A}). \tag{13}$$

This operator satisfies **C**, **SVC** *and* **R**[G], *but violates* **L**. *This expression is very simple. It would be a very relevant expression if the aggregation operators were purely ordinal like a weighted min or max operator. However, we are interested in models U having some linearity property. Yet any linearity in U is lost in (13), which can make its expression more complex to understand for the user.*

Example 7 *Consider*

$$[U]_{-A}^{\mathrm{G},B}(x_{-A}) = \sum_{y_A \in Y_A} U(y_A, x_{-A}). \tag{14}$$

This operator satisfies **L**, **SVC** *and* **R**[G], *but violates* **C**. *This relation is not normalized so that $[U]_{-A}^{\mathrm{G},B}$ in (14) is not of the same order of magnitude as for U.*

Example 8 *Consider*

$$[U]_{-A}^{\mathrm{G},B}(x_{-A}) = \sum_{y_A \in Y_A} \left(\prod_{i \in A} \gamma_i^B(y_i) \right) U(y_A, x_{-A}),$$

where $\gamma_i^B(y_i) = \frac{2}{|Y_i|+2}$ if $y_i \in \overline{Y}_i$, and $\gamma_i^B(y_i) = \frac{1}{|Y_i|+2}$ else. This operator satisfies **L**, **C** *and* **R**[G], *but violates* **SVC**. *This expression does not treat in a similar way all elements in Y_i.*

Example 9 *Let us define* $[U]_{-A}^{\mathrm{G},B}(x_{-A}) = \frac{1}{|Y_i|}\sum_{y_i\in Y_i} U(y_i, x_{-i})$ *if $\exists i \in N$ s.t. $A = \{i\}$, and $[U]_{-A}^{\mathrm{G},B}(x_{-A}) = 0$ else. This operator satisfies* **L**, **C** *and* **SVC**, *but violates* **R**[G]. *This expression is not intuitive as what is done for subsets of at least two elements has nothing to do with what is done for singletons.*

5 Monotone Preference Models

We now consider monotone models $U \in \mathcal{U}_{\mathrm{M}}(B)$.

Quasi-Linearity

Axiom **L** cannot be strictly speaking satisfied for functions in $\mathcal{U}_{\mathrm{M}}(B)$, as $[\beta'\, U' + \beta''\, U'']_{-i}^{\mathrm{M},B}$ shall be a monotone and normalized aggregation function, which is clearly not the case of $\beta'\, [U']_{-i}^{\mathrm{M},B} + \beta''\, [U'']_{-i}^{\mathrm{M},B}$ when $\beta' + \beta'' \neq 1$ or $\beta', \beta'' < 0$. More precisely, $[\beta'\, U' + \beta''\, U'']_{-i}^{\mathrm{M},B}$ and $\beta'\, [U']_{-i}^{\mathrm{M},B} + \beta''[U'']_{-i}^{\mathrm{M},B}$ shall correspond to two equivalent interval scales. Hence one shall have the next axiom defined for $\mathrm{T} \in \{\mathrm{M}, \mathrm{C}\}$.

Axiom Quasi-Linearity (QL[T]): *For $B \subseteq N$, $i \in B$ with $B \neq \{i\}$, $U', U'' \in \mathcal{U}_{\mathrm{T}}(B)$, and $\beta', \beta'' \in \mathcal{U}_{\mathrm{G}}(B \setminus \{i\})$ with $\beta'(t), \beta''(t) \geq 0\ \forall t \in Y_{B\setminus\{i\}}$,*

$$\left([\beta'\, U' + \beta''\, U'']_{-i}^{\mathrm{T},B}\right) \propto_{B\setminus\{i\}} \left(\beta'\, [U']_{-i}^{\mathrm{T},B} + \beta''\, [U'']_{-i}^{\mathrm{T},B}\right).$$

Proposition 4 *Under* **QL**[T]*, for $B \subseteq N$ and $i \in B$ with $B \neq \{i\}$, there exists coefficients $\{\gamma_i^B(y_i)\}_{y_i \in Y_i}$ s.t. $[U]^{\mathrm{T},B}_{-i}(x_{-i}) = \frac{G_i^B(x_{-i}) - G_i^B(\perp_{-i})}{G_i^B(\top_{-i}) - G_i^B(\perp_{-i})}$ for all $x_{-i} \in Y_{B\setminus\{i\}}$, where $G_i^B(x_{-i}) = \sum_{y_i \in Y_i} \gamma_i^B(y_i)\, U(y_i, x_{-i})$.*

Sketch of the proof: The proof uses the fact that if $f \propto_B g$ (with $f \in \mathcal{U}_\mathrm{T}(B)$, $g \in \mathcal{U}_\mathrm{G}(B)$), then for all $x_B \in Y_B$, $f(x_B) = \frac{g(x_B) - g(\perp_B)}{g(\top_B) - g(\perp_B)}$. ■

Axioms **C** and **SVC** do not apply to $\mathcal{U}_\mathrm{M}(B)$. However, assume that U takes the form: $U(x) = U_i(x_i) \times U_{-i}(x_{-i})$, with $U_i \in \mathcal{U}_\mathrm{M}(Y_i)$ and $U_{-i} \in \mathcal{U}_\mathrm{M}(Y_{-i})$. By Proposition 4, $G_i^B(x_{-i}) = \left[\sum_{y_i \in Y_i} \gamma_i^B(y_i)\, U_i(y_i)\right] U_{-i}(x_{-i})$. If the term in bracket is non null, then $[U]^{\mathrm{M},B}_{-i}(x_{-i}) = \frac{U_{-i}(x_{-i}) - U_{-i}(\perp_{-i})}{U_{-i}(\top_{-i}) - U_{-i}(\perp_{-i})} = U_{-i}(x_{-i})$ as U_{-i} is normalised.

Symmetry

By Prop. 4, the restriction operator on i is a normalized average over terms $U(y_i, x_{-i}) - U(y_i, \perp_{-i})$. The next axiom is a symmetry property saying that $[U]^{\mathrm{T},B}_{-i}$ shall not be modified if we permute these differences over the elements of Y_i. We can also multiply these differences by a constant, without any consequence on the restriction.

Axiom Symmetry (S): *Let $B \subseteq N$, $i \in B$ with $B \neq \{i\}$, a permutation π on the values of Y_i, $U, U' \in \mathcal{U}_\mathrm{M}(B)$ and $C > 0$ s.t. for all $y_i \in Y_i$ and $x_{-i} \in Y_{B\setminus\{i\}}$*

$$\begin{aligned} &U(y_i, x_{-i}) - U(y_i, \perp_{-i}) \\ &= C\,(U'(\pi(y_i), x_{-i}) - U'(\pi(y_i), \perp_{-i})). \end{aligned} \tag{15}$$

Then $[U]^{\mathrm{T},B}_{-i}(x_{-i}) = [U']^{\mathrm{T},B}_{-i}(x_{-i})$.

Without the presence of constant C, there does not exists in general $U, U' \in \mathcal{U}_\mathrm{M}(B)$ satisfying (15), and thus axiom **S** would be void.

Adding **R**[M], we obtain an axiomatization of $[\cdot]^{\mathrm{M},\cdot}_{-\cdot}$.

Theorem 2 *$[\cdot]^{\mathrm{M},\cdot}_{-\cdot}$ satisfies* **QL**[M], **S** *and* **R**[M]*, if and only if for all $B \subseteq N$, $A \subset B$ and all $x_{-A} \in Y_{B\setminus A}$,*

$$[U]^{\mathrm{M},B}_{-A}(x_{-A}) = \frac{\sum_{y_A \in Y_A} (U(y_A, x_{-A}) - U(y_A, \perp_{-A}))}{\sum_{y_A \in Y_A} (U(y_A, \top_{-A}) - U(y_A, \perp_{-A}))}.$$

For every $U \in \mathcal{U}_\mathrm{M}(B)$, this expression is well-defined and belongs to $\mathcal{U}_\mathrm{M}(B \setminus A)$.

This results shows that there is a unique restriction operator fulfilling the three axioms. The restriction operator is now a normalized sum over all values of the missing attributes.

Sketch of the proof: Under **S**, the coefficients $\gamma_i^B(y_i)$ are the same for all $y_i \in Y_i$. To this end, one just needs to show that for all $U \in \mathcal{U}_\mathrm{M}(B)$, there exists $U' \in \mathcal{U}_\mathrm{M}(B)$ satisfying (15).

The end of the proof is similar to that of Theorem 1. ■

Example 10 (Ex. 3 cont.) *The restriction of U to the second attributes is $[U]^{\mathrm{M},\{1,2\}}_{-1}(t) = \frac{F(t)}{F(c)}$, where $F(t) = (U(a,t) - U(a,a)) + (U(b,t) - U(b,a)) + (U(c,t) - U(c,a))$. Then $[U]^{\mathrm{M},\{1,2\}}_{-1}(a) = 0$, $[U]^{\mathrm{M},\{1,2\}}_{-1}(b) = \frac{1}{2}$, $[U]^{\mathrm{M},\{1,2\}}_{-1}(c) = \frac{3}{4}$, $[U]^{\mathrm{M},\{1,2\}}_{-1}(d) = 1$. On average, the utility of c (resp. b) lies half way in-between b and d (resp. a and d).*

6 Monotone Idempotent Preference Model: Case of the Choquet Integral

We now consider elements of $\mathcal{U}_\mathrm{C}(B)$, that is Choquet integrals. These are monotone idempotent aggregation functions.

6.1 Axioms and Characterization Result

Quasi-Linearity

We assume property **QL**[C]. For $U \in \mathcal{U}_\mathrm{C}(B)$, $[U]^{\mathrm{C},B}_{-i}$ shall belong to $\mathcal{U}_\mathrm{C}(B \setminus \{i\})$. This implies a stronger result than Proposition 4. More precisely, the sum is only performed on the extreme points 0 and 1, as shown by the following result.

Proposition 5 *Under* **QL**[C]*, there exists coefficients $\{\gamma_i^N(y_i)\}_{y_i \in \overline{Y}}$ such that*

$$[U]^{\mathrm{C},B}_{-i}(x_{-i}) = \frac{\sum_{y_i \in \overline{Y}} \gamma_i^B(y_i)\,(U(y_i, x_{-i}) - U(y_i, 0_{-i}))}{\sum_{y_i \in \overline{Y}} \gamma_i^B(y_i)\,(U(y_i, 1_{-i}) - U(y_i, 0_{-i}))}.$$

Sketch of the proof: One can show that the expression of Proposition 4 applied on a Choquet integral fulfills idempotency iff $\gamma_i^B(y_i) = 0$ for $y_i \in Y \setminus \overline{Y}$. ■

Symmetry

Symmetry axiom **S** says that all values of attribute Y_i are treated symmetrically. By Proposition 5, only the two extreme values 0 and 1 count. The next axiom restricts **S** to these extreme values, but is applied only to a subset of cardinality at least 3.

Axiom Symmetry (S'): *For $B \subseteq N$ and $i \in B$ with $|B| \geq 3$, consider $U, U' \in \mathcal{U}_\mathrm{C}(B)$ s.t. $\forall x_{-i} \in Y_{B\setminus\{i\}}$*

$$\begin{aligned} U(1_i, x_{-i}) - U(1_i, 0_{-i}) &= U'(0_i, x_{-i}) - U'(0_i, 0_{-i}) \\ U(0_i, x_{-i}) - U(0_i, 0_{-i}) &= U'(1_i, x_{-i}) - U'(1_i, 0_{-i}). \end{aligned}$$

Then $[U]^{\mathrm{C},B}_{-i}(x_{-i}) = [U']^{\mathrm{C},B}_{-i}(x_{-i})$.

Proposition 6 *Under* **QL**[C] *and* **S'***, for all $x_{-i} \in Y_{B\setminus\{i\}}$*

$$[U]^{\mathrm{C},B}_{-i}(x_{-i}) = \frac{\sum_{y_i \in \overline{Y}} (U(y_i, x_{-i}) - U(y_i, 0_{-i}))}{\sum_{y_i \in \overline{Y}} (U(y_i, 1_{-i}) - U(y_i, 0_{-i}))}.$$

Proof : Consider two capacities v and v' depending only on three attributes i, j, k and $x_j < x_k$. Then $C_v(1_i, x_{-i}) - C_v(1_i, 0_{-i}) = x_j\,(v(\{i,j,k\}) - v(\{i,k\})) + x_k\,(v(\{i,k\}) - v(\{i\}))$ and $C_v(0_i, x_{-i}) - C_v(0_B) = x_j\,(v(\{j,k\}) - v(\{k\})) + x_k\, v(\{k\})$. Then if $v(\{i,j,k\}) - v(\{i,k\}) = v'(\{j,k\}) - v'(\{k\}) =: d_1^j$, $v(\{j,k\}) - v(\{k\}) = v'(\{i,j,k\}) - v'(\{i,k\}) =: d_0^j$, $v(\{i,k\}) - v(\{i\}) = v'(\{k\}) =: d_1^k$ and $v(\{k\}) = v'(\{i,k\}) - v'(\{i\}) := d_0^k$, then the conditions of **S** apply for C_v and $C_{v'}$. Then one shall have by Proposition 5

$$\frac{x_j\left(\gamma_i^B(1)\, d_1^j + \gamma_i^B(0)\, d_0^j\right) + x_k\left(\gamma_i^B(1)\, d_1^k + \gamma_i^B(0)\, d_0^k\right)}{\gamma_i^B(1)\left(d_1^j + d_1^k\right) + \gamma_i^B(0)\left(d_0^j + d_0^k\right)}$$

$$= \frac{x_j\left(\gamma_i^B(1)\, d_0^j + \gamma_i^B(0)\, d_1^j\right) + x_k\left(\gamma_i^B(1)\, d_0^k + \gamma_i^B(0)\, d_1^k\right)}{\gamma_i^B(1)\left(d_0^j + d_0^k\right) + \gamma_i^B(0)\left(d_1^j + d_1^k\right)}$$

where $[C_v]^{\mathrm{C},B}_{-i}(x_{-i})$ (resp. $[C_{v'}]^{\mathrm{C},B}_{-i}(x_{-i})$) is equal to the left hand side (resp. right hand side) of this relation. As this shall hold for every $d_1^j, d_0^j, d_1^k, d_0^k$ (small enough), and $x_j < x_k$, we conclude that $\gamma_i^N(1) = \gamma_i^N(0)$. ■

Theorem 3 *$[\cdot]^{\mathrm{C},\cdot}_{\cdot}$ defined for $B \subseteq N$, $A \subset B$ and $U \in \mathcal{U}_{\mathrm{C}}(B)$ satisfies* **QL**[C], **R**[C] *and* **S'** *if and only if for all $B \subseteq N$, $A \subset B$ with $A \neq B$, and all $U \in \mathcal{U}_{\mathrm{C}}(B)$*

$$[U]^{\mathrm{C},B}_{-A}(x_{-A}) = \frac{F_A^B(x_{-A}) - F_A^B(0_{-A})}{F_A^B(1_{-A}) - F_A^B(0_{-A})} \tag{16}$$

where $F_A^B(x_{-A}) = \sum_{D\subseteq A} U(1_D, 0_{A\setminus D}, x_{-A})$. Moreover for every function in $\mathcal{U}_{\mathrm{C}}(B)$, (16) is well-defined and belongs to $\mathcal{U}_{\mathrm{C}}(B \setminus A)$.

This results shows that there is a unique restriction operator fulfilling the three axioms. The restriction operator is a normalized sum over the extreme points 0 and 1 of the missing attributes.

Proof : One can easily verify that (16) satisfies **QL**[C], **R**[C] and **S'**. The "only if" part of the theorem is provided by combining Proposition 6, and proceeding as in Theorem 1.

Consider $U \in \mathcal{U}_{\mathrm{C}}(B)$, $A \subset B$ and $k \in B \setminus A$. By (3), there exists $D \subseteq B \setminus (A \cup \{k\})$ such that $U(1_D, 0_{B\setminus(D\cup\{k\})}, 1_k) > U(1_D, 0_{B\setminus(D\cup\{k\})}, 0_k)$. By monotonicity of U, this yields

$$U(1_D, 0_{A\setminus D}, 1_{B\setminus A}) > U(1_D, 0_{A\setminus D}, 0_{B\setminus A}).$$

Hence the denominator of (16) is always non-zero, so that (16) is well-defined for every function in $\mathcal{U}_{\mathrm{C}}(B)$. ■

6.2 Interpretation

The next result provides the expression of the restriction operator for a two-additive Choquet integral.

Theorem 4 *For a two additive Choquet integral U defined on B with importance and interaction indices v_i and $I_{i,j}$ respectively, $[U]^{\mathrm{C},B}_{-A}$ is also a two-additive Choquet integral with the following importance and interaction terms:*

$$v_i^{B\setminus A} = \frac{v_i}{\sum_{k\in B\setminus A} v_k} \quad \textit{and} \quad I_{i,j}^{B\setminus A} = \frac{I_{i,j}}{\sum_{k\in B\setminus A} v_k}. \tag{17}$$

Expressions (17) are very intuitive and simpler to compute than (16). The relative importance between criteria in $B \setminus A$ remain the same. Hence, in order to get an idempotent model, the weight v_i is divided by $\sum_{k\in B\setminus A} v_k$. The interaction index between two criteria in $B \setminus A$ is simply the previous interaction divided by the same factor. The other interactions are merely discarded, as expected.

Example 11 (Ex. 4 cont.) *Attributes $x^{\mathrm{RA}}_{\mathrm{PA}}$ and $x^{\mathrm{RA}}_{\mathrm{C}}$ are unknown (see Ex. 1). By Theorem 4, the restriction $[U]_{-\mathrm{RA}}$ of U (defined in Ex. 4) to the remaining two attributes is*

$$[U]_{-\mathrm{RA}}(x^{\mathrm{CA}}_{\mathrm{C}}, x^{\mathrm{CA}}_{\mathrm{PE}}) = \frac{2x^{\mathrm{CA}}_{\mathrm{C}}}{3} + \frac{x^{\mathrm{CA}}_{\mathrm{PE}}}{3} - \frac{1}{3}\frac{|x^{\mathrm{CA}}_{\mathrm{C}} - x^{\mathrm{CA}}_{\mathrm{PE}}|}{2}.$$

Comparing to the expression of U, we keep only the terms for which the variables are known (namely $x^{\mathrm{CA}}_{\mathrm{C}}$ and $x^{\mathrm{CA}}_{\mathrm{PE}}$). As for U, C is twice as more important than PE, and there is a strong positive interaction between C and PE. Expression $[U]_{-\mathrm{RA}}$ thus completely makes sense.

7 Discussion and Perspectives

This paper addresses the problem of how to handle missing data interpreted as non-relevant, in a preference model. More specifically, given a model defined on the full set of attributes, we have defined a restriction operator returning a preference model defined on a subset of attributes. Axiomatic characterizations are proposed for three classes of models. For general quantitative models, the restriction operator is characterized by linearity, recursivity and particular cases of constant models and models depending only on one variable. We obtain the average of the value of the utility over all possible values of the missing attributes. The second class is the set of monotone quantitative models satisfying normalization conditions. The linearity axiom is changed to fit with these conditions. Adding recursivity and symmetry, we obtain a normalized average of the utility over all values of the missing criteria. The last class is the Choquet integral w.r.t. non-degenerate capacities. With the same axiom as before, we obtain a simpler expression averaging only on the extreme values of the missing criteria. Finally, a very intuitive interpretation is provided for two-additive Choquet integrals.

Our approach has advantages over the existing ones. A robust approach is too pessimistic. In Ex. 1, it would penalize a system when the values of the two attributes are missing. There is no reason to do so as there is no RA in the scenario. The drawback of the probabilistic approach is on the assumptions made for the completion of the missing data. However, it is interesting to note that it yields exactly the same expression as in our approach for the non-monotone model (see Theorem 1). The main asset of our approach is that it brings a strong axiomatic justification.

If a naive approach works well for the non-monotone case, it is not the case for the other models. We have shown that for monotone models, one shall use a normalized average rather than a simple average to get a consistent result, and that for the Choquet integral, we shall only consider the extreme points. These major points are derived by the axiomatic approach and the closed form of the restriction operator.

The paper can be extended in several directions. Theorems 2 and 3 are not strictly speaking expected utility. It would be interesting to investigate whether it can be put under the form of a conditional expectation, with the denominator being the expected utility of the conditioned event. The idea of removing missing data can be adapted to other fields such as machine learning or data fusion. Applying a learned (classification/regression) model when attributes are missing while still requiring a precise prediction requires to make strong assumption about the missingness process, such as MAR, NMAR. Our current approach provides an alternative way to handle missing attributes by simply removing it from the model, allowing one to make predictions without assumptions about the missingness process.

Acknowledgments

This work has been supported by the European project SESAR H2020 PJ14.04.03 EECNS: "Essential and Efficient Communication Navigation and Surveillance Integrated System".

References

[Bacchus and Grove, 1995] Fahiem Bacchus and Adam Grove. Graphical models for preference and utility. In *Conference on Uncertainty in Artificial Intelligence (UAI)*, pages 3–10, Montreal, Canada, July 1995.

[Benabbou *et al.*, 2017] Nawal Benabbou, Patrice Perny, and Paolo Viappiani. Incremental elicitation of Choquet capacities for multicriteria choice, ranking and sorting problems. *Artificial Intelligence*, 246:152–180, 2017.

[Boutilier *et al.*, 2004] Craig Boutilier, Ronen Brafman, Carmel Domshlak, Holger Hoos, and David Poole. CP-nets: A tool for representing and reasoning with conditional ceteris paribus preference statements. *Journal of Artificial Intelligence Research (JAIR)*, 21:135–191, 2004.

[Choquet, 1953] Gustave Choquet. Theory of capacities. *Annales de l'Institut Fourier*, 5:131–295, 1953.

[Dominiak, 2006] Cezary Dominiak. Multicriteria decision aid under uncertainty. *Multiple Criteria Decision Making*, 1:63–82, 2006.

[Durbach and Stewart, 2012] Ian Durbach and Theodor Stewart. Modeling uncertainty in multi-criteria decision making. *European Journal of Operational Research*, 223:1–14, 2012.

[ESASSP, 2012] ESASSP. EUROCONTROL Specification for ATM Surveillance System Performance (volume 1). Technical report, EUROCONTROL-SPEC-0147; ISBN: 978-2-87497-022-1, 2012.

[Fishburn, 1967] Peter Fishburn. Interdependence and additivity in multivariate, unidimensional expected utility theory. *International Economic Review*, 8:335–342, 1967.

[Fishburn, 1970] Peter Fishburn. *Utility Theory for Decision Making*. J. Wiley & Sons, 1970.

[Galand *et al.*, 2013] Lucie Galand, Julien Lesca, and Patrice Perny. Dominance rules for the choquet integral in multiobjective dynamic programming. In *23rd International Joint Conference on Artificial Intelligence (IJCAI 2013)*, pages 538–544, Beijing, China, August 2013.

[Garca-Laencina *et al.*, 2010] Pedro Garca-Laencina, José-Luis Sancho-Gmez, and Anibal Figueiras-Vidal. Pattern classification with missing data: a review. *Neural Computing and Applications*, 19(2):263–282, 2010.

[Gaspars-Wieloch, 2015] Helena Gaspars-Wieloch. A decision rule for uncertain multicriteria mixed decision making based on the coefficient of optimism. *Multiple Criteria Decision Making*, 10:32–47, 2015.

[Goodfellow *et al.*, 2016] Ian Goodfellow, Yoshua Bengio, and Aaron Courville. *Deep learning*. MIT press, 2016.

[Grabisch and Labreuche, 2010] Michel Grabisch and Christophe. Labreuche. A decade of application of the Choquet and Sugeno integrals in multi-criteria decision aid. *Annals of Operation Research*, 175:247–286, 2010.

[Grabisch *et al.*, 2000] Michel Grabisch, Toshiaki Murofushi, and Michio Sugeno. *Fuzzy Measures and Integrals. Theory and Applications (edited volume)*. Studies in Fuzziness. Physica Verlag, 2000.

[Konczak and Lang, 2005] Kathrin Konczak and Jérôme Lang. Voting procedures with incomplete preferences. In *IJCAI-05 Multidisciplinary Workshop on Advances in Preference Handling*, 2005.

[Krantz *et al.*, 1971] David H. Krantz, Duncan Luce, Patrick Suppes, and Amos Tversky. *Foundations of measurement*, volume 1: Additive and Polynomial Representations. Academic Press, 1971.

[Lahdelma *et al.*, 1998] Risto Lahdelma, Joonas Hokkanen, and Pekka Salminen. SMAA – Stochastic Multiobjective Acceptability Analysis. *European Journal of Operational Research*, 106:137–143, 1998.

[Li *et al.*, 2015] Xihua Li, Fuqiang Wang, and Xiaohong Chen. Support vector machine ensemble based on Choquet integral for financial distress prediction. *International Journal of Pattern Recognition and Artificial Intelligence*, 29, 2015.

[Little and Rubin, 2002] Roderick Little and Donald Rubin. *Statistical Analysis with Missing Data*. Wiley Publishers, 2002.

[Lundberg and Lee, 2017] Scott Lundberg and Su-In Lee. A Unified Approach to Interpreting Model Predictions. In *31st Conference on Neural Information Processing Systems (NIPS 2017)*, pages 4768–4777. Long Beach, CA, USA, 2017.

[McCoy *et al.*, 2018] John McCoy, Steve Kroon, and Lidia Auret. Variational autoencoders for missing data imputation with application to a simulated milling circuit. *IFAC (International Federation of Automatic Control)*, 51:141–146, 2018.

[Saar-Tsechansky and Provost, 2007] Maytal Saar-Tsechansky and Foster Provost. Handling missing values when applying classification models. *Journal of Machine Learning Research*, 8:1623–1657, 2007.

[Smieja *et al.*, 2018] Marek Smieja, Lukasz Struski, Jacek Tabor, Bartosz Zieliński, and Przemyslaw Spurek. Processing of missing data by neural networks. In *32nd Conference on Neural Information Processing Systems (NeurIPS 2018)*, pages 2724–2734, Montréal, Canada, 2018.

[Tehrani *et al.*, 2012] Ali Fallah Tehrani, Weiwei Cheng, Krzysztof Dembczynski, and Eyke Hüllermeier. Learning monotone nonlinear models using the Choquet integral. *Machine Learning*, 89:183–211, 2012.

[Wilson, 2014] Nic Wilson. Preference inference based on lexicographic models. In *Twenty first European Conference on Artificial Intelligence (ECAI'14)*, pages 921–926, Amsterdam, The Netherlands, August 2014.

[Yeh *et al.*, 2017] Raymond Yeh, Chen Chen, Teck Yian Lim, Alexander Schwing, Mark Hasegawa-Johnson, and Minh Do. Semantic image inpainting with deep generative models. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 5485–5493, 2017.