Non linear preference models: why, how?

Olivier Cailloux[1] and Sébastien Destercke[2]

[1] LAMSADE
olivier.cailloux@dauphine.fr
[2] HDS

Abstract. Literature involving preferences of artificial agents or human beings often assume their preferences are linear, that is, the preference can be represented using a complete transitive antisymmetric binary relation. Much has been written however on more complex, or more interesting, models of preferences. In this article we review some of the reasons that have been put forward to justify more complex modeling, and review some of the techniques that have been proposed to obtain models of such preferences. (Optional: we connect to various related literature about argumentation, …)

1   Introduction

Here is a possible approach we might want to use.

- What's a model of preference? Example with weighted sum. But: forbids some possibly interesting evaluations; need to recode evaluations and hence is dependent of irrelevant alternatives; w1 > w2 unclear.
- Utility: is holistic. Forbids incomparability. Mandates transitivity.
- Models of preferences can have (at least) two interpretations: normative or descriptive. Goals and (sometimes) tools to model preferences and reasonable hypothesis that can be postulated about the shape of the preferences vary according to these two lines.
- (Reasonable) hypothesis about descriptive preferences. Why it's more difficult than can be thought naïvely.
- Reasonable hypothesis about normative preferences. Why it's more difficult than can be thought naïvely. Talk about incompleteness. Applying normative approaches: talk about prescription a la Roy…
- Focus on incompleteness rather than non-transitivity. Can be lack of information or intrinsic.
- Present usual models (Savage? MAVT?) and alternative ones (belief functions? Electre?)?
- Talk about choice under uncertainty wrt incompleteness.
- Impact of relaxations on problems of choice, ranking, classification.
- How to obtain information under incompleteness hypothesis, meaning and interpretation.
- (If space permits) Links to argumentation theory: build argumentative models that explain / justify a decision.?
- (If space permits) Path to validation of normative models. Talk about empirical social choice; reflective equilibrium. Use the resulting indeterminacy (lability). Maximizing without transitivity or completeness?

In this paper, we consider choice situations (and others): given a Decision Maker (DM) and a set of alternatives $\mathscr{A}$, we wish to model the preferences of the DM... (To be described.)

Because complexity requires justification, we will first consider the simplest approach to preference modeling, namely the weighted sum. We will briefly discuss its shortcomings. Relaxing it progressively, we will try to convince the reader that even more basic assumptions on preferences such as completeness and transitivity are not always justified. This will motivate the introduction of more interesting approaches. We do not claim novelty in these arguments and approaches: we want to point out and summarize the various arguments from the literature that readers non-expert in preference modeling may ignore, and underline the richness and diversity of the existing approaches.

2   Model of preference: the weighted sum

When considering a choice situation, it may be useful to try to obtain a formal model of her preferences. This model can be used to describe the way she chooses, or to recommend an option. (More on the uses of models later.) Let us see first what a simple model may look like. We start by defining a multiple-criteria choice decision situation (MCDS), one of the situations considered in this paper.

We consider a set $\mathscr{A}$ called the set of "alternatives", that represents the possible options to choose from. Alternatives are here considered as mutually exclusive and the set $\mathscr{A}$ is exhaustive: the choice situation mandates to choose exactly one alternative from the set. Let us further assume some structure on $\mathscr{A}$: assume each alternative is evaluated on multiple criteria. The finite set $\mathcal{G}$ supposedly represents all the criteria, or aspects, that the DM may wish to consider when choosing. Each criterion $g \in \mathcal{G}$ comes with an evaluation scale X_g, and an evaluation function $f_g : \mathscr{A} \to X_g$ that evaluates each alternative. To simplify notations, we will denote the evaluation function f_g simply by g, thus equate the criterion with its evaluation function. Finally, it may be useful to apply the decision problem to different subsets of alternatives (either as a thought experiment, or because the problem can happen again in the future). Let $A \subseteq \mathscr{A}$ denote a subet of alternatives, that represents the alternatives to choose from in a given occurence of the decision problem. An MCDS $(\mathscr{A}, \mathcal{G})$ is a choice situation which can be described using such exhaustive sets of alternatives and criteria.

For example, say the DM must choose what to plant in her garden. The set of alternatives $\mathscr{A}$ are all possible vegetables, the criteria $\mathcal{G}$ measure the taste, quantity, and price of each vegetable, and A are the vegetables that are available this year for planting. In our example, we could have $X_{g_1} = \{A, B, C, D\}$, a set of labels, with $g_1(a)$ representing the taste of the vegetable $a \in \mathscr{A}$ as considered by the DM (A is the best taste, D the worst), $X_{g_2} = [0, 100]$, with $g_2(a)$ representing the number of meals that the DM would enjoy if deciding to plant a, and $X_{g_3} = \mathbb{R}$, where $g_3(a)$ indicates the price to pay for planting a.

Given an MCDS $(\mathscr{A}, \mathcal{G})$, we are interested in designing an aggregation function $f : \mathscr{P}(A) \to \mathscr{P}(A) \setminus \varnothing$, such that $f(A) \subseteq A$, that represents the best alternatives from the point of view of the DM. We allow the function to return a subset of alternatives to allow for ex-æquo. (The domain of the function is sometimes restricted to the finite subsets $A \subseteq \mathscr{A}$.) It will be interesting to parameterize the function with preference parameters, that represent the subjectivity of the DM. Let

Ω denote the set of possible preference values for the preference parameters. An aggregation scheme is such a set Ω and a mechanism for deriving an aggregation function f_ω given $\omega \in \Omega$.

When facing an MCDS, the weighted sum approach comes to mind very naturally. Assume for simplicity that $X_g \subseteq \mathbb{R}$ and that the usual ordering $>$ on $\mathbb{R}$ reflects goodness on all scales. Given a finite set $A \subseteq \mathscr{A}$, define for each criterion the worst evaluation in that set, $\min_{a \in A} g(a) = m_g \in X_g$ and the best evaluation, $\max_{a \in A} g(a) = M_g \in X_g$. The weighted sum aggregation scheme defines Ω as the set of all possible weights that sum to one, thus $\omega = \left(w_g \in \mathbb{R}^+ \right)_{g \in \mathcal{G}}$, with $\sum_{g \in \mathcal{G}} w_g = 1$, and given $\omega \in \Omega$ and a finite set A, defines $f_\omega(A) = \arg\max_{a \in A} \sum_{g \in \mathcal{G}} w_g \frac{g(a) - m_g}{M_g - m_g}$ (the denominator is replaced with 1 if $M_g = m_g$).

To illustrate, consider our previous example restricted to only two criteria, g_2 and g_3, with the evaluations given in table 1 (price is set negative so that greater evaluations are better). Given $A = \{a_1, a_2, a_3\}$, the weighted sum approach considers that a choice can be found by associating to each alternative the value $v(a) = w_{g_2} \frac{g_2(a) - 20}{15} + w_{g_3} \frac{g_3(a) + 160}{60}$, for some weights w_{g_2} and w_{g_3}, and picking the alternatives having highest values.

	g_2	g_3
a_1	35	−120
a_2	20	−100
a_3	23	−160

Table 1. Evaluations of some vegetables

This approach is simple, but may not be innocuous. One of its problem is that it fails to satisfy the Independence of irrelevant alternatives. This property of an aggregation scheme requires that for all preferential parameters ω, whenever some a_1 is chosen by f_ω and some a_2 is not, given some set A, changing other alternatives in A may not reverse the decision. Formally, Independence of irrelevant alternatives mandates that, $\forall \omega \in \Omega$: $a_1 \in f_\omega(A), a_2 \notin f_\omega(A) \Rightarrow \nexists A' \subseteq \mathscr{A} \mid a_1, a_2 \in A', a_2 \in f_\omega(A'), a_1 \notin f_\omega(A')$.

In our example, pick $w_{g_2} = w_{g_3} = \frac{1}{2}$. Observe that $v(a_1) = \frac{1}{2} 1 + \frac{1}{2} \frac{2}{3} = \frac{5}{6}$, $v(a_2) = \frac{1}{2}$, $v(a_3) = \frac{1}{10}$. Thus $f_\omega(\{ a_1, a_2, a_3 \}) = \{a_1\}$. Now replace a_3 by a_3' with evaluations $g_2(a_3') = 10$ and $g_3(a_3') = -110$. We obtain $v'(a) = w_{g_2} \frac{g_2(a) - 10}{25} + w_{g_3} \frac{g_3(a) + 120}{20}$, $v'(a_1) = \frac{1}{2} 1 + \frac{1}{2} 0 = \frac{1}{2}$, $v'(a_2) = \frac{7}{10}$, $v'(a_3) = \frac{1}{4}$, thus a_2 is now selected while a_1 is not. This feature is unwanted if the DM considers that the relative positions of a_1 and a_2 do not depend on the evaluations of some third alternative a_3.

Incidentally, the infamous Shanghaï ranking exhibits such unfortunate property [TODO refs].

To solve this problem, the normalization procedure may be modified: define for each criterion m_g and M_g once and for all, independently of A. But the choice of these extreme points will have a strong influence on the aggregation function, as just illustrated, and it may be unclear how to define them. Relatedly, the natural interpretation of the weights parameter may not correspond to what the DM has in mind. For example, it may be natural for the DM to state that all criteria have

equal importance, in which case the analyst may be tempted to set all weights to an equal value. The effect of this will depend on the extreme points m_g, M_g: changing M_g to $M'_g = 2M_g - m_g$, without changing the weight w_g, will half the effective importance of the related criterion. Similarly, the DM might state that g is "more important" than g'. This should not necessarily be interpreted as meaning that $w_g > w_{g'}$.

+ talk about the impossibility to pick some interesting alternatives with a weighted sum?

3   MAVT

A more principled approach is useful to make sure we pick an aggregation scheme with suitable properties. Principled here means that we want to introduce sufficient properties on the way the DM chooses, that suffice to guarantee that an aggregation function that represents the DM's choices exists within the aggregation scheme.

The most usual approach to MCDS is to attribute a value to each alternative and assume that the choice function picks the maximal alternatives according to v. This starts as in the weighted sum approach, with the important difference that v may not depend on A, only on $\mathscr{A}$ and $\mathcal{G}$.

With such a value function underhand, we may reason with a binary preference relation $\succeq$ on the set of alternatives. Define: $a \succeq b \Leftrightarrow v(a) \geq v(b)$.

It is customary to consider $\succeq$ as a relation that is naturally defined, known by the DM, and with no need of explanation. This, with this approach, we assume we can observe $\succeq$ and wonder in which cases we can define a suitable v.

1. $\succeq$ is complete. 2. $\succeq$ is transitive.

Theorem 1. *Assume $\mathscr{A}$ is finite. There exists v such that $[v(a) \geq v(b)$ iff $a \succeq b]$ iff $\succeq$ is a complete transitive binary relation.*

If $\mathscr{A}$ is infinite, the implication from left to right still holds. The other direction, from existence of $\succeq$ complete and transitive to existence of a suitable v, requires a supplementary technical condition (there must be some subset $A \subseteq \mathscr{A}$ such that, for all $a \succ c$ with $a, c \in \mathscr{A}$, $\exists b \in A \mid a \succeq b \succeq c$ [?, Ch. 2, Th. 2, p. 40]).

Thus, those conditions require that $\succeq$ be a weak-order.

When some supplementary conditions are fulfilled (such as separability, to be described succinctly here), these conditions allow to use the Multiple Attribute Value Theory to solve the aggregation problem. When the supplementary conditions do not hold, other approaches may be suitable such as Choquet integral, … We refer the reader to the numerous surveys about those approaches (give refs) as we choose rather to focus on the reasons to question more fundamental conditions, namely completeness and transitivity of $\succeq$. Before doing this, we want to give another (important!) example of a preference model, that will be of use when discussing which conditions hold.

Before doing this however, we need to say a word about what those preferences really represent and what the goal of modeling those may be. This is required to give meaning to discussion about which conditions are reasonable to postulate about $\succeq$, and how to check whether they hold.

4 Different approaches to preference modeling

In the descriptive approach to preferences, the goal of the model is to reflect the normal behavior of a DM. Typically, we would have collected a database of sample choices of the DM, say, of choices of food products in his favorite store, and we would try to obtain the model that best reflects his choice attitude. Such a model may be used to predict his behavior, e.g. for marketing or regulation purposes. It may also be applied with the aim of replacing the DM by automating his decision procedure, although it is debatable whether the normative approach would better suit this use case.

Under the normative approach, the goal is to reflect on how the DM ought to choose. Either according to external norms that are somehow collectively decided as representing rational behavior, or using some norms that the DM himself accepts after careful reflexion. Hence, the decision outcome using such approach may differ from that the DM would have chosen in his daily life. Consider as an example a responsible person in an enterprise who wants to think about the procedure used to recruit employees. After having collected the figures, it may appear that, for some unconscious reason, the recrutement is biased against some particular socio-economic or gender category, even considering equal competences. It may be, further, that the DM himself is not happy about this situation. Thus, he may try to find a strategy of selecting employees that would avoid such biases, therefore actively trying *not* to build a perfectly descriptive model of his normal selection attitude.

When the focus is on providing decision help to a DM by letting him think about and adopt the norms he prefers, rather than considering them as external norms that he ought to follow under threat of being considered irrational, the approach is often termed prescriptive, or constructive. There are important differences between these terms, and different authors use them somewhat differently. (Give refs.) The focus here being not on those differences, we refer to ... and will use the general normative term as an umbrella.

When building recommender systems in the literature in artificial intelligence, the focus is often on descriptive approaches. This is usually left implicit, with no discussion about possible alternatives. We think however that an interesting path is offered by normative approaches. Descriptive approaches will, by design, reflect cognitive limitations exhibited by us, normal human beings. Those limitations are numerous and sometimes obviously not in agreement with what a more thoughtful and knowledgeable person would do, as is well known and will be illustrated in this article (although there is debate about how such a sentence is to be interpreted exactly, more about this later). Providing (more) normative-based automatic recommendations might help provide sound advices, help increase serendipity, and possibly reduce incitations to merchants to exploit imperfections on the human reasoning abilities by using marketing techniques that may lead to choices that the DM himself would possibly reject when thinking more carefully.

> OC : This should be moved to the conclusion, probably.

5 Transitivity and completeness in descriptive approaches

When conditions about the $\succeq$ preference relation are assumed to bear on what DMs normally do, it makes it very easy to test them. Very numerous laboratory

experiments have tackled this question.

We must first be again more specific about what $\succeq$ means and how it should be observed. To make sure it is really the behavior of an individual that is captured with this relation, many experiments place subjects under real choices, presenting to them two objects among which the individual has to pick exactly one. This choice then represents his preference, by definition: here preference is meant to reflect the object the individual would choose when facing a choice.

When working under utility theory, it is common to ask questions about probability equivalent loteries, or about …