The problem of minimum connivance set for preferences elicitation

Introduction

Preferences elicitation is one of the most interesting and challenging applications of decision theory to the recommendation; it arises in many subset evaluations related problems and consists, having a set of alternatives $\mathcal{A}$, in using a function f that represents the preferences of the user over the subsets of alternatives $2^{\mathcal{A}}$ to incrementally query new preferences and then to fine-tune the recommendations.
Since many functions could represent the same preferences, we generally assume the function's structure we use.

To ensure that this function is expressive enough to represent the user's preferences, we make it depend on parameters θ that affect the recommendations provided.
For instance, in this article, we will model the user's preferences with a binary relation $\succeq$ defined on the subsets such that $A \succeq B$ means that the user prefers the subset A to the subset B.
We will also suppose that this relation is representable with an additive function f that depends on a set of parameters θ each parameter representing the utility of having a subset of alternatives, such that f is defined as follows.

$$f_\theta(A) = \sum_{\omega \in \theta} x_\omega^A u_\omega$$

Where x_ω^A is a binary indicator of $\omega \in A$,.We say that the model is k-additive if θ does not contain a subset ω such that $\omega > k$.
Knowing a set R of preferences $R^{\succeq} = \{(A,B); A, B \in 2^{\mathcal{A}}; A \succeq B\}$ we define $\theta_{R^{\succeq}}$ as the set of compatible parameters :

$$\theta \in \theta_{R^{\succeq}} \iff \forall (A,B) \in R^{\succeq}; f_\theta(A) \geq f_\theta(B)$$

The objective of the elicitation process is to reduce the size of θ_R by querying a new preference from the decision-maker at each iteration. The procedure stops when we find out a necessarily optimal solution S such that $\forall \theta \in \theta_{R^{\succeq}}$, $\forall X \in 2^{\mathcal{A}}; f_\theta(S) \geq f_\theta(X)$.
Several difficulties may arise when using an approach such as the one described in this article, for example, that the decision-maker is not infallible or rational. However, the most evident one is that an n-additive function contains many parameters that grow exponentially with the number of alternatives.
In practice, we assume that the function is k-additive for a k that does not depend on the number of alternatives, the problem is that the value of k represents a trade-off between the expressiveness of the model and its simplicity, so a model with a too-small k may fail to represent the user's preferences by reducing θ_R to the empty set. However, while a too large k increases the difficulty of adjusting the parameters and the number of queries needed to find a necessarily optimal solution.
We could think of starting with a 1-additive model and increasing k each time the model fails at representing the preferences of the user; the problem is that if, for example, the preferences of the user are 1-additive for all the alternatives except three of them, which are only worthwhile taken together we are obliged to use a 3-additive model. However, all the u_ω for $\omega \geq 2$ except those concerning these three alternatives will always be at 0 and still complexify the problem.
This paper proposes extending θ only when needed and only with interaction terms that capture

the exact interaction that made the model fail at representing the user's preferences.

To do so, we will start by presenting a proposition formalised by Fishburn [?] that allows us to find for a profile of preferences R a sufficient and necessary condition for the existence of a function f_θ that represents the preferences.

After that, we will prove that the problem of verifying this condition is NP-Complete and propose a mixed-integer linear formulation for it. Finally, we will propose a procedure of elicitation that works by extending the model while trying to fit its parameters by detecting the subsets of alternatives that interacts without the model considering it.

Representability of a subset of preferences

In this part we will investigate the problem, having a set R of preferences of determining if we can found an additive function f_θ for a given θ that represents it.

Formal definition

Let's start by defining the problem formally. Having a θ fixed, we say that R a subset of preferences is θ-representable if we can found $u = \{u_{\omega_0}, \ldots, u_{\omega_k} | \omega_0, \ldots, \omega_k \in \theta\}$ such that the function $f_\theta(A) = \sum_{\omega \in \theta} x_\omega^A u_\omega$ where x_ω^A is an indicator of $\omega \in A$ satisfies:

$$\forall (A, B) \in R; f_\theta(A) \geq f_\theta(B)$$

This is equivalent to:

$$\forall (A, B) \in R; f_\theta(A) - f_\theta(B) \geq 0$$

$$\forall (A, B) \in R; \sum_{\omega \in \theta} x_\omega^A u_\omega - \sum_{\omega \in \theta} x_\omega^B u_\omega \geq 0$$

and by putting: $d_\omega^{(A,B)} = x_\omega^A - x_\omega^B$ it became :

$$\forall (A, B) \in R; \sum_{\omega \in R} d_\omega^{(A,B)} u_\omega \geq 0$$

if we note $D^{R^\succeq}$ the matrix (m, k) with $|R^\succeq| = m$ and $|\theta| = k$ defined by the coefficients $d_\omega^{(A,B)}; \forall \omega \in \theta, \forall (A, B) \in R^\succeq$ the problem of θ-representability is the problem of finding out if the polyhedron $P = \{U, D^R U \geq 0\}$ is empty.

Now we will slightly change our model to take into account the indifferences, from now on we define $R^\sim$ as $R^\sim = \{(A, B); A, B \in 2^A; A \sim B\}$ with $\sim$ the binary relation that describe the indifference i.e $A \sim B \equiv \neg(A \succeq B) \wedge \neg(B \succeq A)$, and we note $D^{R^\sim}$ the coefficients $d_\omega^{(A,B)}; \forall \omega \in \theta, \forall (A, B) \in R^\sim$.

Finally, In the following we will note R the set of all the preferences, $R = R^\succeq \cup R^\sim$, so that the problem of θ-representability is now finding out if the polyhedron became :
$P = \{U; D^{R^\succeq} U \geq 0; D^{R^\sim} U = 0\}$ is empty.

Fishburn condition

In his article, Fishburn [?] formullated a necessary and sufficient condition for the 1-representability and the 2-representability, we will now generalize it to the θ-representability.

A set R of preferences is θ-representable if and only if for each two subsets of subsets $A = \{A_0, \ldots, A_j\} \subset 2^{2^A}$ and $B = \{B_0, \ldots, B_m\} \subset 2^{2^A}$, such that:

$$A \approx_\theta B \equiv \sum_{A_i \in A} \sum_{\omega \in \theta} x_\omega^{A_i} = \sum_{B_i \in B} \sum_{\omega \in \theta} x_\omega^{B_i}$$

the following holds:

$$\exists i \leq m; A_i \succeq B_i \longrightarrow \exists j \leq m; B_j \succeq A_j$$

In other terms, a certificate of non θ-representability is a subset $R' \subset R^\succ$; $R' = \{(A_i, B_i); A_i \succ B_i \vee A_i \sim B_i\}$ such that $\cup_i A_i \approx_\theta \cup_i B_i$.

Regarding the matrix $D^{R^\succ}$, since each line represents a preference $r_i = (A_i \succ B_i)$, selecting a subset of constraints is equivalent to building two subsets of subsets $A = (A_0, \ldots, A_j)$, $B = (B_0, \ldots, B_j)$ such that $\forall i \leq j; A_i \succ B_i$, and if all the columns of the subset of constraints sums to 0 since each sum represents $\sum_{A_i \in A} \sum_{\omega \in \theta} x_\omega^{A_i} - \sum_{B_i \in B} \sum_{\omega \in \theta} x_\omega^{B_i}$ that means that the problem of finding out if R is θ-representable is equivalent to the problem of finding a subset of rows in the matrix $D^{R^\succ}$ where each column sums to 0.

It's a necessary condition because if we find a subset like R', that means that $f_\theta(\cup_i A_i) = f_\theta(\cup_i B_i)$ because $f_\theta(\cup_i A_i) - f_\theta(\cup_i B_i) = 0$ but for $\forall A_i \in A, \forall B_i \in B; f_\theta(A_i) > f_\theta(B_i)$ and it gives us a contradiction.

And to prove that it's a sufficient condition we should use the theoreme of alternatives formulated by Fishburn [??], this theorem states that having a set of n inequalities and m equalities made with k variables and represented respectively by the two matrices D, D' such that:

$$DU \leq 0$$
$$D'U = 0$$

Exactly one of the two following holds:

- The system has a solution U.

- There are non negative integers $a_1, a_2, \ldots a_n$ at least one of which is positive, and integers $b_1, \ldots, b_m$ such that:

$$\forall j \in \mathbb{N}, j \leq k; \sum_{i=1}^{n} a_i D_j^i + \sum_{i=1}^{m} b_i D_j^i = 0$$

To prove that the Fishburn condition is a sufficient condition we apply the theoreme of alternatives with $D = R^\succeq$ and $D' = R^\sim$.

Let's suppose that we does'nt have a solution U to the système, with the integers $(a_i)_{i \leq n}$ and $(b_i)_{i \leq m}$ we can build two subsets A, B such that the Fishburn condition fails, to see that let's start by ommit in both the nul integers, and for each $b_i < 0$ we replace $A \sim B$ by $B \sim A$ in the matrix $R^\sim$.

Now let's explain how to build the two subsets A and B for which the Fishburn condition does'nt hold.

- Since each positive integer a_i is associated to a constraint $A_i \succeq B_i$, for each positif integer we add to A a_i time the subset A_i and we add to B a_i times the subset B_i.
- Since each positive integer b_i is associated to a constraint $A_i \sim B_i$, for each positif integer we add to A b_i time the subset A_i and we add to B b_i times the subset B_i.

We obtain two subsets for which we have $\forall i; A_i \succeq B_i \vee A_i \sim B_i$ and since we have:

$$\forall \omega \in \theta; \sum_{i=0}^{n} a_i D_\omega^i + \sum_{i=0}^{m} b_i D_\omega^i = 0$$

It means that each $\omega \in \theta$ is present with the same number in A and in B and thus that $A \approx_\theta B$ and so the Fishburn condition does'nt hold.

The problem of finding a non θ-representable subset of a set of preferences

In this part we will show that the problem that consists in finding a non θ-representable in R a set of preferences is NP-Complete.

To do so let's starts by showing off that it's a NP problem by considering a certificate consisting in the subset of preferences R' that is non θ-representable, computing the matrices $D'^{\succeq}$ and $D'^{\sim}$ is done by iterating on it and takes at most $O(|\theta|(n+m)|)$ where n is the nombre of strict preferences and m the number of indifferences, and then it's sufficient to verify that the sum of each column is 0.

After that, in a matrix of elements that can be either 0,1 or -1, we will reduce the problem of finding a bound subset (without repetitions) of lines that sums to 0 for each column into the **EXACT SUM** problem.

Let's start by defining the **EXACT SUM** problem

- **Entry:** a set of elements $a_0, a_1, \ldots, a_n$ a set of subsets $\mathcal{S} = \{S_1, S_2, \ldots S_n\}$ each one containing a subset of the elements.
- **Decision** we try to find a set of subsets $\mathcal{X} \subset \mathcal{S}$ such that each element a_i is present exactly once in $\cup_{S_i \in \mathcal{X}} S_i$

And now we will show that it every instance of this problem could be transformed into an instance of the bound variant of our problem, the transformation is simple, we just have to build a matrix that have one column per element $a_0, a_1, \ldots, a_n$ and one line for each subset of $\mathcal{S}$, after that make each cell of the matrix $D_{a_i}^{S_j}$ an indicator if $a_i \in S_j$.

We add a fictive line at the end that is composed of -1 $\forall a_i$ and thus to build a subset $\mathcal{X}$ of $\mathcal{S}$ where each element is taken exactly once we should choose a subset of lines in the matrix where each element is taken 1 time and then select the last line to have a subset of lines where each column sums at 0, it's the only way of solving it if we consider that any S is non empty which is not restrictive since all the empty S_i should be removed in a preprocessing.

So we conclude by saying that the **EXACT SUM** problem is at least as difficult as the bound version of our problem θ-**REPRESENTABLE**.