Roundtable: towards a definition of interaction?

C. Labreuche

Thales Research & Technology
Palaiseau, France
email: christophe.labreuche@thalesgroup.com

workshop “Interacting criteria in MCDA”, Bruxelles, 2017

Models of interest

Focus: utility-based models (representing interaction)

- Choquet integral
- Multi-linear extension
- GAI model

But not the Sugeno integral.

Interaction indices: Case of a capacity

Case of a capacity

Interaction index [Murofushi, Soneda]

$$I_{i,j}(v) = \sum_{S \subseteq N \setminus \{i,j\}} \frac{|S|!(n-|S|-2)!}{(n-2)!} \left[v(S \cup \{ij\}) - v(S \cup \{i\}) - v(\{j\}) + v(S)\right]$$

Assertion

How to generalize this formula to any $U : X \to \mathbb{R}$?

Approach 1 – Suggestion from [Greco, Mousseau, Slowinski'12]

[Greco, Mousseau, Slowinski'12]

- GAI model, where the terms u_S (with $|S| = 2$) can take both positive and negative signs.
- Idea: Following the intuition on the Möbius transform, the sign of the interaction is given by the sign of the term $u_{i,j}$

Example

- Consider $U(x_1, x_2) = 2\,x_1 + x_2 - \max(x_1, x_2)$, which has a negative term.
- From the relation, $\min(x_1, x_2) + \max(x_1, x_2) = x_1 + x_2$,
- Equivalent expression: $U(x_1, x_2) = x_1 + \min(x_1, x_2)$.

Approach 1 – Suggestion from [Greco, Mousseau, Slowinski'12]

[Greco, Mousseau, Slowinski'12]

- GAI model, where the terms u_S (with $|S| = 2$) can take both positive and negative signs.
- Idea: Following the intuition on the Möbius transform, the sign of the interaction is given by the sign of the term $u_{i,j}$

Example

- Consider $U(x_1, x_2) = 2\, x_1 + x_2 - \max(x_1, x_2)$, which has a negative term.
- From the relation, $\min(x_1, x_2) + \max(x_1, x_2) = x_1 + x_2$,
- Equivalent expression: $U(x_1, x_2) = x_1 + \min(x_1, x_2)$.

Conclusion

- We have two equivalent expressions with opposite signs.
- So the sign of $u_{i,j}$ does not give information on the nature of interaction.

Theorem [Grabisch, Labreuche'2016]

Let us consider a GAI model U (with $\mathcal{S}$ restricted to singletons and pairs) that is monotone and non-negative. Then there exists non-negative and monotone functions u_S s.t. for all $x \in X$

$$U(x) = \sum_{1 \leq i \leq n} u_i(x_i) + \sum_{1 \leq i < j \leq n} u_{i,j}(x_i, x_j).$$

Approach 2 – Let us start with importance indices: Following Marichal view

Back to capacities

Importance index can be rewritten [Marichal]

$$\phi_i(v) = \int_X \frac{\partial C_v}{\partial x_i}(x)\, dx$$

Def 1 [Ridaoui, Grabisch, Labreuche ECSQARU 2017]

$$\phi_i(U) = \int_X \frac{\partial U}{\partial x_i}(x)\, dx$$

Example

For a single-peaked functions such as

- a triangle function $U(x) = \|x - z\|$,
- a Gaussian function,

then $\phi_i(U) = 0$, even though criteria have some importance.

Approach 2 – Let us start with importance indices: Following Marichal view

Analysis of the example

- In the definition of the Shapley value, the capacity is monotone w.r.t. each criterion
- But U might be non-monotone w.r.t. some attributes [... normalization role of the utility function u_i]
- So $\frac{\partial U}{\partial x_i}(x)$ is not necessarily non-negative

Def 2 [Ridaoui, Grabisch, Labreuche ADT 2017]

$$\phi_i(U) = \int_X \left| \frac{\partial U}{\partial x_i}(x) \right| \, dx$$

Approach 2 – Case of the interaction

Following Marichal view: Back to capacities

Interaction index can be rewritten [Marichal]

$$I_{i,j}(v) = \int_X \frac{\partial^2 C_v}{\partial x_i \partial x_j}(x)\, dx$$

Def 1 [Ridaoui, Grabisch, Labreuche ECSQARU 2017]

$$I_{i,j}(U) = \int_X \frac{\partial^2 U}{\partial x_i \partial x_j}(x)\, dx$$

$\Longrightarrow$ indicates the average sign of interaction (complementarity vs. substitutability)

Def 2 [Ridaoui, Grabisch, Labreuche ADT 2017]

$$I_{i,j}(v) = \int_X \left| \frac{\partial^2 U}{\partial x_i \partial x_j}(x) \right| dx$$

$\Longrightarrow$ indicates the average intensity of interaction (strong interaction vs. weak interaction)

Approach 3 – Sobol indices

ANOVA

Any multivariate function can be decomposed in the following way (ANOVA decomposition):

$$Y = F(Z) = F_\varnothing + \sum_{i=1}^{n} F_i(Z_i) + \sum_{i<j} F_{ij}(Z_i, Z_j) + \cdots + F_N(Z) = \sum_{S \subseteq N} F_S(Z_S),$$

with

$$\begin{aligned}
F_\varnothing &= \mathbb{E}[Y] \\
F_i(Z_i) &= \mathbb{E}[Y|Z_i] - F_\varnothing \\
F_{ij}(Z_i, Z_j) &= \mathbb{E}[Y|Z_i, Z_j] - F_i(Z_i) - F_j(Z_j) - F_\varnothing \\
&= \mathbb{E}[Y|Z_i, Z_j] - \mathbb{E}[Y|Z_i] - \mathbb{E}[Y|Z_j] + E[Y] \\
\vdots &= \vdots \\
F_S(Z_S) &= \mathbb{E}_{Z_{-S}}[Y|Z_S] - \sum_{T \subset S} F_T(Z_T) = \sum_{T \subseteq S} (-1)^{|S \setminus T|} \mathbb{E}_{Z_{-T}}[Y|Z_T] \\
\vdots &= \vdots \\
F_N(Z) &= \sum_{T \subseteq N} (-1)^{|N \setminus T|} \mathbb{E}_{Z_{-T}}[Y|Z_T].
\end{aligned}$$

Approach 3 – Sobol indices

Sobol indices

Variance of Y:

$$\mathrm{Var}[Y] = \sum_{\emptyset \neq S \subseteq N} \mathrm{Var}[F_S(Z_S)].$$

First-order Sobol' indices:

$$\frac{\mathrm{Var}[F_S(Z_S)]}{\mathrm{Var}[Y]}$$

Comparison of Approaches 2 and 3

Comparison Shapley-like interaction and Sobol indices

- Sensitivity analysis:
  - Interaction indices = average value of the partial derivative of C_v w.r.t. its components.

    It is thus the mean value of a local sensitivity analysis.
  - Sobol' indices = sensitivity analysis based on variance.

    They are used to identify which factors shall be fixed in order to reduce as much as possible the variance on the output variable.
- Commensurability among terms
  - Interaction indices: if $\phi_1 = 2\ I_{1,2}$, does it mean that the importance of criterion 1 is twice more important than the interaction between variables 1 and 2?
  - Such a comparison is possible with the Sobol' indices as the variance of the output variable is decomposed into the variance of each variable individually, each pair of variables, and so on.

    If $\mathrm{Var}[F_1] = 2\ \mathrm{Var}[F_{1,2}]$, then one can say that variable 1 alone is twice more influential than the interaction between variables 1 and 2.