Coalitional decompositions of quantities of interest: an input-centric point of view

M. Il Idrissi[†,1,2,3], N. Bousquet[§,1,3,4], F. Gamboa[§,2], B. Iooss[§,1,2,3], J-M. Loubes[§,2]

[†] PhD student [marouane.il-idrissi@edf.fr]. [§] PhD supervisor
PhD expected duration: March. 2021 - March. 2024

[1] EDF Lab Chatou, 6 Quai Watier, 78401 Chatou, France
[2] Institut de Mathématiques de Toulouse, 31062 Toulouse, France
[3] SINCLAIR AI Lab., Saclay, France
[4] Sorbonne Université, LPSM, 4 place Jussieu, Paris, France

Abstract

Coalitional decompositions of quantities of interest (QoI) are at the cornerstone of many machine learning interpretability methods [8] and global sensitivity analysis (GSA) [3]. Whenever one wishes to study the behavior of a black-box, input-output model, these decompositions allow to extract insights, and to quantify importance in a particular way. Essentially, they allow to decompose a statistical parameter related to a random output, which is assumed to be relevant to the underlying uncertainty quantification study, into elements related to each coalition (i.e., subset) of inputs.

Traditionally, such decompositions are achieved using a "model-centric" approach: given a random vector of $d \in \mathbb{N}^*$ inputs denoted $X = (X_1, \dots, X_d)^\top \sim P_X$, and a black-box model G, one seeks to decompose the random output $G(X)$ as the following sum:

$$G(X) = \sum_{A \in \mathcal{P}(D)} G_A(X_A), \tag{1}$$

where $\mathcal{P}(D)$ denotes the power-set (i.e., the set of subsets) of $D := \{1, \dots, d\}$. A classical example of QoI is the variance of the output, denoted $\mathbb{V}(G(X))$. Provided that the inputs are mutually independent, it leads to the variance coalitional decomposition:

$$\mathbb{V}(G(X)) = \sum_{A \in \mathcal{P}(D)} V_A, \quad \text{where} \quad V_A = \sum_{B \subseteq A} (-1)^{|A|-|B|} \mathbb{V}(\mathbb{E}[G(X) \mid X_A]),$$

which is nothing more than the Sobol-Hoeffding variance decomposition [6, 9].

However, one can adopt an "input-centric approach", leveraging tools from combinatorics. It relies on the fact that $(\mathcal{P}(D), \subseteq)$ forms a partially ordered set (poset), with a very particular algebraic structure: it is isomorphic to a Boolean lattice [4]. Leveraging Rota's extension of the Möbius inversion formula [7] to posets leads to the following result, which can be seen as a generalized Inclusion-Exclusion principle.

Corollary 1 (Möbius inversion formula on power-sets) *Let φ and ψ be functions from $\mathcal{P}(D)$ to an abelian group $\mathbb{A}$. Then the following equivalence holds:*

$$\varphi_A = \sum_{B \subseteq A} \psi_B, \quad \forall A \in \mathcal{P}(D) \quad \iff \quad \psi_A = \sum_{B \subseteq A} (-1)^{|A|-|B|} \varphi_B, \quad \forall A \in \mathcal{P}(D). \tag{2}$$

This result can be leveraged for coalitional QoI decompositions in the following way:

- The QoI, represented by φ_D, can be valued in an abelian group (and not necessarily $\mathbb{R}$) ;
- Let φ_A, $A \in \mathcal{P}(D)$, be given. Then define ψ_A as the RHS of (2) ;
- Since the both sides of (2) are equivalent, it leads, in particular, to the coalitional decomposition of φ_D as:
$$\varphi_D = \sum_{A \in \mathcal{P}(D)} \psi_A.$$

This approach is analogous to cooperative game theory [1], where the chosen value function is φ, and the $(\psi_A)_{A \in \mathcal{P}(D)}$ represent the Harsanyi dividends [5] of the cooperative game.

This input-centric approach allows to define decompositions for a broader range of QoIs and does not require the inputs to be independent. Adopting this point of view leads to some remarkable observations: "ANOVA-like" decompositions (e.g., variance, covariance matrix for multivariate outputs, MMD-ANOVA [2]) from the GSA litterature can be recovered, and proved to hold even if the inputs are not independent.

During this talk, the input-centric approach to coalitional decompositions is presented and discussed. Conditions for obtaining unambiguous and interpretable decompositions of very general QoIs are presented. Its link with the traditional model-centric approach is discussed as well, which paves the way towards the definition of theoretically suitable coalitional decompositions whenever the inputs are not independent.

Short biography (PhD student)

After graduating from ENSAI and Rennes 1 University in 2020, I started a CIFRE PhD track in March 2021 at EDF R&D and Institut de Mathématiques de Toulouse, working on the development of interpretability methods for ML models. My research interests are at the crossroads between sensitivity analysis and explainable artificial intelligence methods, with and emphasis on their theoretical foundations.

References

[1] J. M. Bilbao. *Cooperative Games on Combinatorial Structures*, volume 26 of *Theory and Decision Library*. Springer US, Boston, MA, 2000.

[2] S. Da Veiga. Kernel-based ANOVA decomposition and Shapley effects - Application to global sensitivity analysis. working paper or preprint, 2021.

[3] S. Da Veiga, F. Gamboa, B. Iooss, and C. Prieur. *Basics and Trends in Sensitivity Analysis. Theory and Practice in R.* SIAM, 2021.

[4] B. A. Davey and H. A. Priestley. *Introduction to Lattices and Order*. Cambridge University Press, Cambridge, 2 edition, 2002.

[5] J. C. Harsanyi. A Simplified Bargaining Model for the n-Person Cooperative Game. *International Economic Review*, 4(2):194–220, 1963.

[6] W. Hoeffding. A class of statistics with asymptotically normal distribution. *Annals of Mathematical Statistics*, 19(3):293–325, 1948.

[7] J. P. S. Kung, G-C. Rota, and C. Hung Yan. *Combinatorics: the Rota way*. Cambridge University Press, New York, 2012. OCLC: 1226672593.

[8] C. Molnar. *Interpretable Machine Learning*. Lean publishing edition, 2020.

[9] I.M Sobol'. Global sensitivity indices for nonlinear mathematical models and their monte carlo estimates. *Mathematics and Computers in Simulation*, 55(1):271–280, 2001.