Cooperative game theory and importance quantification

Marouane Il Idrissi[123], Nicolas Bousquet[234], Fabrice Gamboa[1], Bertrand Iooss[123], Jean-Michel Loubes[1]

[1]*Institut de Mathématiques de Toulouse, Université Toulouse III.*

[2]*EDF R&D, Lab Chatou.*

[3]*SINCLAIR AI Laboratory.*

[4]*LPSM, Sorbonne Université.*

Abstract

Being able to quantify the importance of random inputs of an input-output black-box model is at the cornerstone of the fields of sensitivity analysis (SA) and explainable artificial intelligence (XAI). To perform this task, methods such as *Shapley effects* and *SHAP* have received a lot of attention. The former offers a solution for output variance decomposition with non-independent inputs, and the latter proposes a way to decompose predictions of predictive models. Both of these methods are based upon the *Shapley values*, an allocation mechanism from the *cooperative game theory*.

This presentation aims a shedding light on the underlying mechanism behind the paradigm of cooperative games for input importance quantification. To that extent, a link is drawn an extension of the *Möbius inversion formula* to boolean lattices leading to coalitional decompositions [1]. Allocations can be seen as aggregations of such decomposition, leading to a more general view of the importance quantification problem.

This generalization is leveraged in order to solve a problem in the context of global SA with dependent inputs. The Shapley effects are known not to be able to detect exogenous inputs (i.e., variables not in the model). Using a different allocation, namely the *proportional values*, lead to interpretable importance indices with the ability to identify such inputs [2].

These indices are illustrated on a classical problem of surrogate modeling of a costly numerical model: the transmittance performance of an optical filter. It allows for clear and interpretable decision rules for feature selection and dimension reduction.

Keywords: Uncertainty quantification ; Sensitivity analysis ; Interpretability ; Importance quantification

AMS subject classifications: Primary: 62P30 ; Secondary: 91A12, 62H99

References

[1] Il Idrissi, M., Bousquet, N., Gamboa, F., Iooss, B., and Loubes, J. M. (2023). *On the coalitional decomposition of parameters of interest.* arXiv preprint arXiv:2301.02539.

[2] Herin, M., Il Idrissi, M., Chabridon, V. and Iooss, B. (2022). *Proportional marginal effects for global sensitivity analysis.* arXiv preprint arXiv:2210.13065.