



Contribution ID: 43

Type: **not specified**

Cooperative Game Theory and Importance Quantification

Thursday, 16 May 2024 11:00 (25 minutes)

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 cooperative game theory.

This presentation aims to shed light on the underlying mechanism behind the paradigm of cooperative games for input importance quantification. To that extent, a link is drawn with the Möbius inversion formula to boolean lattices leading to coalitional decompositions of quantities of interest. 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, leads to interpretable importance indices with the ability to identify such inputs.

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.

Type of presentation

Invited Talk

Primary author: IL IDRISSE, Marouane (Institut de Mathématiques de Toulouse)

Co-authors: IOOSS, Bertrand (EDF R&D); GAMBOA, Fabrice (Institut de Mathématiques de Toulouse); LOUBES, Jean-Michel (Institut de Mathématiques de Toulouse); Mr BOUSQUET, Nicolas (EDF R&D)

Presenter: IL IDRISSE, Marouane (Institut de Mathématiques de Toulouse)

Session Classification: Invited session

Track Classification: Spring Meeting