Contribution ID: **15** Type: **not specified**

Explaining numerical models with the Generalized Hoeffding decomposition

Additive functional decomposition of arbitrary functions of random elements, under the form of *high-dimensional model representations* is crucial for global sensitivity analysis and more generally understanding black-box models. Formally, for random inputs $X = (X_1, \dots, X_d)^\top$, and an output $G(X)$, it amounts to finding the unique decomposition
\begin{equation}
G(X) = \sum_{A \in D} G_A(X_A) \quad \quad \quad(1)
\end{equation}
where $D = \{1, \dots, d\}$, D is the set of subsets of D, and $G_A(X_A)$ are functions of the subset of input $X_A = (X_i)_{i \in A}$. Whenever the X_i are assumed to be mutually independent, such a decomposition is known as *Hoeffding's decomposition*. It is well known to allow the derivation of meaningful Sobol' indices for the analysis of the output variance, among others. Whenever the inputs are not assumed to be mutually independent, several generalizing approaches have been proposed in the literature, but at the price of imposing restrictive assumptions on the correlation structure or lacking interpretability.

By viewing random variables as measurable functions, we prove that a unique decomposition such as (1), for square-integrable black-box outputs $G(X)$, is indeed possible under fairly reasonable assumptions on the inputs:

1. Non-perfect functional dependence;
2. Non-degenerate stochastic dependence.

Novel sensitivity indices based on this generalized decomposition can be proposed, along with theoretical arguments to justify their relevance. They can disentangle effects due to interactions and due to the dependence structure. Such indices will be discussed, in light of recent results obtained for numerical models with multivariate Bernoulli inputs, used in numerous applications.

Special/ Invited session

Explainability_FR

Classification

Both methodology and application

Keywords

Hoeffding decomposition;sensitivity;interactions

Primary author: Mr FERRERE, Baptiste (Institut de Mathématiques de Toulouse)

Co-authors: Prof. GAMBOA, Fabrice (Institut de Mathématiques de Toulouse); Prof. LOUBES, Jean-Michel (Institut de Mathématiques de Toulouse); MURÉ, Joseph (EDF); Prof. BOUSQUET, Nicolas (EDF R&D)

Presenter: MURÉ, Joseph (EDF)

Track Classification: Explainability (by FR)