Hoeffding HDMR, Sobol' HDMR and the Shapley Value

Thierry, A., Mara
PIMENT EA 4518, University of Reunion, Le Tampon, Reunion

Let $y = f(\boldsymbol{x})$ be the quantity of interest, function of some predictors $\boldsymbol{x} = (x_1, \dots, x_d)$, also called features or inputs. We denote $\mathcal{D} = (1, \dots, d)$, $\mathcal{D}_{+i}$ any subset of $\mathcal{D}$ that contains the label $i \in \mathcal{D}$. We aim at assessing how sensitive is y to the x variables. To do so, it is convenient to refer to some high-dimensional model representation (hdmr) of f to assess the contribution of each input variable to y.

The first hdmr we consider is the one of W. Hoeffding [1], that is,

$$f(\boldsymbol{x}) = f_0^H + \sum_{i_1=1}^{d} f_{i_1}^H(x_{i_1}) + \sum_{i_2>i_1}^{d} f_{i_1,i_2}^H(x_{i_1}, x_{i_2}) + \dots + f_{1\dots d}^H(x_1, \dots, x_d) \tag{1}$$

where, $f_0^H = \mathbb{E}\left[f(\boldsymbol{x})\right]$ and $f_{\boldsymbol{\alpha}}^H = \sum_{\beta \subseteq \alpha} (-1)^{|\boldsymbol{\alpha}|-|\boldsymbol{\beta}|} \mathbb{E}\left[f(\boldsymbol{x})|\boldsymbol{x}_\beta\right]$, $\boldsymbol{\alpha} \subseteq \mathcal{D}$. Hoeffding hdmr is always unique but the summands are only orthogonal if the x-variables are independent of each other. Otherwise, it is not obvious to infer how the input variables contribute to $f(\boldsymbol{x})$ or to its variance $\mathbb{V}\left[f(\boldsymbol{x})\right]$ given that they can contribute alone or mutually (due to correlations and interactions). L. Shapley [5] derived some statistic $\phi_i(\boldsymbol{x})$ to assess the fair contribution of x_i to y. By fair, it is meant that mutual contributions are equally shared among the cooperating variables. It results that,

$$f(\boldsymbol{x}) = f_0^H + \sum_{i=1}^{d} \phi_i(\boldsymbol{x}) \tag{2}$$

and it can be shown that $\phi_i(\boldsymbol{x}) = \sum_{\alpha \subseteq \mathcal{D}_{+i}} \dfrac{f_\alpha^H(\boldsymbol{x}_\alpha)}{|\alpha|}$. Besides, by denoting $\boldsymbol{\phi} = (\phi_1, \dots, \phi_d)$ and $C = \mathrm{Cov}(\boldsymbol{\phi})$ the covariance matrix of $\boldsymbol{\phi}$, one obtains the variance-based Shapley value [3] as follows, $Sh_i = \sum_{j=1}^{d} C_{i,j}$.

The second hdmr we consider is the one of I.M. Sobol' [6], that stipulates that for any $\boldsymbol{u} \sim \mathcal{U}\left(0,1\right)^d$, we can write,

$$g(\boldsymbol{u}) = g_0 + \sum_{i_1=1}^{d} g_{i_1}(u_{i_1}) + \sum_{i_2>i_1}^{d} g_{i_1,i_2}(u_{i_1}, u_{i_2}) + \dots + g_{1\dots d}(u_1, \dots, u_d) \tag{3}$$

with the summands orthogonal to each other by imposing that $\int_0^1 g_\alpha(\boldsymbol{x}_\alpha) \mathrm{d}u_{i_k} = 0$, $\forall i_k \in \boldsymbol{\alpha}$. The hierarchical Rosenblatt transformation (RT) [4] provides the link between $\boldsymbol{u}$ and $\boldsymbol{x}$, as follows,

$$\begin{cases} u_{i_1} = F_{i_1}(x_{i_1}) \\ u_{i_2} = F_{i_2|i_1}(x_{i_2}|x_{i_1}) \\ \vdots \\ u_{i_d} = F_{i_d|\sim i_d}(x_{i_d}|\boldsymbol{x}_{\sim i_d}) \end{cases} \tag{4}$$

where $(i_1, \dots, i_d)$ is an arbitrary ordering of the set $(1, \dots, d)$, F_{i_1} is the marginal cumulative density function (cdf) of x_{i_1}, $F_{\alpha|\beta}$ is the conditional cdf of $\boldsymbol{x}_\alpha$ on $\boldsymbol{x}_\beta$ with $\alpha \cap \beta = \emptyset$. Obviously, $y = f(\boldsymbol{x}) = g(\boldsymbol{u})$, but neither the RT is unique (unless the variables be independent of each other) and nor the Sobol' hdmr.

We note that only $g_0 = f_0^H$ and $g_{i_1}(F_{i_1}^{-1}(u_{i_1})) = f_{i_1}^H(x_{i_1})$ when the inputs are not independent. From the Sobol' hdmr in Eq.(3) it is possible to compute the following variance-based sensitivity indices [2],

$$S_{x_{i_1}} = \frac{\mathbb{V}\left[\mathbb{E}\left[y|u_{i_1}\right]\right]}{\mathbb{V}\left[y\right]} = \frac{\mathbb{V}\left[g_{i_1}(u_{i_1})\right]}{\mathbb{V}\left[y\right]}, \tag{5}$$

$$ST^{ind}_{x_{i_d}} = \frac{\mathbb{E}\left[\mathbb{V}\left[y|\boldsymbol{u}_{\sim i_d}\right]\right]}{\mathbb{V}\left[y\right]} = \frac{\sum_{\alpha\subseteq\mathcal{D}_{+i_d}} \mathbb{V}\left[g_\alpha(u_\alpha)\right]}{\mathbb{V}\left[y\right]}, \tag{6}$$

$$S^{ind}_{x_{i_d}} = \frac{\mathbb{V}\left[\mathbb{E}\left[y|u_{i_d}\right]\right]}{\mathbb{V}\left[y\right]} = \frac{\mathbb{V}\left[g_{1\ldots d}(u_1,\ldots,u_d)\right]}{\mathbb{V}\left[y\right]}, \tag{7}$$

$$ST_{x_{i_1}} = \frac{\mathbb{E}\left[\mathbb{V}\left[y|\boldsymbol{u}_{\sim i_1}\right]\right]}{\mathbb{V}\left[y\right]} = \frac{\sum_{\alpha\subseteq\mathcal{D}_{+i_1}} \mathbb{V}\left[g_\alpha(u_\alpha)\right]}{\mathbb{V}\left[y\right]}, \tag{8}$$

$S_{x_{i_1}}$ is the amount of variance explained by x_{i_1} alone including its cooperative contribution due to its dependence on $\boldsymbol{x}_{\sim i_1}$ while $S^{ind}_{x_{i_d}}$ is the one of x_{i_d} that does not account for the mutual contribution. $ST_{x_{i_1}}$ is the amount of variance explained by x_{i_1} including all its mutual contributions (i.e. interactions+correlations) with $\boldsymbol{x}_{\sim i_1}$ while $ST^{ind}_{x_{i_d}}$ does not take into account contributions due to the dependences of x_{i_d} on $\boldsymbol{x}_{\sim i_d}$.

In my talk I will discuss the pros and the cons of the different approaches to analyze model responses or any given dataset and I will show some examples.

References

[1] W. Hoeffding. A class of statistics with asymptotically distribution. *Annals of Mathematics and Statistics*, 19:293–325, 1948.

[2] T. A. Mara and W. E. Becker. Polynomial chaos expansionfor sensitivity analysis of model output with dependent inputs. *Reliability Engineering and System Safety*, 214:107795, 2021.

[3] A. B. Owen and C. Prieur. Sobol' indices and shapley value. *J. of Uncertainty Quantification*, 2:245–251, 2014.

[4] M. Rosenblatt. Remarks on the multivariate transformation. *Annals of Mathematics and Statistics*, 43:470–472, 1952.

[5] L. S. Shapley. *A value for n-person games*, chapter Contribution to the Theory of Games II (Annals of Mathematics Studies 28), pages 307–317. Princeton University Press,, Princeton, NJ, 1953.

[6] I. M. Sobol'. Sensitivity estimates for nonlinear mathematical models. *Math. Mod. and Comput. Exp.*, 1:407–414, 1993.

[Thierry Mara; University of Reunion, 15 Avenue Rene Cassin, CS 92003, 97400 Saint-Denis, Reunion]
[thierry.mara@univ-reunion.fr – https://piment.univ-reunion.fr]