$\mathbb{L}_2$-Boosting on Generalized Hoeffding Decomposition for Dependent Variables - Application to Sensitivity Analysis

M. Champion, G. Chastaing, S. Gadat, C. Prieur

SIAM Conference on Uncertainty Quantification

March 31, 2014

Hierarchical Orthogonal Functional Decomposition

Model : $Y = f(X) + \varepsilon$, where f can be written as a sum of increasing dimension functions (ANOVA decomposition) :

$$\begin{aligned} f(X) &= \underbrace{f_\emptyset}_{\text{constant term}} + \sum_i \underbrace{f_i(X_i)}_{\text{main effects}} \\ &+ \sum_{1 \leq i < j \leq p} \underbrace{f_{ij}(X_i, X_j)}_{\text{interaction effects}} + \ldots + \underbrace{f_{1,\ldots,p}(X)}_{\text{residual}}. \end{aligned}$$

Theorem : HOFD

Under some assumptions on the density function of X, this decomposition is unique as soon as $\forall u, f_u \in H_u^0$, where :

$$H_u^0 = \{f_u, \langle f_u, f_v \rangle = 0, \forall v \subset u, \forall f_v\}.$$

How to determine the sparse HOFD ?

Algorithm 1: Greedy HOFD

Input: Orthonormal system $(\phi^i_{l_i})_{l_i=0}^L$ for $i \in [\![1, p]\!]$, i.i.d observations $\mathcal{O} := (X_j, Y_j)_{j=1\dots n}$.

Initialization : Split $\mathcal{O}$ in a partition $\mathcal{O}_1 \cup \mathcal{O}_2$ of size (n_1, n_2).

- For any $u \in S$, use a HOGS procedure with observations $\mathcal{O}_1$ to construct the approximation

$$\hat{H}_u^{L,0,n_1} := \text{Span}\{\hat{\phi}^u_{1,n_1}, \cdots, \hat{\phi}^u_{L_u,n_1}\} \text{ of } H_u^{L,0}.$$

- Use an $\mathbb{L}_2$-Boosting algorithm on $\mathcal{O}_2$ with the random dictionary

$$\mathcal{D} = \{\hat{\phi}^1_{1,n_1}, \cdots \hat{\phi}^1_{L,n_1}, \cdots, \hat{\phi}^u_{1,n_1}, \cdots, \hat{\phi}^u_{L_u,n_1}, \cdots\}$$

  to obtain the Sparse Hierarchically Orthogonal Decomposition.

Hierarchically Orthogonal Gram-Schmidt procedure

- To simplify, we only consider the interaction effects of order lower than 2.
- The procedure aims at constructing a basis of $H_u^{L,0}$ where $L \in \mathbb{N}^*$.

1. Initialization : $(\phi_{l_i}^i)_{l_i=0}^L$ basis of L^2. Then :

$$H_i^L = \text{Span}\{1, \phi_1^i, ..., \phi_L^i\}.$$

Hierarchically Orthogonal Gram-Schmidt procedure

- To simplify, we only consider the interaction effects of order lower than 2.
- The procedure aims at constructing a basis of $H_u^{L,0}$ where $L \in \mathbb{N}^*$.

2. Let $u = \{i, j\}$. To construct a basis for $H_{ij}^{L,0}$ which satisfies the hierarchical orthogonal constraints, we are looking for a basis of the form :

$$\phi_{l_{ij}}^{ij}(X_i, X_j) = \phi_{l_i}^{i}(X_i) \times \phi_{l_j}^{j}(X_j) + \sum_{k=1}^{L} \lambda_{k,l_{ij}}^{i} \phi_k^i(X_i) + \sum_{k=1}^{L} \lambda_{k,l_{ij}}^{j} \phi_k^j(X_j) + C_{l_{ij}},$$

where the constants are determined by resolving the constraints, that lead to a linear system :

$$A^{ij}\lambda = D^{l_{ij}}.$$

Empirical procedure

In practice, we use an empirical version of the HOGS procedure which consists in substituting the inner product by its empirical version :

$$\langle h, g \rangle_{n_1} = \frac{1}{n} \sum_{s=1}^{n_1} h(x^s) g(x^s).$$

The empirical version of the basis is then denoted as

$$\forall u, \;\; \hat{H}_u^{L,0,n_1} = \mathrm{Span}\{\hat{\phi}^u_{1,n_1}, \cdots, \hat{\phi}^u_{L_u,n_1}\}.$$

Empirical procedure

Assumptions on the model $(\mathbf{H_b})$:

$(\mathbf{H_b^1})$ $M := \sup_{i\in[\![1,p]\!], l_i\in[\![1,L]\!]} \left\|\phi^i_{l_i}(X_i)\right\|_\infty < +\infty$,

$(\mathbf{H_b^2})$ The number of variables p_n satisfies $\frac{\log p_n}{n} \underset{n\to+\infty}{\longrightarrow} 0$

$(\mathbf{H_b^{3,\vartheta}})$ The Gram matrices of the linear system satisfy :

$$\exists C > 0 \quad \forall (i,j) \in [\![1, p_n]\!]^2 \qquad det(A^{ij}) \geq Cn^{-\vartheta}.$$

Theorem

Under Assumptions $(\mathbf{H_b})$ *the sequence* $\left(\hat{\phi}^u_{l_u,n_1}\right)_u$ *satisfies :*

$$\sup_{u,l_u} \left\|\hat{\phi}^u_{l_u,n_1} - \phi^u_{l_u}\right\| \xrightarrow[n\to+\infty]{\mathbb{P}} 0.$$

Greedy selection of sparse HOFD

- f can be approximated by :

$$f(X) \simeq \bar{f}(X) = \sum_{u} \sum_{l_u} \beta^u_{l_u} \hat{\phi}^u_{l_u,n_1}(X_u).$$

- To estimate the coefficients $(\beta^u_{l_u})_{l_u}$, we use $\mathbb{L}_2$-Boosting algorithm on the second sample $(y^s, x^s)_{s=1,\dots,n_2}$. Denote $\mathcal{D}$ the dictionary of functions

$$\mathcal{D} = \{\hat{\phi}^1_{1,n_1}, \cdots \hat{\phi}^1_{L,n_1}, \cdots, \hat{\phi}^u_{1,n_1}, \cdots, \hat{\phi}^u_{L_u,n_1}, \cdots\},$$

and $G_k(\bar{f})$ the approximation of $\bar{f}$ at step k as a linear combination of elements of $\mathcal{D}$.

Algorithm 2: The $\mathbb{L}_2$-Boosting

Input: Observations $\mathcal{O}_2 := (y^s, x^s)_{s=1,\cdots,n_2}$, shrinkage parameters $\gamma \in]0,1]$ and number of iterations $k_{up} \in \mathbb{N}^*$.

Initialization : $G_0(\bar{f}) = 0$.

for $k = 1$ *to* k_{up} **do**

- Select $\hat{\phi}^{u_k}_{l_{u_k},n_1} \in \mathcal{D}$ such that

$$\left|\langle Y - G_{k-1}(\bar{f}), \hat{\phi}^{u_k}_{l_{u_k},n_1}\rangle_{n_2}\right| = \max_{\hat{\phi}^{u}_{l_u,n_1} \in \mathcal{D}} \left|\langle Y - G_{k-1}(\bar{f}), \hat{\phi}^{u}_{l_u,n_1}\rangle_{n_2}\right|.$$

- Compute the new approximation of $\bar{f}$ as

$$G_k(\bar{f}) = G_{k-1}(\bar{f}) + \gamma\langle Y - G_{k-1}(\bar{f}), \hat{\phi}^{u_k}_{l_{u_k},n_1}\rangle_{n_2} \cdot \hat{\phi}^{u_k}_{l_{u_k},n_1}.$$

end

Output: $\hat{f} = G_{k_{up}}(\bar{f})$.

Consistency of the algorithm

Other assumptions on the model :

($\mathbf{H}_{\varepsilon,\mathbf{q}}$) $\mathbb{E}(|\varepsilon|^q) < \infty$, for one $q \in \mathbb{R}_+$.

($\mathbf{H_s}$) The true parameter β^0 satisfies $\left\|\beta^0\right\|_1 < \infty$.

Theorem

Assume that functions $(\hat{\phi}^u_{l_u,n_1})_{l_u,u}$ *are estimated from the first sample* $\mathcal{O}_1$ *under* ($\mathbf{H_b}$)*. Then,* $\hat{f}$ *is defined on* $\mathcal{O}_2$ *as*

$$\hat{f}(X) = G_{k_n}(\bar{f}), \quad \text{with } \bar{f} = \sum_{u,l_u} \beta^{u,0}_{l_u} \hat{\phi}^u_{l_u,n_1}(X_u).$$

If we assume that ($\mathbf{H_s}$) *and* ($\mathbf{H}_{\varepsilon,\mathbf{q}}$) *are satisfied, then there exists a sequence* $k_n := C \log n$ *such that*

$$\|\hat{f} - \tilde{f}\| \xrightarrow[n \to +\infty]{\mathbb{P}} 0.$$