

TITRE DU SUJET DE RECHERCHE : Techniques d'Intelligence Artificielle pour l'analyse de sensibilité globale avec entrées dépendantes.

Nom du laboratoire : Institut Camille Jordan – ICJ - UMR 5208
Site web : <http://math.univ-lyon1.fr/>

Nom de l'équipe de recherche : Probabilité, statistique, physique mathématiques
Site web : <http://math.univ-lyon1.fr/spip.php?article12>

Nom du directeur de thèse : Véronique Maume-Deschamps, <http://math.univ-lyon1.fr/~maume>

Adresse Email : veronique.maume@univ-lyon1.fr

Etablissement : Université Claude Bernard Lyon 1

Clémentine Prieur, <https://ljk.imag.fr/membres/Clementine.Prieur/>, Université Grenoble Alpes, clementine.prieur@univ-grenoble-alpes.fr.

Ecole doctorale : ED 512 InfoMaths, <http://edinfomaths.universite-lyon.fr/l-ecole-doctorale-infomaths-ed-512--314590.kjsp>

Présentation du sujet : L'analyse de sensibilité globale est indispensable pour comprendre et optimiser des codes de calcul complexes; ceux-ci peuvent provenir de l'industrie, des sciences de l'environnement ... Les entrées de ces modèles ne sont pas entièrement spécifiées / connues et les incertitudes sur les entrées se répercutent sur l'incertitude de la sortie. Un des buts de l'analyse de sensibilité globale est de quantifier l'impact des incertitudes des entrées sur la sortie du modèle. Par exemple, on peut étudier comment la température et la pression à l'entrée d'un procédé industriel de production de gaz influe sur la pureté de gaz produit. Des mesures de sensibilité sont alors introduites permettant de mesurer l'impact de la température, de la pression ou de l'interaction des deux sur la pureté. En pratique bien sûr le nombre de paramètres d'entrée n'est pas limité à deux et peut même devenir très grand. Les indices de Sobol sont utilisés lorsque les entrées incertaines (dans l'exemple la température et la pression) peuvent être modélisées par des variables aléatoires indépendantes.

Dans de nombreuses applications industrielles, on ne peut pas considérer les entrées comme étant indépendantes. En effet, les entrées du système peuvent être le résultat d'une calibration ou d'une optimisation qui les rend dépendantes ou, comme dans l'exemple ci-dessous, être liées par des relations physiques. Quand les entrées du modèle complexe sont modélisées par des variables aléatoires dépendantes, les indices de Shapley fournissent des indices de sensibilité pertinents. Ils proposent une répartition équilibrée de la variance totale de la sortie entre les différentes entrées. D'un autre côté, l'analyse de sensibilité orientée quantiles (QOSA) s'intéresse à l'impact des variables d'entrées sur les quantiles de la variable de sortie.

Le calcul ou l'estimation des indices de Shapley et des indices QOSA sont coûteux en temps de calcul, essentiellement car il faut calculer / estimer des lois conditionnelles. Ainsi des méthodes de calcul efficaces doivent être développées, aussi bien pour les indices de Shapley que pour les indices QOSA.

Le but de cette thèse est de proposer de nouvelles méthodes d'estimation de ces indices, basées sur l'apprentissage statistique et l'intelligence artificielle (essentiellement des forêts aléatoires et des réseaux de neurones) afin d'obtenir des méthodes d'estimation

efficaces de ces indices. Les défis scientifiques sont aussi bien théoriques (consistance des estimateurs, comportement asymptotique, intervalles de confiance) que computationnels et orientés vers des applications concrètes.

La candidate ou le candidat devra avoir un niveau de master en mathématiques avec des compétences en statistique et probabilité. Il est aussi attendu des compétences et un goût pour la programmation (par exemple en R, Python, Julia ou C/C++ ...).

Références bibliographiques :

Athey S., Tibshirani J., Wager, S. *Generalized random forests*. ArXiv: 1610.01271.

Benoumechiara, N., Elie-Dit-Cosaque, K. (2019). *Shapley effects for sensitivity analysis with dependent inputs: bootstrap and kriging-based algorithms*. ESAIM: Proceedings and Surveys, 65, 266-293.

Breiman, L. (2001). *Random forests*. Machine learning, 45(1):5-32.

Devroye, L., Györfi, L., and Lugosi, G. (2013). *A probabilistic theory of pattern recognition*, volume 31. Springer Science & Business Media.

B. Iooss and C. Prieur (2019), *Shapley effects for sensitivity analysis with correlated inputs: comparisons with Sobol' indices, numerical estimation and applications*. To appear in International Journal for Uncertainty Quantification.

Maume-Deschamps, V. and Niang, I. (2018). *Estimation of quantile oriented sensitivity indices*. Statistics & Probability Letters, 134:122-127.

Meinshausen, N. (2006). *Quantile regression forests*. Journal of Machine Learning Research, 7(Jun):983-999.

Owen, A. B. (2014). Sobol' indices and Shapley value. *SIAM/ASA Journal on Uncertainty Quantification*, 2(1), 245-251.

Owen, A. B. and Prieur, C. (2017) *On Shapley value for measuring importance of dependent inputs*. SIAM/ASA Journal on Uncertainty Quantification, 5 (1), p.986-1002.

Scornet, E., Biau, G., & Vert, J. P. (2015). *Consistency of random forests*. The Annals of Statistics, 43(4), 1716-1741.

Shapley, L. S. (1953). Stochastic games. *Proceedings of the national academy of sciences*, 39(10), 1095-1100.

Sobol, I. M. (1993). *Sensitivity estimates for nonlinear mathematical models*. Mathematical Modelling and Computational Experiments, 1(4):407-414.

Mots-clés : Analyse de sensibilité globale, dépendance stochastique, apprentissage statistique.