Sujet de stage de fin d'études / M2 – Année universitaire 2020/2021

Étude de nouvelles métriques d'importance relative pour l'interprétabilité en apprentissage statistique

CONTEXTE :

Au sein d'EDF R&D, le département PRISME a pour mission de proposer des solutions innovantes pour une exploitation plus performante des différents moyens de production du groupe EDF : maîtrise des risques, amélioration de la sûreté, optimisation des performances technico-économiques, maîtrise de la durée de vie des installations et des composants. Au sein de PRISME, le groupe « Gestion d'Actifs, Incertitudes et Apprentissage statistique » participe avec TOTAL et Thalès au laboratoire commun SINCLAIR (*Saclay Industrial Collaborative Laboratory for Artificial Intelligence Research*), à des activités de R&D sur l'explicabilité des algorithmes couramment utilisés en IA (modèles simulés ou modèles appris). En effet, les modèles de *machine learning* (ML) sont souvent considérés comme des boîtes noires fournissant des prédictions difficiles à interpréter, ce qui peut limiter l'acceptabilité de ces nouveaux outils numériques. Les premiers travaux de SINCLAIR ont porté sur les mesures d'importance globale qui consistent à mesurer l'impact des variables d'entrée (ou *features*) sur les sorties d'un modèle ML[1]. Le verrou de la prise en compte des dépendances statistiques entre les entrées a été particulièrement étudié avec ses solutions basées sur des métriques issues de la théorie des jeux comme les indices de Shapley, dont les problématiques d'estimation ont été approfondies[2].

OBJECTIF DU STAGE :

Les indices de Shapley ne satisfont pas toutes les propriétés préconisées par certains auteurs, en particulier le principe d'exclusion qui veut que la mesure d'importance d'une entrée qui n'est pas présente dans le modèle ML soit nulle[3]. L'objectif du stage sera de développer des métriques permettant de pallier cette insuffisance. Ce stage s'articulera selon plusieurs phases menées éventuellement en parallèle :

1) Etude bibliographique : état de l'art sur les mesures d'importance (indépendantes du type de modèle ML) basées sur la variance (en particulier les indices de Shapley) : définitions, propriétés, algorithmes d'estimation, codes disponibles ;
2) Proposition et étude des propriétés (théoriques et numériques) d'une mesure d'importance alternative aux indices de Shapley et fondée sur le concept de « Proportional Marginal Decomposition »[3]. Etude de différents estimateurs possibles ;
3) Développement de scripts (R et/ou Python) et validation numérique sur des cas-test académiques et des jeux de données publics;
4) Application à des données EDF en apprentissage supervisé.

En fonction de l'avancement du stage et de l'appétence du stagiaire pour des aspects plutôt théoriques ou numériques, des études complémentaires mêlant d'autres formulations de métriques, des stratégies d'estimation avancées ou le développement des liens avec les techniques d'analyse de sensibilité de modèles numériques pourront être envisagées.

[1] Molnar, C. *Interpretable Machine Learning* (2019). https://christophm.github.io
[2] Broto, B. (2020). https://tel.archives-ouvertes.fr/tel-02976702
[3] Grömping, U. (2017). https://www.jstatsoft.org/article/view/v017i01

PROFILS :

Etudiant(e) de M2 maths appliquées/probabilités/statistiques ou d'école d'ingénieurs.

COMPETENCES SOUHAITEES :

▷ De solides compétences en statistiques, analyse de données et analyse numérique seraient appréciées.

▷ Aisance en programmation informatique (R ou Python).

▷ Aisance dans la communication, orale et écrite, en Français et/ou en Anglais.

ENVIRONNEMENT INFORMATIQUE :

▷ OS Linux ou Windows.

▷ R, Python.

▷ LaTeX.

CONTACTS :

bertrand.iooss@edf.fr

vincent.chabridon@edf.fr

DUREE ENVISAGEE :

5 mois à partir de mars ou avril 2021

LIEU :

EDF R&D – EDF Lab Chatou

Département Performance, Risque Industriel et Surveillance pour la Maintenance et l'Exploitation (PRISME)

6, Quai Watier - 78401 Chatou