Apprentissage artificiel pour classification non-équilibrée, avec applications à la cybersecurité

Numéro de la fiche : OPR-1469

Sommaire

DIRECTION DE RECHERCHE

Toby Dylan Hocking, Professeur - Département d'informatique

RENSEIGNEMENTS

toby.dylan.hocking@usherbrooke.ca

UNITÉ(S) ADMINISTRATIVE(S)

Faculté des sciences
Département d'informatique

CYCLE(S)

2e cycle
3e cycle

LIEU(X)

Campus de Sherbrooke

Description du projet

Dans le traitement de données volumineuses, il arrive souvent d'avoir plusieurs classes, dont certaines sont très minoritaires, ce qu'on appelle un problème non-équilibré. Par exemple, en classement des documents (par ex : données avec 99% courriel / 1% pourriel), en médicine (3% malades, 97% non), en traitement d'image (par ex : ImageNet avec 1000+ classes), ou grands modèles de langage (par ex : 50 000 tokens = classes). Les algorithmes d'apprentissage tel que réseaux de neurones peuvent alors être très lents à entraîner, ou bien donner des mauvais résultats pour les classes minoritaires. Nous avons proposé d'améliorer l'apprentissage dans ce cas, grâce à une nouvelle fonction de perte [Hillman and Hocking, 2023]. La méthode dans ce papier marche seulement pour le cas de 2 classes, et dans ce projet de doctorat nous proposons un projet de recherche sur une extension pour le cas de plusieurs classes (3 ou plus), et pour d'autres applications dans la cybersecurité.

Référence : Jonathan Hillman and Toby Dylan Hocking. Optimizing ROC curves with a sort-based surrogate loss for binary classification and changepoint detection. Journal of Machine Learning Research, 24(70):1–24, 2023. https://jmlr.org/papers/v24/21-0751.html

Demarche pour candidater. Lire et suivre les indications sur https://tdhock.github.io/blog/2024/application/

Financement offert

Oui

25 000 $ / an pour maîtrise, 35 000 $ / an pour doctorat

Partenaire(s)

Intact , Sherweb

La dernière mise à jour a été faite le 3 novembre 2026. L'Université se réserve le droit de modifier ses projets sans préavis.