

Stage de Master 2 : Impact de la configuration automatique multi-objectif de réseaux de neurones profonds sur l'interprétabilité

Les réseaux de neurones profonds (Deep Learning), qui sont efficaces en classification (prédiction des achats clients, classification des activités d'une société,...), bénéficient de moyens de calculs de plus en plus importants (Nvidia Tensor Core). La conception d'un réseau de neurones requiert de nombreux hyperparamètres qu'il faut adapter soigneusement au problème traité : nombre de nœuds, type de connexion entre les nœuds, fonction d'activation,... Actuellement, une recherche croissante est dirigée vers l'automatisation du paramétrage de ces algorithmes afin d'obtenir la meilleure efficacité. De plus, les réseaux de neurones adoptent une approche de type « boîte noire » difficile à interpréter. Ceci les rend difficilement compatibles avec une utilisation médicale selon le rapport du CCNE (Comité consultatif national d'éthique pour les sciences de la vie et de la santé) et pose un vrai défi d'explicabilité. Dans ce contexte, l'XAI (eXplainable Artificial Intelligence) s'intéresse à la création d'outils pour améliorer la compréhension de ce type d'approches.

Dans ce stage, nous nous intéresserons à 2 aspects.

1/ Le premier aspect concerne l'**hyper-paramétrage automatique de réseaux de neurones profonds**. Nous proposons d'utiliser des méthodes de configuration automatique d'algorithmes tels que paramILS [Hutter2009] pour déterminer la configuration d'hyperparamètres idéale. Il s'agira de générer des réseaux de neurones dont les caractéristiques sont adaptées au problème et à son environnement.

2/ Le second aspect concerne l'**interprétabilité de l'intelligence artificielle**. Dans un premier temps l'objectif est d'étudier les approches émergentes telles que BreakDown [Staniak2019] qui permettent d'augmenter l'interprétabilité des réseaux de neurones. Dans un deuxième temps, en étudiant grâce à l'approche proposée précédemment comment les hyper-paramètres influent sur la qualité de l'interprétation. Cela permettra d'élaborer une nouvelle version de l'approche proposée, qui maximise à la fois la qualité de l'interprétation et la performance du réseau.

Ce stage pourra éventuellement donner suite à une thèse. Il se déroule au sein de l'équipe ORKAD du laboratoire CRISAL de l'université de Lille (<https://www.cristal.univ-lille.fr/equipes/orkad/>).

Pour postuler, merci d'envoyer un courriel à julie.jacques@univ-catholille.fr avec les éléments suivants : CV, lettre de motivation, relevés de notes des deux dernières années d'études.

Références

[Blot2017a] *A. Blot, L. Jourdan and M-E. Marmion* : **Automatic Design of Multi-Objective Local Search Algorithms**, GECCO 2017, ACM pp. 227-234, 2017.

[Hutter2009] *F. Hutter, H.H. Hoos, K. Leyton-Brown, T. Stuetzle*, **ParamILS: An Automatic Algorithm Configuration Framework**, Journal of Artificial Intelligence Research. 36 (2009) 267–306.

[Jacques2013-a] *J. Jacques, J. Taillard, D. Delerue, L. Jourdan, and C. Dhaenens* : **MOCA-I: discovering rules and guiding decision maker in the context of partial classification in large and imbalanced datasets**, Learning and Intelligent OptimizationN, Lecture Notes in Computer Science (LNCS), 2013

Contacts

- Julie Jacques, Enseignant-Chercheur, équipe ORKAD (CRISAL / Université Catholique de Lille) Julie.jacques@univ-catholille.fr
- Clarisse DHAENENS, Professeur des universités (HDR), équipe ORKAD (CRISAL / Université de Lille) clarisse.dhaenens@univ-lille.fr