

Stage M2 au CEA : Nouvelles approches par réseaux de neurones profonds pour le traitement des données de spectrométrie de masse en santé

Contexte : La métabolomique par spectrométrie de masse (LC-MS) est une technologie puissante en santé pour découvrir de nouveaux marqueurs permettant de prédire la survenue d'une maladie ou la réponse à un traitement. Les données LC-MS, composées de deux dimensions (masse et temps de rétention des molécules), sont de grande taille, parcimonieuses et bruitées. Leur prétraitement (détection et quantification des signaux) est une étape clé. Les algorithmes existants (par transformée en ondelette continue dans les deux dimensions successivement [1]), génèrent de nombreux faux positifs et négatifs. Ces dernières années, l'apparition des approches d'apprentissage profond ouvre de nouvelles perspectives de traitement des données MS plus robustes directement à partir des données 2D [2].

Sujet : L'objectif du stage est de développer de nouvelles architectures profondes pour la détection et la quantification de l'ensemble des signaux et motifs présents dans les données d'un échantillon biologique.

Travail demandé : La première partie (traitement du signal et des images) sera consacrée à la construction de la base d'entraînement pour étiqueter les signaux connus dans les données disponibles au laboratoire. Des approches de simulation réalistes seront mises en œuvre pour augmenter la base d'apprentissage. La seconde partie (apprentissage profond) portera sur le développement d'architectures multitâches pour la segmentation de nuages de points, la détermination de contours, et le regroupement de signaux associés (e.g. par des mécanismes d'attention). L'ensemble des algorithmes seront implémentés en Python et appliqués au traitement des données de cohortes cliniques.

Mots clés : apprentissage profond, traitement d'image, détection d'objets, nuages de points, segmentation, spectrométrie de masse, santé

Références :

[1] Tautenhahn, R. *et al.* (2008) Highly sensitive feature detection for high resolution LC/MS. *BMC Bioinformatics*.

[2] Skarysz *et al.* (2018) Convolutional neural networks for automated targeted analysis of raw gas chromatography-mass spectrometry data. *International Joint Conference on Neural Networks (IJCNN)*.

Profil : Nous recherchons un.e candidat.e avec un très bon dossier en mathématiques appliquées (traitement d'images, statistiques, informatique), et motivé.e par les applications multidisciplinaires (chimie, physique, biologie). Le master pourra se prolonger par une thèse.

Lieu du stage : Le stage s'effectuera au centre CEA de Saclay, dans l'équipe science des données métabolomiques (Unité Médicaments et Technologies pour la Santé ; <https://odisce.github.io>), en partenariat avec l'équipe imagerie génétique (NeuroSpin ; <https://brainomics.org>).

Durée du stage : 6 mois à partir de mars 2024

Date limite de candidature : 30 novembre 2023

Contact : Merci d'adresser votre candidature (CV détaillé, lettre de motivation, notes de licence et master et nom de deux référents) à :

Etienne Thévenot (etienne.thevenot@cea.fr)

Equipe sciences des données (<https://odisce.github.io>)

Unité Médicaments et Technologies pour la Santé

CEA, Centre de Saclay, F-91191 Gif sur Yvette, France