

Support avancé « MD ConvLSTM » (Mars 2022 - Septembre 2022)

Nom du code : MD_ConvLSTM

Partenaire : Pierre-Yves Libouban, *Institut de Chimie Organique et Analytique, CNRS - UMR7311, Orléans*

Personnel IDRIS : Maxime Song, Camille Parisel

Description du code : MD_ConvLSTM est un projet de recherche sur l'amélioration des algorithmes de prédiction automatique de l'affinité de liaison entre une protéine et un ligand. L'intérêt est de pouvoir profiter des méthodes IA pour accélérer considérablement la découverte de nouvelles molécules susceptibles de devenir des médicaments.

Il existe déjà des modèles de Deep Learning capables d'interpréter une représentation 3D du couple protéine – ligand. Ce projet a pour objectif d'ajouter une dimension temporelle qui permettra d'apprécier l'évolution du comportement d'un ligand à proximité d'une protéine.

Côté responsable de projet, une part importante de la collaboration a consisté en la génération de simulations temporelles du couple protéine – ligand. En effet, plus la donnée est complexe (ici, l'ajout de la dimension temporelle à la dimension spatiale), plus il est nécessaire d'avoir des échantillons nombreux et représentatifs de la problématique à résoudre.

Travail effectué et résultats obtenus :

- Portage d'un code de référence pour la prédiction d'affinité à base d'images 3D afin de travailler avec des bibliothèques et une architecture modernes (passage de Tensorflow 1.2 à Pytorch 1.8, améliorations localisées de couches du modèle). Cela a servi de prise en main et de référence pour les performances des modèles plus complexes.
- Implémentation et évaluation de différentes architectures pour les données spatio-temporelles – motivées par la difficulté à générer un nombre conséquent de simulations.
- Adaptation du code pour être compatible avec un support multi-gpu (réduisant ainsi les temps d'entraînement).
- Intégration de Hydra et MLFlow : le premier facilite le choix des hyper paramètres pour les expériences (i.e. les entraînements), le second permet le suivi et la comparaison des résultats d'expériences.

Les codes produits servent actuellement à continuer les expériences destinées à évaluer les bons hyper paramètres et à valider les jeux de données. Des développements supplémentaires suite à ces résultats sont prévisibles.