Stage M2/Ingénieur : World model pour des données tabulaires

Sujet :

Il s'agit de développer des algorithmes en apprentissage automatique pour prédire l'état d'un patient à partir de données tabulaires. Les « modèles du monde » (« world models » en anglais) [1], en particulier l'architecture JEPA (Joint Embedding Predictive Architecture), sont très prometteurs. Ce sujet de stage propose d'étudier l'adaptation de l'architecture JEPA aux données tabulaires.

Objectif :

Ce sujet de stage s'intéresse à l'utilisation des modèles auto-supervisés du type World Model pour l'analyse de données tabulaires et, par extension, de données radiomiques. La radiomique est une discipline basée sur le calcul de descripteurs numériques à partir d'images médicales et de leur analyse pour prédire, par exemple, la nature d'une lésion.
Le but fondamental du stage est d'adapter l'utilisation du modèle JEPA aux données tabulaires. Une donnée tabulaire est une information structurée sous forme de tableau, organisée en lignes et en colonnes. Il s'agit de l'un des formats les plus répandus dans le monde professionnel. Une solution récente nommée T-JEPA [2] est très pertinente mais elle nécessite d'être étudiée plus attentivement. En particulier, l'approche T-JEPA s'appuie sur une représentation vectorielle (« embedding » en anglais) contextuelle des données qui mériterait d'être davantage interprétée, explicitée et modélisée. Les méthodes dérivées de T-JEPA comme Var-T-JEPA [3] seront également étudiées durant le stage.
Les architectures JEPA qui se seront avérées les plus pertinentes seront appliquées aux données radiomiques. Plusieurs jeux de données radiomiques sont disponibles en ligne [4, 5]. Il sera essentiel de produire des résultats expérimentaux clairs et bien documentés qui pourront être transmis à des médecins ou à des biologistes. D'autres données tabulaires pourront être étudiées selon les préférences du stagiaire.

Contexte :

Le pôle « Signal, Images, Systèmes (SIS) » du laboratoire I3S collabore sur ce sujet avec le Centre Hospitalier Universitaire (CHU) de Nice. Le stagiaire pourra donc bénéficier de l'expertise d'un professeur gastro-entérologue et hépatologue.

Profil recherché :

Formation niveau M2/Ingénieur en mathématiques appliquées et/ou en informatique.
Ces travaux requièrent les compétences suivantes :
- Programmation informatique : les développements informatiques seront réalisés en Python et Pytorch
- Notions de bases en machine learning.

Lieu du stage : campus SophiaTech (Sophia Antipolis).

Encadrant du stage : M. Lionel Fillatre (professeur des universités)
Courriel : lionel.fillatre@i3s.unice.fr
Site web : https://lionel-fillatre.github.io/post/

Durée : de 4 à 6 mois (démarrage en mars/avril 2027 selon les contraintes du stagiaire).

Rémunération : environ 650 euros par mois.

Bibliographie :

[1] M. Assran et al. Self-supervised learning from images with a joint-embedding predictive architecture. CVPR, 2023.
[2] H. Thimonier et al. T-JEPA: Augmentation-free self-supervised learning for tabular data. ICLR, 2025.
[3] M. Gögl and C. Yau. Var-JEPA: A Variational Formulation of the Joint-Embedding Predictive Architecture – Bridging Predictive and Generative Self-Supervised Learning. ICML, 2026.
[4] https://www.kaggle.com/c/glioma-radiomics/overview
[5] https://wiki.cancerimagingarchive.net/display/Public/NSCLC-Radiomics-Genomics