

Sujet de stage d'ingénieur et/ou master II

Titre du stage :

Deep Learning pour la santé

Objectif :

Le Deep Learning (ou apprentissage profond) est une technique d'analyse de données qui permet d'obtenir des résultats de classification très performants quel que soit le domaine applicatif. L'apprentissage profond est classiquement décomposé en deux étapes : le calcul des descripteurs profonds et la classification de ces descripteurs. Plus précisément, le réseau de neurones profonds transforme les données d'entrées (signaux, sons, images,...) en descripteurs complexes puis ces descripteurs sont exploités par un algorithme de classification (généralement un classifieur « softmax ») qui affecte une classe aux données traitées en fonction des descripteurs profonds calculés.

Ce sujet de stage s'intéresse à l'utilisation du Deep Learning pour l'analyse des données médicales. Il s'agit plus particulièrement de décider si un patient a développé ou non certaines maladies du foie en exploitant des données cliniques et des données biologiques. La stéato-hépatite non alcoolique du foie (connue du grand public sous le nom « maladie du foie gras » [1]) est la maladie dont la détection sera privilégiée pendant le stage. Dans ce but, nous disposons d'un jeu de données réelles étiquetées qui contient de nombreux attributs médicaux et plusieurs centaines de patients. Les étiquettes décrivent l'état du foie du patient. Le but du stage est d'utiliser un algorithme de Deep Learning supervisé [2] pour estimer l'état du foie du patient. Il s'agit en particulier de rechercher les transformations mathématiques non-linéaires induites par le réseau de neurones profond (« Deep Neural Network » [3]) qui favoriseraient cette estimation. Le stagiaire devra identifier les forces et les faiblesses, tant théoriques que pratiques, de cette approche pour le problème de classification considéré.

Contexte :

L'équipe MediaCoding du laboratoire I3S collabore sur ce sujet avec le Centre Hospitalier Universitaire (CHU) de Nice. Le stagiaire disposera donc de données médicales et cliniques de grande qualité. Il pourra également bénéficier de l'expertise d'un professeur gastro-entérologue et hépatologue.

Travail à réaliser :

L'étudiant aura différents travaux à réaliser :

1. Étudier les possibilités d'une architecture « Deep Learning » pour modéliser des relations non-linéaires entre les données médicales et l'état du foie.
2. Proposer un algorithme d'apprentissage supervisé qui exploite l'architecture analysée dans l'étape précédente.
3. Étudier l'efficacité de l'algorithme sur des données simulées qui reposent sur un modèle non-linéaire connu.
4. Tester l'efficacité de l'algorithme sur le jeu de données réelles.

Principaux livrables attendus :

1. Chaîne complète d'analyse des données en Python.
2. Description technique et mathématique de la chaîne complète d'analyse.
3. Évaluation quantitative des performances de la chaîne complète d'analyse.
4. Rapport de stage.
5. Diaporama qui résume tous les aspects du stage.

Les développements informatiques seront réalisés en Python avec les bibliothèques Scikit-Learn [4] (pour utiliser des algorithmes de machine learning) et Pytorch [5] (pour le Deep Learning).

Profil recherché :

Ces travaux requièrent les compétences suivantes :

- Programmation informatique (connaître préalablement Python n'est pas nécessaire mais souhaitable)
- Notions de bases en machine learning (réseau de neurones, SVM, arbre de décision, etc.)

Lieu du stage : campus SophiaTech (Sophia Antipolis).

Encadrant du stage :

- M. Lionel Fillatre (professeur des universités)
Courriel : lionel.fillatre@i3s.unice.fr

Durée : de 4 à 6 mois (démarrage en février ou mars 2020).

Rémunération : environ 550 euros par mois.

Bibliographie :

[1] <https://www.the-nash-education-program.com/fr/quest-ce-que-la-nash/>

[2] <https://see.stanford.edu/Course/CS229/47>

[3] <http://www.deeplearningbook.org>

[4] <https://scikit-learn.org/stable/>

[5] <https://pytorch.org>