

STAGIAIRES
JUSQU'À L'ÉTÉ 2022

Charles GRELLOIS

<http://www.grellois.fr>

charles.grellois@univ-amu.fr

1 Encadrement de stagiaires

1.1 Emilie Genty : premiers pas en machine learning, analyse de données médicales

Emilie (L1 puis L2 Mathématiques Physique Chimie Informatique (MPCI), Aix-Marseille Université) a été ma première stagiaire, pendant 5 semaines à l'été 2020. Je l'ai faite travailler sur les bases du machine learning avec Python : machine learning "standard" avec scikit-learn (SVM, Naive Bayes, Random Forest...) et sur du deep learning avec Keras, sur la base d'exemples trouvés sur Internet. C'était l'occasion pour elle de découvrir la discipline, et pour moi de m'y plonger un peu plus. Nous avons utilisé, entre autres, un jeu de données médicales sur le cancer du poumon, pour faire de la classification.

Elle devait faire son stage en 2021 dans un laboratoire de mathématiques et informatique pour l'oncologie en Afrique du Sud, qui a été interdit par notre université au dernier moment à cause de la situation sanitaire. Elle m'a donc recontacté et nous avons décidé de travailler à nouveau ensemble. Entretemps, j'avais réussi via ma collaboration naissante avec des médecins marseillais à récupérer des données médicales venant du projet MARLEN. Nous avons donc étudié le devenir cognitif d'enfants soignés pour un cancer du cerveau, le médulloblastome. Emilie a dû nettoyer les données, nous avons défini des métriques pour quantifier le devenir cognitif, puis nous avons croisé les données (avec Excel et Python) à la recherche de corrélations significatives. Des méthodes de clustering se sont montrées utiles. Ce stage a été une étape importante pour le projet, auquel je continue aujourd'hui à collaborer.

1.2 Pauline Bonnet : logique déontique

A l'été 2021, Pauline (L3 ENS Paris-Saclay) a fait un stage de 8 semaines avec Nicola Olivetti (LIS, Marseille) et moi sur des questions de logique modale. Il s'agissait d'étudier une logique déontique constructive basée sur les notions d'obligation et de permission : on peut formuler des énoncés logiques et, via des modalités, les rendre obligatoires ou simplement permis pour un agent. Ces questionnements modaux rejoignent des lignes de recherche en philosophie; cependant la partie technique relève des mathématiques et de la logique informatique. Nous avons donné une axiomatisation, un calcul de séquents et une sémantique pour cette logique, puis nous avons commencé à réfléchir à des extensions avec des opérateurs d'interdiction par exemple. Ce travail est toujours en cours avec Tiziano Dalmonte et Nicola Olivetti.

1.3 Rémy Citérin et Louis Jalouzot : machine learning et synthèse de programmes

Rémy Citérin (L3 ENS Ulm) et Louis Jalouzot (L3 ENS Lyon) ont effectué à l'été 2021 un stage de respectivement 8 et 6 semaines avec Nathanaël Fijalkow (LaBRI, Bordeaux) et moi, sur la synthèse de programmes en lien avec le challenge ARC. Les problèmes ARC ont été récemment posés par François Chollet [Cho19] comme une sorte de "test de QI" pour intelligence artificielle. L'approche développée durant le stage a suivi la ligne de recherche de Nathanaël Fijalkow : on a travaillé à générer des programmes résolvant quelques-uns des problèmes d'ARC. Il s'est agi d'identifier les "briques de base" pour construire des programmes-solution, puis de générer un grand nombre de problèmes afin d'entraîner un modèle de deep learning, qui sera ensuite utilisé pour guider les résolutions des problèmes ARC. Cette approche est bien entendu très préliminaire, le challenge étant très difficile.

1.4 Théo Esposito : initiation au machine learning

Théo (L3 MPCI, Aix-Marseille Université) a effectué un stage de 4 semaines avec moi à l'automne 2021. Il a découvert le machine learning : les aspects "classiques" avec scikit-learn puis le deep learning avec Keras, à l'aide du livre de François Chollet [Cho17]. Il a travaillé sur des données médicales issues de la littérature (classification sur la maladie de Parkinson), puis sur des images à classifier (via un défi Kaggle). Ensuite, il a travaillé sur de véritables données médicales que j'ai pu obtenir via mes relations avec les médecins marseillais. Il s'agit de données sur un cancer dermatologique, le mélanome : on a six données biologiques, et on sait s'il y a eu ou non récurrence après un certain temps. Théo a obtenu via du deep learning des modèles de la récurrence de précision et d'AUC intéressants.

1.5 Aimé Jean : mu-calcul et automates alternants à parité

Aimé (L3 MPCI, Aix-Marseille Université) a effectué un stage de 4 semaines avec moi à l'automne 2021. Il souhaitait faire un stage théorique, nous avons donc travaillé sur le lien entre le mu-calcul modal et les automates alternants à parité. Ces deux modèles sont équivalents, dans le sens où une formule de mu-calcul modal se traduit en un automate qui n'accepte que les arbres satisfaisant la formule ; et réciproquement tout automate se code en une formule équivalente. Mais il n'existe pas à ma connaissance de preuve *directe*, avec une traduction *directe*, entre les deux modèles. Aimé a su travailler vite et très bien, et nous avons ainsi une preuve d'équivalence qui, avec un peu de cosmétique, me semble présentable à un workshop.

1.6 Hadil Bechar et Romain Prokopp : un jeu multijoueur sans serveur

Je prévois d'encadrer ces deux étudiants de L3 MIASHS pour un mois à partir du 11 avril 2022. Ils souhaitent développer un jeu pour mettre en application leurs apprentissages en programmation. Je pense leur proposer un jeu multijoueur sans serveur, utilisant des notions d'algorithmique distribuée. Si nous en avons le temps, j'aimerais qu'ils implémentent une IA, même minimale, mais décentralisée et équitablement répartie. Je les ferai travailler ensemble, selon des méthodes de gestion de projet dont ils auront à rendre compte dans leur rapport.

2 Encadrement de projet

J'ai encadré cette année pendant quatre mois le projet de Master 1 de Robin Gauthey (M1 MIAGE). Celui-ci n'ayant pas choisi de faire une alternance, il devait travailler trois jours par semaine pendant ces quatre mois sur un sujet en collaboration avec moi.

J'ai repris des problématiques issues de ma collaboration avec les médecins marseillais, sous un angle plus technique que scientifique. Il s'est donc agi de créer un site Web qui soit une interface pour un "backend" Python effectuant des calculs complexes, puis une application mobile du même style. Robin s'est également essayé au machine learning sur un jeu de données fournies par les médecins.

Globalement, sur le site comme sur l'application mobile, l'idée est d'avoir un formulaire permettant au médecin d'entrer des données médicales. Puis celles-ci sont "transférées" à Python, qui réalise des calculs (évaluation avec un modèle de machine learning ou d'équations différentielles). Et l'utilisateur reçoit ensuite, sur le Web ou sur l'application, les résultats du calcul mis en forme (diagnostic, courbe d'évolution...).

Cette approche est préliminaire ; il reste encore du travail de réflexion avec les médecins autour de la conception de l'outil, mais nous avons maintenant tous les éléments techniques pour avancer rapidement.