Ecole Doctorale des Sciences Fondamentales

Titre de la thèse : Détection d'anomalies dans les grands relevés astronomiques

Directeur de thèse : GANGLER Emmanuel
Unité de rattachement : UMR 6533
Equipe : LSST
Etablissement de rattachement :
Courriel et téléphone : gangler@clermont.in2p3.fr 0473407842
Co-encadrant éventuel : DE OLIVEIRA ISHIDA Emille Eugenia
Unité de rattachement : LPC
Etablissement de rattachement : UCA CNRS/IN2P3

Résumé : Les données produites par les relevés astronomiques ont connu une augmentation constante de volume et de complexité. Le Zwicky Transient Facility (ZTF), en cours, a favorisé le développement de nombreuses techniques d'exploration et d'apprentissage automatiques de données. Celles-ci représentent des jalons importants pour répondre à nos besoins pour les prochaines années, mais il reste encore beaucoup à faire. À titre d'exemple, le Vera Rubin Observatory Legacy Survey of Space and Time (LSST), qui devrait commencer ses opérations en 2022, produira un ensemble de données plus vaste et plus complexe que n'importe lequel de ses prédécesseurs. Le relevé produira une image de l'ensemble du ciel austral tous les quelques jours pendant 10 ans, ce qui nous permettra d'étudier en détail le ciel transitoire (objets astronomiques dont la luminosité change en fonction du temps). Chaque nuit, le LSST devrait détecter 10 millions de candidats transitoires qui doivent être classés avant de pouvoir être utilisés pour produire des résultats scientifiques. Dans ce nouvel environnement Big Data, les anciennes méthodes d'analyse des données ne peuvent pas être utilisées. Il est nécessaire d'employer les techniques modernes d'apprentissage automatique qui peuvent aider le chercheur à identifier les événements scientifiquement intéressants.

En France, nous construisons actuellement l'infrastructure qui nous permettra d'accomplir cette tâche. Le broker Fink est un environnement logiciel capable de digérer le grand nombre de candidats produits par LSST et conçu pour identifier les objets intéressants. Cependant, la définition de ce qui est intéressant change en fonction de l'intérêt scientifique de chaque chercheur. L'infrastructure de Fink permet à chaque utilisateur de construire son propre filtre et d'identifier, dans le flux de données émergeant du télescope, quels sont les candidats qui répondent à ses intérêts scientifiques personnels. Souvent, ces filtres sont basés sur des techniques d'apprentissage automatique de pointe.

L'objectif de la thèse est d'étudier comment nous pouvons intégrer la connaissance du domaine dans le processus d'apprentissage automatique dans le contexte de la détection active d'anomalies. Les efforts seront consacrés sur 2 fronts principaux : comment concevoir et sélectionner des caractéristiques astrophysiques significatives à partir du flot des alertes ?

Comment adapter les techniques de détection active d'anomalies pour faire face aux données manquantes et aux erreurs de mesure, qui sont des caractéristiques intrinsèques des données astronomiques ? L'étude sera réalisée à l'aide de simulations (pour les contrôles de santé) et de données réelles de ZTF et plus tard sur LSST (si disponibles à la fin de la période de doctorat).

Le candidat retenu doit maîtriser les outils numériques et statistiques, ainsi que la programmation d'algorithmes spécifiques en Python. Une formation en techniques traditionnelles d'apprentissage automatique est également requise. Pour le contenu astronomique, une compréhension de base des systèmes photométriques, des sources transitoires et des processus d'observation est souhaitable.

Directeur : Pr Yanick HEURTEAUX
tél. 04.73.40.50.64
E-mail : yanick.heurteaux@uca.fr

Secrétariat : Lucie NAUD
tél. 04.73.40.53.76
E-mail : edsf.drv@uca.fr