

Classification de série temporelle de durée hautement variable dans un contexte astrophysique

Frédéric Comby¹, Jérôme Pasquet^{2,3}, Nancy Rodriguez¹, Marc Chaumont¹, Johanna Pasquet⁴, Dominique Fouchez⁴

Contacts : Marc Chaumont (marc.chaumont@lirmm.fr) / Jérôme Pasquet (jerome.pasquet@univ-montp3.fr)

1. LIRMM : 161 rue Ada, 34392 Montpellier. marc.chaumont@lirmm.fr, frederic.comby@lirmm.fr, nancy.rodriguez@lirmm.fr

2. TETIS : Maison de la télédétection, 500 rue Jean-François Breton 34093 Montpellier. jerome.pasquet@univ-montp3.fr

3. AMIS : Université Paul Valéry, Montpellier, France

4. CPPM : Campus Universitaire de Luminy, 163 Avenue de Luminy, 13009 Marseille Johanna Pasquet : pasquet@cprm.in2p3.fr – Dominique Fouchez : fouchez@cprm.in2p3.fr

Mots clés : *Traitement du signal, Classification, Machine-learning, Deep-Learning, Cosmologie.*



Fig.1: Représentation visuelle de ce à quoi pourrait ressembler une super nova (<https://www.supernovae.net/>)

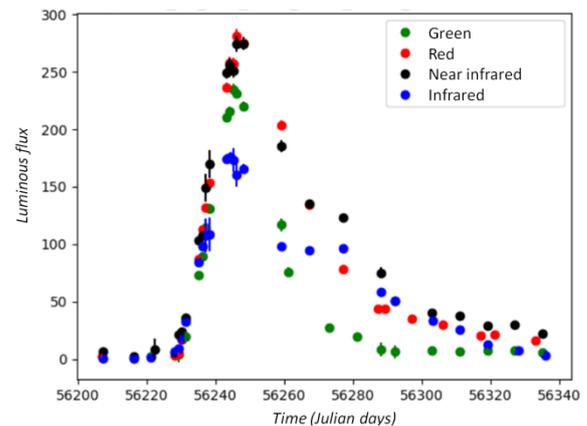


Fig.2 : Courbes de lumière simulée d'une supernovae

La cosmologie tend à comprendre l'origine, la composition et l'évolution de notre Univers. Pour cela il est essentiel d'identifier et de classer les différentes sources lumineuses dans le ciel. Cette tâche historiquement réalisée manuellement sur des plaques photos est désormais grandement améliorée par l'utilisation d'instruments optiques et électroniques perfectionnés ainsi que de méthodes du traitement de l'image efficaces pour extraire le flux des objets dans le ciel. Plus récemment, des algorithmes de classification automatique à base d'apprentissage profond (*deep-learning*) ont révolutionné la tâche de classification des objets astrophysiques. Ils peuvent traiter directement des images du ciel acquises dans plusieurs bandes spectrales et à différents instants, ou directement la série temporelle représentant la variation du flux d'un objet au cours du temps pour une bande spectrale donnée (voir sur la Figure 2, la série temporelle d'une supernova illustré en Figure 1). La série temporelle, très utilisé en astrophysique est appelé courbe de lumière, et permet de caractériser les objets astrophysiques ayant une variabilité temporelle.

La classification des séries temporelles est un véritable défi en astrophysique. Plusieurs méthodes ont déjà été testées. La plupart d'entre elles sont basées sur l'extraction de caractéristiques judicieusement choisies suivies par un apprentissage à base d'algorithmes de classification bien connus tels que les forêts aléatoires. En 2018, un challenge Kaggle (PLAstiCC) a été lancé par la communauté astrophysique afin de trouver le meilleur algorithme pour la classification de courbes de lumière. Les données proposées étaient des courbes de lumière simulées d'objets présentant une variabilité temporelle courte (moins d'un jour) et longue (plus d'une année).

Dans un premier temps le stagiaire devra effectuer un état de l'art des méthodes actuelles utilisées pour la classification en astrophysique [1], [2], [3]... Nous étudierons notamment la solution gagnante de PLAstiCC [4]. Les différentes architectures seront comparées entre elles, et en particulier avec une architecture spécifique basée sur le deep learning tel que PELICAN [5] développée par une partie de l'équipe encadrant.

Dans un second temps le stagiaire proposera des améliorations de la classification en utilisant des approches basées sur le deep learning. Cela pourra être une extension de l'approche PELICAN à des objets dont la variabilité temporelle est longue (PELICAN s'est classé 10ème au challenge PLAstiCC). Cela pourra être également l'utilisation de réseaux récurrents [5].

Notons que ce travail s'inscrit dans le cadre d'une collaboration entre Montpellier (LIRMM/TETIS) et Marseille (CPPM : Centre de Physique des Particules de Marseille) financé pour une durée de 4 ans par l'ANR. Une thèse en lien avec ce sujet débutera en septembre/octobre 2020.

Profil recherché : Master (M2) ou Ecole d'Ingénieur (3ème année) ayant une bonne maîtrise de la programmation (C++, Python, Tensorflow, PyTorch, etc), et une ou plusieurs connaissances en traitement des images, fouille de données / indexation / classification, architecture des machines/installation d'OS, Anglais écrit scientifique. **Aucune connaissance en astrophysique n'est requise.**

Modalité de candidature : Envoyez un CV, une lettre de motivation ainsi que votre relevé de notes de M1 le plus tôt possible. Après pré-sélection des candidatures, des entretiens téléphoniques ou en personne seront planifiés.

Contacts : Marc Chaumont (marc.chaumont@lirmm.fr) / Jérôme Pasquet (jerome.pasquet@univmontp3.fr)

Lieu du stage : Le stage se déroulera au LIRMM équipe ICAR (campus St Priest), Montpellier, France.

Période du stage : 1er semestre 2018 (5-6 mois).

Gratification de stage : environ 550€ mois.

Bibliographie:

[1] FATS: Feature Analysis for Time Series - I. Nun - 2015

[2] Photometric Supernova Classification with Machine Learning - M. Lochner et al. - APJs 2016

[3] A CNN adapted to time series for the classification of Supernovae, Anthony Brunel, Johanna Pasquet, Jérôme Pasquet, Nancy Rodriguez, Frédéric Comby, Dominique Fouchez, Marc Chaumont, EI'2019, in Proceedings of Color Imaging XXIV: Displaying, Processing, Hardcopy, and Applications - joint sessions dealing with color in astronomy / astrophysics, Part of IS&T International Symposium on Electronic Imaging, Burlingame (suburb of San Francisco), California USA, 13 - 17 January, 2019.

[4] Avocado: Photometric Classification of Astronomical Transients with Gaussian Process Augmentation - K. Boone 2019

[5] PELICAN: deep architecture for the Light Curve Analysis - J. Pasquet et al. - A&A 2019

[6] Long Short-Term Memory - S. Hochreiter - Neural Comput. 1997