MACHINE LEARNING FOR ASTROPHYSICS
2ND EDITION CATANIA, 8-12 JULY, 2024

Contribution ID: **215** Type: **not specified**

A semi-supervised ”cluster-then-label” scheme for photometric classification of evolved stars

Monday, 8 July 2024 11:50 (20 minutes)

Evolved stars are the main drivers of Galactic evolution from a chemo-dynamical perspective. Therefore, a comprehensive census of these sources is essential for the study of stellar feedback. However, identifying evolved stars is a time-consuming task that often involves spectroscopic observations. With the advent of next-generation telescopes (Roman, LSST) capable of delivering deep, all-sky photometric surveys at an unprecedented scale, the need for automated, photometry-based methods becomes apparent.

Deep learning offers an opportunity to leverage the vast amount of data available in astronomical archives. In particular, supervised classification techniques have proven to be an efficient solution for multiple science cases, though suffer from the need for (1) large annotated datasets (often incomplete, unreliable and time-expensive); and (2) from intrinsic class imbalances, that reflect the varied sizes of the different stellar populations.

In this talk, we propose a “”cluster-then-label”” scheme to mitigate these problems. Initially, we employ a small annotated dataset to explore the underlying structure of the data space using density-based clustering. Subsequently, new unlabelled data points are projected into this space and assigned labels based on their location relative to clusters of labelled objects. The annotated data points serve as a ground truth to inform a classifier, which then identifies the optimal decision boundaries to separate meaningful groups. We will present the preliminary results of this work, which has the potential to spot new evolved star candidates within a sample of unclassified sources.

Presenter: BORDIU, Cristobal

Session Classification: Past and future multiwavelength all-sky surveys