



Contribution ID: 113

Type: **Poster Presentation**

Classification of Evolved Stars with (Unsupervised) Machine Learning

Monday, 30 May 2022 14:56 (3 minutes)

The next-generation of observational astronomy instrumentation is expected to generate massively large and high complexity data volumes (big data) at rates of several gigabytes per second. Such enormous volumes impose extremely challenging demands on traditional approaches for data processing and analysis. Machine learning algorithms are playing an increasingly important role in detecting and classifying celestial objects in big data volumes. Our work is focused on analysing the effectiveness of unsupervised machine learning algorithms for classification of evolved stars based on multi-wavelength photometric measurements. The foundation is a custom made reference dataset compiled from available stellar catalogues for target sources - Asymptotic Giant Branch, Wolf Rayet, Luminous Blue variable and Red Supergiant stars. The dataset is composed of approximately 16,000 sources and features 8 independent colours retrieved from photometric catalogues - Wise, 2MASS and Gaia, spectral features were not considered within the dataset. Our experimental results indicate that the clustering algorithm HDBSCAN can utilise colours effectively to classify these sources, with the highest result having attained 65% accuracy. We further investigated the application of feature extraction methods to the dataset, including autoencoders and manifold learning algorithms UMAP and T-SNE. Our results show that these methods significantly improve clustering performance, most notably separating oxygen-rich and carbon-rich AGB stars, despite exhibiting very similar temperatures. Our best result was achieved by combining UMAP and HDBSCAN, attaining accuracy of 86%. We envisage that our findings can be replicated across other datasets containing photometric data, towards achieving even higher accuracies - to this extent we plan to perform a future systematic experimentation. We are also planning to make our ML pipeline available within the NEANIAS cloud-based science gateway to provide an easy-to-use interactive testbed environment, inviting domain scientists to design, realise, evaluate and optimise customised classification workflows for evolved stars.

Main Topic

Supervised/Unsupervised/Semi-supervised Learning

Secondary Topic

Classification and regression

Participation mode

In person

Primary authors: BORDIU, Cristobal (Istituto Nazionale di Astrofisica (INAF)); WELSH, Jamie (Student); SCI-ACCA, Eva (Istituto Nazionale di Astrofisica (INAF)); BUFANO, Filomena (Istituto Nazionale di Astrofisica (INAF)); TAN,

Jiacheng (Senior Lecturer, University of Portsmouth); KROKOS, Mel (Senior Lecturer, University of Portsmouth)

Presenter: WELSH, Jamie (Student)

Session Classification: Poster Session Day 1