Astronomical classification of objects using machine learning techniques

Walter A. Santos[a 1], Marcus V. C. Duarte[a], Alberto M. Benito[a], Laura M. Sampedro-Hernandez[a], Laerte Sodre Jr.[a], and Claudia Mendes de Oliveira[a]

[a]Instituto de Astronomia, Geofisica e Ciencias Atmosfericas, Universidade de Sao Paulo, Sao Paulo, SP, Brazil

Abstract

Any astronomical survey needs to face the challenge of characterizing the observed sources in terms of its astronomical nature; i.e., among stars, galaxies, quasars, planetary nebulae, supernovae,... or artifacts. Although deep, pencil-like surveys are mostly dominated by extragalactic sources (due to the cosmic volume they cover), the new generation of full-sky local Universe photometric surveys like S-PLUS (Mendes de Oliveira et al. 2016, in prep.) will be completely dominated instead by galactic objects, due to the photometric-depth they aim to achieve. Moreover, the ever increasing amount of sources/objects in astronomical surveys makes it harder for researchers to date mine, for themselves, the objects they are actually interesting in. Classification of objects provided in survey catalogs speed up data mining studies considerably, as evidenced by the widely successful Sloan Digital Sky Survey (SDSS) and the myriad of queries and samples made out of it.

In this work, we will present results on classification of objects in terms of separating star from galaxies (or point-like sources from extended), using all the relevant information available (their colors, brightness, relative size, etc) in the aforementioned S-PLUS survey. Special care has been taken while constructing representative training sets (from both empirical and simulated data) in order to apply machine learning techniques, notably, in this case, random forests.

Using independent, blind test samples, we prove that the methodology used can achieve great high rates of completeness and purity in the retrieved samples, ranging from 80% to practically 100% of completeness in the case of bright objects. In the near future, we intend to follow this procedure to not only separate galaxies from stars, but also to classify other types of astronomical objects, such as planetary nebulae, as well as sub-classifying different types of objects (such as stellar types) in unsupervised learning/categorization approach.

Keywords: object classification, machine learning, computational astrophysics.

[1]E-mail Corresponding Author: walter.augusto@gmail.com