The Role of the Human Expert in the Era of Big Data

Emille E. O. Ishida

CNRS/Laboratoire de Physique de Clermont (LPC), Université Clermont-Auvergne (UCA), Clermont Ferrand, France
emilleishida@gmail.com

The full exploitation of the next generation of large scale photometric surveys depends heavily on our ability to provide a reliable early-epoch classification based solely on photometric data. In preparation for this scenario, there have been many attempts to apply different machine learning algorithms to a series of classification problems in astronomy. Although different methods present different degrees of success, text-book machine learning methods fail to address the crucial issue of lack of representativeness between spectroscopic (training) and photometric (target) samples. In this talk I will show how Active Learning (or optimal experiment design) can be used as a tool for optimizing the construction of spectroscopic samples for classification purposes. I will present results on how the design of spectroscopic samples from the beginning of the survey can achieve optimal classification results with a much lower number of spectra and show how this strategy is being applied to the current ZTF alert stream by the Fink broker. I will also describe how such strategies have proven to be effective also in search for scientifically interesting anomalies within the efforts of the SNAD collaboration.