Machine learning: lessons learnt with the QUBRICS survey

Francesco Guarneri*[2,1], Giorgio Calderone[2], Stefano Cristiani[3,2], Andrea Grazian[4], Konstantina Boutsia[5], Guido Cupani[2], Fabio Fontanot[6], and Valentina D'odorico[7,2]

[2]Astronomical Observatory of Trieste – Italy
[1]Dipartimento di Fisica [Trieste] – Italy
[3]Istituto Nazionale di Fisica Nucleare, Sezione di Trieste – Italy
[4]INAF - Osservatorio Astronomico di Padova – Italy
[5]Las Campanas Observatory – Chile
[6]Astronomical Observatory of Trieste (INAF-OATs) – Italy
[7]Scuola Normale Superiore di Pisa – Italy

Abstract

Astronomical datasets are rapidly growing both in size and complexity: current and future surveys, such as Gaia, DESI, SKA, Euclid or the Rubin Observatory will generate an unprecedented amount of data, far beyond the feasibility of human review and analysis; automatic tools are needed to extract information hidden in these datasets.
Bright, high-redshift QSOs are examples of gems waiting to be dug out: especially in the Southern Hemisphere, their number is still relatively scarce. Since 2019, the QUBRICS survey (QUasars as BRIght beacons for Cosmology in the Southern hemisphere) aims to identify such sources based on current and future available photometric catalogs applying state-of-the-art machine learning techniques. After spectroscopic confirmation, over 400 new, bright ($i < 18$) and high-redshift ($z > 2.5$) QSOs have been identified in the last two years, using various techniques (e.g. CCA, PRF, XGB). This talk will describe the QSO selection algorithms, their performances and the current state of QUBRICS, highlighting a number of lessons, typical of Astronomical Surveys, but of general relevance for the field of machine learning, that we derived looking for the needles in the QUBRICS haystacwagnerk. These lessons are shaping our prospects in the hunt for cosmic beacons.
Slides: in PDF
Video: https://youtu.be/J3rouymZ1T8

*Speaker