Galaxy Evolution and Feedback across Different Environments
Proceedings IAU Symposium No. 359, 2020
T. Storchi-Bergmann, W. Forman, R. Overzier & R. Riffel, eds.
doi:10.1017/S1743921320001829

Classification and photometric redshift estimation of quasars in photometric surveys

L. M. Izuti Nakazono[1], C. Mendes de Oliveira[1], N. S. T. Hirata[2], S. Jeram[3], A. Gonzalez[3], S. Eikenberry[3], C. Queiroz[4], R. Abramo[4] and R. Overzier[5]

[1]Instituto de Astronomia, Geofísica e Ciências Atmosféricas da U. de São Paulo, Cidade Universitária, 05508-900, São Paulo, SP, Brazil

[2]Departamento de Ciência da Computação, Instituto de Matemática e Estatística da USP, Cidade Universitária, 05508-090, São Paulo, SP, Brazil

[3]Department of Astronomy, University of Florida, 211 Bryant Space Center, Gainesville, FL 32611, USA

[4]Departamento de Física Matemática, Instituto de Física, Universidade de São Paulo, SP, Rua do Matão 1371, São Paulo, Brazil

[5]Observatório Nacional / MCTIC, Rua General José Cristino 77, Rio de Janeiro, RJ, 20921-400, Brazil

Abstract. We present a machine learning methodology to separate quasars from galaxies and stars using data from S-PLUS in the Stripe-82 region. In terms of quasar classification, we achieved 95.49% for precision and 95.26% for recall using a Random Forest algorithm. For photometric redshift estimation, we obtained a precision of 6% using k-Nearest Neighbour.

Keywords. methods: statistical, catalogs, quasars: general

1. Introduction

Several techniques have been developed to classify extended and point-like sources on astronomical images. However, the classification of stars and quasars remains a challenge, as both are unresolved sources. Some techniques have been applied to perform star/quasar separation on photometric data, such as: colour-colour cuts (e.g. Pâris *et al.* 2018); Bayesian Statistics (e.g. Yang *et al.* 2017), and Machine Learning algorithms (e.g. Jin *et al.* 2019). In this work we use the data from 336 deg^2 of The Southern Photometric Local Universe Survey (S-PLUS) first data release (DR1) to create a new quasar catalog. This survey will cover 9300 deg^2 with the Javalambre optical filter system consisting of 7 narrow bands and 5 bands similar to the Sloan Digital Sky Survey (SDSS) broad-bands.

2. Methodology

The quasar sample from S-PLUS DR1 has 13 683 sources with near-infrared counterpart from Wide-field Infrared Survey (WISE) and their spectra were obtained from SDSS DR14 (Pâris *et al.* 2018). We retrieved from SDSS DR15 a total of 52 914 stars and 84 723 galaxies with infrared counterpart from AllWISE catalog. With these samples, we trained a Random Forest algorithm that classifies any S-PLUS observation in quasar, star, or galaxy. The performance of this classifier was evaluated through 4-Fold cross-validation in terms of precision/purity, recall/completeness and F-measure (harmonic mean of precision and recall). We tested the following feature spaces: (i) 12 S-PLUS bands;

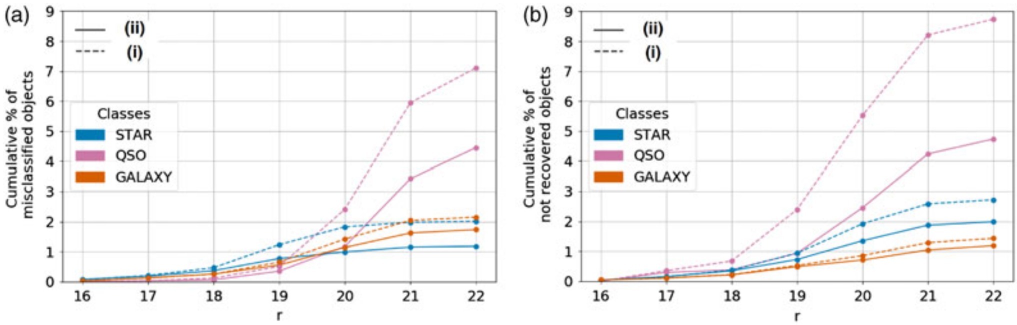

Figure 1. Cumulative percentages of (a) misclassified and (b) not recovered objects per magnitude and per class. The models referring to the feature spaces (ii) and (i) are represented by a solid and dashed curve, respectively.

(ii) 12 S-PLUS bands + 2 WISE bands; (iii) 5 S-PLUS broad-bands and (iv) 5 S-PLUS broad-bands + 2 WISE bands. We also included four morphological parameters: full width at half maximum (FWHM), major semi-axis, minor semi-axis and Kron radius. For the redshift estimation, we trained a k-Nearest Neighbours (kNN) algorithm. The performance of this model was evaluated through Holdout cross-validation in terms of normalized median absolute deviation (σ_{NMAD}). We considered the same feature spaces (i), (ii), (iii) and (iv) without the morphological parameters and with addition of 4 more spaces based on colors rather than magnitudes.

3. Results

For the classification problem, the best performances came from the feature spaces (ii) and (i). We achieved a macro-averaged F-measure of 97.44% (95.95%), with WISE (without WISE). In terms of quasar classification, we achieved 95.49% (92.83%) for precision and 95.26% (91.23%) for recall (Fig. 1). In terms of star classification, we achieved a precision of 98.82% (97.98%) and recall of 98% (97.26%). Finally for galaxies, a precision of 98.26% (97.83%) and a recall of 98.8% (98.56%) were obtained. The best precision for the photometric redshift estimation achieved is $\sigma_z = 6.56\%$ considering $k = 5$ when training kNN within the feature space 12 S-PLUS bands + 2 WISE bands with colors. The narrow-bands improved the star/quasar/galaxy classification, especially when no WISE information is available. We conclude that the S-PLUS optical system is a powerful tool for finding new quasars and for estimating their photometric redshifts.

References

Moore, J. A., Pimbblet, K. A., & Drinkwater, M. J. 2006, *PASA*, 23, 135
Pâris, I., Petitjean, P. Aubourg, É., *et al.* 2018, *A&A*, 613, A51
Heintz, K. E., Fynbo, J. P. U., Høg, E., *et al.* 2018, *A&A*, 615, L8
Yang, Q., Wu, X., Fan, X., *et al.* 2017, *AJ*, 154, 269
Jin, X., Zhang, Y., Zhang, J., *et al.* 2019, *MNRAS*, 485, 4539
Mendes de Oliveira, C., Ribeiro, T. , Schoenell, W., *et al.* 2019, *MNRAS*, 489, 241

https://doi.org/10.1017/S1743921320001829 Published online by Cambridge University Press