Boletim da Sociedade Astronômica Brasileira, **32**, no. 1, 50-51

Photometric redshifts for S-PLUS using machine learning techniques

E. Vinicius-Lima, L. Sodré Jr., & the S-PLUS collaboration.

[1] Instituto de Astronomia, Geofísica e Ciências Atmosféricas, USP; e-mail: erik.vini@usp.br,laerte.sodre@iag.usp.br

Abstract. The distance to celestial objects is a fundamental quantity for studies in astronomy and cosmology. Until recently, the only way to obtain this information was via spectroscopy, but the increasingly bigger surveys, with enormous amounts of data, made this approach infeasible. Therefore a new way to estimate the distance to objects, based on photometry, was developed. Photometric redshifts can be acquired for many objects in a time-inexpensive manner. In this work, the objective is to investigate how machine learning methods perform when using the 12 filter system of S-PLUS. Also, a comparison with the currently used method for this purpose (BPZ) is done. The results show that S-PLUS has potential to acquire accurate photometric redshifts using machine learning techniques

Resumo. A distância para objetos celestiais é uma quantidade fundamental para estudos em astronomia e cosmologia. Até recentemente, a única forma de obter esta informação era via espectroscopia, mas o desenvolvimento de mapeamentos cada vez maiores, com enormes quantidades de dados, fez com que esta abordagem se tornasse inviável. Portanto, uma nova forma de estimar a distância à estes objetos, baseada em fotometria, foi desenvolvida. Redshifts fotométricos podem ser adquiridos para muitos objetos e forma rápida. O objetivo deste trabalho é investigar a performance de códigos de aprendizado de máquina quando é utilizado o sistema de 12 filtros do S-PLUS. Também é feita uma comparação com o método atualmente utilizado para este fim (BPZ). Os resultados mostram que o S-PLUS tem potencial para obter redshifts fotométricos de alta precisão com técnicas de aprendizado de máquina.

Keywords. Galaxies: distances and redshifts – Surveys – Techniques: photometric – Galaxy surveys

1. Introduction

Given the importance of precise measurements of redshift to constrain cosmological parameters, probe the evolution and formation of galaxies and study the large-scale structures of the universe, a number of different surveys were developed including SDSS. More recently, stronomical projects reached a gigantic scale with the DES and LSST surveys, capable of probing hundreds of millions of galaxies. Due to bservational constraints, the acquisition of spectroscopic redshifts is impossible for all these objects. As such, new techniques were eveloped to obtain the redshift based only on photometry. Nowadays there are two main approaches to determine this quantity. One is via emplate-fitting (TF) and the other is with machine learning (ML). The focus of this work is on machine learning techniques. The data used omes from a cross-match between S-PLUS DR1 and SDSS/WISE catalogues.

2. Methodology

Three different ML methods were used in this work, but each one is based on a different approach. They are compared to the TF code BPZ.

ANNz2: Artificial Neural Networks for photometric redshifts (2.0) is based on shallow neural networks and boosted decision/regression trees (BDTs). Only BDTs were used in this research and they work by making splits of a training dataset, in order to narrow down the target variable (i.e redshift).

GPz: Gaussian Processes (GPs) for photometric redshifts is a method that uses sparse GPs which describe the training dataset using a limited number of basis functions. This way the regression target can be generated by a linear combination of non-linear functions plus a noise term.

Keras: Keras is a Python API (application programming interface) that provides ease of access to low-level machine learning languages such as TensorFlow, Theano or CNTK. This method is more commonly applied to deep-learning problems.

BPZ: Bayesian Photometric Redshifts is a template fitting code, which is based on a comparison between an objects' spectral energy distribution and a library of model spectra. Then, using some sort of optimization metric, such as a minimization, it is possible to determine the best photometric redshift of the object.

3. Datasets

The S-PLUS Data-Release 1 catalogue (Mendes de Oliveira et al. 2019) contains photometry on the twelve bands using three different apertures. The "_auto" magnitudes are total restricted magnitudes, containing most the the light from the galaxy while retaining a high signal-to-noise. For this reason, they are ideal for photometric redshift determinations (Molino et al. 2017).

The S-PLUS photometric redshifts (Molino et al. 2019) are obtained using the second version of the template-fitting algorithm BPZ (Benítez 2000), which is not publicly available. This code is capable of using a luminosity-based prior to weight the likelihood-function obtained from the comparison of data and templates, and also uses optimized SED models for redshift determination for the local universe..

The SDSS - Data Release 15 (Aguado et al. 2019) is the latest data release of the project, and includes all previous releases. The main addition to the SDSS data in this release comes from the MaNGA survey (Bundy et al. 2015), designed to obtain spectra for nearby galaxies using integral field units. DR15 also had an update regarding the flux calibration schemes (details can be found on the project website). The spectroscopic redshifts and classifications used in this work are obtained using the "spec1d" pipeline (Bolton et al. 2012).

The WISE survey (Wright et al. 2010) is a project whose main mission was executed between 2009 and 2011, and which

Table 1. Metrics, denotations and definitons that are used to analyse the results of this work.

Metric	Denoted by	Equation
Delta	δz	$z_{\text{phot}} - z_{\text{spec}}$
Scatter	σ_{NMAD}	$1.48 \times \text{median}\left(\frac{\lvert\delta z - \text{median}(\delta z)\rvert}{1 + z_{\text{spec}}}\right)$
Bias	μ	$\overline{\delta z}$
Outlier Fraction	η	$\frac{\lvert\delta z\rvert}{1 + z_{\text{spec}}} > 0.15$

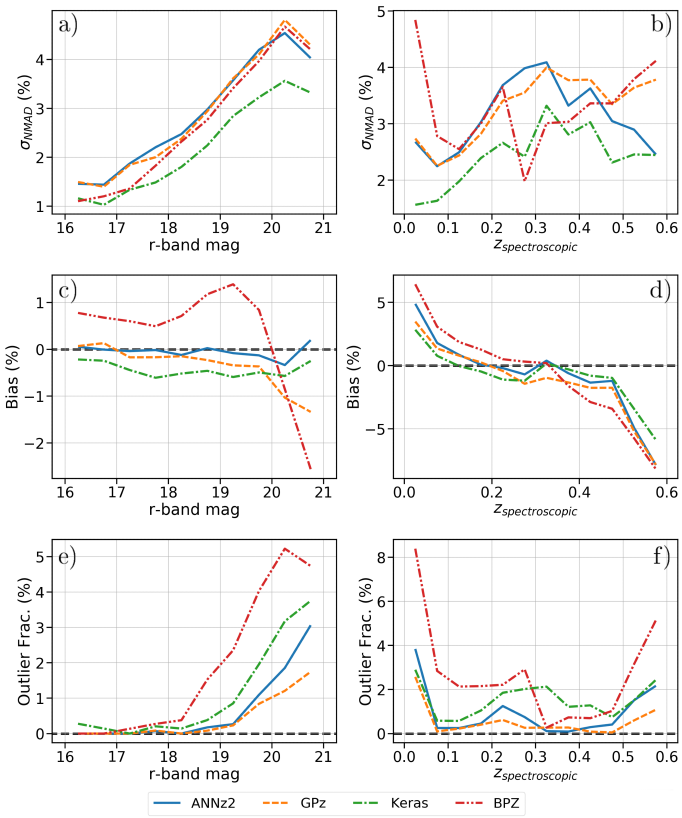

Figure 1. Comparison of results obtained for different methods.

objectives were to study asteroids, stars and luminous galaxies. The project is space-based with an infrared telescope able of imaging the whole sky. The filters used are called W1, W2, W3 and W4, and their effective wavelengths are 3.4, 4.6, 12.1 and 22.2 micrometers, respectively.

4. Results

It is important to define the metrics that will be used to evaluate the models. The metrics chosen are the same presented in Molino et al. (2019). In order to provide a fair comparison between the ML methods and BPZ, the following metrics were adopted:

As can be seen in Figure 1, the machine learning methods had a better overall performance than BPZ in both magnitude and redshift bins. Specially when the bias and fraction of outliers is analyzed. Also, the results obtained with the deep-learning code Keras were the most accurate (lowest σ_{NMAD}), with a difference as high as 1% for fainter objects.

5. Conclusions

Surveys with narrow-band filters, such as S-PLUS, J-PLUS and J-PAS provide an excellent opportunity to acquire accurate results using machine learning methods. Also, comparatively, the methods presented in this work provided better results than those obtained with the template-fitting based BPZ for the same testing sample.

Acknowledgements. The author acknowledges the financial support provided by CNPq (Conselho Nacional de Desenvolvimento Científico e Tecnológico) during this research.

References

Mendes de Oliveira, C. et al., *MNRAS*, p. 2048 , 2019.
Molino, A. et al., *MNRAS*, vol. 470, no. 1, pp. 95–113 , 2017.
Molino, A. et al., *arXiv e-prints* , 2019.
Benítez, N., *ApJ*, vol. 536, pp. 571–583 , 2000.
Aguado, D. S. et al., *ApJS*, vol. 240, no. 2, 23 , 2019.
Bundy, K. et al., *ApJ*, vol. 798, 7 , 2015.
Bolton, A. S. et al., *AJ*, vol. 144, 144 , 2012.
Wright, E. L. et al., *AJ*, vol. 140, pp. 1868–1881 , 2010.