Classifying Quasars and Predicting their Photometric Redshifts using DECaLS and WISE Data

Rahul Gupta and Varvara Semenova; Supervisor: Boris Leistedt; Astrophysics

Imperial College London

Introduction & Context

Quasi-stellar objects (quasars; QSOs) are some of the brightest objects in the Universe. They hold clues to processes on the largest cosmological scale, e.g. large-scale structure formation, black hole formation and dynamics of the primordial Universe. Spectroscopically, quasars can easily be distinguished from stars. However, with the advent of new cosmological surveys (e.g. Euclid, LSST), we need a less expensive and time-consuming method.

We use measurements of emitted flux in broad **optical bands** (g, r, z) from DECaLS and **infrared bands** (W1 and W2) from WISE instruments that constitute the largest and deepest surveys in those frequencies at present.

After data cleaning, we model the star and quasar populations using **Gaussian Mixture Models** [1]. We build a **Binary Bayesian Classifier** [2] capable of identifying quasars in a test set. We evaluate the **purity** and **completeness** of the classifier. Then, we use GMMs to predict **photometric redshift** (photo-z) of quasars which is a bonus of this approach [3, 4, 5].

$$\text{purity} = \frac{\text{no. of correctly identified QSOs}}{\text{no. of objects classified as QSOs}}$$

$$\text{completeness} = \frac{\text{no. of correctly identified QSOs}}{\text{total no. of QSOs in the test sample}}$$

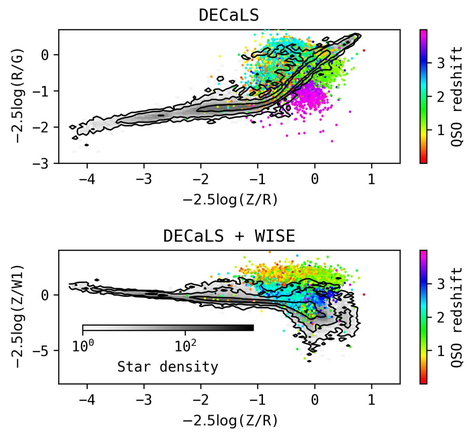

Figure 1: Flux values for 50000 randomly selected stars (black density plot) and quasars (coloured dots). The significant overlap in star and quasar loci in optical flux space (top) makes it difficult to differentiate objects. Adding WISE bands (bottom) helps to separate them but still a probabilistic approach is needed.

Methodology

1. Data Cleaning
2. Fit **GMMs** to stars and quasar data (incl. redshift for quasars)
3. Build a binary **classifier** to find probability of an object being a star/quasar
4. Estimate **photometric redshift** for quasars using conditional GMMs

Investigate classifier performance against redshift, r-band flux, position on the sky, etc.

Evaluate performance against observed spectroscopic redshift (spectro-z)

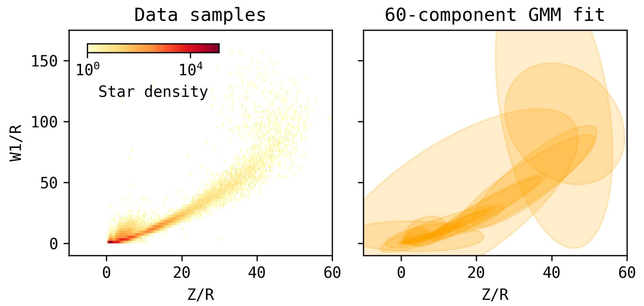

Figure 2: Flux ratios Z/R and W1/R for all 312k star samples (left) and the 60 Gaussian components fitted to the data (right).

Classifier Performance

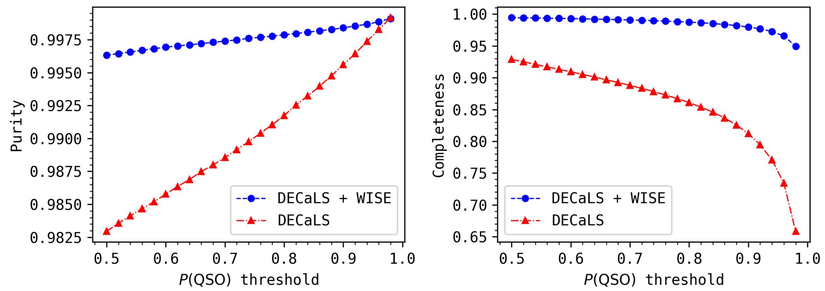

Figure 3: Purity (left) and completeness (right) as a function of posterior quasar probability threshold using photometry from DECaLS + WISE (blue) and DECaLS only (red). In both cases, adding WISE photometry improves the performance of the classifier.

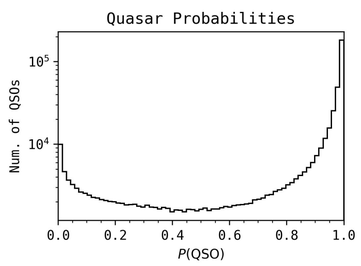

Figure 4: The probability of a spectroscopically confirmed quasar in the test sample being a quasar according to our classifier.

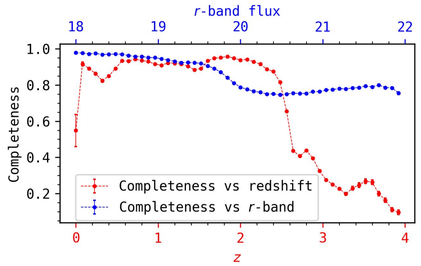

Figure 5: Completeness as a function of r-band flux (blue) and redshift (red). Completeness decreases sharply at z = 2.2 (where quasar and star loci overlap most), and at r-flux ~ 19.7 (maximum PSF depth, see Fig. 6).

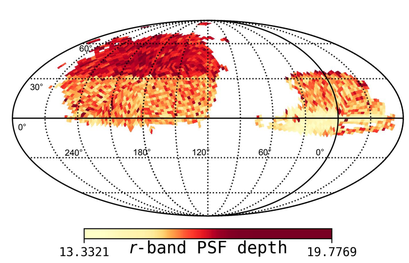

Figure 6: Depth of r-band of the survey shown on the sky. Maximum r-band PSF (point spread function) depth is ~19.78.

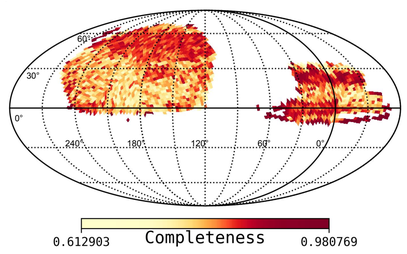

Figure 7: Completeness of the classifier on the sky. In certain regions, completeness correlates with the depth of r-band. However, there are patches with shallow r-band and high completeness. This can potentially be explained by difference in imaging programs in the survey.

Photometric Redshift

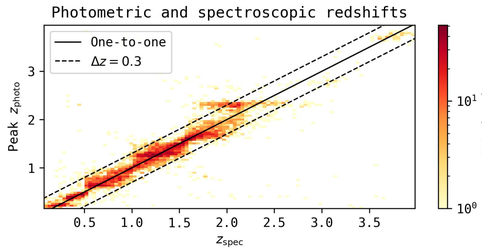

Figure 8: Comparison of spectro-z (SDSS catalogue) with photo-z for a sample of 10000 quasars. Perfect estimates lie on the diagonal. Majority of object lie within Δz of 0.3 (dashed).

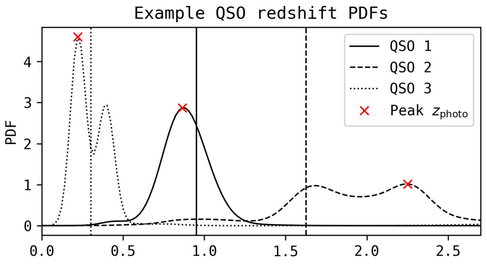

Figure 9: Example of photo-z PDFs for 3 quasars: QSO 1 - well defined redshift. QSO 2 - poorly defined redshift distribution, QSO 3 - "double-peaked" distribution. Vertical lines show respective spectro-z.

Results & Conclusions

- Using GMMs with DECaLS and WISE data produces a classifier with precision/completeness better than previously used hard color cuts.
- Trends in purity and completeness with redshift, r-band, and position on the sky can be explained from knowledge about the survey parameters.
- Predicting more accurate photo-z using GMMs is possible however better priors are required.

Relation to existing research

1. This works gives valuable insights into the trade-off between complexity of the model and performance of GMM compared with other popular methods (such as hard colour cuts).
2. GMM classification/photo-z estimation has not yet been used on the combination of DECaLS and WISE.
3. Our classification methods can be combined with existing work for quasar target selection for the spectroscopic DESI survey.

Implications for novel research

1. New clean and complete catalogue of quasars (and their redshifts) for the use in further cosmological studies.
2. Understanding of the classifier performance will be useful for Euclid and LSST studies over the next decade to effectively classify objects in those upcoming surveys.

References [1] Bovy J., Hogg D. W., Roweis S. T., 2011, Ann. App. Stat., 5, 1657; [2] Bovy J. et al., 2011, ApJ, 729, 141; [3] Bovy J. et al., 2012, ApJ, 749, 41; [4] DiPompeo, M. A., Bovy, J., Myers, A. D., Lang, D. 2015, MNRAS, 452, 3124; [5] Yang Q., et al., 2017, AJ, 154, 6