Title: LOR The Delight PZ Estimator

Contributors: Sylvie Dagoret-Campagne, John Y. H. Soo

Co-signers: John Y. H. Soo, Jean_Eric Campagne

0. Summary Statement

The Delight PZ Estimator [1][2] is Hierarchical Bayesian Photometric Redshift Estimator [3] combining a Flux-Template Fitting method with a Machine Learning (ML) method based on Gaussian Processes which average and kernel are particularly designed for the physics of estimating the photometric redshift for each extragalactic LSST Object.

1. Scientific Utility

The Delight PZ Estimator encapsulates the physical knowledge of the Flux-Redshift relation based on a chosen set of SED templates with the high flexibility of learning relative Flux-Redshift corrections with the ML technique. Doing so, it is not necessary to train the estimator with a huge sample covering the whole LSST Flux-redshift space but it possible to rely on a limited low redshift biased sample expecting the extrapolating power of the templates toward high redshift (from spectroscopic survey). The goal of the Gaussian Process ML is to find similar spectrum correction features in target galaxies compared to those in the training galaxies (which could be attributed to emission lines or continuum effects like dust absorption).
Moreover, the estimator is able to handle Flux-Redshift biases (Eddington-Malmquist biases) by introducing a nuisance parameter which will be marginalized over.
This estimator is running fast (O(N), where N is the number of training samples) once a small set of hyperparameters has been optimized on the training sample.

2. Outputs

The goal is to produce a redshift probability density function (z-pdf) from the measured calibrated band fluxes which is nothing but the redshift posterior of training galaxies taken as the prior. Then from the z-pdf, any metrics such as the average, the mode, the variance,… can be computed depending on the requirements of the cosmological probe using it.

3. Performance

Studies can prove over a training set representative of LSST data (For example DC2 simulated data) that performances comply with the Science Requirements Document (ls.st/srd) over the redshift range from 0 to 3. The performance has already been demonstrated with real data, at low redshift with the PAU survey (redshift 0 -1) [4].

4. Technical Aspects

Scalability - Will probably meet. The Delight PZ Estimator's run time depends on the size of the training set. Pre-run will find the few hyperparameters by optimizing the marginal likelihood on the training set or by inter-calibrating with another PZ estimator.

Inputs and Outputs - Will meet. The Delight PZ Estimator only requires galaxy calibrated fluxes and their uncertainties as inputs, which will already exist in the LSST Object catalog. The output point estimates, errors, and binned PDFs are all consistent with the PZ-related Object elements defined in the DPDD.

Storage Constraints - Will meet. The Delight PZ Estimator requires a training set of galaxies with Fluxes and their uncertainties and "true" redshifts. In addition, during the learning phase, the estimator writes pre-computed elements of the average and Kernel for each training galaxy.

External Data Sets - Will probably meet. The Delight PZ Estimator has been run on both simulated and real data with training sets that may match more or less the test set in terms of flux and redshift distributions. In order to use the Delight Estimator for LSST Objects, a more careful assessment of the impact of incomplete/impure training sets needs to be done. It is not yet clear that such an external data set exists to provide PZ results of adequate quality.

Estimator Training and Iterative Development - Will meet. As mentioned above, more work is needed to validate the use of a spectroscopic training set and suitable SED template sets, but once the correct training set is identified there would be no further iterative development.

Computational Processing Constraints - Will meet. The Delight PZ Estimator does not require that a large amount of data is held in memory at any given time. The Delight PZ Estimator is designed to be parallelizable. For the moment it heavily relies on writing/reading files, not on computer memory. Concerning speed, the trained model could be accelerated via a machine learning emulator.

Implementation Language - Will meet. The Delight PZ Estimator is written in Python 3.

References

[1] S.J. Schmidt, A.I. Malz, J.Y.H. Soo, I.A. Almosallam, M. Brescia, S. Cavuoti, J. COHEN-TANUGI, A.J. Connolly, J. DeRose, P.E. Freeman, M.L. Graham, K.G. Iyer, M.J. Jarvis, J.B.Kalmbach, E. Kovacs, A.B. Lee, G. Longo, C.B. Morrison, J.A. Newman, E. Nourbakhsh,E. Nuss, T. Pospisil, H. Tranin, R.H. Wechsler, R. Zhou, and R. Izbicki. Evaluation of probabilistic photometric redshift estimation approaches for LSST. Monthly Notices of the Royal Astronomical Society, 499(2):2, 2020.

[2] Boris Leistedt and David W. Hogg. Data-driven, Interpretable Photometric Redshifts Trained on Heterogeneous and Unrepresentative Data.ApJ, 838(1):5, March 2017.

[3] Boris Leistedt, David W. Hogg, Risa H. Wechsler, and Joe DeRose. Hierarchical modeling and statistical calibration for photometric redshifts. The Astrophysical Journal, 881(1):80, aug 2019.

[4] John Y H Soo, Benjamin Joachimi, Martin Eriksen, Malgorzata Siudek, Alex Alarcon, LauraCabayol, Jorge Carretero, Ricard Casas, Francisco J Castander, Enrique Fernandez, JuanGarcia-Bellido, Enrique Gaztanaga, Hendrik Hildebrandt, Henk Hoekstra, Ramon Miquel,Cristobal Padilla, Eusebio Sanchez, Santiago Serrano, and Pau Tallada-Crespi. The PAUS survey: narrow-band photometric redshifts using Gaussian processes. Monthly Notices of the Royal Astronomical Society, 503(3):4118–4135, 03 2021