Optimising the LSST Observing Strategy for Supernova Light Curve Classification with Machine Learning

Tarek Allam Jr.*, Rahul Biswas§, Renée Hložek¶ Michelle Lochner†, Jason D. McEwen*, Hiranya Peiris§, Christian Setzer§,
* Mullard Space Science Laboratory, University College London, Holmbury Hill Rd, Dorking RH5 6NT, UK
† African Institute for Mathematical Sciences, 6 Melrose Road, Muizenberg, 7945, South Africa
‡ South African Radio Astronomy Observatory, The Park, Park Road, Pinelands, Cape Town 7405, South Africa
§ The Oskar Klein Centre for Cosmoparticle Physics, Stockholm University, AlbaNova, Stockholm, SE-106 91, Sweden
¶ Department of Astronomy and Astrophysics, University of Toronto, 50 St. George St., Toronto, ON M5S 3H4, Canada

***Abstract*—The cadence that is eventually decided upon for the Large Synoptic Survey Telescope (LSST) will play an important part in our ability to do Type Ia Cosmology. With the expected number of Supernovae that will be detected being far larger than previous surveys, it is not feasible for all to be spectroscopically followed up. Therefore, being able to photometrically classify supernova well will allow us to leverage the power of the datasets LSST will provide and further constrain cosmological parameters. Presented here is a machine learning approach to comparatively study different proposed observing strategies of the LSST to determine the optimal cadence suited for classification of supernova light curves.**

I. Introduction

The aim of this investigation is to analyse the affect a particular cadence has on one's ability to photometrically classify supernova. This investigation has been carried out by the developers of `snmachine` who work in the Supernova Working Group under the umbrella of the Dark Energy Science Collaboration (DESC). `snmachine` is a DESC product that is used as a photometric classification pipeline [1].

The motivation for this work comes from the desire to identify as many Type Ia Supernovae as possible in order to help constrain the nature of Dark Energy. LSST will observe more Supernovae than ever before, at a rate that is not feasible for all transients to be spectroscopically followed up and classified. Thus to handle the deluge of data and the challenge of classifying objects photometrically, machine learning methods are required.

In order to conduct this analysis, use of `SNANA`[2], has been employed to generate the latest light curves that correspond to different cadences runs from `OpSim`[3] outputs.

By interpolating the sampled light curved with Gaussian processes and then applying a wavelet decomposition to these interpolated light curves, one obtains features that could be provided to a classifier, in this case a Random Forest algorithm. For performance, the dimensions of these features were reduced further with a principle component analysis and then these reduced features were provided as inputs to the algorithm. To ensure a controlled test, for each cadence run, a classifier was trained on 2000 light curves only and then tested on the remaining set of light curves that were in the corresponding dataset produced from `SNANA`, in relation to specific `OpSim` cadence simulation. The results for which are shown in Figure 1. The performance of the interpolation is directly affected by the amount of samples one has on the light curve. More samples improve the reliability of the Gaussian processes and thus provide better features via the wave decomposition.

Therefore, it can be understood that in order to classify transients, short sampling of a light curve is important. This is particularly crucial for early classification leading to possible spectroscopic follow up.

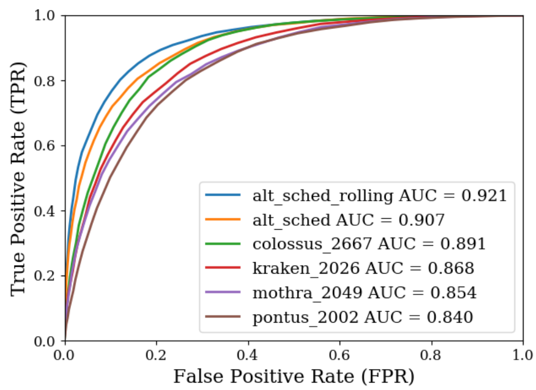

Fig. 1. auROC performance of various `OpSim` cadence runs for WFDY10

Figure 1 shows the comparative classification performances between 6 cadences for the Wide-Fast-Deep observations over the 10 year survey (WFDY10). The area under the Receiver Operating Characteristic (auROC) curves were chosen as the metric to evaluate the performance.

The results for the WFDY10 case show a significant difference in classification performance, with `alt_sched_rolling` being most favourable. This cadence has a higher number of points on a light curve in r-band, thus providing better sampling along the light curve. Further studies are being carried out to explore which other specific cadence properties result in better classification performance.

Further analysis was also done on Deep-Drilling-Fields (DDF) cadence runs with work in progress for models trained on DDF and then tested on WFD. Work is ongoing to formulate the optimal cadence with results being published soon following the call for white papers, and to be incorporated with the PLAsTiCC competition[1].

References

[1] M. Lochner, J. D. McEwen, H. V. Peiris, O. Lahav, and M. K. Winter, "Photometric supernova classification with machine learning," *The Astrophysical Journal Supplement Series*, vol. 225, no. 2, p. 31, 2016.

[2] R. Kessler, J. P. Bernstein, D. Cinabro, B. Dilday, J. A. Frieman, S. Jha, S. Kuhlmann, G. Miknaitis, M. Sako, M. Taylor *et al.*, "Snana: A public software package for supernova analysis," *Publications of the Astronomical Society of the Pacific*, vol. 121, no. 883, p. 1028, 2009.

[3] P. Marshall, T. Anguita, F. B. Bianco, E. C. Bellm, N. Brandt, W. Clarkson, A. Connolly, E. Gawiser, Z. Ivezic, L. Jones *et al.*, "Science-driven optimization of the lsst observing strategy," *arXiv preprint arXiv:1708.04058*, 2017.

[1] https://www.kaggle.com/c/PLAsTiCC-2018