RNAAS RESEARCH NOTES OF THE AAS

Approximating Stellar Metallicity Using Photometric Machine Learning

To cite this article: Rik Ghosh and Soham Saha 2022 *Res. Notes AAS* **6** 57

Draft version March 17, 2022
Typeset using LaTeX **RNAAS** style in AASTeX631

Approximating Stellar Metallicity using Photometric Machine-Learning

Rik Ghosh and Soham Saha

ABSTRACT

Stellar metallicity is an important metric in analyzing stellar evolution. Measuring metallicity (e.g. [Fe/H]) usually requires spectroscopic data, but difficulties associated with gathering spectra for distant objects severely reduces the number of stars for which metallicity can be calculated. The Sloan Expedition for Galactic Understanding and Exploration (SEGUE) spectroscopic surveys are one of the most abundant public-sources of objects with spectra. Despite cataloging over 10^6 objects, the SEGUE data makes up only 0.4% of the data in SDSS photometric surveys. To overcome the limited spectra, this machine-learning method can approximate [Fe/H] from the 5 SDSS photometric filters with a root-mean-square-error (RMSE) of 0.277 dex. The RMSE from this method is similar to the scatter expected in [Fe/H] measurements from low-resolution spectra. Therefore, this method achieves similar accuracy to low-resolution spectra but can be applied to a few orders of magnitude more stars than what the current spectroscopic surveys allow.

Keywords: Metallicity(1031) — Photometric Systems(1233) — Stellar Evolution(1599) — Astronomy Software(1855) — Random Forests(1935)

1. INTRODUCTION

Several surveys over the years have gathered astrometric, photometric, and spectroscopic data for stellar targets. Despite the technological improvements in the instruments for recording data, there is an ever-growing gap between the amount of spectral data available for every object with photometric information. Despite the unprecedented volume of spectra gathered by recent surveys, spectroscopic instruments are not capable of observing each of the photometrically cataloged stars. The Sloan Digital Sky Survey (SDSS) Photometric Catalog has cataloged over one billion photometric sources but the SEGUE 1 and SEGUE 2 spectroscopic surveys only have about 2 million objects with optical spectra (Alam et al. 2015). Since case-by-case analysis for each data point becomes impossible, data-driven machine-learning methods can prove to be extremely helpful. The data-driven nature of machine-learning models and the technological advancements in computation time allows these models to process large datasets and derive desired stellar quantities in a relatively small time.

Many studies in the past have used machine-learning models to study stellar quantities (Debosscher et al. 2007) (Dubath et al. 2011), but recently the focus has shifted to studying fundamental physical properties (Miller et al. 2015). In the case of [Fe/H], a star with enhanced metal content produces less flux in the visible blue wavelength ($\sim$ 4500 Å) of the optical spectrum. Therefore, surveys with blue filters (u' and g' in SDSS) can be used to approximate metallicity via photometric colors of the star. The best estimates for photometric methods to compute [Fe/H] produce a scatter of $\sim$0.3 dex (Kerekes et al. 2013).

Using the 5 SDSS broadband photometric filters (u', g', r', i', z'), a Random Forest (RF) Regressor can be trained to approximate [Fe/H]. The model is trained with data from$\sim$140,000 stars cataloged by the SEGUE 1 and SEGUE 2 surveys of SDSS that have reliable estimates of [Fe/H]. The final model is capable of estimating [Fe/H] with a low error rate.

2. DATA

The training set for the RF model is constructed from a sample of stars with existing SDSS optical spectra. The data comes from SEGUE 1 and SEGUE 2 surveys, which have around 400,000 stellar observations. The optical spectra are analyzed by the SEGUE Stellar Parameters Pipeline (SSPP), which are optimized to estimate [Fe/H] for stellar sources along with corresponding uncertainties (Lee et al. 2008). For high signal-to-noise ratio spectra, SSPP measures [Fe/H] has a typical uncertainty of 0.29 dex. The pipeline also flags spectra for which reliable estimates cannot be provided.

Therefore, the training set includes only stars (376073) that did not raise any flags during SSPP processing (211247) and had at least 2 SSPP readings with no duplicates (182408). These filters on the raw data ensure that both the photometric and spectroscopic uncertainties are small. With a $75-25$ train-test-split this creates a training sample of ∼140,000 stars with reliable measurements of [Fe/H].

During the pre-processing stage, extreme outliers (x < Q1 - 1.5 IQR or x > Q3 + 1.5 IQR) are also removed for each SDSS passband to make the data reliable. The training sample is further reduced to around 139,850 stars.

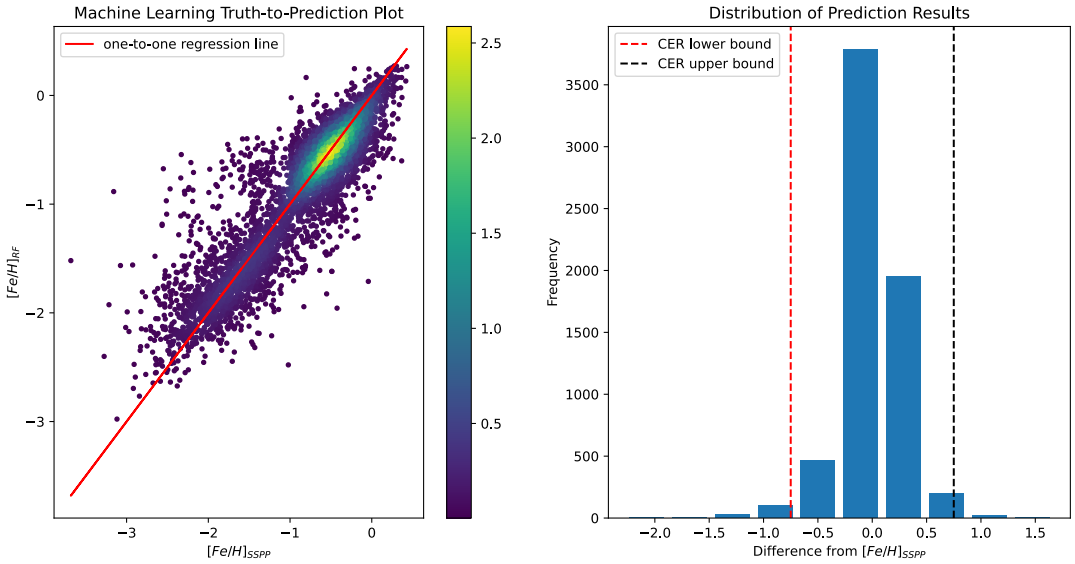

Figure 1. Optimized result of running the model on test set

3. RESULTS

The 5 SDSS broadband photometric filters are used to construct 4 color channels (u' $-g$', g' $-r$', r' $-i$', i' $-z$'), which form the input features for the supervised machine-learning model. To perform regression between photometric colors and [Fe/H], an RF Regressor is used. This Regressor builds multiple decision trees from random samples of the training set. The random sampling at each node of the tree ensures that the RMSE is minimized in the resulting branches. The average of the outputs from each tree, therefore, gives a robust estimate of [Fe/H].

After training the model and optimizing it by tuning the hyperparameters, the cross-validated RMSE on the training set turned out to be 0.271 dex. The result of applying the model on a test set produced an RMSE of 0.277 dex. The catastrophic error rate (CER), which is the probability of a prediction being 0.75 dex away from the true value, for the test set was 0.024 (2.4%). The results of running the model on the test set can be seen in Figure 1, plot (1). The model has a tight scatter around the one-to-one regression line. Since the SSPP estimates of [Fe/H] have a typical uncertainty of $\sim$ 0.24 dex, this RF model produces a scatter that is similar to that of a low-resolution spectrum. But, this model only depends on photometric input, and therefore can be used to calculate [Fe/H] values of stars

whose spectrum couldn't be obtained previously. This expands the catalog for [Fe/H] values to close to a billion stars (virtually all stars with photometric data), approximately 3 orders of magnitude larger than the spectroscopic data available.

4. CONCLUSION

Metallicity is a fundamental parameter for all stars. This machine-learning model is capable of estimating [Fe/H] with a typical scatter of ~ 0.27 dex. This method is also fast and can be applied in batches of $10,000$ stars with a relatively small computational cost. This method thus provides metallicity measurements for about 3 orders of magnitude more stars than the current spectroscopic surveys. This expands the potential applications of studying metallicity. It can help with the search for rare class of extremely metal-poor stars (Schlaufman & Casey 2014) (although the training set will need to be enhanced to include more of the metal-poor stars). Furthermore, it can aid the study of stellar evolution, and help understand the formation of the Milky Way.

We would like to thank Dr. Shyamal Mitra, the leader of the Geometry of Space research group at The University of Texas at Austin, for his help in organizing this line of inquiry.

REFERENCES

Alam, S., Albareti, F. D., Prieto, C. A., et al. 2015, The Astrophysical Journal Supplement Series, 219, 12, doi: 10.1088/0067-0049/219/1/12

Debosscher, J., Sarro, L. M., Aerts, C., et al. 2007, Astronomy & Astrophysics, 475, 1159, doi: 10.1051/0004-6361:20077638

Dubath, P., Rimoldini, L., Süveges, M., et al. 2011, Monthly Notices of the Royal Astronomical Society, 414, 2602, doi: 10.1111/j.1365-2966.2011.18575.x

Kerekes, G., Csabai, I., Dobos, L., & Trencséni, M. 2013, Astronomische Nachrichten, 334, 1012, doi: 10.1002/asna.201211983

Lee, Y. S., Beers, T. C., Sivarani, T., et al. 2008, The Astronomical Journal, 136, 2022, doi: 10.1088/0004-6256/136/5/2022

Miller, A. A., Bloom, J. S., Richards, J. W., et al. 2015, The Astrophysical Journal, 798, 122, doi: 10.1088/0004-637x/798/2/122

Schlaufman, K. C., & Casey, A. R. 2014, The Astrophysical Journal, 797, 13, doi: 10.1088/0004-637x/797/1/13