EPSC Abstracts
Vol. 13, EPSC-DPS2019-588-1, 2019
EPSC-DPS Joint Meeting 2019

EPSC
EPSC-DPS Joint Meeting 2019
15–20 September 2019 | Geneva, Switzerland

Big Data and Machine learning for Exoplanets and Astrobiology: Results from NASA Frontier Development Lab

Daniel Angerhausen (1,2) for the NASA Frontier Development Lab Exoplanet and Astrobiology Teams
(1) Center for Space and Habitability, University of Bern (2) Blue Marble Space Institute of Science, Seattle, United States
(daniel.angerhausen@csh.unibe.ch)

Abstract

We present results from NASA's Frontier Development Lab 2018, an Artificial Intelligence/Machine Learning incubator tackling challenges in various fields of space sciences. Here we focus on the results of the Exoplanet and Astrobiology teams: a project on planet candidate classification in TESS data and modeling and retrieval of exoplanet atmospheres and spectra in the context of life detection. A particular focus will be on two data sets produced at FDL 2018: a set of 3 million exoplanet spectra calculated with the GSFC Planetary Spectrum Generator (PSG) and a set of 150.000 exoplanet atmospheres computed with the ATMOS code.

1. Introduction

For the third time since 2016 machine learning researchers and space science domain experts spent the summer in Silicon Valley to work on some of humanity's most important present day challenges in space. The 8 week long program developed in partnership with NASA's Ames Research Center - NASA frontier development lab (FDL)[1] - at the SETI institute faced problems ranging from predicting solar storms and localization on the moon to finding life in space. The program aims to apply AI technologies to challenges in space exploration by pairing machine learning expertise with space science and exploration researchers from academia and industry. These interdisciplinary teams address tightly defined problems and the format encourages rapid iteration and prototyping to create outputs with meaningful application to the space program and humanity. What makes FDL unique is its close collaboration with industry stakeholders like such as *Intel, Google CLoud, Kx Systems, IBM* and *NVIDIA* and key players in private space such as *SpaceResourcesLu, Lockheed Martin, KBR-Wyle and XPRIZE.*

[1] https://www.frontierdevelopmentlab.org/

2. Exoplanet challenge

The Exoplanet team used state-of-the-art deep learning models to automatically classify Kepler transit signals as either exoplanets or false positives. Their *Astronet* code expanded upon work of Shallue and Vanderburg (SV18) by including additional scientific domain knowledge into the network architecture and input representations to significantly increase overall model performance (Fig. 1). Notably, they achieved

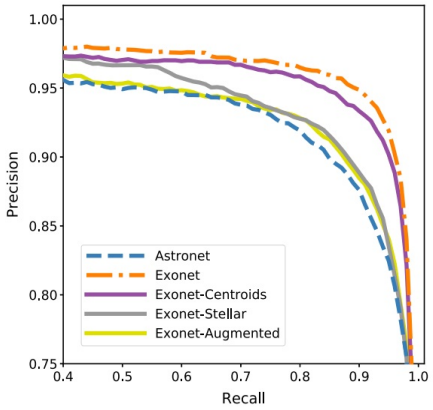

Figure 1: From Ansdell et al., 2018, FDL 3.0 Exoplanet team: Precision-recall curve of the team's Astronet compared to those of the benchmark Exonet (SV18) with different additions of scientific domain knowledge to show the individual contributions to increases in model performance.

15 - 20 percent gains in recall for the lowest signal-to-noise transits that can correspond to rocky planets in the habitable zone. They input CCD pixel centroid time-series information derived from Kepler data and key stellar parameters taken from the catalogue

into the network and also implement data augmentation techniques to alleviate model over-fitting. These improvements allowed them to drastically reduce the size of the model, while still maintaining improved performance. These smaller models are better for generalization, for example from Kepler to TESS data. Their work illustrates the importance of including expert domain knowledge in even state-of-the-art deep learning models when applying them to scientific research problems that seek to identify weak signals in noisy data. This classification tool will be especially useful for current and upcoming space-based photometry missions focused on finding small planets, such as TESS and PLATO. A first application to TESS data was published in Osborn et al. 2019.

3. Astrobiology challenges

The Astrobiology team 1 project demonstrated how cloud computing capabilities can accelerate existing technologies and map out previously neglected parameter spaces (Bell et al., 2019). They succeeded in modelling tens-of-thousands of potential atmospheres over a few days, using software (*ATMOS*) that was originally intended for use in single run applications. The full atmospheric composition data set that was generated will become a useful resource for the community to understand distributions of habitability parameters and will enable better interpretations of future observations of exoplanet atmospheres and potential biosignatures. The software product created during FDL has the potential to significantly improve the accessibility of *ATMOS* for a wide community of researchers. The parameter search approach, can be adopted to simulate more atmospheres and/or modified to rapidly iterate on other problems that utilize *ATMOS*. In Soboczenski et al., 2018 and Cobb/Himes at al., 2019 the Astrobiology 2 team presented a Machine-Learning-based retrieval framework called *Intelligent exoplaNet Atmospheric RetrievAl (INARA)* that consists of a Bayesian deep learning model for retrieval and a data set of 3,000,000 synthetic rocky exoplanetary spectra generated using the NASA Goddard Planetary Spectrum Generator (PSG). This work represents the first ML retrieval model for rocky, terrestrial exoplanets and the first synthetic data set of terrestrial spectra generated at this scale (Zorzan et al, in prep.; Himes et al., in prep.).

4. Summary and Conclusions

We present results from the NASA FDL Exoplanet and Astrobiology challenges 2018: an improved method to classify Exoplanet candidates from photometric surveys, a new deep learning approach for spectral retrieval of Exoplanet atmospheres and containerized software of Astrobiology legacy codes, which we used to provide large data sets for the community. We demonstrate that the NASA frontier development lab format is a successful blueprint for other potential industry/academia collaborations in the future.

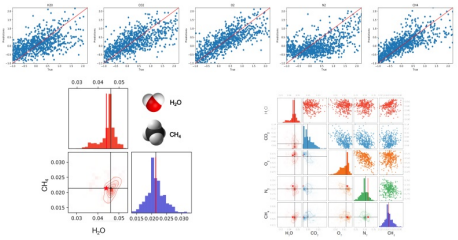

Figure 2: From Soboczenski et al., 2018, FDL 3.0 Astrobiology 2 team: exoplanet spectral retrieval predictions of H_2O, CO_2, O_2, N_2 and CH_4 (top), predictive joint distributions for a random planet (bottom, true value: red star).

Acknowledgements

The work presented here, are the result of the 2018 NASA Frontier Development Lab (FDL) exoplanet and astrobiology challenges. The teams want to thank the organisers, mentors, and sponsors for providing this opportunity. The results are based on work supported by *Google Cloud, Nvidia* and *Kx Systems*.

References

[1] Bell, A., et al. 2018, *NIPS 2018 CiML workshop*, ciml.chalearn.org

[2] Soboczenski, F., et al. 2018, *NeurIPS Workshop on Bayesian Deep Learning*, arXiv:1811.03390

[3] Ansdell, M., et al. 2018, ApJL, 869 (1), L7

[4] Shallue, C. J., & Vanderburg, A. 2018, AJ, 155, 94

[5] Cobb A., Himes M. et al. 2019, AJ, in review

[6] Osborne H. et al. 2019, A&A, accepted, arXiv:1902.08544