National Aeronautics and Space Administration

Evaluating the power of PU (positive and unlabeled data) learning to characterize exoplanet-host stars

Principal Investigator: Yasuhiro Hasegawa (326); Co-Investigators: Umaa Rebbapragada (398)

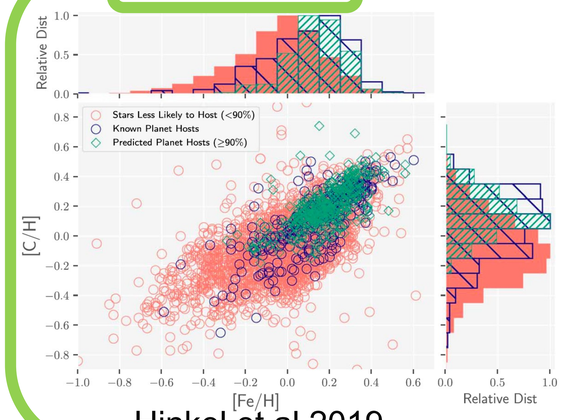
Program: FY21 R&TD Innovative Spontaneous Concepts

Objectives

Apply PU (positive and unlabeled data) learning to stellar elemental abundances that are listed in the Hypatia Catalog (Hinkel et al. 2016)

Evaluate the power of PU learning for reliably determining the list of stellar elemental abundances that correlate with the presence of exoplanets

Background



Hinkel et al 2019

Stellar properties may dictate planet formation

The Fe abundance is widely used for efficiently detecting giant exoplanets

Machine learning may be useful for listing up all the key elements

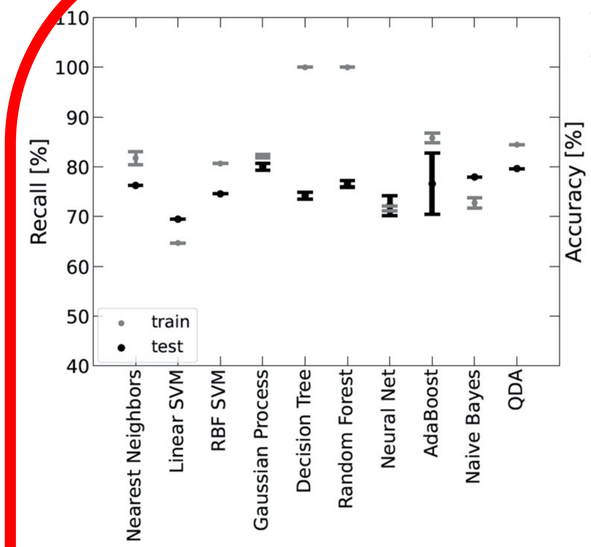
Significance/Benefits to JPL and NASA

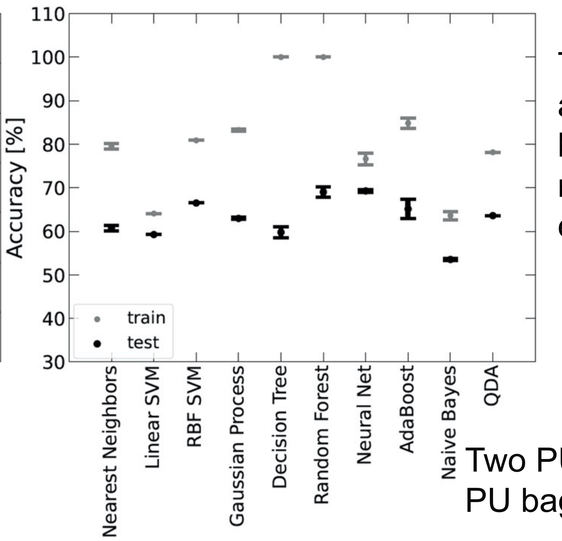
This work provided novel insights about the power of PU learning on the relation between stellar elemental abundances and the detectability of exoplanets

Our finding determined the future direction of research, where stellar elemental abundances will/should be estimated from observationally unbiased stellar spectra and uniform/consistent data analyses.

A new bridge between exoplanet science and data-driven science was created

Approach and Results





	Recall Rate (%) Train	Recall Rate (%) Test	Accuracy (%) Train	Accuracy (%) Test
Standard Random Forest	100	72 +/- 6	100	74 +/- 2
PU Bagging	100	77 +/- 4	78 +/- 2	74 +/- 2
Two Step Approach w/o iteration	100	97 +/- 2	100	22 +/- 3
Two Step Approach w/ iteration (Neg. = Pos.)	100	92 +/- 3	100	43 +/- 3
Two Step Approach w/ iteration (Neg. = 2 Pos.)	100	88 +/- 4	100	53 +/- 3
Two Step Approach w/ iteration (Neg. = 4 Pos.)	100	83 +/- 7	100	63 +/- 3
Two Step Approach w/ iteration (Neg. = 6 Pos.)	100	78 +/- 5	100	68 +/- 3

The recall rate and accuracy are two important measures because the main task is to recover/predict the exoplanet-host stars

The performances of all 10 classifiers are comparable

Two PU learning methods are used: PU bagging and Two-step approach

The number of the reliable negative samples is increased in the two-step approach (the last three rows)

PU bagging achieved slightly better performance

The current data quality is not good enough to reliably list up the key stellar elemental abundances for predicting the presence of exoplanets, even applying the PU learning methods

Publications

Hansen, Byun, Rebbapragada, 2021, AJ, in prep

Reference: N. Hinkel, et al., 2016, ApJS, 226, 66
N. Hinkel, et al., 2019, ApJ, 880, 40

National Aeronautics and Space Administration

Jet Propulsion Laboratory
California Institute of Technology
Pasadena, California

www.nasa.gov

PI/Task Mgr Contact
Email: Yasuhiro.Hasegawa@jpl.nasa.gov

Clearance Number:
RPC/JPL Task Number: R21234