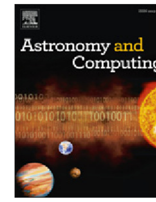
Astronomy and Computing 23 (2018) 141–150

Contents lists available at ScienceDirect

Astronomy and Computing

journal homepage: www.elsevier.com/locate/ascom



Full length article

Theoretical validation of potential habitability via analytical and boosted tree methods: An optimistic study on recently discovered exoplanets

S. Saha [a], S. Basak [a], M. Safonova [c], K. Bora [b,*], S. Agrawal [a], P. Sarkar [a], J. Murthy [d]

[a] *Department of Computer Science and Engineering, PESIT-BSC, Bangalore, India*
[b] *Department of Information Science and Engineering, PESIT-BSC, Bangalore, India*
[c] *M. P. Birla Institute of Fundamental Research, Bangalore, India*
[d] *Indian Institute of Astrophysics, Bangalore, India*

ARTICLE INFO

Article history:
Received 18 September 2017
Accepted 15 March 2018
Available online 26 April 2018

ABSTRACT

Seven Earth-sized planets, known as the TRAPPIST-1 system, was discovered with great fanfare in the last week of February 2017. Three of these planets are in the habitable zone of their star, making them potentially habitable planets (PHPs) a mere 40 light years away. The discovery of the closest potentially habitable planet to us just a year before — Proxima b and a realization that Earth-type planets in circumstellar habitable zones are a common occurrence provides the impetus to the existing pursuit for life outside the Solar System. The search for life has two goals essentially: looking for planets with Earth-like conditions (Earth similarity) and looking for the possibility of life in some form (habitability). An index was recently developed, the Cobb–Douglas Habitability Score (CDHS), based on Cobb–Douglas habitability production function (CD-HPF), which computes the habitability score by using measured and estimated planetary parameters. As an initial set, radius, density, escape velocity and surface temperature of a planet were used. The proposed metric, with exponents accounting for metric elasticity, is endowed with analytical properties that ensure global optima and can be scaled to accommodate a finite number of input parameters. We show here that the model is elastic, and the conditions on elasticity to ensure global maxima can scale as the number of predictor parameters increase. K-NN (K-Nearest Neighbor) classification algorithm, embellished with probabilistic herding and thresholding restriction, utilizes CDHS scores and labels exoplanets into appropriate classes via feature-learning methods yielding granular clusters of habitability. The algorithm works on top of a decision-theoretical model using the power of convex optimization and machine learning. The goal is to characterize the recently discovered exoplanets into an "Earth League" and several other classes based on their CDHS values. A second approach, based on a novel feature-learning and tree-building method classifies the same planets without computing the CDHS of the planets and produces a similar outcome. For this, we use XGBoosted trees. The convergence of the outcome of the two different approaches indicates the strength of the proposed solution scheme and the likelihood of the potential habitability of the recently announced discoveries.

1. Introduction

With discoveries of exoplanets pouring in hundreds, it is becoming necessary to develop some sort of a quick screening tool – a ranking scale – for evaluating habitability perspectives for the follow-up targets. We have proposed a novel inductive approach, inspired by the Cobb–Douglas model from production economics, to verify theoretical conditions of global optima of the functional form to model and to compute the habitability score of exoplanets — the Cobb–Douglas Habitability Score (CDHS; Bora et al., 2016). The discovery of an exoplanet, Proxima b (Anglada-Escudé, 2016), orbiting the nearest star (Proxima Centauri), generated a lot of stir in the news (Witze, 2016) because it is located in the habitable zone and its mass is in the Earth's mass range: 1.27–3 $M_{\oplus}$, making it a potentially habitable planet (PHP) and an immediate destination for the Breakthrough Starshot initiative (Breakthrough Starshot, 2016). A few months after the announcement of Proxima b, another family of terrestrial-size exoplanets – the TRAPPIST-1 system – was discovered (Gillon et al., 2017).

* Corresponding author.
E-mail addresses: snehanshusaha@pes.edu (S. Saha), k_bora@pes.edu (K. Bora).

https://doi.org/10.1016/j.ascom.2018.03.003

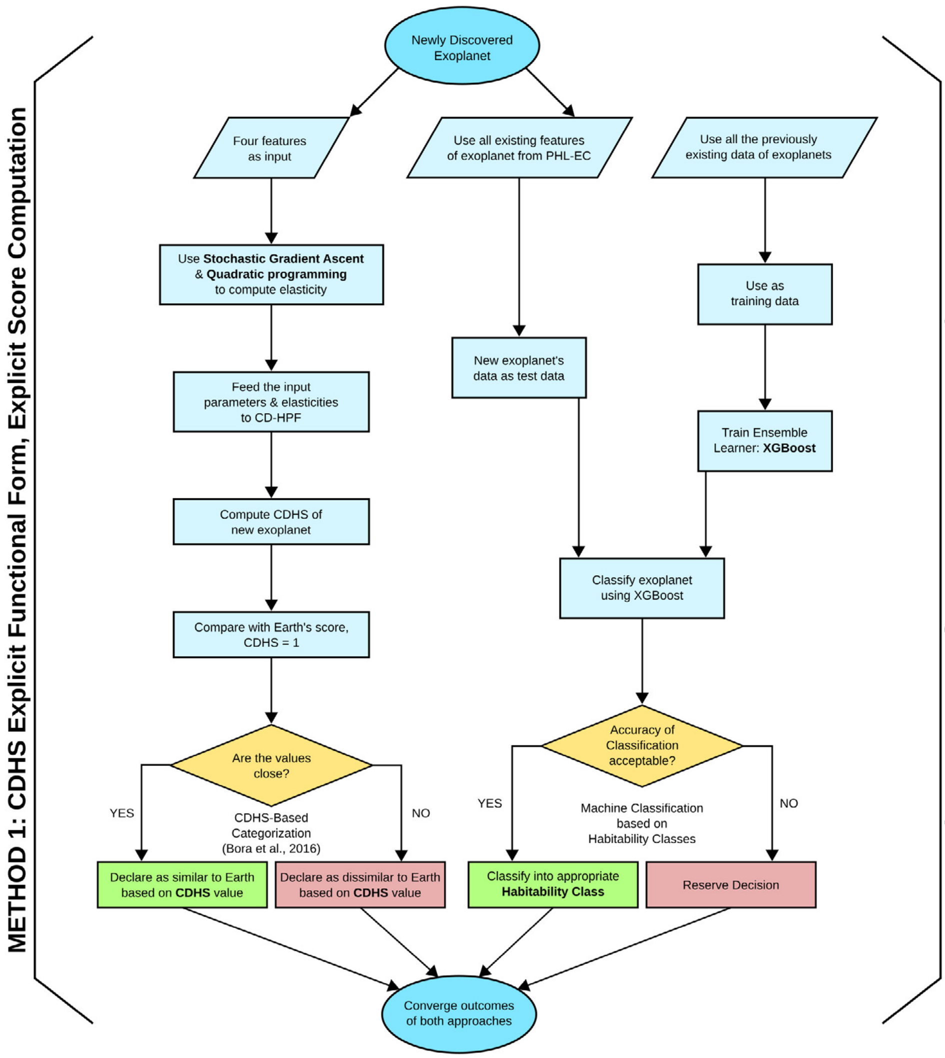


Fig. 1. The convergence of two different approaches in the investigation of potential habitability. The outcome of the explicit scoring scheme i.e. Method 1 (based on CDHS) placed Proxima b in the "Earth League" (Earth-Similarity classification of exoplanets) and is empirically synonymous to being classified into a potentially habitable class of exoplanets via Method 2 (Not based on Earth-Similarity classification, rather supervised feature based machine classification, independent of CDHS). The convergence in outcome is remarkable but not accidental.

This work is motivated by testing the efficacy of the suggested model, CDHS, in determining the habitability score, the proximity to the "Earth-League", of the recently discovered exoplanets. Therefore, it was natural to test whether the model can also classify Proxima b and the planets in the TRAPPIST-1 system as potentially habitable by computing their habitability scores. This could indicate whether the model may be extended for a quick check of the potential habitability of newly discovered exoplanets in general. As we see from the results of our work, this is indeed the case with the TRAPPIST-1 planets.

The flowchart in Fig. 1 summarizes our new approach to the habitability investigation of exoplanets (on the example of Proxima b and TRAPPIST-1 system). This approach is based on the combination of two methods. The outcome of classification of exoplanets based on the CDHS (Method 1) is tallied with another machine classification method which discriminates samples (exoplanets) into classes based on the features/attributes of the samples (Method 2). The similar outcome from both approaches (the exoplanets being characterized for their potential habitability), is markedly different in structure and methodology, fortifies the growing advocacy of using machine learning in astronomy.

Our habitability score model considers four parameters/features, namely mass, radius, density and surface temperature of a planet extracted from the PHL-EC (Exoplanet Catalog hosted by the Planetary Habitability Laboratory (PHL)).[1] Though the catalog contains 68 observed and derived stellar and planetary parameters, we have considered only four for the CDHS model as a lot of parameters are redundant. We show that the CDHS model is scalable, i.e. capable of accommodating more parameters (see Section 4 on model scalability and Section 3 in supplementary file). Therefore, we may use a greater number of parameters in the future to compute the CDHS.

PHL classifies exoplanets into five categories based on their thermal characteristics: non-habitable, and potentially habitable: psychroplanet, mesoplanet, thermoplanet, and hypopsychroplanet. Proxima b and the TRAPPIST-1 system are amongst the recent additions to the catalog with recorded features. Here, we employ a non-metric classifier (Gradient Boosted Trees, XGBoost) to predict the class label of the recently discovered exoplanets. We compute the accuracy of our classification method, and aim to converge the result with the habitability score of the recently discovered exoplanets, which may suggest its proximity to the

[1] The latest updated (November 2017) dataset can be downloaded from the PHL website: http://phl.upr.edu/projects/habitable-exoplanets-catalog/data/database.