EPSC Abstracts
Vol. 13, EPSC-DPS2019-817-1, 2019
EPSC-DPS Joint Meeting 2019

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

Correcting Transiting Exoplanet Light Curves for Stellar Spots: A Machine Learning Challenge for the ESA Ariel Space Mission

Nikolaos Nikolaou (1), Ingo Waldmann (1), Subhajit Sarkar (2), Angelos Tsiaras (1), Billy Edwards (1), Mario Morvan (1), Kai Hou Yip (1) and Giovanna Tinetti (1)
(1) UCL, Department of Physics & Astronomy, UK, (2) University of Cardiff, School of Physics & Astronomy, UK

Abstract

The field of exoplanet discovery and characterisation has been growing rapidly in the last decade. However, several big challenges remain, many of which could be addressed using machine learning methodology. For instance, the most successful method for detecting exoplanets, transit photometry, is very sensitive to the presence of stellar spots. The current approach is to identify the effects of spots visually and correct for them manually or discard the data. As a first step to automate this process, we are organising a competition for the 2019 European Conference of Machine Learning (ECML) on data generated by ArielSim, the simulator of the European Space Agency's upcoming Ariel mission, whose objective is to characterise the atmosphere of 1000 exoplanets. The data consist of pairs of light curves corrupted by stellar spots and the corresponding clean ones, along with auxiliary observation information. The goal is to correct light curves for the presence of stellar spots (multiple signal denoising). This is a yet unsolved problem in the community. In this talk we will discuss the problem, the impact of a solution, introduce the basics of machine learning and present the outline of the competition as well as initial baseline solutions.

1. The Ariel Mission

Exoplanet characterisation is at the very forefront of astronomical research. With the first extrasolar planet discovered in 1995, the number of known planets has grown exponentially in the last two decades and over 3900 planets are known today. Whilst the last decades were more concerned with the task of finding them, the field is now embarking on the challenge of characterising them. Questions such as "What are they made of?", "how did they form?", "why is our own solar system so different from other planetary systems?" and of course, "is there life elsewhere?" are some of the major drivers of the field. By observing the atmospheres of these foreign worlds, we will be able to analyse their chemistry, thermal structure and even cloud coverage. This information is crucial to answering the fundamental questions above. The goal of the European Space Agency M4 Ariel mission [1], to be launched in 2028, is to observe the atmosphere of 1000 transiting exoplanets and allow large scale population studies.

2. Ariel Data Challenges

In the context of promoting and communicating the goals of Ariel to a more general audience and encouraging a wider collaboration around it with broader the European and international research community —especially with respect to data analysis— several challenges related to analysing the data that will be collected from Ariel have been released to the public. Researchers from both the field of astrophysics and astronomy, but also from data analysis related fields, like statistics and machine learning, are invited to contribute solutions (methodologies, algorithms or models) to problems of interest to the mission. These problems revolve around correction of systematics, isolating signals from multiple sources of noise or refining the results of atmospheric retrievals.

Beyond the goals already stated —and the obvious one: solving problems that will further our capability to interpret and analyse the observations of Ariel— this approach allows for a diverse range of solutions to these problems to be explored in parallel, leveraging both domain knowledge and computational statistics methodologies, thus greatly expanding the resources (both human and computational) dedicated to analyzing Ariel data, with minimal cost. Successful solutions or identifications of interesting followup problems can lead to publications relevant to Ariel and exoplanetary

[1] https://ariel-spacemission.eu

science in general.

3. Correcting for Stellar Spots

Among the challenges we have released is the task of correcting transiting exoplanet light curves for the presence of stellar spots and faculae. The problem itself is important to solve, not only for Ariel but for the analysis of transit photometry and spectroscopy data in general. Solving it will mean improving our understanding of the characteristics of currently confirmed exoplanets, potentially recognising false positive or false negative detections and improving our ability to analyse new observations –primarily but not limited to those expected from Ariel– without the need to equip new telescopes with additional instruments with all the extra costs this implies. The current approach is to identify the effects of spots and faculae visually and correct for them manually or discard the data. Our goal is to investigate if a fully automated solution to the problem is feasible and if it is, to have it implemented within the next years and ready to be applied to data from Ariel.

3.1. Leveraging Machine Learning

As many problems in the field of exoplanetary science, this too is characterised by a unique set of challenges, including: (i) a large amount of data to process, (ii) a very low signal to noise ratio, (iii) an underlying pattern which is non-linear and whose parametric form is a-priori unknown, (iv) the available information comes in multiple forms (the lightcurves are time-series, but we can also have additional observation information) and finally (v) a high degree of degeneracy.

Currently, most data analysis in the field is performed either by visual inspection, or by applying simple transformations to the observations. As human capabilities for detecting faint patterns in multidimensional data are limited, and with the ever increasing data volume, the entire procedure needs to become fully automated.

All these challenges make the field of exoplanetary science a fertile ground for the application of machine learning and data mining methodology, for automating pattern discovery, simplifying existing models and refining current evaluation methods. We would like Ariel to fully benefit from these computational solutions to processing large amounts of data, so our aim is to increase interaction with the machine learning community.

To this end, the stellar spot removal data challenge is specifically targeting the machine learning and data science research community as an audience. The field of Machine Learning is primarily conference- driven and the European Conference on Machine Learning (ECML) is the annual European venue of research dissemination for the machine learning community, attracting an international audience of researchers from both academia and industry. As such, it offers a great opportunity for interacting with the community and establishing a presence in it.

Our data challenge was proposed as a competition to be organised in the context of the upcoming ECML 2019[2] and has been accepted by the conference organisers as one of the 3 hosted competitions. The special session of ECML dedicated to our competition will take place on May 20 2019[3]. We have advertised the competition extensively and created a dedicated website for this and future Ariel data analysis challenges[4]. Through the website, participants can be informed of the competition's rules and timeline, the description of the data and the problem. They can submit solutions (in the form of predictions of their models on held out data) which will automatically be scored and displayed on a leaderboard.

We have already produced some initial baseline solutions using machine learning approaches on small scale versions of the problem with very encouraging results. These were based on Deep Neural Networks and Denoising Autoencoders and will be discussed in detail in the talk, along with the winning solution, the specifics of the dataset and the challenges of organising a competition of this scale.

4. The Next Steps

Our ultimate aim with this competition is to bridge the two communities. Exoplanetary science offers a large amount of interesting data and challenging problems which need to be solved in order for the field to advance. Machine learning offers a large arsenal of techniques for using data to answer interesting questions and needs large amounts of data and interesting questions to showcase its latest advances. Our future plans include organising workshops on the analysis of exoplanetary science data or even dedicated sessions in future iterations of ECML or other big machine learning conferences.

[2] http://ecmlpkdd2019.org

[3] The competition launched on April 15 2019 and will close on August 15 2019. At the time of the submission of this abstract, there were almost 70 participants.

[4] http://ariel-datachallenge.space