Machine Learning for Exoplanet Detection

Dr David Armstrong, d.j.armstrong@warwick.ac.uk (Astrophysics)
Dr Theo Damoulas, t.damoulas@warwick.ac.uk (Computer Science)

Background: A key method for the detection of new planets outside our Solar System is searching for *transits*, eclipses as the planet blocks some of the light from its parent star. Satellites such as NASA's Kepler monitor hundreds of thousands of stars for extended periods, measuring their brightness over time. Finding transits in this data is complex, and involves significant human input. Real planets are obscured by false positives, signals which can mimic a planet but in fact arise from other sources. Here we will use available data from the Kepler satellite, which contains 2300 known planets, 2300 known false positives, and 1800 candidate signals containing among them as yet unknown planets. This project will work on classifying the candidate signals, separating the true planets from the false positives using a variety of machine learning techniques. Established methods such as random forests or neural nets can be implemented, or novel methods developed by the student.

Prospective Deliverables: The target for this project is to deliver an algorithm capable of classifying candidate planetary signals, which will be used to discover new exoplanets outside our Solar System. In particular, we hope that the algorithm will be capable of classifying terrestrial, small and often hard to detect planets.

Real World application: Machine learning techniques and 'Big Data' are rapidly growing areas across sections of industry, forming the core of the current drive towards automation. In terms of planet detection, such algorithms are crucial for understanding the statistical populations of planets in the galaxy, and for determining how common Earth-like planets are. To explore the underlying population of planets, we need a complete understanding of the efficiency of our detection algorithms. This is greatly aided by automation, allowing tests to be performed with simulated planets and completeness statistics to be derived.

Extension to PhD: The project can be extended to a full PhD, by considering the full process of planet detection and applying it to future missions. The PhD would begin by moving beyond separating planets and false positives to automating the complete detection process. We would then apply the new tools to data from the upcoming NASA TESS satellite, launching this year, which will observe millions of stars and ~20000 planets (over an order of magnitude more than are already known). The goal of the PhD will be to expose the populations of planets in this dataset, finding key habitable planets for future research and exploring the impact of planet formation on the planet population as a whole.

We are in contact with Google about providing external mentors for this PhD program. This would allow the student to learn about state of the art algorithms used in industry, make connections with a high-level industrial partner, as well as use cloud computing in dealing with the large data volumes that will be available.