Analyzing TESS data with a neural network - Proposal

Alexander Evitt

Objective

Purpose

The purpose of this research is to use machine learning techniques to sort through the Transiting Exoplanet Survey Satellite's light curve data and identify notable stars for further study. Ideally, the neural network would identify black holes near stars, although the exact case that makes a star noteworthy is still being looked into.

Expected Outcome

It is expected that this research will result in the creation of a neural network that will process parts of the TESS dataset and flag stars of interest for researchers to analyze, most likely stars in unique black hole systems.

Justification

This research is important because it allows faster analysis of data. Identifying stars of interest is very important to the mission of TESS, as the spacecraft will analyze over 200,000 stars. No human has the capability to analyze all the data returned, so automatic systems to analyze it are important.

Description

Explanation and Techniques

The analysis of the dataset will be done through a neural network. This neural network will be created by manually analyzing light curve data and then using it to train the network. This will be done until a sufficient accuracy is reached, as determined by testing it on light curve data. All the light curves will be supplied from the MAST website.

Variables and Constraints

In the light curve data, the independent variable is time and the dependent variable is the amount of light emitted from the star. However, this research focuses on data analysis. The independent variables for this research are the parameters of the neural network (weights and biases), while the dependent variable is the accuracy of the system. This research is constrained by the limitations of computing power and accuracy that are imposed upon lightweight neural networks.

Feasibility Study

	Available Resources	Needed Resources
Personnel	Myself and a partner (Elizabeth Sellers)	n/a
Equipment	School computers	Cloud computing (likely Amazon) for computational power
Supplies	n/a	n/a
Knowledge/Skills	Basic linear algebra knowledge	Machine learning knowledge, with multivariable calculus

Proposed Budget

Very little in the way of supplies is needed to complete this project. The only thing needed is a large amount of computing power, which is necessary for machine learning research. Cloud computing is necessary to conduct this project, and some expense may be involved, but exactly how much depends on where the cloud computing comes from, and the entire process is still under-development by myself, my research mentor, and several other research students.

Risk Assessment

Very little risk is inherent in the research. Pure mathematics, data analysis, and computing carry little to no risk.

Alternatives

Alternative research is possible, although the same result likely will not be achieved. Classification of large datasets is a classic machine learning problem, and completing it without the use of machine learning is near impossible. However, some alternatives exist in the case of minor issues. A lack of computational power, for example, could be resolved by running a sized-down version of the neural network at the expense of efficiency and accuracy. While not optimum, this could accomplish similar goals. The neural network plans to look through the dataset and identify special cases of black hole based systems, but, in the event that the specific type looked for is deemed infeasible or already researched, the case could be changed. For example, the TESS dataset is designed to allow the identification of exoplanets, but researchers have already used machine learning to look for exoplanets. This research could look at similar special cases, although they would have to be identifiable through the use of light curve data.