Applications of Machine Learning in Detecting Exoplanets

Torsha Majumder

October 8, 2020

Abstract

Analyzing the astronomical data and preparing the data set is an essential part of developing a machine learning model for detecting exoplanets or any other stellar objects. Today, we can train machine learning models in an unbiased way that can differentiate between planet candidates and astronomical false positives with greater accuracy. But the models that are developing still need some human oversight to make the classification powerful. Every space mission has its objective to perform in observing the different regions of the sky, and therefore in a different galactic environment. Thus, the models developed can only identify some of the specific patterns in the given data, and its ability to recognize patterns beyond its training data is limited. Some machine learning classifiers have proven to be helpful as automated-vetting systems. One of the most successful classes of machine learning algorithms - deep learning models - has now become a powerful tool for pruning away the obvious false positives. The major drawback of these deep learning models is its inability to identify the planet candidates precisely since these models are highly dependent on the training data set and the meta-parameters; it is difficult to understand the underlying constraints to make these models generalized, and hence more reliable. This article will focus on discussing - some of the popular methods used in generating the astronomical data sets like citizen science space projects, human-vetting systems, automated-vetting systems; the deep learning models - AstroNet, AstroNet-K2 - developed to analyze the Kepler and K2 data set; and the downside of deep neural networks in astronomical data sets.

Introduction

Machine learning has now become a pivotal tool in data analysis of the citizen science space projects and anomaly detection in spacecraft telemetry. The citizen science projects utilize the traditional method of crowdsourcing the human insight for labeling the astronomical data with a simple yes/no answer; the aggregation of information in a group often results in a decision which is better

than any single member in that group. This approach is based on a successful method called "The Wisdom of Crowds" by James Surowiecki. The citizen science space projects help in data annotation, which is essential in supervised learning and validating the machine learning systems. One of the major applications of these machine learning models is in the exoplanet discovery. The supervised machine learning models learn to recognize patterns from these annotated data to differentiate among planet candidates, astronomical false positives, and non-transiting phenomenon. Data annotation helps the model to learn features that are essential in identifying the planet transits. Thus, the trained machine learning model can be applied to a new data set to automate the vetting process. But this is not true in most cases because the predictive power of the machine learning models is limited to the data set, which is being used to train the model. It can only identify patterns specific to the training data, and its prediction fails when incorporated in a completely new data set.

The citizen science space projects have widened the field of classification tasks by introducing these data to a larger group of human volunteers. The classifications made by the volunteers are further inspected laboriously by a small science team. Some of the popular and successful citizen science space projects are:

- Galaxy Zoo 3D: It has helped in identifying the internal structures of the galaxies by MaNGA (Mapping Nearby Galaxies at Apache Point Observatory) survey. This project got completed, where it has made a total of 112,637 classifications.

- Planet Hunters TESS: The data set is provided by NASA's Transiting Exoplanet Survey Satellite (TESS) mission to look for planets outside of our solar system, including planets that could support life. This is one of the popular on-going projects with a total classification size of 175,190.

- Disk Detective: The data set is provided by NASA's WISE mission and other surveys to search for dusty debris disks similar to our asteroid belt and gas-rich primordial disks, which are the birthplaces of planets. This project is still on-going with a total classification size of 142,678.

- Planet Patrol: The data set is provided by NASA's Transiting Exoplanet Survey Satellite (TESS) mission to look for variable stars, eclipsing binary stars, blended stars, glitches in the data. This is one of the popular and recently launched projects with a total classification size of 86,619.

In exoplanet discoveries, the training data set consists of planet signals known as TCEs or Threshold Crossing Events. These are periodic signals which are generated by an algorithm - transit search - designed to search for transiting exoplanets in a light curve. Thus, the light curve production is a vital stage in the process of generating these TCEs. The next step after the production of TCEs is to label them. This is a two-step process; (1) triage: this is a quick and efficient way of labeling the TCEs to narrow down the data set into

a subset; (2) vetting leads to carefully considering the subset that resembled the possible transit signals generated after triage. In addition to this, machine learning models are introduced as an important tool to automate this vetting process and to reduce the human effort.

Machine Learning Applications

Machine Learning applications are broadening its horizon in the field of space science. The research areas which include its applications are anomaly detection, spacecraft telemetry, and data analysis. The citizen science projects proposed to differentiate meaningful data from noise, whereas anomaly detection focuses on searching for bad signals (anomalies) from the data; The main challenge is to classify the true anomalies that are neither noise nor an outlier. One of the obvious applications here is to classify the astronomical false-positive planet candidates from the given data sets. Spacecraft telemetry contains data in a time-series format, which requires state-of-the-art feature engineering techniques to transform the raw telemetry data into useful data to train state-of-the-art machine learning models. Some of the applications of machine learning include - predicting the thermal power consumption in the Mars Express (MEX) spacecraft that optimizes the scientific operations of MEX by estimating the power consumption of the thermal subsystem efficiently, and an online learning system that helped in reducing the error in predicting International Space Station battery charge variation by 25%. Again, data normalization and augmentation has also proven to be successful in the analysis of the astronomical data.

This article will focus on discussing some of the popular automated vetting systems developed using machine learning techniques in recent years for the detection of exoplanets. The autovetter system uses the random forest approach for classifying the input data set that consists of TCEs (mostly labeled by humans) and a set of attributes associated with each TCEs. It categorizes the TCEs into three classes: PC (Planet Candidate), AFP (Astrophysical False Positive), and NTP (Non-Transiting Phenomenon). The autovetter learns to map between the attributes and predicted class from the training data set. This mapping is then applied uniformly and consistently to all TCEs to produce a catalog of planet candidates. There was another set of a catalog of planet candidates generated by the robovetter system, an expert system designed to automatically classify the TCEs. The autovetter and robovetter followed independent methodology to automate the process of human classification of planet candidates. The key difference between these two machine learning techniques lies in the 'expert system' approach in the robovetter system where the decision rules are generated explicitly. But the autovetter's decision rules are learned autonomously from the data.

Another class of machine learning model introduced for automatically vetting the Kepler TCEs is the deep neural network model called AstroNet. This is one of the most successful convolutional neural network architectures that has achieved higher accuracy in differentiating between planet candidates and as-

tronomical false positives in the Kepler data. The data set consists of labeled TCEs from the autovetter catalog, and the classes got binarized into "planet" (PC) and "not planet" (AFP/NTP) for training the AstroNet model. This model further exploits the model aggregation technique "ensemble learning" to improve the overall model's performance and to reduce the model's variance. AstroNet-K2 is a modified version of AstroNet architecture for utilizing the K2 data set. The key differences in its architecture are in modifying the global and the local views in the AstroNet model, interpolating the data, and the addition of several scalar features into the neural network architecture. Several machine learning techniques have been developed to analyze the telescopic data. Recently, a logistic regression algorithm has been developed to attempt to detect non-transiting hot Jupiter. But amongst these, deep learning models are of greater interest due to its robust behavior.

Conclusion

So far, many machine learning models got developed which were able to classify classes based on some specific astronomical data. It is challenging to train a model that will perform equivalently on all data sets without human intervention. As we can see that the model's architecture is modified to make it adaptive to a different data set to have a robust outcome. One of the achievements of these models is that it can identify the astronomical false positives with greater precision, if not the potential planet candidates. As a result, these models help to weed out the least planet-like signals and allowing the astronomers to spend some more time in scrutinizing the potential planet candidates.

References

1. https://arxiv.org/pdf/1809.00542.pdf

2. https://ieeexplore.ieee.org/document/9132719

3. https://arxiv.org/abs/1903.10507

4. https://arxiv.org/abs/1712.05044

5. https://archive.stsci.edu/kepler/manuals/KSCI-19091-001.pdf

6. https://www.zooniverse.org/projects/klmasters/galaxy-zoo-3d

7. https://www.zooniverse.org/projects/nora-dot-eisner/planet-hunters-tess/

8. https://www.zooniverse.org/projects/ssilverberg/disk-detective

9. https://www.zooniverse.org/projects/marckuchner/planet-patrol