Kepler Light Curve Classification Using Deep Learning and Markov Transition Field (Student Abstract)

Shane Donnelly, Ayan Dutta
University of North Florida, 1 UNF Drive, Jacksonville, Florida 32224 USA
n01380643@unf.edu, a.dutta@unf.edu

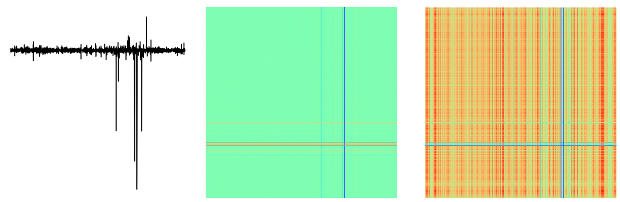
Abstract

An exoplanet is a planet, which is not a part of our solar system. Whether life exists in one or more of these exoplanets has fascinated humans for centuries. NASA's Kepler Space Telescope has discovered more than 70% of known exoplanets in our universe. However, manually determining whether a Kepler light curve indicates an exoplanet or not becomes infeasible with the large volume of data. Due to this, we propose a deep learning-based strategy to automatically classify a Kepler light curve. More specifically, we first convert the light curve time series into its corresponding Markov Transition Field (MTF) image and then classify it. Results show that the accuracy of the proposed technique is 99.39%, which is higher than all current state-of-the-art approaches.

Introduction

Search for life on other planets of our solar system or in a different solar system far away from us is a challenging problem. A planet or an exoplanet needs to be at a favorable distance from its star(s) (called the Goldilocks zone) to sustain life. Astrophysicists and astrobiologists have used powerful telescopes to collect data about these far objects and have analyzed them for years. One such equipment is NASA's Kepler Space Telescope which has collected a large amount of data (e.g., light curves) between 2009 and 2013.

The Kepler telescope was used to identify Earth-sized exoplanets orbiting nearby stars. This was done by measuring the brightness of stars within its field of view and then this data was analyzed to find periodic dimming within the brightness, presumably caused by an exoplanet passing in front of its host star. The telescope looked at 530,506 stars and found 9564 Kepler Objects of Interest (KOI), of which, 6554 were labeled as "Confirmed" or "False Positive" (as of May 2023). We have collected the data used in this paper from the NASA Exoplanet archive. A light curve, or a time series of flux (brightness) measurements over time, can be generated for each KOI and we are attempting to classify the object as an "exoplanet" or a "false positive". This is significant as more missions are conducted and larger datasets are collected on different stars and planets, having the most accurate classifier could drastically reduce the time needed for manual verification of such exoplanets.



a) Standard b) GADF c) MTF (our input)

Figure 1: Three tested input types: two baselines and ours.

Given the success of deep learning in recent years, scientists have used deep learning-based classification techniques to identify exoplanets (Dattilo et al. 2019; Yu et al. 2019). `AstroNet` is one of the most popular deep learning models for this (Shallue and Vanderburg 2018). Instead of inputting raw time-series data to 1D convolutional layers, in this paper, we propose to first convert the light curve into a Markov Transition Field (MTF) image. An MTF is an image of transition probabilities from a discrete time series. For a signal that is periodic, a trend within the MTF would presumably show itself, as the probability of transitioning to a certain periodic value would be higher than transitioning to any random value. Given that an exoplanet has an orbital period, this means that a dip in brightness being measured from its host star should also be periodic. A method that can highlight periodic values would then be best for identifying an orbiting exoplanet and ignoring noise or other causes of brightness change. MTF has previously been successfully used for converting various time-series data.

To this end, we propose a novel neural network architecture that, instead of using the raw light intensity curves, uses corresponding MTF images as inputs. We have compared our proposed technique against two baselines – standard light curve plot (represents brightness over time, specifically standardized lux over days) and the corresponding Gramian Angular Difference Field (GADF) (a popular form of converting time-series data into images). Note that the brightness was initially measured with multiple different exposure times including fast, short, and long times. We have only used the long exposures here. One such confirmed exoplanet data point is shown in Fig. 1 in three different test formats.

Methodology

The standard light curves were acquired using the Lightkurve Python library. All long exposures for any individual star were downloaded and stitched together. Outliers were then removed using a sigma clipping technique and gaps were filled with randomly distributed Gaussian noise, dependent on the light curve being worked on. The curve was then flattened with a Savitzky-Golay filter and the time series was plotted. To generate both the MTF and GADF, we have used the `pyts` library. For an MTF, the time series was processed with a uniform strategy, with bin size 8. The resulting field was then down-scaled by a factor of 10 along each axis. For the GADF, the time series was processed as a standard Gramian Angular Field. The resulting field was then down-scaled by a factor of 10 along each axis.

Our proposed convolutional neural network architecture consists of 13 layers including three sets of convolutional layers, pooling layers, and drop-out layers. This is followed by a flattening layer and then two dense layers. The three convolutional layers have 16, 32, and 64 filters, a stride of 5, and the ReLU activation function. The pooling layer has a pool size of 2 and the dropout layer has a rate of 0.1. The fully connected dense layers have sizes of 64 and 16 respectively with the ReLU activation function. Finally, the output layer is placed with the Sigmoid activation function. For MTF and GADF (RGB images), the input shape is (577, 577, 3) whereas the standard input (gray-scale image) has the shape of (577, 577, 1). The batch size is set to 16 and the number of epochs is 75. We have used the binary cross-entropy loss function along with the Adam optimizer. The learning rate is set to 0.00001.

The total number of data in our dataset was 6552 with 1913 "confirmed" and 4639 "false positives". The validation of the training was done using the holdout method with a 90/10 split. After every epoch, the model was evaluated and the corresponding accuracy/F-score values are shown in Fig. 2. The numbers reported in Table 1 are from the final epoch.

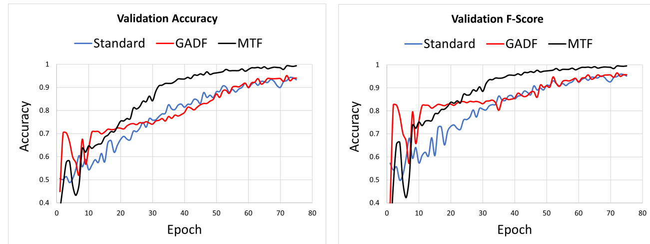
Results and Discussion

Although we do not have access to the data used in earlier studies, we mention a few benchmark studies relevant to ours and their reported statistics. (Ansdell et al. 2018) reported an accuracy of 95.8% with Astronet, which again, is convincingly outperformed by our proposed approach. Similarly, (Dattilo et al. 2019) proposed a modified version of Astronet, which yielded an accuracy of 98%.

The summary of our results is presented in Table 1. As can be seen, our proposed MTF-based classification achieves the highest accuracy compared to the other baselines. Surprisingly, the standard data provided almost 6% lower accuracy than ours. GADF provided the second-based performance in terms of accuracy, recall, and F-score. For all key metrics, we have comprehensively outperformed these baselines. Our results open the door for astrophysicists and computer scientists to use corresponding MTF images of the flux intensity data for future classification tasks.

Input Type	Accuracy	Precision	Recall	F-score
Standard	93.35	98.14	92.35	95.16
GADF	94.06	97.59	93.87	95.69
MTF (ours)	**99.39**	**99.68**	**99.46**	**99.57**

Table 1: Overall statistics (%)



a) Accuracy b) F-Score

Figure 2: Convergence of accuracy and F-score metrics during the validation.

Conclusion and Future Work

Light intensity data captured by NASA's Kepler Space Telescope is a significant data source for astrophysicists to identify exoplanets. Manual identification can be biased and infeasible due to the large volume of the dataset. To this end, we have proposed a deep learning-based classification technique that first converts the raw light curve data into Markov Transition Field images, which are then passed through our designed convolutional neural network. Numerical results show that our proposed technique outperforms two baselines by up to 6.04% in terms of classification accuracy. In the future, we plan to investigate fast and short exposure data types and will perform a more comprehensive comparative analysis against existing approaches.

References

Ansdell, M.; Ioannou, Y.; Osborn, H. P.; Sasdelli, M.; Smith, J. C.; Caldwell, D.; Jenkins, J. M.; Räissi, C.; Angerhausen, D.; et al. 2018. Scientific domain knowledge improves exoplanet transit classification with deep learning. *The Astrophysical journal letters*, 869(1): L7.

Dattilo, A.; Vanderburg, A.; Shallue, C. J.; Mayo, A. W.; Berlind, P.; Bieryla, A.; Calkins, M. L.; Esquerdo, G. A.; Everett, M. E.; Howell, S. B.; Latham, D. W.; Scott, N. J.; and Yu, L. 2019. Identifying Exoplanets with Deep Learning II: Two New Super-Earths Uncovered by a Neural Network in K2 Data. *arXiv preprint arXiv:1903.10507*.

Shallue, C. J.; and Vanderburg, A. 2018. Identifying exoplanets with deep learning: A five-planet resonant chain around kepler-80 and an eighth planet around kepler-90. *The Astronomical Journal*, 155(2): 94.

Yu, L.; Vanderburg, A.; Huang, C.; Shallue, C. J.; Crossfield, I. J.; Gaudi, B. S.; Daylan, T.; Dattilo, A.; Armstrong, D. J.; Ricker, G. R.; et al. 2019. Identifying exoplanets with deep learning. III. Automated triage and vetting of TESS candidates. *The Astronomical Journal*, 158(1): 25.