Boletim da Sociedade Astronômica Brasileira, **31**, no. 1, 67-68

Automatic detection of active stars in the Kepler satellite database

Edgar Sgroi Rocha[1], I. F. Santos[1], & Adriana Valio[2]

[1] Mackenzie Presbyterian University — Computing and Informatics Faculty
e-mail: edgar.sgroi@gmail.com, learsi.isr@gmail.com
[2] Mackenzie Presbyterian University — Center for Radio Astronomy and Astrophysics at Mackenzie
e-mail: adrivalio@gmail.com

Abstract. This work presents the application of neural networks for the identification of active stars in the Kepler database. Many of the discovered planets are in orbit of stars similar to the Sun. Some are bigger, some smaller, younger or older than our star. The study of solar-type stars of different ages is a powerful tool for understanding the Sun's behavior in the past and also what happen in the future. An important feature of stars is its magnetic activity. This activity manifests itself in the form of dark spots on the surface of stars. As the star rotates, these spots cause modulation in the lightcurve of the star.

Resumo. Este trabalho apresenta a aplicação de redes neurais para a identificação de estrelas ativas na base de dados do Kepler. Assim, muitos planetas descobertos se encontram em órbita de estrelas similares ao Sol. Algumas são maiores, outras menores, mais jovens ou mais velhas do que nossa estrela. Seu estudo é uma ferramenta poderosa para entendermos o comportamento do Sol no passado e também como será no futuro. Uma característica importante das estrelas é a sua atividade magnética. Esta atividade se manifesta na forma de manchas escuras na superfície da estrela. Conforme a estrela rotaciona, as manchas causam uma modulação na sua curva de luz.

Keywords. stars – active stars – neural network

1. Introduction

The Kepler mission has monitored approximately 160,000 stars in the direction of the Cygnus Constellation to detect exoplanets through the technique of planetary transits. Thus, considering the large volume of data in the mission base, it is necessary to use computational techniques for the organization and processing of data to facilitate the manipulation and use of such information.

Currently, more than 2,300 planets have already been discovered by the Kepler mission, many of these around active stars. Such activity can impact the habitability of planets very close to their host star. To infer this impact, automatic detection of active stars is necessary. This is done by measuring the light curve modulation of this star, due to the presence of dark starspots on its surface.

The objective of this work is the development of a software layer for the identification and classification of active stars in the Kepler database through the use of Artificial Neural Networks (RNA).

2. Active Stars

Young stars present fast rotation, producing a more efficient magnetic dynamo and therefore more spots on their surfaces, as shown in Figure 1. Because they have more spots, the peak-to-peak brightness variation is more intense for younger stars, up to 10% (Figure 1a). In the case of older stars such as the 4.6 billion-year-old Sun, the modulation is well under 1% (Figure 1c), according to Silva (2006).

Stellar activity, detected through modulation in its light curve, may have a fatal impact on the habitability of planets very close to its star (Estrela & Valio 2018). Automatic detection of active stars is necessary given the large amount of data.

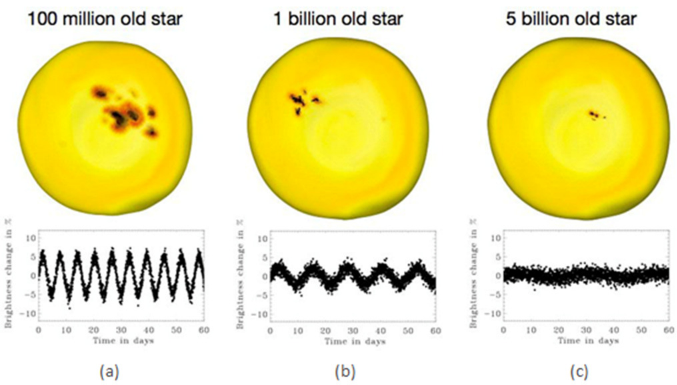

Figure 1. Light curve modulation due to stellar activity. Figure taken from http://amazingnotes.com/2011/06/05/how-astronomer-calculate-the-age-of-a-star.

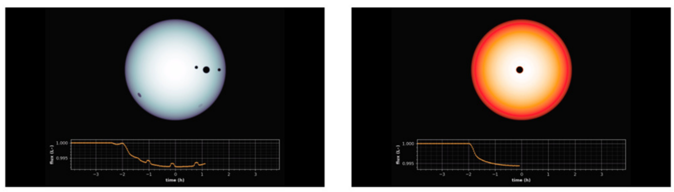

Figure 2. CRAAM-SVO — Planetary transit simulator.

3. Planetary Transit

When a planet eclipses its host star, there is a small decrease in the intensity of the stellar brightness, usually less than 1 %, depending on the radius of the planet. This periodic phenomenon is called planetary transit.

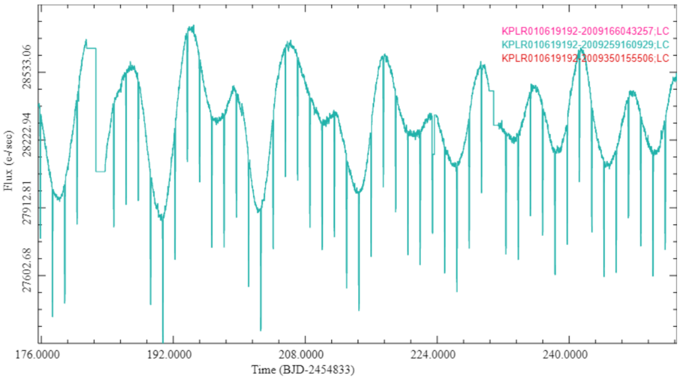

FIGURE 3. Example of an active star light curve. Figure taken from `http://archive.stsci.edu/cgi-bin/kepler/dataset_lookup`.

A model that simulates the passage of the planet, and its exomoons, in front of the star is shown in Figure 2 (Santos et. al. 2017). The model yields the stellar flux as a function of time, that is light curve, with the transits shown in the bottom panel. The shape of the transit depends on parameters of the planet and its orbit such as planet radius, orbital angle of inclination, and the semi-major axis of the orbit. Small signatures caused by transit over starspots can also be identified in the transit light curve (Silva 2003).

4. Artificial neural networks

Artificial Intelligence has different areas, such as deep learning that consists in the creation of algorithms to favor that a certain machine can evidence characteristics for which it was previously trained. for this, a set of data is presented, and from these data, the correlation between them are explored, discovering patterns, applying algorithms, and generating models that can be generalized to a specific task. Figure 3 presents a stream of data to be analyzed to classify whether the star is active.

Supervised learning consists of labeled data, which has the corresponding correct outputs, that is, the model output is compared to the expected output, by adjusting the parameters to an acceptable threshold. Unsupervised learning consists of presenting unlabelled data, so the algorithm seeks similarities between the data to form groups.

5. Backpropagation

Backpropagation is one of the forms of machine learning in which the model is trained by input data and expected output data according to Russel & Norvig (2016).

The backpropagation algorithm consists of two phases, the first is forward pass, where the inputs are passed through the network and the output predictions obtained (Russel & Norvig 2016). The second phase is a backward pass, where we calculate the loss function gradient in the final layer, *i.e.*, the network forecast layer, and use this gradient to recursively apply the chain rule in order to update the network weights.

Model creation consists of model building, model training, model testing and model evaluation (Russel & Norvig 2016). Thus, the construction of the model depends on the algorithm of machine learning. After constructing the model, it needs to be trained, as presented in 1. At this stage, the model is trained using training data and expected output for this data.

Once the model has been trained, it is possible to test the model. During this phase, a second set of data is loaded. This dataset has never been seen by the model and therefore true accuracy will be verified.

Upon completion of the model training, and if the model shows the correct result, it can be saved. Finally, the saved model can be used in the real world. Thus this model can be used to evaluate new data. This project is in the test phase of the model.

```
batch_size = 256

# configuration use for training
train_datagen = ImageDataGenerator(
        rescale=1./255,
        shear_range=0.2,
        zoom_range=0.2,
        horizontal_flip=True)

# configuration use for testing:
# only rescaling
test_datagen = ImageDataGenerator(rescale=1./255)

# this is a generator that will read pictures
# found in subfolers of 'data/train', and
# indefinitely generate batches of augmented image data
train_generator = train_datagen.flow_from_directory(
        'flux',  # this is the target directory
        target_size=(300, 300),  # resized to 300x300
        batch_size=batch_size,
        class_mode='binary')  # binary_crossentropy loss

# this is a similar generator, for validation data
validation_generator = test_datagen.flow_from_directory(
        'validation',
        target_size=(300, 300),
        batch_size=batch_size,
        class_mode='binary')
```

Code 1: Model Test.

6. Conclusions

This study is based on the Artificial Neural Networks (RNA) strategy for pattern recognition by learning machines to automatically identify active stars in the Kepler mission database. The results point to the feasibility of automatic detection of active stars considering the large volume of data obtained by the Kepler mission. Stellar activity, detected through modulation in its light curve, may have a fatal impact on the habitability of planets very close to its star.

Acknowledgements. We thank MackPesquisa, the research funding program of the Mackenzie Presbyterian Institute and FAPESP (São Paulo Research Foundation)

References

Estrela, R., & Valio, A. 2018, *Astrobiology*, 18, 1414-1424
Russell, S. J., & Norvig, P. 2016. Artificial intelligence: a modern approach. Malaysia: Pearson Education Limited
Silva, A. 2003, *ApJ (Letters)*, 585, L147
Silva, A. V. 2006. Nossa Estrela: O Sol. São Paulo: Livraria da Fisica
Santos, V. N., Santos, I. F., & Valio, A. 2017, In: Proceedings da XLI Reunião Anual da Sociedade Astronômica Brasileir, Boletim da Sociedade Astronômica Brasileira, 30, 50