Boletim da Sociedade Astronômica Brasileira, **31**, no. 1, 139-140

Study of quasar OV236 variability with scalable tree boosting system

S. B. Soltau[1] & L. C. L. Botti[2]

[1] CAGE/UPM. Postgraduate Program in Science and Geospatial Applications at Mackenzie Presbyterian University. UNIFAL-MG. Federal University of Alfenas. e-mail: sbsoltau@gmail.com

[2] CRAAM, Center for Radio Astronomy and Astrophysics Mackenzie, Engineering School, Mackenzie Presbyterian University. DAS/CEA/INPE/MCTIC, National Institute for Space Research. ROI/INPE, Itapetinga Radio Observatory. e-mail: luizquas@yahoo.com.br

Abstract. One of the mostly interest aspects of compact radio sources is the variability. This preliminary study examines the possibility of using a machine learning method called XGBoost to explore OV236 (PKS 1921-293) periodicity, through flux density treatment and analysis in light curves, on data from the University of Michigan Radio Astronomy Observatory (UMRAO), acquired at 4.8 GHz, 8.0 GHz, and 14.5 GHz, between 1975 to 2011 years. EXtreme Gradient Boosting (XGBoost) is a scalable end-to-end tree boosting system effective and used machine learning method, which is used widely by data scientists to achieve state-of-the-art results. XGBoost was used both to in the creation of a regularly-spaced temporal series of density flux data, and to identify accented activity ranges, that are outburst-characteristic.

Resumo. Um dos principais aspectos de interesse das radiofontes compactas é a variabilidade. Este estudo preliminar examina a possibilidade de usar um método de aprendizado de máquina chamado *XGBoost* para explorar a periodicidade do OV236 (PKS 1921-293), por meio do tratamento e análise da densidade de fluxo nas curvas de luz, em dados do Observatório de Radioastronomia da Universidade de Michigan (*UMRAO*) obtidos em 4, 8 GHz, 8, 0 GHz e 14, 5 GHz, entre os anos de 1975 e 2011. O EXtreme Gradient Boosting (*XGBoost*) é um sistema de árvore otimizado escalonável de ponta a ponta e eficaz método de aprendizado de máquina, que é amplamente utilizado por cientistas de dados para obter resultados no estado da arte. O *XGBoost* foi usado tanto na criação de uma série temporal regularmente espaçada de dados de fluxo de densidade, quanto na identificação de faixas de atividade acentuadas, que são características de *outburst*.

Keywords. (Galaxies:) quasars: general – Methods: data analysis

1. Motivation & Background

The Catalog of Quasars and Active Galactic Nuclei (AGN) by Véron-Cetty, M.-P., & Véron (2010) shows that these object make up around 7% of all observed galaxies. A quasi-stellar radio source (quasar) is an AGN with broad emission lines at high redshift. Around 5% to 10% of quasars are strong radio sources (D'Onofrio, Marziani & Sulentic 2012).

The radio source OV236 (PKS 1921-293) is a quasar that shows great variability and one of the most compact known Tornikoski (1996), emits in all energy range. It is also classified as a *BL Lacertae* object (Impey, Brand, Wolstencroft & Williams 1982), due to its polarization about 14% and variability 2.2 μm (137 THz, in the mid-infrared range). It has an emission-line spectrum at a redshift $z = 0.3525$ (Wills & Wills 1981).

Multi-frequency studies explore distinct aspects of compact radio sources, in particular, flux density variations, to determine periodicities in light curves (Botti 1990; Gastaldi 2017). Methods for determining periodicities in the radio range include Fourier Transform, Lomb-Scargle Periodogram, Wavelet Transform and Cross Entropy, among others (Cincotta & Nunez 1995).

Advances in Artificial Intelligence have provided machine learning algorithms, such as Neural Networks, Ensemble and Deep Learning, that have entered astrophysical studies and provided computational approaches dissimilar to previous methods, including potential applications for radio source analyses (Witten, Frank, Hall & Pal 2016).

EXtreme Gradient Boosting (XGBoost) is a set of machine learning methods boosted tree based, packaged in a library designed and optimized for the creation of high performance algorithms (Chen & Guestrin 2016).

2. Data source & objective

In this first exploratory study, the aim was to apply XGBoost in the characterization of the radio source OV236, through flux density treatment and analysis in light curves, using data from the University of Michigan Radio Astronomy Observatory UMRAO (2018), acquired at 4.8 GHz, from 1980 to 2011; 8.0 GHz, from 1975 to 2011 and 14.5 GHz, from 1976 to 2011. The radio source OV236 data at UMRAO operating frequencies comprises an irregular sampling time series, due to several factors that influence astronomical data acquisition from terrestrial stations, such as meteorological conditions, maintenance of system, receivers etc. These sampling difficulties produce unevenly spaced time series, which impose limitations on more conventional analysis methods.

3. Methodological procedures & computational screening

In this approach, XGBoost was applied to two tasks. First, in the creation of a regularly-spaced temporal series (Fig. 1), and, second, in the analysis of the temporal evolution of the light curve in each frequency, to discover ranges of accentuated flux density and distinctive outburst features, and, thus, determine the periodicity of these events (Fig. 2). During data processing, by a number of experimentation feature trainings, XGBoost was able to capture light curve behavior, allowing for the circumvention

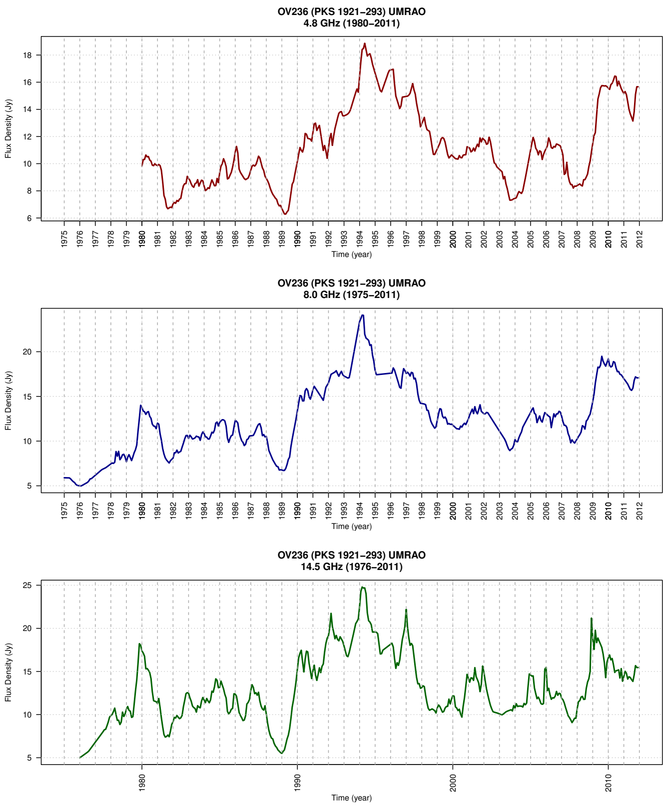

Figure 1. Light curves with 4.8 GHz, 8.0 GHz and 14.5 GHz from UMRAO, after regularly-spaced temporal series process by XGBoost 1st phase computational screening.

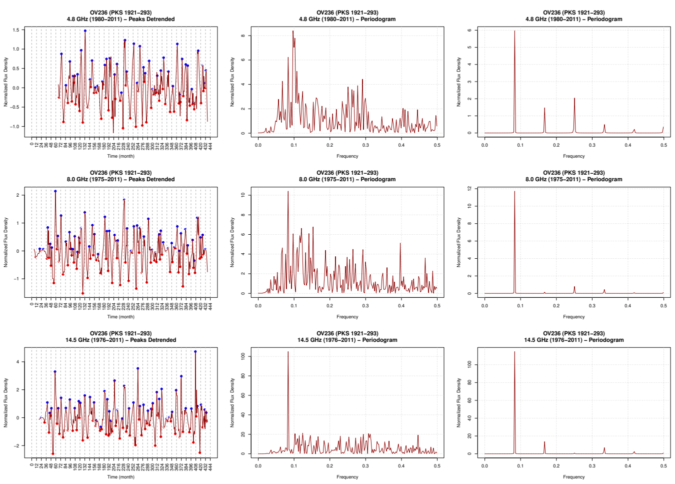

Figure 2. Light curves with 4.8 GHz, 8.0 GHz and 14.5 GHz normalization, detrended and peaks detected and successive improved accuracy by XGBoost 2nd phase computational screening.

of possible distortions and unwanted smoothing, typical of traditional periodicity calculation methods.

4. Preliminary results

The analysis phase, which initially provided training data for the algorithm created with the XGBoost library, was identified data segments that allowed for estimations of the occurrence frequency of events between 14, 36, 62 months in the 1975 to 2011 time series, at 4.8 GHz; 15, 31, 57 months at 8.0 GHz and 15, 53, 76 years at 14.5 GHz, with an accuracy of about 78% calculated by the algorithm (Fig. 1).

5. Future Prospects

This are preliminary and partial results of a study under development. Improvements in the computer codes, refinements in the periodicities, and attach additional frequencies data are some of next steps needed. In order to continue this research, similar analysis carried out for other energy range such as X-rays, to deepen knowledge on periodicity and possible correlations between components of the OV236 radio source in particular, and quasars in general, provided by flux density analyses (Fig. **??**).

XGBoost robustness provided the opportunity to investigated statistical correlations between the data acquisition UMRAO frequencies in 4.8 GHz, 8.0 GHz and 14.5 GHz.

Acknowledgements. I would like to thank my supervisor, Prof. Botti, for the patient guidance and advice he has provided. To CAGE/UPM — Postgraduate Program in Science and Geospatial Applications at Mackenzie Presbyterian University, CRAAM – Center for Radio Astronomy and Astrophysics Mackenzie and Mackenzie Presbyterian University. To my colleagues at Federal University Alfenas, Prof. Humberto Brandão, for share your expertise on XGBoost, Prof. Aníbal Bezerra, for active partnership in Python machine learning programming and Prof. Artur Justiniano by the insights and challengers suggestions. This research has made use of data from the University of Michigan Radio Astronomy Observatory which has been supported by the University of Michigan and by a series of grants from the National Science Foundation, most recently AST-0607523.

References

Botti, L. C. L.1990, Spectrum variability study of radio sources in the 22 to 43 GHz range (Unpublished doctoral thesis). Instituto de Pesquisas Espaciais, S ao José dos Campos (Brazil)

Chen, T., & Guestrin, C. 2016. in Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining - KDD '16

Cincotta, P. M., Mendez, M., & Nunez, J. A. 1995, Astrophysical Journal, Supplement, 449 , 231

D'Onofrio, M. and Marziani, P. & Sulentic, J. W, 2012, Fifty Years of Quasars: From Early Observations and Ideas to Future Research, (New York: Springer)

Gastaldi, M. R. 2017, Uso do Periodograma de Lomb e da transformada Wavelet para detecçÃo de periodicidades em radiofontes extragalácticas (Unpublished doctoral thesis). CAGE/UPM, S ao Paulo (Brazil)

Impey, C. D., Brand, P. W. J. L., Wolstencroft, R. D., & Williams, P. M. 1982, Monthly Notices of the Royal Astronomical Society, 200, 19

Tornikoski, M., et al. 1996, Astronomy and Astrophysics, Supplement, 116, 157

University of Michigan Radio Astronomy Observatory 2018, UMRAO Database Interface. Retrieved from `https://dept.astro.lsa.umich.edu/datasets/umrao.php`

Véron-Cetty, M.-P., & Véron, P. 2010, Astronomy and Astrophysics, 518 , A10

Wills, D., & Wills, B. J. 1981, Nature, 289, 384

Witten, I. H. and Frank, E. and Hall, M. A. and Pal, C. J. 2016, Data Mining: Practical Machine Learning Tools and Techniques, (Cambridge, MA: Elsevier Science)