EPSC Abstracts
Vol. 13, EPSC-DPS2019-130-2, 2019
EPSC-DPS Joint Meeting 2019

EPSC
EPSC-DPS Joint Meeting 2019
15–20 September 2019 | Geneva, Switzerland

Machine learning identification of asteroid families

Valerio Carruba (1), Safwan Aljbaae (2), André Lucchini (1), and Edmilson R. De Oliveira (1)
(1) São Paulo State University (UNESP), School of Natural Sciences and Engineering, Guaratinguetá, SP, 12516-410, Brazil (valerio.carruba@unesp.br)
(2) National Space Research Institute (INPE), Division of Space Mechanics and Control, C.P. 515, 12227-310, São José dos Campos, SP, Brazil.

Abstract

Asteroid families are groups of asteroids that share a common origin. They can be the outcome of a collision or be the result of the fission of a parent body or its satellites. Collisional asteroid families have been identified for several decades using Hierarchical Clustering Methods in proper elements domains. In this method, the distance of an asteroid from a reference body is computed, and, if it is less than a critical value, the asteroid is added to the family list. The process is then repeated with the new object as a reference, until no new family members are found. Recently, new machine-learning clustering algorithms have been introduced for the purpose of cluster classification. Here, we apply supervised-learning hierarchical clustering algorithms for the purpose of asteroid families identification. Values of the areas under the curve (AUC) coefficients below Receiver Operating Characteristic (ROC) curves for the identified families are optimal, consistently above 85%. Overall, we identify 7 new families and 15 new clumps in regions where the method can be applied, that appear to be consistent and homogeneous in terms of physical and taxonomic properties. Machine-learning clustering algorithms can, therefore, be a very efficient and fast tools for the problem of asteroid family identification.

1. Introduction

Asteroid families are groups of asteroids that share a common origin. They can either be the product of a collision, or they may result from the rotational fission of a parent body or their satellites. Collisional asteroid families are usually identified in domains of proper elements that are constants of motion over timescales of Myr [7]. Among the methods used for identifying asteroid families, the hierarchical clustering method (HCM) is one of the most commonly used (see [2] and the method section for a discussion of this approach).

Recently, machine-learning clustering algorithms, available in the *PYTHON* programming language, have been used with great success for problems like clusters identification [8] Methods such as K-means, mean-shift, hierarchical clustering algorithms, are now very commonly used among data scientists, and applied to various different fields, such as biology, paleontology, and, as in this work, astronomy. These algorithms have proven to be efficient, fast, and reliable in problems of supervised learning. Here, we will attempt to use machine-learning hierarchical clustering methods for the purpose of asteroid families identification, in domain of asteroid proper elements $(a, e, \sin i)$.

2. Methods

In the HCM,First, the distance between pairs of objects involving a putative parent body and a family candidate is computed in the domain of proper elements according to a predefined metric. The most commonly used distance metric d is defined as [2]:

$$d = na\sqrt{\frac{5}{2}\times(\frac{\Delta a}{\overline{a}})^2 + 2\times(\Delta e)^2 + 2\times(\Delta \sin i)^2}, \quad (1)$$

where (a, e, i) are the proper semi-major axis, eccentricity and inclination, the symbol Δ is associated with the difference between pairs of proper elements, and $\overline{a}$ is the mean value of the proper semi-major of a given pair of asteroids. If the distance between two objects is less than a value defined as the local cut-off, the family candidate is assigned to the parent body asteroid group. The process is then repeated with the new family member now considered as a parent body until no new members are found. [1] define a nominal distance cut-off d_0 as the average minimum distance between all neighboring asteroids in the same region of the asteroid.

In this work, we implemented scikit-learn [8] hierarchical clustering algorithms for the problem of asteroid family identification. Dendrogram clusters of

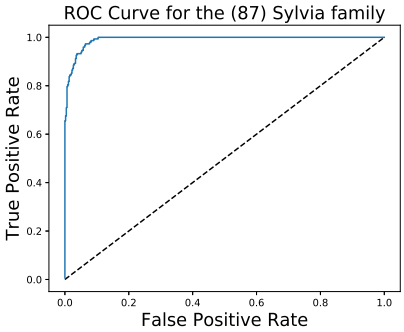

Figure 1: Area under Receiver Operating Characteristic (ROC) curves for the case of the identification of the (87) Sylvia family.

asteroid distances computed using Eq. (1) can be obtained almost instantaneously using Python algorithms such as *linkage*, and asteroid families obtained for different cut-off values can be easily identified.

To study the accuracy of this method, we use the Receiver Operating Characteristic (ROC) curve approach. ROC curves are a well-established tool to quantitatively define the efficiency and accuracy of machine-learning clustering algorithms. A detailed description of how ROC curves are computed and of the rationality behind them can be found in [6]. In this work, we first computed asteroid distances with respect to the alleged parent body using the distance metric given by eq. (1). The sample of family members obtained by the machine-learning hierarchical clustering method is then randomly split into a train sample (60% of the total) and a test sample (40%). A logistic regression algorithm is then applied to the train sample, and the probability of the test sample to belong to the family is then assessed based on the results of the train sample.

Figure (1) show ROC curves for the family of (87) Sylvia, that has an AUC score of 0.991. For the studied families, AUC values were consistently above 0.855, which suggest a very high performance of the method.

3. Summary and Conclusions

Following the approach of [3], families are considered robust when they are recognizable not only at the nominal cutoff, but for a range of at least 10 m/s above and below this value. We considered a group to be a family if there are at least 25 members, and a clump if there are at least 10 members. Overall, we identify 7 new families and 15 new clumps in regions where the method can be applied, i.e., the regions of the Hungaria asteroids, the inner, central and outer main belt at high inclinations, and the region of the Cybele asteroids, for which classical HCM can be applied without incurring in the problem of *chaining*. The new groups low appear to be consistent and homogeneous in terms of physical and taxonomic properties. Machine-learning clustering algorithm can, therefore, be a very efficient and fast tool for the problem of asteroid family identification.

Acknowledgments

We would like to thank the São Paulo State Science Foundation (FAPESP, grant 2018/20999-6) and the Brazilian National Research Council (CNPq, grant 301577/2017-0).

References

[1] Beaugé, C., Roig, F., 2001, Icarus, 153, 391.

[2] Bendjoya, P., Zappalá V., 2002. Asteroids III. Univ. Arizona Press, Tucson, 613.

[3] Carruba, V., 2010b, MNRAS, 408, 580.

[4] Carruba V., Nesvorný, D., Aljbaae, S.,Huaman, M. E., 2015, MNRAS, 451, 4763.

[5] Carruba, V., Aljbaae, S., Lucchini, A., De Oliveira, E. R. 2019, MNRAS, submitted.

[6] Fawcett, Tom (2006). Pattern Recognition Letters. 27, 861

[7] Knežević, Z., Milani, A., 2003, A&A, 403, 1165.

[8] Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V. et al., 2011, JMLR 12, 2825.