Automatic Morphological Classification of Galaxy Images

Salinas J. Z.[a1], García-Torres M.[b], Schaerer C. E.[a], Legal H. [a] and Divina F.[b]

[a]National University of Asunción, San Lorenzo, Paraguay

[b]Pablo de Olavide University, Seville, Spain

Received on August 23, 2016 / accepted on September 14, 2016

Abstract

Galaxy classification is an important task in astronomy in the large scale study of the universe. Although this task traditionally is done manually, astronomy has experienced an explosion of data that require the use of new techniques to deal with this increase of the data volume. In this paper we analyze the performance of several machine learning approaches (naive Bayes, support vector machine and C4.5) on automated classification of galaxy images. We tested the approaches on 82 images taken from the [1] catalog and, following the previous works [2, 3], we considered three (E, S, Irr), five (E, S0, Sa+Sb, Sc+Sd, Irr) and seven (E, S0, Sa, Sb, Sc, Sd, Irr) galaxy types.

The images are standardized to remove noise and the effect of orientation and translation. Features are extracted by morphological appearance (MF), Principal Component Analysis (PCA) and Independent Component Analysis (ICA). The MF are based on the perceived visual characteristics of the galaxy like elongation, form-factor, convexity, bounding-to-fill factor, asymmetry index, horizontal and vertical peaks of histogram. PCA and ICA are extracted from the covariance matrix of the transpose of dataset $C = A^T A$ where a row of A represent an image converted into a 1 dimensional vector.

Experimental results show that increasing the number of galaxy types degrades the model performance. The combination of standardization with PCA leads to classification models with high accuracy (89% with naive Bayes for 3 galaxy types). However, although morphological features leads to models with lower predictive power (84% with naive Bayes and 3 galaxy types), this type of features allows a better understanding of the classification rules.

Despite the promising results, more research is necessary to improve the classification when increasing the number of glaxy types. A possible reason of the poor results achieved for 5 and 7 classes could be the small number of instances for some classes; therefore increasing the dataset could improve the results.

Keywords: Image processing, datamining, machine learning.

[1]E-mail Corresponding Author: jzsalinas@gmail.com

References

[1] Z. Frei, "Automatic morphological classification of galaxies" Institute of Physics, Eötvös University.

[2] J. de la Calleja, O. Fuentes. "Automated Classification of Galaxy Images," Instituto Nacional de Astrofísica, Óptica y Electrónica, Luís Enrique Erro 1. 2004.

[3] S. Kasivajhula, N. Raghavan, H. Shah. "Morphological Galaxy Classification Using Machine Learning".