

# The verification of Machine Learning methods for binary morphological classification of SDSS-galaxies

Tuesday, 24 December 2019 10:55 (20 minutes)

Classification of galaxy types is one of the cornerstones for extragalactic astrophysics and observational cosmology. The most precise method of galaxy classification used by astronomers so far is the manual classification. Its major drawback, however, is the extensive usage of manpower, either from highly skilled professionals, or, in some cases (such as in Galaxy Zoo project) amateur astronomers. Modern or forthcoming galaxy surveys (SDSS, LSST, DES, KiDS, etc.) are able to detect hundreds of millions of galaxies, impossible to classify manually. This exaggerates the interest to use the alternatives in form of various recently developed Machine Learning (including Deep Learning) techniques for automated classification of galaxies.

We present a study on the verification of Machine Learning (ML) methods to be applied for binary morphological classification of galaxies (*E* - Early and *L* - Late types). We used two kinds of supervised learning techniques: the first is classical machine learning classifiers, and the second is deep convolutional neural network classifier.

We want to emphasize that, unlike most other authors, we paid attention to the visual cleaning of the dataset. Our dataset contains of  $\sim 300\,000$  SDSS-galaxies from DR9 with unknown morphological types at  $z < 0.1$  redshift. The training galaxy sample contains of  $\sim 6\,000$  galaxies with the visually inspected morphological types selected randomly with different redshifts and luminosity from the total sample.

We applied the following classical machine learning classifiers: Naive Bayes, Random Forest, Support Vector Machines, Logistic Regression, and k-Nearest Neighbor algorithm. To study the classifier, we used absolute magnitudes, color indices and inverse concentration index to the center. It turned out that Support Vector Machine Classifiers provide a highest accuracy - 96.4 % (96.1 % - *E* and 96.9 % - *L*).

We used the deep convolutional neural network classifier, namely Xception, to provide a classification of (g-r-i) composite images (25 arcsec in each axis in size) of SDSS-galaxies. We provided the data augmentation (horizontal and vertical flips, random shifts on  $\pm 10$  pixels, and rotations within 180 degrees), that was randomly applied to the images during learning. It turned out that deep convolutional neural network provide accuracy - 93.5 % (95 % - *E* and 86 % - *L*).

We compared our results and proposed a method to learn the benefits from both approaches (classical machine learning classifiers and deep convolutional neural network classifier). We show the common mistakes of both algorithms, and propose to stack these two approaches to block these mistakes, with a main goal to increase the overall classification quality of SDSS galaxies.

**Primary authors:** Dr DOBRYCHEVA, Daria (Main Astronomical Observatory of the National Academy of Sciences of Ukraine); Mr VASYLENKO, Maksym (Main Astronomical Observatory of the National Academy of Sciences of Ukraine); Mr KHRAMTSOV, Vlad (Institute of Astronomy, V.N. Karazin Kharkiv National University); Dr VAVILOVA, Irina (Main Astronomical Observatory of the National Academy of Sciences of Ukraine)

**Presenter:** Dr DOBRYCHEVA, Daria (Main Astronomical Observatory of the National Academy of Sciences of Ukraine)

**Session Classification:** Astrophysics and Cosmology

**Track Classification:** Astrophysics and Cosmology