
Beyond the Hubble Sequence - Exploring galaxy morphology with unsupervised machine learning

Ting-Yun Cheng^{*1}, Marc Huertas-Company^{2,3}, Christopher Conselice^{5,4}, Alfonso Aragon-Salamanca⁵, Brant Robertson⁶, and Nesar Ramachandra⁷

¹Durham University – United Kingdom

²Instituto de Astrofísica de Canarias (IAC), Universidad de La Laguna (ULL) – Spain

³Observatoire de Paris – Sorbonne Université, Observatoire de Paris, Université PSL, CNRS, LERMA, Paris, France – France

⁵University of Nottingham – United Kingdom

⁴University of Manchester – United Kingdom

⁶University of California, Santa Cruz – United States

⁷High Energy Physics Division, Argonne National Laboratory – United States

Abstract

Conventionally, galaxy morphological classifications are defined by visual assessment. However, visual classification systems such as Hubble types can be intrinsically biased due to the subjective judgement of human classifiers. Additionally, since morphological "classifications" into types is an important and complementary process, it is not clear if we know what these "best types" are, such as, whether a classification scheme results in relatively unique physical properties of the galaxies or traces the merger history in each class. As most machine learning applications in astronomical studies focus on the improvement of accuracy and efficiency, we apply machine learning to approach a new insight towards galaxy morphological classification. In particular, machines decide a classification system to describe the variation shown in galaxy morphology in the dataset. We explore galaxy morphology from SDSS imaging data with an unsupervised machine learning technique composed of a feature extractor using vector-quantisation variational autoencoder and hierarchical clustering. Our methodology results in 27 machine-defined classes which are physically distinctive from each other in stellar mass, absolute magnitude, physical size, and colour. When we merge these clusters into 2 clusters for binary classification, the unsupervised method provides an accuracy of 87% for separating early-type (ETG) and late-type galaxies (LTG) using a dataset with the morphology distribution of nearby galaxies (i.e., 22.7% ETG and 77.3 LTG).

Poster: in PDF

Video: <https://youtu.be/PwILsBo-djE>

*Speaker