NATIONAL CENTRE FOR NUCLEAR RESEARCH

Abstract

Search for galaxy mergers in big sky surveys

Luis E. SUELVES

During their lifecycle and evolution, galaxies can approach each other and collide, becoming some of the most impressive views in the sky. These are galaxy mergers, as their interaction can result in two or more galaxies merging into one. These galaxies show a high variability of morphological distortions due to the tidal forces arising during the process. Their aspect depends on the relative masses, the stage within the process, the line-of-sight projection in the sky, the brightness, distance, and size of the merging members, and the properties of the observation and instruments.

This thesis covers the studies I carried out to classify galaxy mergers in the big datasets obtained from wide sky observations. The surveys I worked with are the Sloan Digital Sky Survey (SDSS) and the deeper Subaru/Hyper Suprime-Cam (HSC). Both preceded the upcoming Large Survey of Space and Time (LSST), which will be carried by the Vera Rubin Observatory. LSST will capture in one night the same amount of data that SDSS collected over nearly a decade. The steep increase of data size from survey to survey has made it imperative to find automatized techniques to treat the data. For that, I worked on dimensionality reduction methods, on the image calibration, and on Machine Learning (ML) techniques such as Neural Networks (NN) or dimensionality reduction methods.

In this thesis, I describe the discovery and development of a new potential methodology to identify galaxy mergers in large surveys. This method is based on the effect that the surrounding of galaxies has in the sky background analysis of the astronomical image. It was discovered when testing the performance of a NN using only photometric information. We trained it on a class-balanced dataset of mergers and non-mergers built out of SDSS galaxies, classified visually by online volunteers in the Galaxy Zoo Data Release 1 (GZ DR1). Testing multiple combinations of the photometric parameters as NN inputs led us to find how the SDSS DR6 sky background error was capable of tracing galaxy mergers with a training-set accuracy of 92.64 ± 0.15 % test-set accuracy of 92.36 ± 0.21 %. Moreover, studying the sky error plane formed by the g and r bands revealed that a decision boundary line is enough to achieve an accuracy of 91.59%. The interpretation of this result is that the sky background error is tracing low signal-to-noise features around the observed galaxies.

The success of the sky error boundary led us to test its extension to a wider set of galaxies from SDSS DR6. This was in fact the whole GZ DR1 set. We studied the presence of merging galaxies of various types in different regions of the diagram. It was also found that non-merging galaxies with nearby stars and non-interacting galaxies could appear in the regions populated by mergers. In order to avoid this contaminated non-mergers, we built a decision tree that discarded galaxies with nearby stars or galaxies too far to be potentially colliding. This provided a 67.07 % of dirty non-mergers that were successfully discarded, and a 72.44 % percent of mergers that were correctly retained. Thus, further tailoring of our approach with focus on reducing the contamination would make the boundary capable of finding SDSS sources.

Finally, we tested how to implement the sky error method in the deeper HSC images. We used images on the North Ecliptic Pole (NEP), available through the AKARI-NEP

collaboration. In this field, I joined the ML-based search for mergers in the HSC images of the North Ecliptic Pole (NEP) led by my auxiliary supervisor dr. William J. Pearson. We confirmed visually the merger candidates identified by the model and published the first catalogue of mergers in the NEP. The Galaxy Zoo: Cosmic Dawn! program provided a follow-up set of morphological classifications, from which I selected a sample for the HSC sky error extension. Within an aperture around each source, I estimated the background and obtained the distribution of low Signal-to-noise pixels. Because the sky error in SDSS is computed during the sky background measurement, I also created my own data reduction, making sure the Low Surface-Brightness (LSB) features from the tidally striped material around the mergers are not lost. I compared the resulting images with and without subtraction of the dithering-based background model, and concluded it does not eliminate the LSB features in the image. I calculated multiple parameters on the LSB pixels, observed the parameter space through dimensionality reduction methods, and found that they have the potential of creating a new area where mergers can be found.

Overall, in this thesis I demonstrate that galaxy mergers can be identified in large sky surveys by the effect that the low signal-to-noise stripped material has around them in the images.