Contribution ID: **115**

Type: **Talk**

Anomaly detection using local measures of uncertainty in latent representations

Friday, December 1, 2023 5:30 PM (15 minutes)

As upcoming SKA-scale surveys open new regimes of observation, it is expected that some of the objects they detect will be "unknown unknowns": entirely novel classes of sources which are outside of our current understanding of astrophysics. The discovery of these sources has the potential to introduce new fields of study, as it did for e.g. pulsars, but relies upon us being able to identify them within peta- or exascale data volumes. Automated anomaly detection using machine learning offers a promising method to ensure that atypical sources are not lost within the data, but these methods are typically incapable of simultaneously classifying non-anomalous sources and identifying anomalies, resulting in separate models being needed to complete both necessary tasks. In this talk, we discuss the possibility of using uncertainty metrics derived from an image classification model to provide anomaly detection within a classification pipeline.

Fanaroff-Riley (FR) galaxies, a type of radio-loud AGN, are among the sources that are expected to see a drastic increase in known population with upcoming large-scale radio surveys, and provide a useful test population for outlier detection because in addition to two "standard" morphologies (FRI and FRII) there are numerous rare morphological subclasses which are particularly useful for the study of AGN environments and engines. Using the MiraBest dataset of Fanaroff-Riley galaxies, we trained supervised deep learning model on binary Fanaroff-Riley sources, reserving hybrid FR galaxies to serve as a sample of "anomalous" objects that might be mistaken for in-distribution sources. Our model architecture used dropout at test time to approximate a Bayesian posterior on predictions, allowing for uncertainty in a class label to be expressed by calculating predictive entropy.

Highly anomalous out-of-distribution sources were found to be located in sparse regions of latent space and hence were easily identifiable, but hybrid sources could not easily be isolated from binary FR galaxies in either latent space or by entropy value alone. Instead, we created a measure of typical local entropy by calculating the average entropy of the nearest ten training set sources to any given point in latent space; this allowed for objects with atypically high or low entropy relative to nearby sources to be identified regardless of the absolute value of their entropy.

Using a test set of both in-distribution binary FRs and "anomalous" hybrid sources, we find that the in-distribution sources show no significant departure from the training set entropy, but hybrid sources have significantly higher entropy than their surroundings in all regions of latent space except where the local entropy is itself maximal. All sources more than 3σ from the local entropy were found to be hybrids, and flagging using this method alone detected 30% of the hybrid sample; the majority of the remaining hybrid sources were found to have near-maximal entropy, meaning that additionally flagging high-entropy sources would allow for both these and the most uncertainly-labelled in-distribution FR galaxies to be inspected while avoiding unnecessary flagging of low-uncertainty sources.

Primary author: Dr PORTER, Fiona (Jodrell Bank Centre for Astrophysics)

Co-author: Prof. SCAIFE, Anna (Jodrell Bank Centre for Astrophysics)

Presenter: Dr PORTER, Fiona (Jodrell Bank Centre for Astrophysics)

Session Classification: Contributed talks

Track Classification: Paris