Quantifying uncertainty in deep learning classification of radio galaxies

Fiona A. M. Porter*(1) and Anna M. M. Scaife(1)(2)
(1) Jodrell Bank Centre for Astrophysics, University of Manchester, UK
(2) The Alan Turing Institute, Euston Road, London, NW1 2DB, UK

1 Extended Abstract

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 the advent of SKA-scale surveys, providing a wealth of information about AGN and their local environments but necessitating the use of robust automated classification to be identified and labelled. While efforts have been made to create such classifiers for FR galaxies, there remains the issue that even the best-trained models are uncertain regarding some labels, just as human classifiers are.

To investigate the uncertainty properties of the FR population we used a dataset of FR galaxies, labelled both for their FR class and the confidence of a human classifier in the assigned class, and trained a Convolutional Neural Network (CNN) using the LeNet architecture modified to include dropout preceding the output layer in order to approximate a Bayesian posterior on predictions. This model was used to extract two types of uncertainty measure from the CNN: aleatoric (irreducible uncertainty resulting from traits of a source that make its class inherently unclear) and epistemic (reducible uncertainty resulting from the model having a limited quantity of data from which to learn class information).

We used the evaluation metrics of entropy and mutual information [1] to calculate the total predictive uncertainty and epistemic uncertainty, respectively, for each source. We then fitted the distributions of sources in latent space using a Gaussian Mixture Model (GMM), and calculated a GMM score that represented the probability of a source belonging to either distribution, which was also used as a measured of epistemic uncertainty [2].

We find that the GMM fitted shows a region of overlap between clusters for the two FR classes, and that presence in this overlap region correlates well with high entropy. Additionally, while both confidently- and uncertainly-labelled sources are present in this region, the density of uncertainly-labelled sources is greater, and confidently-labelled sources commonly show unusual morphology (e.g. wide-angle tail) or possible contamination by background sources. From this, we conclude that entropy (and hence model uncertainty) correlates with human uncertainty in assigning classes, and that the model typically agrees well with humans as to which sources are difficult to classify.

Additionally, in regions of low GMM score (high epistemic uncertainty), some sources which are labelled by the model with very high softmax probability and very low entropy are nonetheless misclassified. These outlying sources reaffirm that softmax probability is not a reliable method to measure confidence in model classification, and show that sources with high epistemic uncertainty can have entropy of arbitrary value [2]. High values of epistemic uncertainty may hence be used in combination with entropy to flag sources for human inspection to determine whether a source is more likely to simply have ambiguous morphology or to be entirely out of distribution.

2 Acknowledgements

FAMP gratefully acknowledges support from STFC and IBM through the iCASE studentship ST/P006795/1. AMS gratefully acknowledges support from Alan Turing Institute AI Fellowship EP/V030302/1.

References

[1] Y. Gal, "Uncertainty in deep learning", Ph.D. dissertation, Dept. of Engineering, University of Cambridge, Cambridge, 2016. Accessed on: January 13, 2022. [Online]. Available: https://mlg.eng.cam.ac.uk/yarin/thesis/thesis.pdf

[2] J. Mukhoti, A. Kirsch, J. van Amersfoort, P. H. S. Torr, and Y. Gal, "Deterministic neural networks with appropriate inductive biases capture epistemic and aleatoric uncertainty", 2021. [Online]. Available: https://arxiv.org/abs/2102.11582