Machine Learning on the Radio Galaxy Zoo

Matthew Alger Supervisor: Cheng Soon Ong

June 3, 2016

Abstract

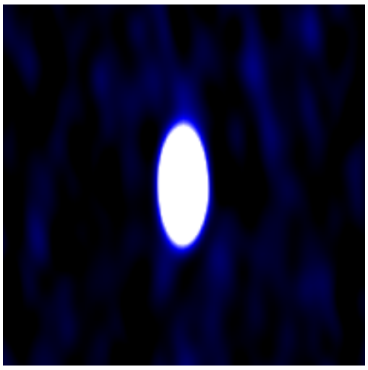
I did something and it kinda worked

1 Introduction

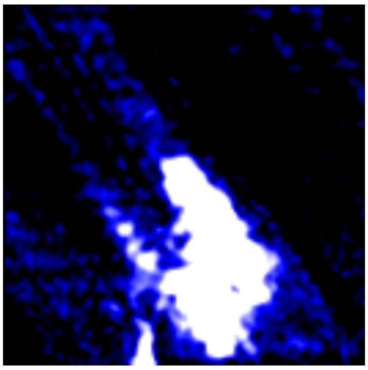
1.1 Cross-identification of Active Galactic Nuclei and Host Galaxies

Radio surveys such as Faint Images of the Radio Sky at Twenty-Centimeters (FIRST) [11, 2] and the Australian Telescope Large Area Survey (ATLAS) [4] are dominated by *active galactic nuclei* (AGNs) [1], radio emissions emitted from the centre of galaxies by supermassive black holes[7]. Galaxies containing an AGN are referred to as *host galaxies*. These galaxies are found in infrared surveys such as the Wide-field Infrared Survey Explorer (WISE) [12] and the SIRTF Wide-area Infrared Extragalactic survey (SWIRE) [10, 5].

Astrophysicists are interested in the properties of both AGNs and their host galaxies, but to investigate either, the AGNs need to be matched to their host galaxies. This is called *cross-identification*. Many AGNs are associated with *compact radio sources*, where the radio emissions directly and simply overlap the host galaxy (Figure 1a), and these AGNs are easy to cross-identify[1]. However, many AGNS are instead associated with *complex radio sources*, where radio emissions can be large, sprawling, and not relate to the host galaxy in any simple way (Figure 1b).



(a) A compact radio source.



(b) A complex radio source.

Figure 1: Example radio emissions.

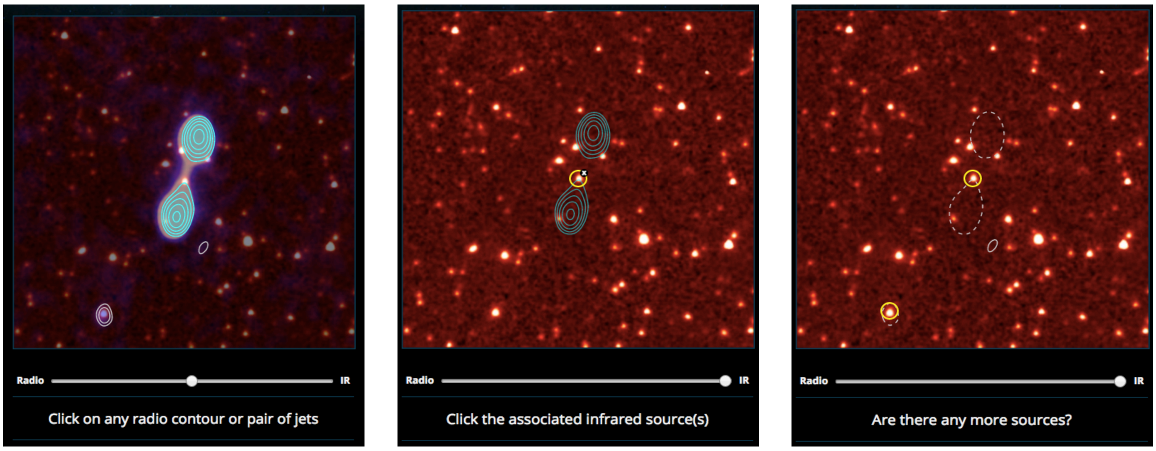
Radio Galaxy Zoo[1] is an online citizen science project that aims to crowdsource the cross-identification problem[1]. Volunteers are presented with a radio image of a small part of the sky (from FIRST or ATLAS) and the corresponding infrared image (from WISE or SWIRE). Each part of the sky presented in this way is called a *subject*, and contains at least one radio emitter. Volunteers are asked to select which radio emissions are part of the same system, and which galaxy in the infrared image contains the corresponding AGN. The workflow is shown in Figure 2.

To increase cross-identification accuracy, each compact radio source is presented to 5 volunteers, and each complex radio source is presented to 20 volunteers[1].

Over 100000 radio sources have been cross-identified by volunteers so far[2] out of the Radio Galaxy Zoo database of around 177000 radio sources, compared to a few thousand classifications by experts[1]. However, new surveys such as the Evolutionary Map of the Universe (EMU) [6] and Westerbork Observations of the Deep APERTIF Northern-Sky (WODAN) [9] are expected to detect over 100 million radio sources[1], making crowdsourcing an intractable solution to the cross-identification problem.

[1] Radio Galaxy Zoo

[2] Based on the data supplied to the author.



(a) Select radio emissions. (b) Locate the AGN. (c) Repeat for all emissions.

Figure 2: Radio Galaxy Zoo volunteer workflow.

In this report, I describe my research into using cross-identifications made by Radio Galaxy Zoo volunteers as a training set for training supervised machine learning algorithms to automatically perform the cross-identification task.

1.2 Related Work

Proctor 2006[8]; Kimball & Ivezić 2008; van Velzen, Falcke, & Körding 2015; Fan et al. 2015[3]

2 Data Sources

2.1 ATLAS

ATLAS is a radio-wavelength survey of the Chandra Deep Field South (CDFS) and the European Large Area ISO Survey – South 1 (ELAIS-S1) fields, chosen as they are areas of the sky covered by the earlier SWIRE survey. This means that the ATLAS observations have corresponding observations in infrared wavelengths[4], which are necessary for cross-identification as the distant galaxies of interest emit infrared radiation.

While the Radio Galaxy Zoo data include classifications of objects in both the ATLAS and FIRST surveys, here I have only focused on the ATLAS observations of CDFS. Reasons[4]:

- ATLAS is small and nice
- We have complete expert classifications of ATLAS
- ATLAS is mostly well-behaved, compact objects
- ATLAS is considered a test run for EMU

ATLAS observations of CDFS consist of a 3.6 deg^2 mosaic of radio images between $3^\mathrm{h}26^\mathrm{m} - 27°00'$ and $3^\mathrm{h}36^\mathrm{m} - 29°00'$. The full ATLAS image of the CDFS field is shown in Figure 3.

Each ATLAS object forms a subject. Each subject consists of a $2' \times 2'$ image patch from Figure 3 and a corresponding image patch from SWIRE centred on the associated ATLAS object.

2.2 SWIRE

SWIRE is an infrared-wavelength survey of seven regions of the sky in seven infrared bands. Of these regions, CDFS and ELAIS-S1 overlap with ATLAS and are hence used in Radio Galaxy Zoo.

SWIRE observations are infrared images in the various fields, and are provided in Radio Galaxy Zoo as $2' \times 2'$ image patches centred on ATLAS subjects. In addition, I make use of the SWIRE CDFS Region Fall '05 Spitzer Catalog[10], which describes all objects detected in the CDFS field in the SWIRE survey.

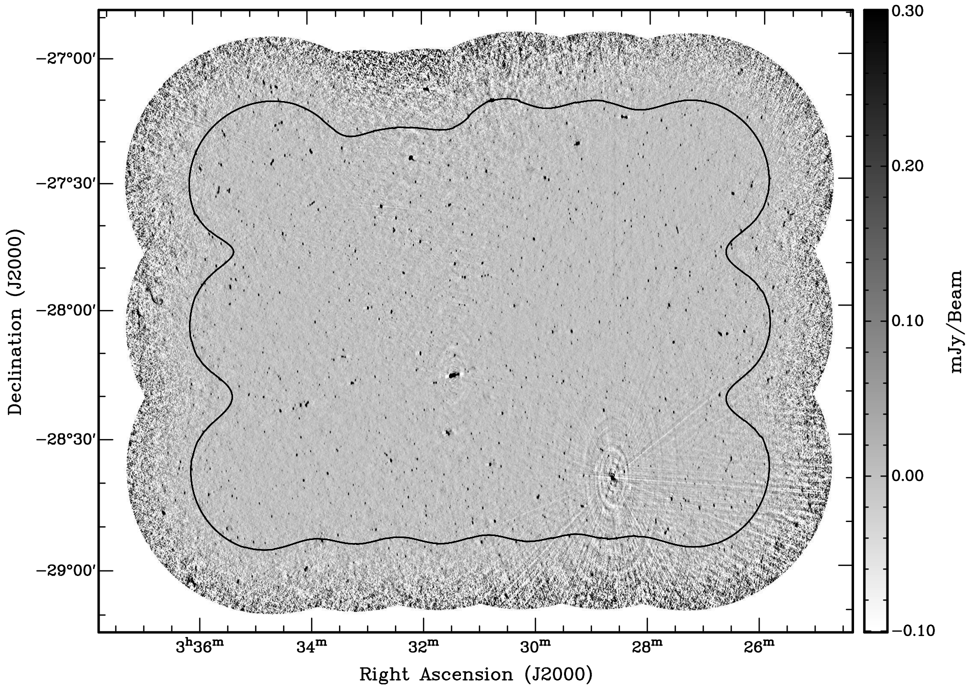


Figure 3: ATLAS observations of CDFS. Reproduced from Franzen et al. [4].
.

For each object in CDFS, the catalogue provides the name, location, infrared fluxes, and stellarity index associated with that object. The location is specified in right ascension and declination. The fluxes are given in five bands (3.6 μJy, 4.5 μJy, 5.8 μJy, 8.0 μJy, and 24 μJy) and describe how bright each object is in the corresponding flux band. Finally, the stellarity index is an indicator of how star-like each object is according to the SExtractor software package, where 0 denotes an object that is totally non-star-like, and 1 denotes an object that is totally star-like[10].

2.3 Radio Galaxy Zoo

3 Cross-Identification as Binary Classification

4 Classification Pipeline

5 Results

6 Discussion

7 Conclusion

References

[1] J. Banfield, O. Wong, K. Willett, R. Norris, L. Rudnick, S. Shabala, B. Simmons, C. Snyder, A. Garon, N. Seymour, et al. Radio Galaxy Zoo: host galaxies and radio morphologies derived from visual inspection. *Monthly Notices of the Royal Astronomical Society*, 453(3):2326–2340, 2015.

[2] R. H. Becker, R. L. White, and D. J. Helfand. The FIRST Survey: Faint Images of the Radio Sky at Twenty Centimeters. *Astrophysical Journal*, 450:559, Sept. 1995. doi: 10.1086/176166.

[3] D. Fan, T. Budavári, R. P. Norris, and A. M. Hopkins. Matching radio catalogues with realistic geometry: application to swire and atlas. *Monthly Notices of the Royal Astronomical Society*, 451(2):

1299–1305, 2015. doi: 10.1093/mnras/stv994. URL http://mnras.oxfordjournals.org/content/451/2/1299.abstract.

[4] T. Franzen, J. Banfield, C. Hales, A. Hopkins, R. Norris, N. Seymour, K. Chow, A. Herzog, M. Huynh, E. Lenc, et al. ATLAS–I. third release of 1.4 GHz mosaics and component catalogues. *Monthly Notices of the Royal Astronomical Society*, 453(4):4020–4036, 2015.

[5] C. J. Lonsdale, H. E. Smith, M. Rowan-Robinson, J. Surace, D. Shupe, C. Xu, S. Oliver, D. Padgett, F. Fang, T. Conrow, et al. SWIRE: The SIRTF wide-area infrared extragalactic survey. *Publications of the Astronomical Society of the Pacific*, 115(810):897, 2003.

[6] R. P. Norris, A. M. Hopkins, J. Afonso, S. Brown, J. J. Condon, L. Dunne, I. Feain, R. Hollow, M. Jarvis, M. Johnston-Hollitt, E. Lenc, E. Middelberg, P. Padovani, I. Prandoni, L. Rudnick, N. Seymour, G. Umana, H. Andernach, D. M. Alexander, P. N. Appleton, D. Bacon, J. Banfield, W. Becker, M. J. I. Brown, P. Ciliegi, C. Jackson, S. Eales, A. C. Edge, B. M. Gaensler, G. Giovannini, C. A. Hales, P. Hancock, M. T. Huynh, E. Ibar, R. J. Ivison, R. Kennicutt, A. E. Kimball, A. M. Koekemoer, B. S. Koribalski, Á. R. López-Sánchez, M. Y. Mao, T. Murphy, H. Messias, K. A. Pimbblet, A. Raccanelli, K. E. Randall, T. H. Reiprich, I. G. Roseboom, H. Röttgering, D. J. Saikia, R. G. Sharp, O. B. Slee, I. Smail, M. A. Thompson, J. S. Urquhart, J. V. Wall, and G.-B. Zhao. EMU: Evolutionary Map of the Universe. *PASA*, 28:215–248, Aug. 2011. doi: 10.1071/AS11021.

[7] B. M. Peterson. *An introduction to active galactic nuclei*. Cambridge University Press, 1997.

[8] D. Proctor. Comparing pattern recognition feature sets for sorting triples in the FIRST database. *The Astrophysical Journal Supplement Series*, 165(1):95, 2006.

[9] H. Röttgering, J. Afonso, P. Barthel, F. Batejat, P. Best, A. Bonafede, M. Brüggen, G. Brunetti, K. Chyży, J. Conway, F. D. Gasperin, C. Ferrari, M. Haverkorn, G. Heald, M. Hoeft, N. Jackson, M. Jarvis, L. Ker, M. Lehnert, G. Macario, J. McKean, G. Miley, R. Morganti, T. Oosterloo, E. Orrù, R. Pizzo, D. Rafferty, A. Shulevski, C. Tasse, I. v. Bemmel, B. Tol, R. Weeren, M. Verheijen, G. White, and M. Wise. Lofar and apertif surveys of the radio sky: Probing shocks and magnetic fields in galaxy clusters. *Journal of Astrophysics and Astronomy*, 32(4):557–566, 2012. ISSN 0973-7758. doi: 10.1007/s12036-011-9129-x. URL http://dx.doi.org/10.1007/s12036-011-9129-x.

[10] J. Surace, D. Shupe, F. Fang, C. Lonsdale, E. Gonzalez-Solares, E. Hatziminaoglou11, B. Siana, T. Babbedge, M. Polletta, G. Rodighiero, et al. The SWIRE data release 2: Image atlases and source catalogs for ELAIS-N1, ELAIS-N2, XMM-LSS, and the Lockman hole. *Spitzer Science Centre, California Institute of Technology, Pasadena, CA*, 2005.

[11] R. L. White, R. H. Becker, D. J. Helfand, and M. D. Gregg. A catalog of 1.4 GHz radio sources from the FIRST survey. *The Astrophysical Journal*, 475(2):479, 1997.

[12] E. L. Wright, P. R. M. Eisenhardt, A. K. Mainzer, M. E. Ressler, R. M. Cutri, T. Jarrett, J. D. Kirkpatrick, D. Padgett, R. S. McMillan, M. Skrutskie, S. A. Stanford, M. Cohen, R. G. Walker, J. C. Mather, D. Leisawitz, T. N. Gautier, III, I. McLean, D. Benford, C. J. Lonsdale, A. Blain, B. Mendez, W. R. Irace, V. Duval, F. Liu, D. Royer, I. Heinrichsen, J. Howard, M. Shannon, M. Kendall, A. L. Walsh, M. Larsen, J. G. Cardon, S. Schick, M. Schwalm, M. Abid, B. Fabinsky, L. Naes, and C.-W. Tsai. The Wide-field Infrared Survey Explorer (WISE): Mission Description and Initial On-orbit Performance. *The Astronomical Journal*, 140:1868-1881, Dec. 2010. doi: 10.1088/0004-6256/140/6/1868.