



Machine learning and future surveys

Jeremy Harwood

Energetics and life-cycles of radio sources ASTRON, March 2018

Collaborators:

R. Morganti (ASTRON) M. Brienza (ASTRON) Rafael Mostert (Leiden) M. Hardcastle (Hertfordshire) M. Jarvis (Oxford) A. Stroe (ESO) I. Heywood (CSIRO) L. Morabito (Oxford) K. Duncan (Leiden)

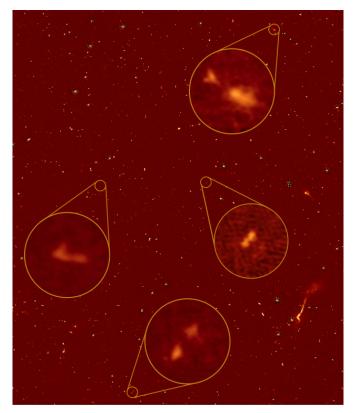






Automated selection to remnant galaxies (and other interesting sources)

- Originally looking to select based on spectral properties e.g. steep spectrum sources
- Many sources aren't as steep as expected at low frequencies (see Marisa's talk for more on this)
- Large surveys are often sparsely sampled i.e. lacking ancillary data for many sources
- May be too high/low to see the spectral break
- Only selecting a small subsample of remnants



Brienza et al. 2017

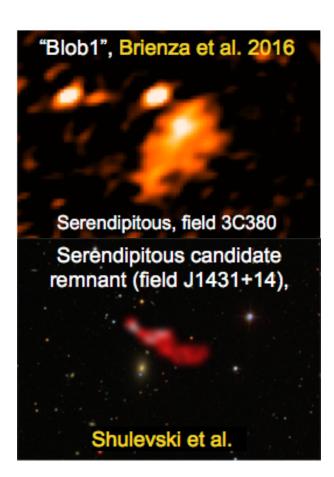






How do we find remnants manually?

- Many well studied remnants found serendipitously
- Morphology
 - Relaxed or prototypical FRI/II?
- Compact features
 - Hotspots/inner jets?
 - Radio core?
- Spectrum
 - Steep or shallow?
 - Sharp cut off?
 - Model comparison (CI on vs off)?
- Machine learning is ideal for this!



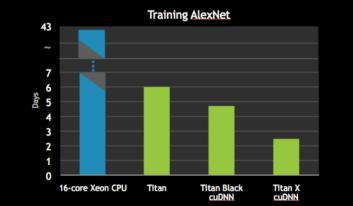






Computer science terminology





Energetics and life-cycles of radio sources Jeremy Harwood – ASTRON, March 2018

Supervised learning

- Aim to assign all input sources to a labelled group
- Used for classification and regression
- Training set requires labels (e.g. remnant vs. active)
- Answer questions such as is this an FRI or FRII?

Unsupervised learning

- Aims to find structure and relationships between groups
- Primarily for clustering (although many other uses)
- Can be used to, for example, cluster similar morphologies

Feature engineering

- What to use and why
- E.g. total flux, peak flux, or the ratio?
- Key to making your method a success!

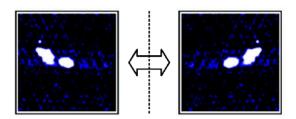




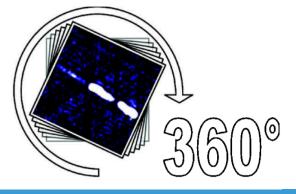


Morphological classification (Rafael Mostert, Ken Duncan et al.)

- Have used unsupervised learning in the form of an artificial neural network
- Uses PINK: a rotation and flip invariant ML algorithm (Polsterer, Gieseke, and Igel, 2015)
- Computationally expensive; requires GPU computing to be completed in a reasonable time
- Created a Self-Organising Map (SOM) for the Lockman Hole
- Can be mapped to other fields in order to identify similar objects (distance from prototype)



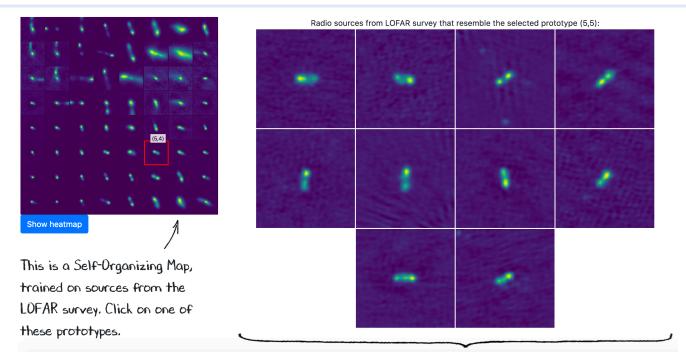
Polsterer, Gieseke, and Igel, 2015











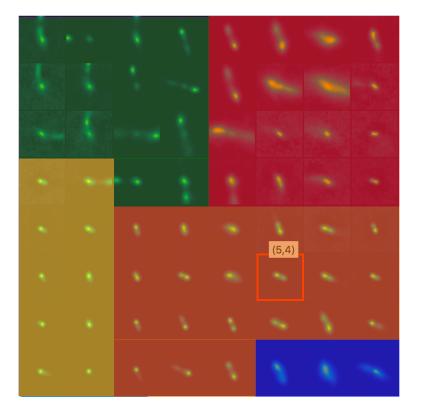
- These prototypical sources can be mapped to other fields (e.g. HETDEX)
- Preliminary tests indicated it might be possible to locate some remnant candidates based on morphology alone
- Likely to only be a specific subset e.g. 3C28 is likely a remnant, but very hard to tell from only its morphology, we need more information!







Clustering and morphological classification



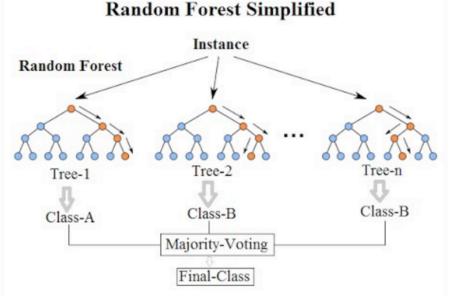
- Clustering is the key to unlocking the power of the self-organised map
- Group the map in to N categories based on their morphology
- Maybe able to classify FRI / IIs but this may be secondary for our purposes
- Compact, extended, relaxed etc. is what we want to know
- Importantly, clustering will provide binary and/or analog parameters for morphology that can feed in to the bigger picture







Random forest algorithm



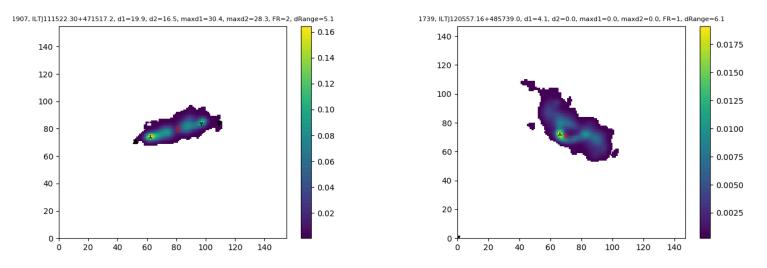
- Need to use all the information we have about sources to unlock the power of ML
- Singular decision trees are fairly common in astronomy
- Simplest (and fastest) solution, but not convinced it is the correct one long term
- Algorithms such as random forest are far more commonly used in the (non-astro) ML world due to reduced over fitting
- This is more complex and requires greater thought wrt to the parameter space







Feature engineering for remnant sources



- What features (observable properties) can/should we use to find remnants?
- Known impact: spectral index, CI on/off model fitting, morphology
- Unknown impact: peak/total flux ratio, redshift, FRI/II classification, environment
- Code such as FRMAL by Beatrix Mingo (pictured above) may be more efficient for FRI/II







🗋 🖕 🖹 🐁 🧮 🕸 : 🕨 📑 🕼 🗣 🗣 : 👭 📫 🚝 🚝 🗰 📓 : 🖬 💥 🥕 🍖 : 🔶 🤌 Users/harwood/projects/machine_learning_relics(code	7.5.5		2 🖿 🕇
Idor - Luarsharwood projects/mathine, learning, wilcs, look, lookmanhole, search py O edds py O lookmanhole, search py O Rs, import py	G G Source Corsole G Object	HED	E 4
1 #//vsr/bln/ewv python3 2 = ++ coding: uft-8 =+			
County Co		Usage	
g suther: harwood		Here you can get help of any object by pressing Cmd+I in front of it, either on the Editor or the Console.	
9 Benort Hungy as to 18 Benort pandas as pol 1 Benort math		Help can also be shown automatically after writing a left parenthesis next to an object. You can activate this behavior in Preferences > Help.	
3) for skian-mole, jestetian joort trais, tet_split 1) for skian-mole, jestetian joort trais, tet_split 1) for skian-mole, jestetian joort trais, tet_split 1) for skian-mole, jestetian joort trais, jestetian joort jestetian joort 1) for skian-mole, joort trais, jestetian, jestetian, jestetian 1) for skian-mole, joort trais, jestetian, jestetian, jestetian 1) for skian-mole, joort trais, jestetian, jestetian, jestetian 1) for skian-mole, joort trais, jestetian, j		New to Spyder? Read our tatonal	
28 28 from astropy.table import Table		Variable explorer File explorer Help	
21 sport astropy.units as u	0.0	Python console	
23 from astroquiry, and import Med 24 from astroquirp, mort coordinates	Car @ Console 1/A @ Console 2		
<pre>About to data</pre>		10 of 02 of 02 of 04 of 1 10 of 02 of 04 of 04 of 1 10 of 02 of 04 of 04 of 04 10 of 02 of 04 of 04 of 04 10 of 04 of 04 of 04 of 04 of 04 10 of 04	
Bilder the sources of its the the fielder 0f_rtiting.set_index(`Source_id', inplatesTrue) 64 65 66 67 68	: if res.id3 -	itting.drop(row.id2, inplace=True)	
66 df_fitie_le(irw, 'remmans') = 1 7 dprint'Icdosr '+ str(Indos) + ' Rowr '+ str(row) + 'ln') 60	: tīp_ df_T	<pre>peak = df_fitting.loc(df_fitting.index = row.id3, 'Peak_flux'].item() itting.drop(row.id3, inplace=True)</pre>	
70 0print(df_fitting.head()) 71 print(df_fitting.loc(df_fitting.index = 52041)	excepts	tmp_peak / tmp_flux	
72 Tor index, rew in df_camidate_multicomp_list.iterrews(): 74 tmg_flux = 0.0 5 tmg_pesk = 0.0	Total_flux Peak_fl	e to find source with an ID of", row.id1, row.id2, row.id3) ux peak_to_total remant	
1/1 (mail > 6) 1/1 (mail > 6) 1/2 (mail > 6)	234 5 1256 0 0.552 Unable 15 126 serve with an Unable to 15 126 serve with an	Eb of 1292 cf84 - 1 Eb of 1292 cf84 - 1 Eb of 320 cf84 cf86 - 1 Eb of 320 cf84 cf86 - 1 Eb of 320 cf84 cf84 - 1 Eb of 1292 cf84 cf84 cf84 cf84 cf84 cf84 cf84 cf84	
lacend O	af 3 0 0 0 Aa H 3		
		History log Python console	

- Development of the code and feature engineering is underway
- Looking at the Lockman hole using the known candidates. This will expand to include a wider variety of remnants as we begin to better understand the parameter space
- Main problem: a small training set! Cross verification is currently not possible
- One possible solution to this is to combined fields with known candidates e.g. H-ATLAS









VLA GHz survey of extragalactic sources in the XMM-LSS field (AGES - XL)



- 30 of the brightest extended sources in the XMM-LSS field
- 97 hours split between L- and C-band at ~1 arcsec resolution
- Large amount of ancillary data available (e.g. LOFAR, XMM, VIDEO, DES)
- One of the MeerKAT MIGHTEE fields, and lower resolution (but wider field) survey by lan Heywood (Oxford)
- Help develop methods for radio galaxies the MIGHTEE surveys







AGES-XL: Science goals and progress

- Primary science drivers:
 - What are a robust set of parameters for models of spectral ageing for the radio galaxy population?
 - If the hotspot/lobe discontinuity common in the population?
 - Are recently revised energetics robust?
- In conjunction with other surveys (e.g. MeerKAT, LOFAR) this will help to solve:
 - What is the characteristic age and lifespan of nearby radio galaxies?
 - How does this age vary as a function of redshift?
- 73/97 hours observed, final observations will be taken summer 2018
- Plenty of opportunity to get involved either with the planned science or your own projects!







Hybrid Morphology Objects (HYMORS) J1315+516 J1315+516 FIRST 1400.0 MH: 1400.0 MHz 51 35 00 51 35 00 \odot 34 45 34 4 0 DECLINATION (J2000 DECLINATION (J2000 15 00 33 45 33 45 13 14 42 13 14 39 38 37 36 RIGHT ASCENSION (J2000) RIGHT ASCENSION = 11.5 mJv/ eam, beam=5.4 x 5.4 arcse = 8.2 mJy/beam, beam=1.70 x 1.57 arcse els = 0.52 mJy x (-1, 1, 2, 4, 8, 16, 32, 64, 128) s = 0.47 mJy x (-1, 1, 2, 4, 8, 16, 32, 64, 128)

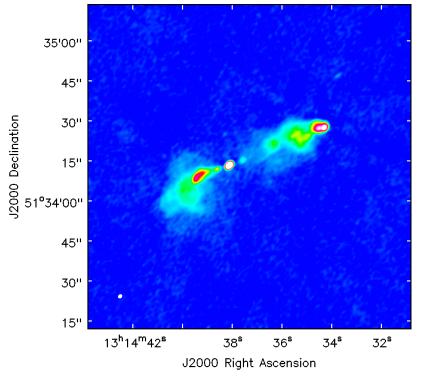
- Discovered by Gopal-Krishna & Wiita (2000) and observed as FR I on one side and FR II on the other
- Very few studies have been performed to date. Recently, only a Galaxy Zoo search by Kapinska et al. and a new object found by de Gasperin (2017)
- No detailed spectral studies of the original samples or any broadband observations that capture both compact and diffuse emission







Hybrid Morphology Objects (HYMORS)



JVLA observation of J1315+516 (z = 0.63) at 6 GHz Full bandwidth between 4 & 8 GHz (C-band) 1.65"x1.25" resolution, combined B&C config.

- 5 examples discovered by Gawronski et al.
 (2006) in the FIRST survey
- 25 hrs JVLA time awarded to observe the sample at L- (1-2 GHz) and C-band (4-8 GHz)
- Observed in multiple configurations give ~1" resolution but also capture the diffuse emission
- First image now complete (left) thanks to Tessa Vernstrom (Dunlap Institute)
- Spectral analysis of the first source soon!







Conclusions

- Machine learning provides a potentially powerful tool for identifying remnant radio galaxies and other sources in large radio surveys
- We cannot rely purely on one parameter (e.g. morphology), if we wish to achieve a robust candidate list and include more than a subset of remnants
- Machine learning can be computationally expensive and often requires specialist setups e.g. GPU rigs plan for this!
- First results should be ready later this year, hopefully expanding the search to the full LoTSS survey (and beyond) as it progresses
- AGES-XL is now almost complete. Let me know if you would like to get involved!
- First images showing combined compact and extended emission of HYMOR produced spectral analysis to follow shortly

