Cognitive MeerKAT

or just computerised telescope operator

The Labyrinth of the unexpected

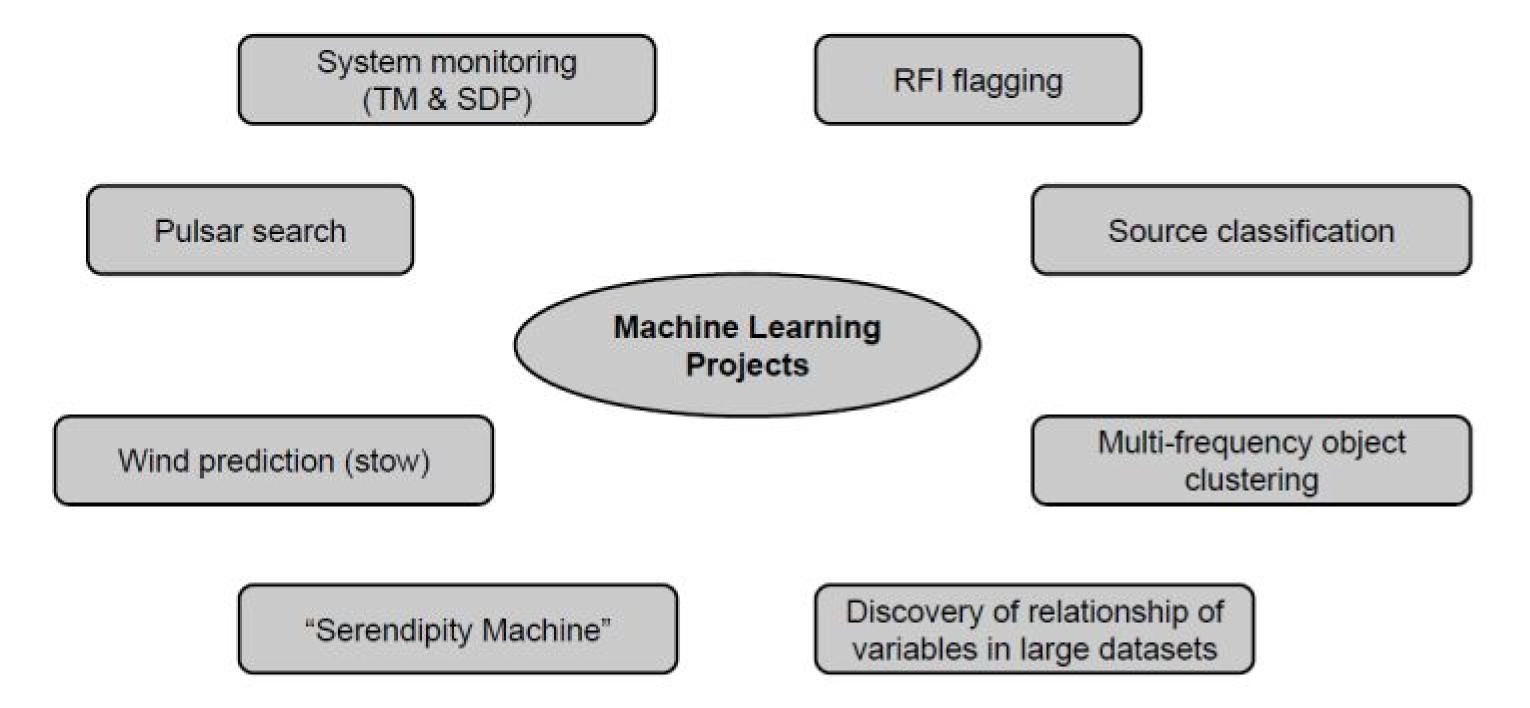
PRESENTER: **Monika Obrocka**
CONTRIBUTORS: A. Aniyan, R. Milenar, K. Ngoasheng, R. Adam, C. van der Merwe, B. Otto

SKA AFRICA
SQUARE KILOMETRE ARRAY

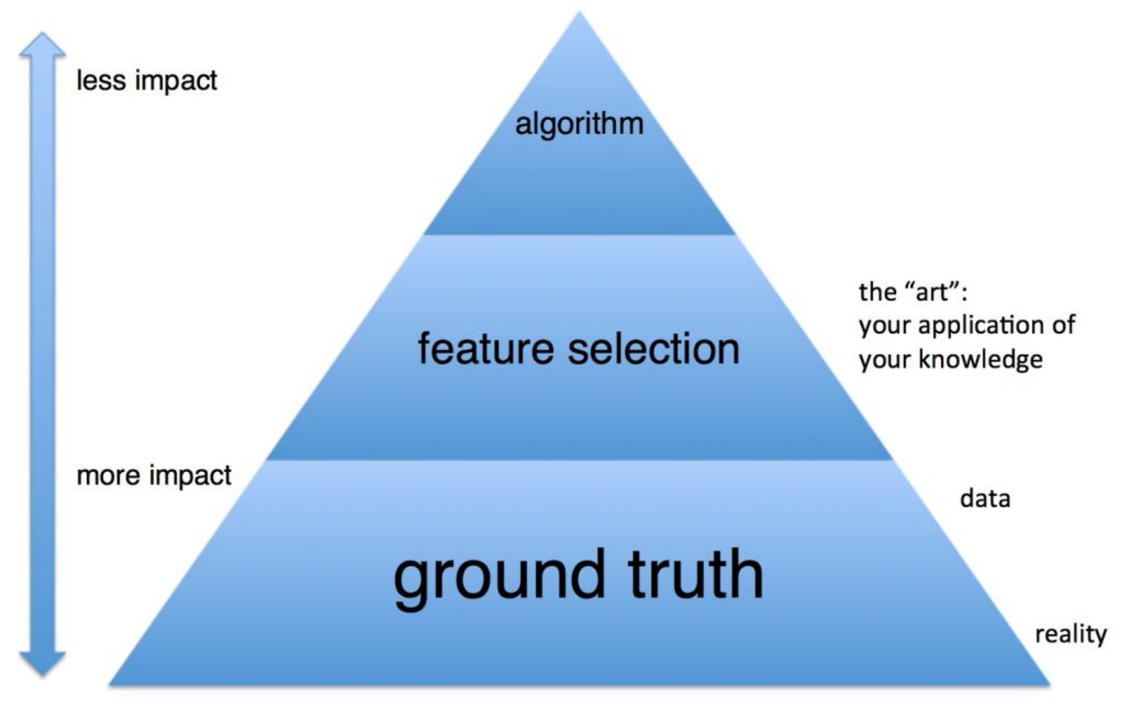
I'M SORRY, DAVE. I'M AFRAID I CAN'T DO THAT.

ML opportunities for MeerKAT

Credit: Rob Adams



The ‘ground truth’ problem



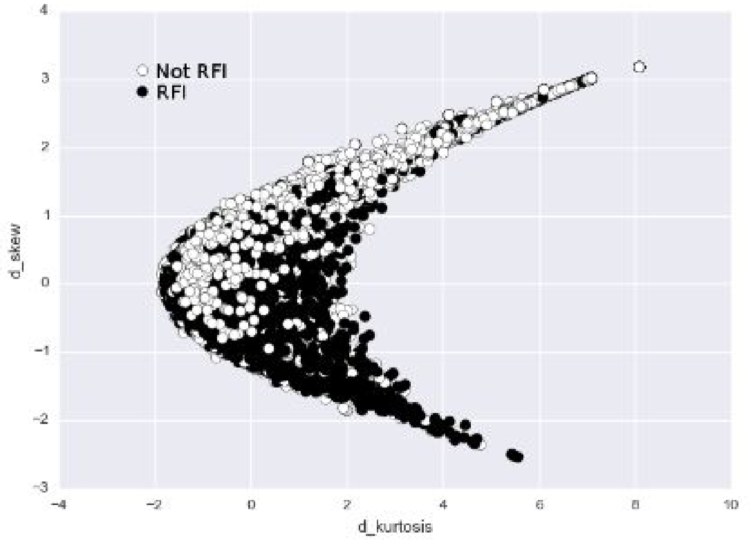
Past/current ML efforts on MeerKAT

- All projects are run by students/postdocs
  - Not part of the official channels == not part of MeerKAT
- All projects use visibility data
- “Ground truth” -> AO flagger results
  - So far unbeaten
- Human flagged data is subjective
  - Depends on the science case
- Models are not transferable
- Flagged RFI is never categorised
  - Are we under or over flagging?
- No beamformer flags on MeerKAT
- Exotic stuff is in BF data

Some examples

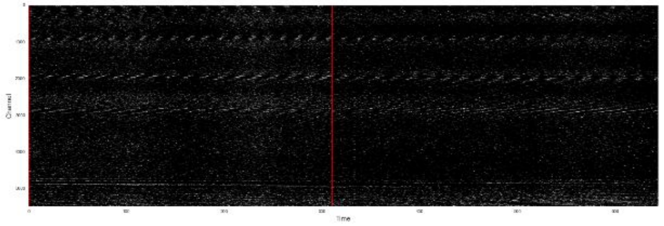
Credit: Arun Aniyan

A RFI vs non-RFI features

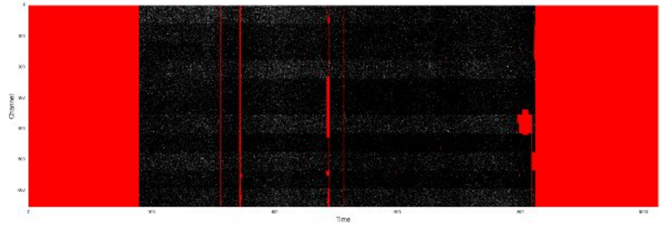


(a) Skew Vs Kurtosis

B Flagging examples



(a) Date set 1



(b) Data set 2


New Chief Scientist for SKA SA
Former Columbia University academic joins SKA South Africa. PAGE 15
SKA SA news
2016 NEWSLETTER FOR SOUTH AFRICA'S SQUARE KILOMETRE ARRAY PROJECT
SKA AFRICA
SKA
SQUARE KILOMETRE ARRAY
Africa
PAGE 10
The search for serendipity
How will we find the unknown unknowns?
&
From 4 of the eventual 64 dishes
MeerKAT produces first remarkable test image
Major milestone towards delivering the SKA
SKA selects final dish design
Ten legacy project teams share
Existing and future survey project teams present plans

arXiv:1611.05570v1 [astro-ph.IM] 17 Nov 2016

Discovering the Unexpected in Astronomical Survey Data

Ray P. Norris[1,2]
[1] *Western Sydney University, Locked Bag 1797, Penrith South, NSW 1797, Australia*
[2] *CSIRO Astronomy & Space Science, PO Box 76, Epping, NSW 1710, Australia*

Abstract: Most major discoveries in astronomy are unplanned, and result from surveying the Universe in a new way, rather than by testing a hypothesis or conducting an investigation with planned outcomes. For example, of the 10 greatest discoveries made by the Hubble Space Telescope, only one was listed in its key science goals. So a telescope that merely achieves its stated science goals is not achieving its potential scientific productivity.
Several next-generation astronomical survey telescopes are currently being designed and constructed that will significantly expand the volume of observational parameter space, and should in principle discover unexpected new phenomena and new types of object. However, the complexity of the telescopes and the large data volumes mean that these discoveries are unlikely to be found by chance. Therefore, it is necessary to plan explicitly for these unexpected discoveries in the design and construction of the telescope. Two types of discovery are recognised: unexpected objects, and unexpected phenomena.
This paper argues that next-generation astronomical surveys require an explicit process for detecting the unexpected, and proposes an implementation of this process. This implementation addresses both types of discovery, and relies heavily on machine-learning techniques, and also on theory-based simulations that encapsulate our current understanding of the Universe to compare with the data.

Keywords: telescopes — surveys

1 Introduction

Popper (1959) described the scientific method as a process in which theory is used to make a prediction which is then tested by experiment. That model, and its principle of "falsifiability", remains the gold standard of the scientific method, and probably drives the majority of scientific progress. Notable recent successes include the discovery of the Higgs boson (ATLAS 2012) and the detection of gravitational waves (Abbott et al. 2016). Conversely, models such as string theory are sometimes criticised (e.g. Woit 2011) for being unfalsifiable, and thus failing to adhere to this Popperian scientific method.

However, the Popperian scientific method is not the only one, and a number of other modes of scientific discovery have been proposed, notably

Russell diagram (Hertzsprung 1908) was an observationally-driven idea of representing data, that led to the development of models of stellar evolution and ultimately nuclear fusion. In another example, the expanding Universe was discovered when Hubble plotted redshifts of galaxies against their brightness (Hubble 1929). More recently, the Hubble Deep Fields (Williams et al. 1996, 2000) were primarily motivated by a desire to explore the early Universe, rather than testing specific models or hypotheses.

1.1 The History of Astronomical Discovery

Astronomical discovery has often occurred as a result of technical innovation, resulting in the Universe being observed in a way that was not previ-

"The most important factor was that discovering the unexpected is harder than expected."

There! I fixed it.


SKA
SKA AFRICA
SQUARE KILOMETRE ARRAY
MeerFunda
'funda' - learning in Zulu
The Labyrinth of the unexpected
PRESENTER: Monika Obrocka
CONTRIBUTORS: A. Aniyan, R. Milenar, K. Ngoasheng, R. Adam, C. van der Merwe, B. Otto

The Business requirement (!)

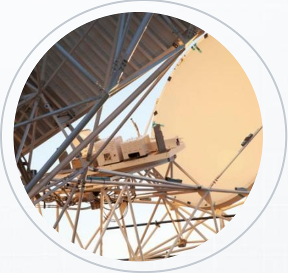
- Organize machine learning capabilities within MeerKAT into a structured, production quality team
- Create a platform that can be expanded towards more complex and larger future projects
- Select a product that is:
  - most likely to be successfully implemented, given time and resource constraints
  - of significance to stakeholders (internal+external)
  - realize this requirement without compromising AR3 delivery

RFI cataloguing and identification

Data availability	MeerKAT data should be available fairly quickly
Training data	Training data not available, but can be automatically flagged
Algorithms	Observation, identification and clustering algorithms will need to be developed
Stakeholder readiness	RFI at MeerKAT site for Beamformer has a receptive audience
Required effort	It is the first time this will be done, but achievable
Required time	It should be ready for testing within 70% of the allocated timeframe
Required resources	There are available resources in MeerKAT for RFI, development and ML, some more resources will be required, but acquirable

MeerKAT RFI monitoring (RFI WG)

(some more mature than others)



RFI sensors

- fixed stations
- mobile vehicles
- handheld sensors

RFI facilities

- reverb and
- anechoic chambers

RFI scans

- MeerKAT
- HERA
- C-BASS

Tracking system

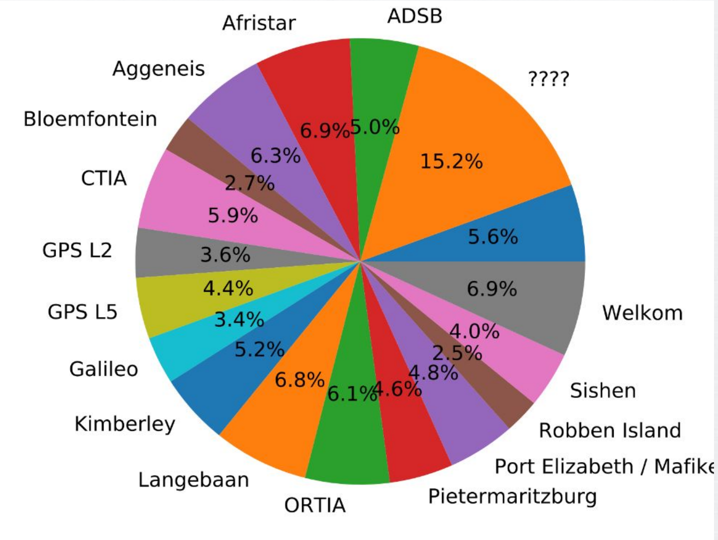
- Aircraft
- Vehicle

How astronomers classify RFI?

Is it RFI?

- No
- Yes:
  - Weak or strong?

Future?



MeerKAT horizon RFI gathering

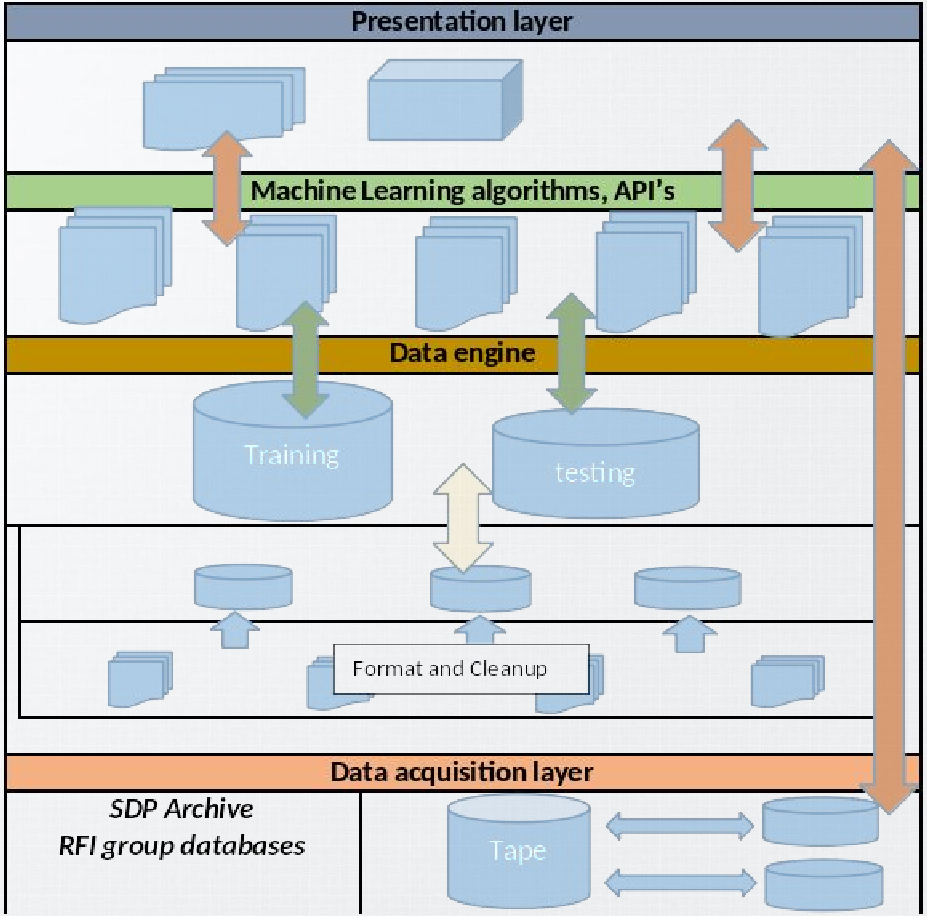
- 3 independent antennas
- 20 minutes observation window
- 60 days observation

Identify known culprits

- Identify RFI & compare to known culprit Database
- Collect and record Unique Identifiers about RFI
- Collate list of known RFI and their features

Learn to identify new culprits

- Apply clustering algorithms across available MeerKAT data
- Human verification of machine-flagged MeerKAT RFI data
- Update Horizon RFI model


Presentation layer
Machine Learning algorithms, API's
Data engine
Training
testing
Format and Cleanup
Data acquisition layer
SDP Archive
RFI group databases
Tape

What we want to find first...

RFI culprits in L band:

Broadcasting band (TV UHF)
GSM 900 Mobile band
Aeronautical Navigation bands
Satellite (GPS etc)

+ everything else RFI WG suggests

Incremental product realization

Increase model dimension

- Sky model
- Flights/planes model
- Satellite model
- Time/period model
- More bands

The team

Monika Obrocka
Arun Aniyan
Isaac Sihlangu

SKA South Africa, a Business Unit of the National Research Foundation.

We are building the Square Kilometre Array radio telescope (SKA), located in South Africa and eight other African countries, with part in Australia. The SKA will be the largest radio telescope ever built and will produce science that changes our understanding of the universe

Contact information

Monika Obrocka

Research Data Scientist

Email: mobrocka@ska.ac.za