SARAO 2022 MSc Project Proposal

Section A

1. Project Title

Detecting Anomalous Transients in MeerTRAP Data

2. Research Area

Science

3. Academic Level

MSc

4. Abstract

With its high sensitivity and large number of antennas, MeerKAT is an excellent telescope for the search for fast transients such as pulsars, rotating radio transients (RRATs) and fast radio bursts (FRBs). The TRAPUM [1] large science project aims to actively search for and discover new pulsars, while the MeerTRAP [2] pipeline rapidly searches MeerKAT observations for new fast transients in a commensal mode. MeerTRAP detects tens of thousands of candidate objects every day, although the vast majority are not real astrophysical sources.

This project proposes to develop anomaly detection techniques and apply them to MeerTRAP data. We will extend the existing Astronomaly [3] framework, allowing the student to focus more on development of new techniques rather than infrastructure work. There is the potential to make interesting new scientific discoveries in the fast transient domain. This project will also develop critical data science skills and essential technology for SKA first light. The project will be done in collaboration with Prof. Ben Stappers, University of Manchester, the PI of both TRAPUM and MeerTRAP.

5. Primary Supervisor's Details

a. **Full name:** Dr. Michelle Lochner
b. **Email address:** mlochner@uwc.ac.za
c. **University:** University of the Western Cape/ South African Radio Astronomy Observatory

6. Co-Supervisor's Details
    a. **Full name:** Prof. Ben Stappers
    b. **University:** University of Manchester

Section B

1. Scientific Merit

The unexpected discovery of new transients such as pulsars and FRBs revolutionised the field of radio astronomy. New telescopes such as MeerKAT have an incredible ability to detect large numbers of these objects. The discovery of pulsars and FRBs were both made by humans looking at the data and noticing something odd. Manually investigating all the data from MeerKAT, particularly when searching for fast transients, is already impossible. With this in mind, the MeerTRAP backend was developed to run in a commensal mode (at the same time as Large Survey Projects and certain other observations), aiming to discover new fast transients. However the data rate is enormous: MeerTRAP delivers tens of thousands of candidates a day, and this only reduces to a few thousand after advanced machine learning is applied.

While the machine learning algorithms can and will be improved, this may come at the risk of only discovering exactly what one is searching for and discarding potential new classes of objects. The framework Astronomaly [3] is an anomaly detection platform, that incorporates an active learning component to include a small amount of human feedback to improve the algorithm. It is capable of sorting through thousands of objects in minutes, prioritising the most interesting ones, thus enabling a semi-automated form of scientific discovery. Astronomaly has been successfully applied to optical fast transient data in the Deeper Wider Faster project, where it was able to detect interesting anomalies in tens of thousands of optical light curves [4].

This project proposes to extend Astronomaly to be able to handle MeerTRAP data (in the form of diagnostic plots such as frequency vs. time waterfall plots and 1d pulse profiles), which involves developing code to read the data and extract features to pass to the rest of the pipeline. There is a possibility of detecting unusual pulsars, new FRBs or even an as yet undiscovered class of transient in the data.

The student will start by writing code to read in the MeerTRAP data. A new feature extraction method will have to be developed to effectively summarise the data for the machine learning. The most promising technique is an autoencoder (a type of deep neural network) but the student will have to investigate and develop the appropriate technique. The student will then apply this technique to the MeerTRAP data, run the rest of the Astronomaly pipeline to detect anomalies and investigate, with the support of fast transient experts such as Prof. Stappers, the nature of the anomalies detected. The algorithm will then be further refined.

2. Feasibility

Data already exists from the MeerTRAP pipeline, some of which is labeled to allow testing of algorithms. While the MeerTRAP data is not public, the PI (Prof. Ben Stappers) has agreed to join this project as co-supervisor and make the data available to the student. Additionally, because the project is making use of the Astronomaly pipeline, the student will be able to focus on key research development as opposed to infrastructure work.

Rough timeline:

- 2022 - Develop new feature extraction and machine learning methods to run on MeerTRAP data
- 2023 - Test on existing MeerTRAP data, possibly also running on a subset of the live MeerTRAP datastream

3. Link to SARAO research priority areas

This project is closely linked to SARAO research priority areas with a focus on MeerKAT. It will develop critical local data science and radio data manipulation skills and pave the way for use of these techniques with the SKA.

4. Student Requirements

Good programming skills are critical for this project, primarily in python. Experience with machine learning is advantageous, but not essential.

References

[1] http://www.trapum.org/
[2] https://www.meertrap.org/
[3] https://arxiv.org/abs/2010.11202 (M. Lochner & B. Bassett, 2020)
[4] https://arxiv.org/abs/2008.04666 (S. Webb, M. Lochner et al. 2020)