Theme 7 - Group 1

Kevin Vinsen, Mingmin Chi, Zhanquan Sun, Weiwei Zhu, Yu Xianchuan

Introduction

Radio Frequency Interference (RFI) is a concern for the highly sensitive radio receivers of modern ground based radio telescopes. Over the years, several techniques have been proposed and implemented for mitigating RFI, either by controlling and eliminating sources of RFI around the radio telescope or its removal at different locations in the receiver chain. Moving the sites of these next generation radio telescopes to remote locations helps reduce RFI from many terrestrial sources, but the increased sensitivity of the instruments can still produce RFI induced artifacts in the observed data.

The original proposal was for the group to investigate whether RFI sources such as: satellite downlinks, power lines, radar reflections, GPS, cameras, and self-RFI could be learnt and characterised using big data and deep learning systems. RFI is a generic problem for all types of radio astronomical receiver, single dishes and interferometers all experience RFI from their surrounds. Single dish systems such as: FAST and Parkes experience it directly, whilst interferometers will see local RFI more strongly on the short baselines close to the RFI source. Whilst more generic RFI will be seen on all baselines.

After discussion amongst the group the original plan was revised to "*To develop a joint Sino-Australian proposal to ERIDANUS (波江座) for RFI mitigation in Radio Astronomy data using Big Data approaches*".

Proposal Plan

The group decided to concentrate of developing a proposal that covers two major SKA Science Data Processor (SDP) use cases, namely the imaging and non-imaging pipelines.

- Imaging pipeline: we will take imaging data from MWA, ASKAP, ATCA and VLA as one use case, and investigate the narrow and broadband RFI that appears on the shorter baselines and longer baselines.
- Non-image pipeline: we will take pulsar detections as the other use case, and investigate single pulse and periodic RFI mitigation in Parkes and FAST data at first, and generalize our solution to SKA in the future.

To alleviate the problems of RFI, big data techniques will be investigated to automatically detect RFI and mitigate/remove it.

We separated the work that should be undertaken as part of the proposal into the following seven sections:

1. Building a RFI knowledge base.
2. Building a reference radio telescope data set.
3. Developing pre-trained feature/parameter models for detecting RFI in astronomy images (similar to ImageNet data).
4. Software detection techniques for RFI detection.
5. Parallel computing environment.
6. Designing unified architecture.
7. Science use cases based on imaging/non-imaging pipelines.

Description of the proposal goals

Building RFI knowledge base

RFI can be generated from many different sources; it can be man made or natural. It is possible to classify the sources to different categories, i.e., by people, such as by local plant equipment, pseudo-random sources (e.g., military, local planes, etc.), regular sources (e.g., planes/satellites, etc.) and by nature causes, such as environmental/space noise (e.g., lightning, reflection, etc.). These are summarised in the table below. For RFI generated by regular sources, it is possible to build a knowledge base for the same or similar RFI and characterise it so the pattern will be visible in the various file formats used by radio astronomy.

<table>
<tr><th colspan="2">RFI Sources</th><th>Specific mitigation techniques</th><th>Common Techniques</th></tr>
<tr><td rowspan="3">Man made interference</td><td>Local plant equipment</td><td>Knowledge base</td><td rowspan="4">Pattern recognition, machine learning, esp. deep learning (parameter/feature models)</td></tr>
<tr><td>Pseudo-random sources (e.g., military, local planes, cameras etc.)</td><td>Anomaly detection, data mining</td></tr>
<tr><td>Regular sources (e.g., planes, satellites, etc.)</td><td>Knowledge base
Pattern recognition</td></tr>
<tr><td>Natural interference</td><td>Environmental / space noise (e.g., lightning, the sun, mice, etc.)</td><td>Data mining
Knowledge base</td></tr>
</table>

Building a reference radio telescope data set

To build a knowledge base and a pre-trained feature/parameter model, a large volume and variety of radio telescope data will need to be collected, e.g., the data from interferometer telescopes, such as: ATCA, LOFAR, MWA, MeerKat, ASKAP, and VLA; and single dish telescopes such as: FAST, Parkes, and Arecibo. Our dataset will be called the Radio Telescope dataset and will be used to characterise RFI and train the various models.

Developing pre-trained feature/parameter models for detecting RFI in astronomy images

In the computer vision community, a pre-trained feature model (VGG-16/19) based on the 1000-class ImageNet dataset can be widely used for image classification task. Similarly, as described above, it is possible to collect the data from a variety of radio telescopes. Using the dataset described above, deep neural networks (Convolutional Neural Networks (CNN), Recurrent Neural Networks (RNN), and combinations of CNN+RNN) will be built to generate general feature models for a general RFI detection task. This will require the use of CNN, RNN and the integration of CNN and RNN; and will need to be carefully designed for a number of platforms of the following existing platforms: Keras, Tensorflow, Theano, MXNET, Torch, Caffe, etc.

Detection techniques for RFI detection

One of the most important goals of this work is to design an automatic RFI detection system by using big data techniques, including but not limited to: machine learning, pattern recognition, data mining, signal processing etc. We will collect the testing algorithms (including supervised learning, unsupervised learning, semi-supervised learning, active learning, deep learning, and crowdsourcing) and make them available as an API that can be used in imaging and non-imaging pipelines as well as by existing tool sets

Parallel computing environments

The computational requirements of the various big data analysis approaches are very high. Many parallel computing environments exist that could aid our work. These include (but are not limited to): Spark, Hadoop, TensorFlow, DALiuGE, and Inside analysis methods. Our work will look to exploit parallelism at every opportunity.

Designing unified architecture.

The long term goal of the project is to provide a unified architecture for RFI mitigation which can work with the common data formats and the imaging and non-imaging use cases. This will allow the astronomer to:

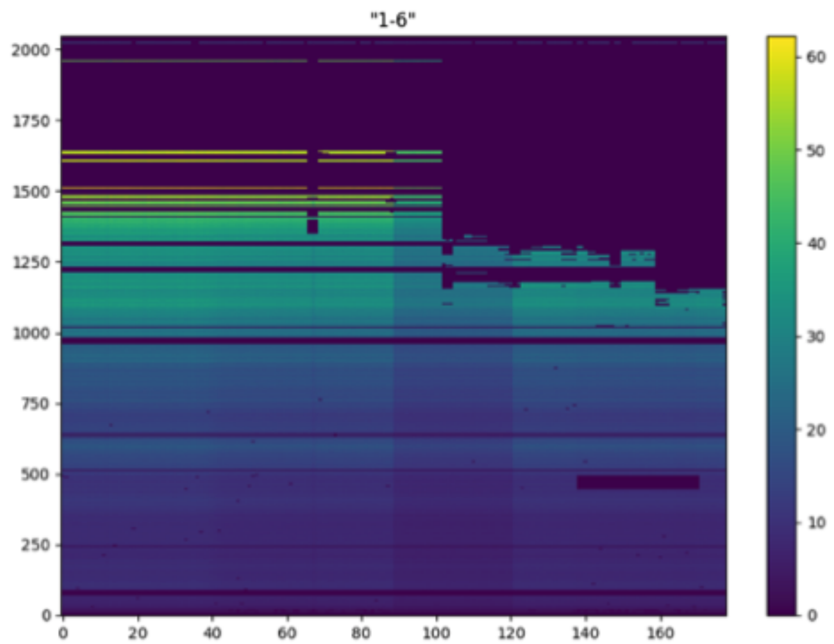
- select which telescope the data came from,
- what type of observation it was,
- which big data tool set they wish to use and
- provide the format they require the results in.

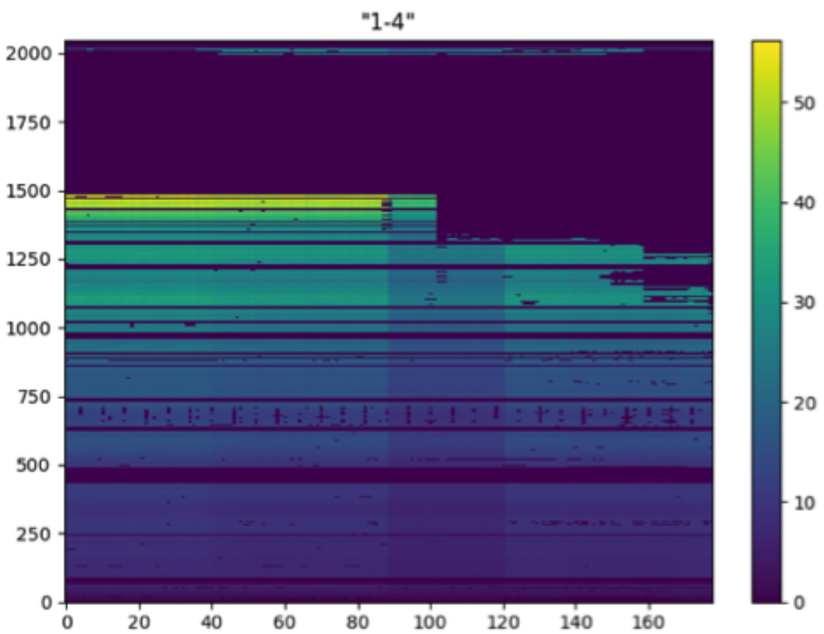
The unified architecture will allow the astronomer to choose the correct models from the knowledge base, based on the telescope used and the use case (non-imaging or imaging). The architecture will all the chosen approaches be integrated seemlessly into the astronomy pipeline.

Science use cases based on imaging and non-imaging pipelines

Imaging pipeline

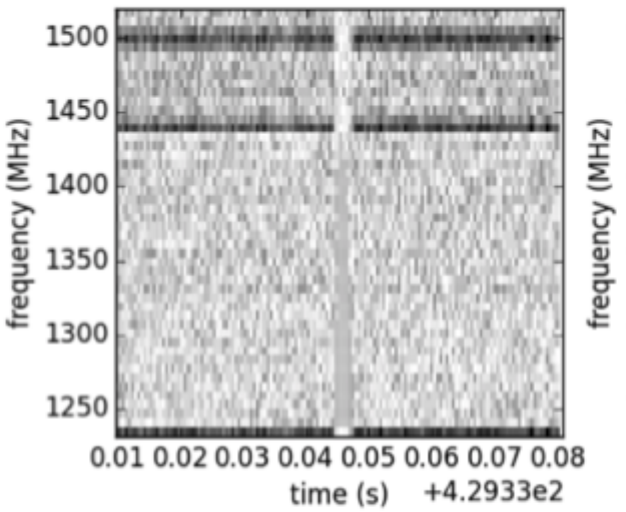
Interferometers tend to be sensitive to local RFI on the shorter baselines. The following two images show data taken from ATCA. In the channels between 650 and 750 nothing is flagged on the longest baseline, 1-6; but a clear cyclic pattern of flagged data is shown on the shortest baseline, 1-4. This type of pattern is characteristic of local RFI. If more data is left unflagged the quality of the image will be improved.

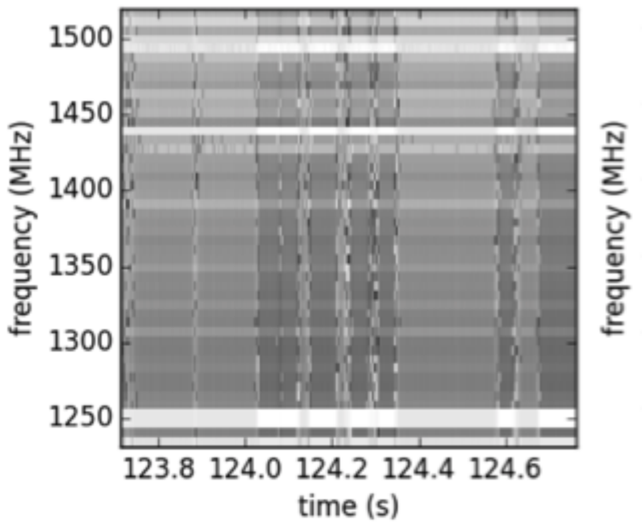




Non-imaging pipeline

The following two images show narrow and wideband RFI in pulsar data. If the RFI can be identified and removed the number of false detections will decrease and true detections will increase.





Future Work

The following deliverables are envisaged over the next few months:

- The development of a database to be hosted at Fudan University to characterise RFI
- Overview paper of the existing approaches to RFI and why our approach will deliver better results for the community

Members of the group will then develop a joint Sino-Australian proposal, based on the 7 themes listed above, to be submit to a body, such as Eridanus, for the development of the ideas described in this document.