URSI GASS 2020, Rome, Italy, 29 August - 5 September 2020

RaFIDe: A Machine Learning based RFI free observation planner for the SKA Era

Shashank Sanjay Bhat[1], Prabu T[2], Snehanshu Saha[3]
(1) PES University
(2) Raman Research Institute
(3) CS and IS, Anuradha and Prashanth Palakurthi Centre for Artificial Intelligence Research (APPCAIR), BITS PILANI K K Birla Goa Campus

Abstract

Signal anomalies in astronomical data mainly come from Radio Frequency Interference (RFI). Radio Frequency Interference (RFI) has plagued the field of radio astronomy. RFI can be either internal (generated by instruments) or external that originates from intentional or unintentional radio emission generated by human activity. Radio Telescopes are known to generate massive amounts of astronomical data. With the huge amount of data being available, a clustering technique can be applied to detect RFI. The quality of the incoming radio signal will be determined by the clustering technique. This will enable us to detect the anomalies in the signal at a particular instant of time. This effort will further enable us to build a database and subsequently apply reinforced time-series machine learning models to predict the quality of the signal. This paper proposes a machine learning approach to study the signal quality over the recent past and make use of this knowledge to plan the near-future observation slots in frequency spectrum and time.

1 Introduction

Radio Astronomy observations are carried out using large Radio Antennas which are known as Radio Telescopes. Low Frequency Radio Telescope observations suffer from time varying interference (RFI) from the neighbourhood. For an effective planning of a radio astronomical observation, it will be useful if a prediction can be made about how much likely an observation planned in a future time slot is going to be free from interference. The project uses unsupervised and Reinforced Time series Machine Learning techniques and aims to help an observer determine the quality of the selected band in the near future.

Figure 1. A Radio Telescope from GMRT

2 Statistical Measures for a signal

A typical Radio Telescope signal without being affected by interference or bias would appear to be Gaussian distributed. An interfering signal tend to alter the measured statistical properties of the celestial signal in a radio telescope.

In this work, our statistical study of the radio telescope signal time series will be based on a histogram analysis. Statistical Measures such as Kurtosis and Skew are taken into consideration for detecting interference in the radio signal.

3 Overall Project Architecture

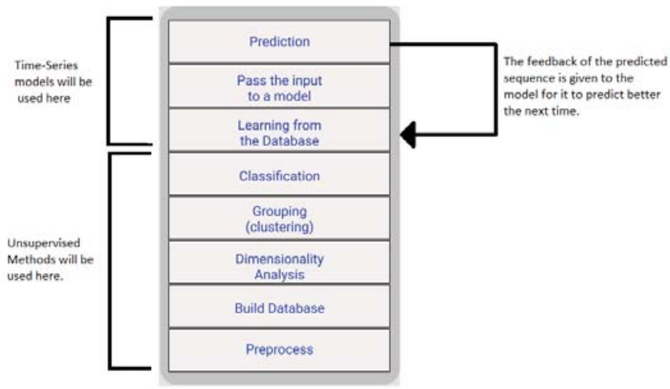

Figure 2. Overall Project Architecture

The project is divided into two phases. The unsupervised phase and the reinforced time-series phase. The unsupervised phase will deal with the data collection, pre-processing, clustering and prediction of the signal quality from the clustering.

The reinforced time-series phase deals with the prediction of the future observation slots for a particular frequency and at a particular time instant.

4 Effect of Interference On Radio Telescope Signals

An ideal Radio Telescope signal without interference or bias would appear with a Gaussian distribution. An interference tends to alter the statistical properties of the signal. Statistical properties such as Kurtosis and Skew gets altered in the Radio signal.

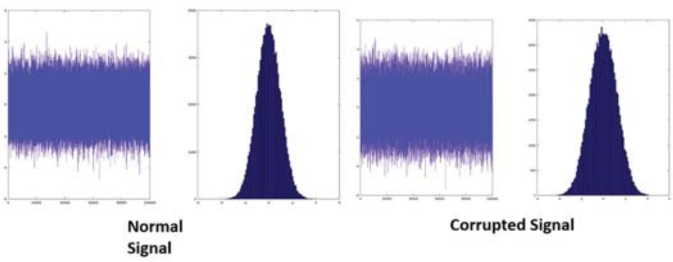

Figure 3. The figure shows a Radio Telescope Signal before Interference (left) and Radio Telescope Signal After Interference (right). A sinusoidal wave having amplitude between +1 and -1 is used for the corruption of the Radio signal.

5 Statistical analysis

5.1 An overview

The analysis is performed on data collected from a low frequency radio telescope at Gauribidanur. Only a short spell of data is considered for the illustration. Four different frequencies are being taken into consideration and the signal is captured from two different antennas.

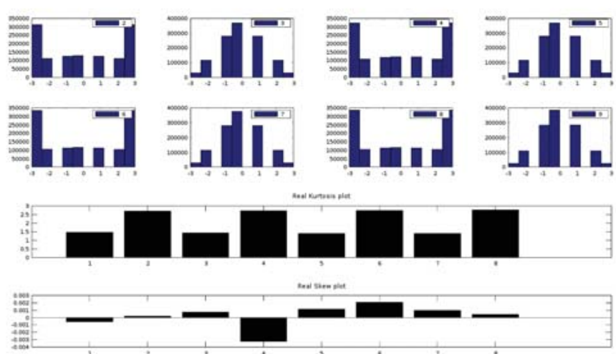

Figure 4. Kurtosis and Skew overview of one second data taken from a telescope. Four different frequencies from two antennas are being taken into consideration here for the analysis.

It is observed that the histogram plots from one of the antennas form a bucket shaped structure. This is due to the clipping of values which were performed as a part of signal preprocessing.

5.2 Detailed Analysis

The data which is taken for analysis contains four different frequencies and from two different antennas. The detailed analysis is performed on two hours of data. The two hours of data is split into 600 meta data files. A random sample of 100 files are taken for the detailed analysis.

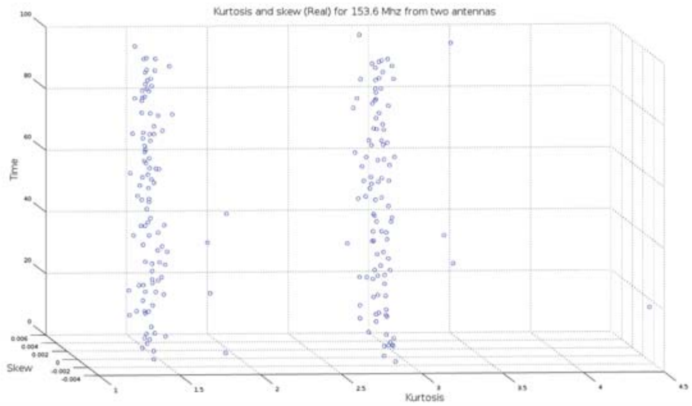

Figure 5. The scatter plot shows the random sample of data from two different antennas.

The kurtosis values from one of the input in this illustration is around a value of 1.5, where as we expect this value to be around 3, if it is not affected by any other artefacts or due to extreme RFI (such as saturation of the inputs). We consider the second antenna signals in the further study.

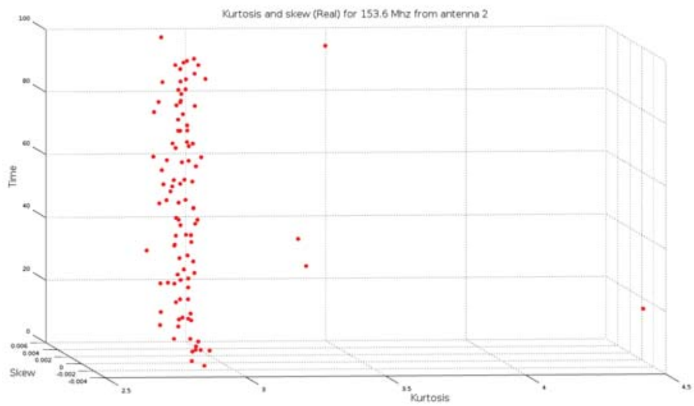

Figure 6. The scatter plot shows the random sample of data from one of the antennas.

The outliers present in the 3D scatter plot is shown in Fig.6. These outliers can be taken as the signals with Radio frequency interference. The time tags corresponding to these outliers will be extracted for finding out the time dependent patterns in the incoming data stream.

6 Application of K-means clustering

To differentiate the good radio signal from the bad one a clustering mechanism is used. K-means clustering is used for binary classification. The number of clusters is limited to 2.

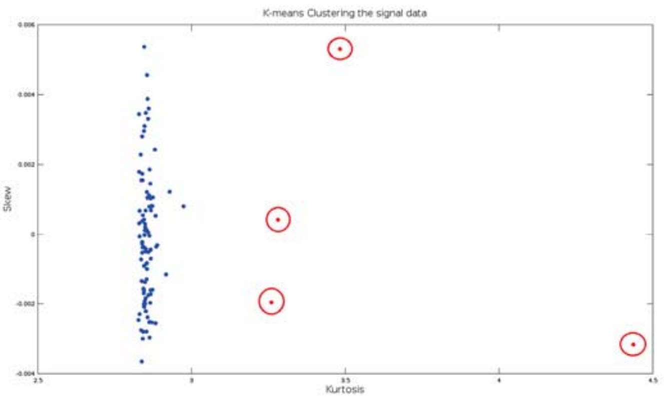

Figure 7. The K-means plot. The previous scatter plot is taken and the Z component is removed and is made 2D. The number of clusters is limited to 2

Four anomalies were detected by the k-means algorithm. It can be safely concluded that these four have RFI.

7 Database Structure

The signal has been assigned a quality factor between 0 and 1. As the quality approaches 1 the signal is said to have more interference. As it approaches 0 the signal can be told to have less interference. The time format followed is 24 hour format

Table 1. The database structure

Frequency	Time	Date	Signal Quality
154 MHz	12:44	19th September	0.97
	12:47	19th September	0.94
	12:50	19th September	0.96
	12:53	19th September	0.97
	12:56	19th September	0.98
	13:00	19th September	0.82
	13:10	19th September	0.68
	13:15	19th September	0.33
	13:25	19th September	0.27
	13:35	19th September	0.15
	13:50	19th September	0.10

8 Prediction of the Radio Signal

The database structure shown above is a short spell of the database built. The database will have the radio signal quality recorded at a much higher periodicity. We will be collecting recordings once every second and for each day. The process will be continued for several days until there is enough data for the training of the machine learning model.

8.1 The Time-Series model

The signal prediction model will be a time series machine learning model. This will be a real time system as it is dependent on the previous recordings of the signal. The performance of the model and the actual readings will be compared. Based on this a feedback mechanism will be built for the model which will aim at making the model predict better.

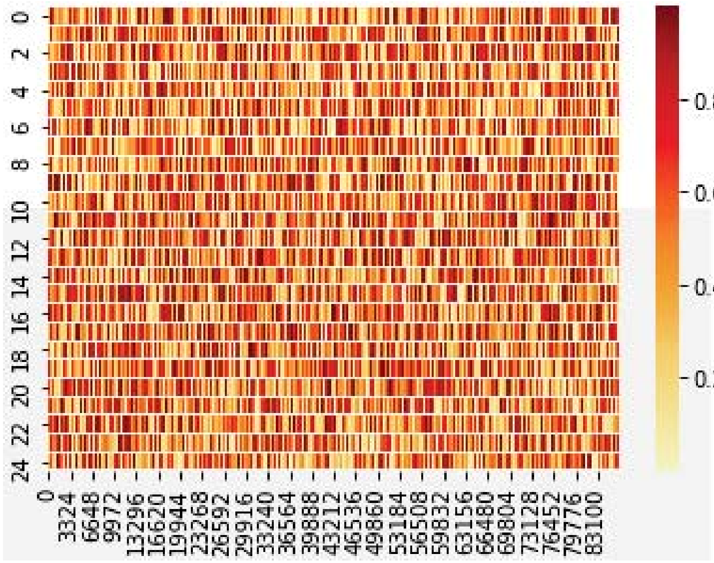

Figure 8. A sample data which is generated for the prediction model. The columns represent the periodicity in which the data was captured (1 second of data captured for the entire day). The rows represent the number of days the data was captured (25 days of data was generated in this case).

8.2 The Prediction

For the prediction a strip of data will be taken from the main data array. This will be given as an input to the time-series model. Based on the pattern which is present in the strip the quality of the radio signal will be generated.

9 Conclusion and Future Work

In this paper we have proposed a method to find RFI free observation slot based on the past history of the RFI affected time and frequency information. We proposed using Machine Learning methods such as clustering methods to identify RFI. We then proposed a reinforced Time Series models to make the prediction and continuously learn from the incoming data stream. We have made initial efforts to classify the data. We are moving to the next phase of the work very soon which is building a real time prediction system. We anticipate that this work will be of much use in the upcoming low frequency telescope era such as SKA.

10 Acknowledgements

The authors would like to thank PES University and RRI for their various supports and encouragements in this project.

References

[1] Anil K Jain. Data clustering: 50 years beyond k-means. 31:15, Jul 2010.

[2] Naomi Altman Jake Lever, Martin Krzywinski. Principal component analysis. *Nature Methods*, 14:3, 2017.

[3] Nadeem Oozeer Olorato Mosiane and Bruce A. Bassett. Radio frequency interference detection using machine learning. *IEEE*, 2016.

[4] S Scholl. Classification of radio signals and hf transmission modes with deep learning. 2019.

[5] Niels Skou Sten Schmidl Sobjaerg Sidharth Misra, Steen Savstrup Kristensen. Cosmos: Performance of kurtosis algorithm for radio frequency interference detection and mitigation. 2008.

[6] Rachael Kroodsma Sidharth Misra, Christopher Ruf. Detectability of radio frequency interference due to spread spectrum communication signals using the kurtosis algorithm. 2009.