Radio Frequency Interference Detection Using Machine Learning

Abhishek Nair, Mohammed Khandwawala, and Mohit Shrivastava
EE16B060, EE16B117, EE16B157

***Abstract*—We explore both traditional and machine learning based methods to identify and mitigate radio frequency interference (RFI) in radio astronomy data-sets. We have implemented a novel image processing based technique for RFI detection which identifies RFI in real time without any expert interaction required. For RFI mitigation, we use non-negative matrix factorization (NMF) in order to accurately reconstruct astronomical data affected by RFI.**

***Index Terms*—Radio Frequency Interference, Mitigation, Machine Learning**

I. Introduction

Radio Astronomy is the study of celestial bodies that give off radio waves. It is a powerful technique used by astronomers to glean information about astronomical phenomena that are invisible or hidden in the other parts of the electromagnetic spectrum. However in recent times, more and more astronomical observations are being contaminated by radio frequency interference. There are multiple possible causes for this including both natural phenomena and man-made sources. Natural phenomena such as the Aurora Boreolis (the Northern Lights) emit wideband radio emissions which can decrease the SNR of astronomical observations. At the same time man-made sources such as cellular communications often introduce narrow-band interference which completely overshadows the astronomical data in that particualar band. In this report we present techniques to both detect and mitigate radio frequency interference. Our detection method does not require the presence of domain experts and can be implemented in real time. At the same time, the technique we have used for mitigation is also able to outperform currently used methods. We have performed our analysis on both synthetic data and actual astronomical data (obtained from the Raman Research Institute).

II. Data

We have used data provided by the Raman Research Institute to perform our analysis. This data corresponds to the CAS A supernova in the Cassiopeia constellation. The SWAN (Sky Watch Array Network) telescope has been used to collect the data. The receiver had a two-sided bandwidth of 16.5 MHz centered at 159 MHz.

Thanks to Prof. Balaji Srinivasan and Prof. Uday Khankhoje from the Indian Institute of Technology Madras. Special thanks to Yash Bhusare (Dept. of Physics) for providing data

III. Detection

A. Traditional Methods

An important feature radio interference signals is that they are typically orders of magnitude stronger than actual astronomical observations. This fact allows us to easily identify radio frequency interference using simple statistical tests. One such statistical test is a simple thresholding operation. We compute the average intensity observed for each frequency channel and plot it. Frequencies that suffer from RFI contain spikes of intensity that can easily be picked out. In the plot below we have chosen the threshold value to be twice the standard deviation of the data. However it is important to note that this method is heuristic in nature and requires domain experts to validate the choice of the threshold. Also if the interference signal is similar to a chirp signal, i.e. it varies in both time and frequency, this method loses it's effectiveness since RFI will not affect any one channel and will therefore not be caught by the thresholding operation.

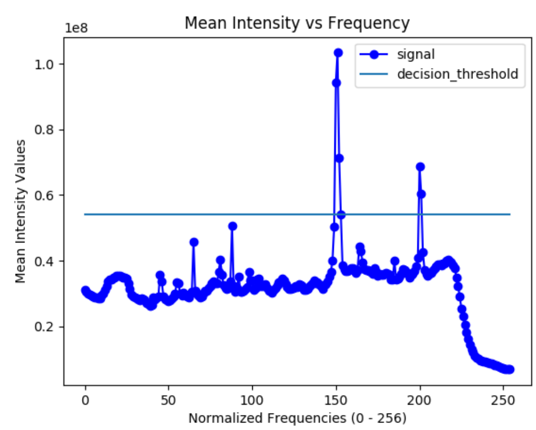

Fig. 1. Detection using traditional method

B. Image Processing Method

Image processing techniques were used to overcome the shortcomings of the traditional methods. The Canny edge detection algorithm was used to detect RFI by treating the spectrogram of the astronomical data as an image. Typically, the level of RFI contamination will vary with both frequency and time. Therefore after edge detection, binary thresholding is used to clearly demarcate RFI and non-RFI data. Next we apply a dilation kernel to make the edges more prominent. At this step in the processing pipeline, our algorithm picks up salt and pepper noise as well. We solve this issue with the use of a contour finding kernel. This kernel replaces each pixel value with the average of all it's neighbours. After this step the detected values are considered RFI and are removed.

Figure 2 shows the spectrogram of synthetic RFI affected data and Figure 3 shows the detected RFI after the dilation step. Image 4 shows the spectrogram of the data after RFI removal.

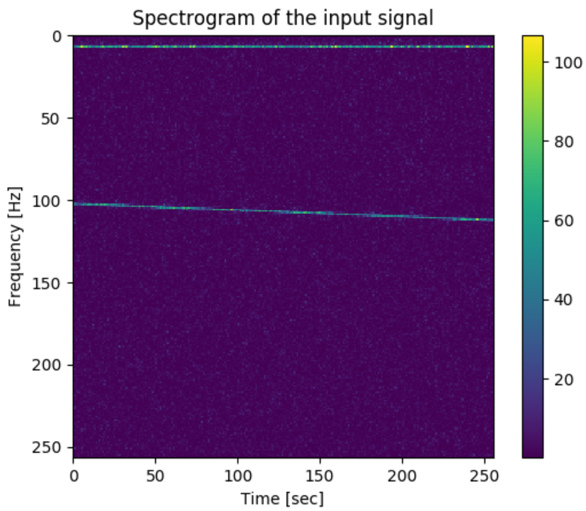

Fig. 2. Spectrogram of synthetic data

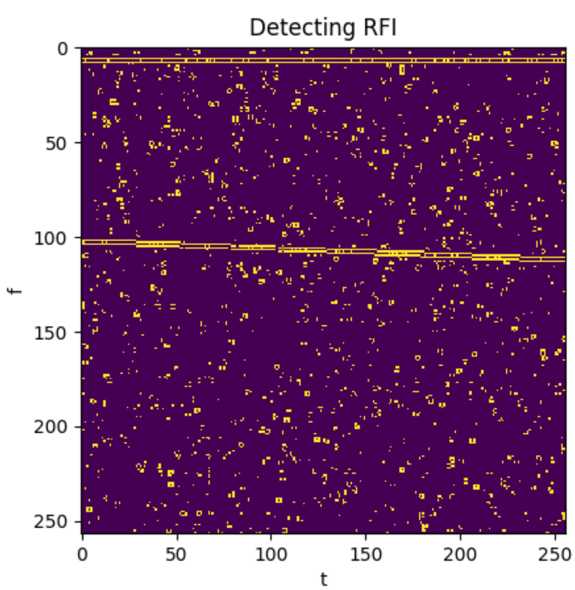

Fig. 3. Detection using image processing method on synthetic data

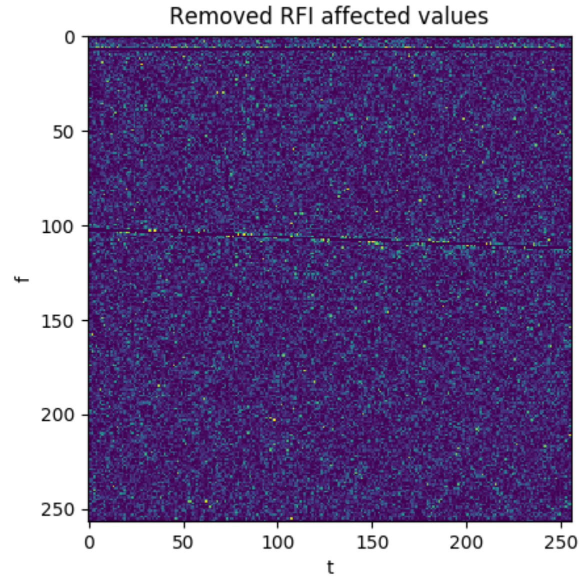

Fig. 4. Detection using image processing method on synthetic data

Figure 5 is the spectrogram of CAS-A data while figure 6 shows the RFI detected image and the image in figure 7 shows the spectrogram after RFI removal.

(In image 5 and 7 nothing has changed other than blanking of RFI affected values. The image looks different because python auto-scales the colours according to the average intensity of the image. Previously, the image was dark as the RFI intensity was relatively high, and when it was removed the noise in the image becomes visible)

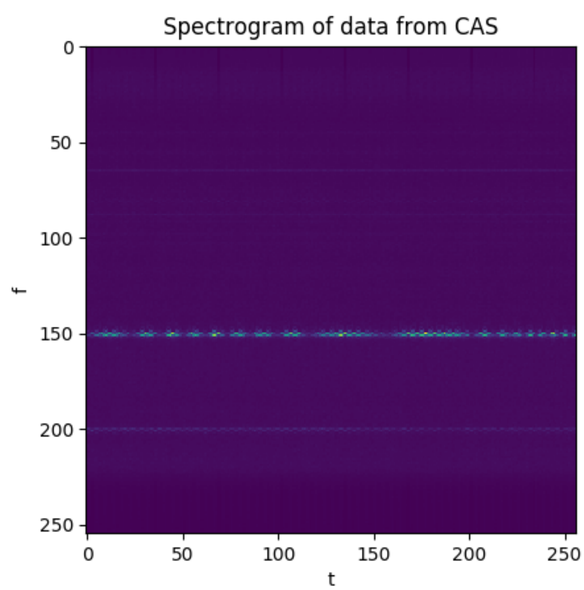

Fig. 5. Spectrogram of CAS-A data

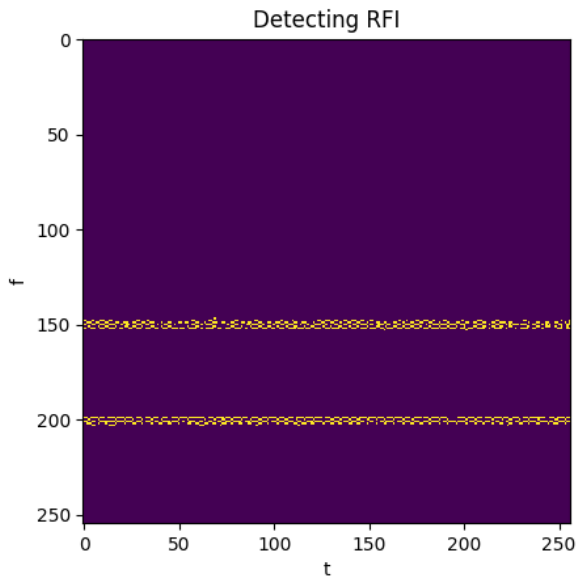

Fig. 6. Detection using image processing method on CAS-A data

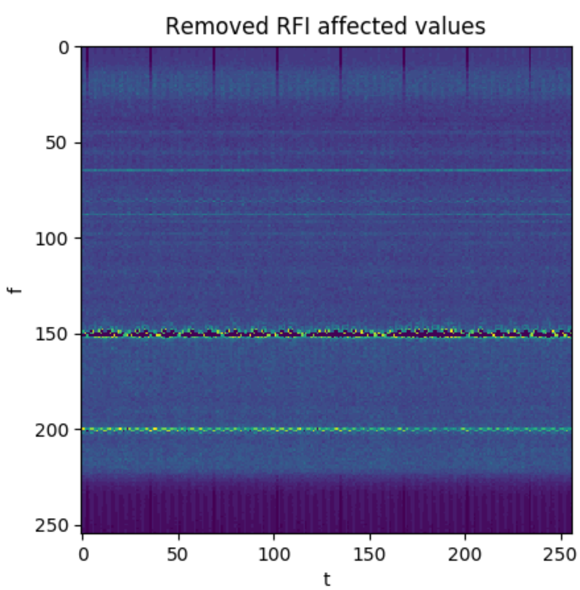

Fig. 7. Detection using image processing method on CAS A data

IV. Mitigation

A. Blanking

Blanking is a very simple mitigation technique and is effective if RFI and useful data occur in separate frequency channels. This method consists of detecting the frequency channels contaminated by RFI and blanking (setting to zero) them. This ensures that the disproportionately large values occurring in these frequency channels does not affect computations performed on other frequency channels. However it is important to note that this method will not work if the actual data and RFI is interleaved, as blanking will discard the actual data as well.

B. Localized Average

This method consists of replacing the RFI contaminated data with the average intensity of all frequency channels at that time instant. This has the same advantages as blanking but also additionally ensures that the statistics of the data are not affected due to the mitigation process. This ensures that statistical methods can still be relied upon to process the radio astronomy data.

C. ***Non-negative Matrix Factorization*** *and its comparison with other mitigation techniques*

Non-negative Matrix Factorization (NMF) is a state of the art feature extraction algorithm. Using NMF it is possible to predict missing values in a matrix using the ones that are present. For our analysis we have used 64 feature vectors for decomposing a 256*256 size data matrix. To compare the three mitigation techniques used we calculate the normalized mean squared error with respect to the original (RFI absent) data for the three cases. (Note : This analysis has been performed on the generated synthetic data.) The blanking technique gave a normalized mean squared error in reconstruction of 6.550% , averaging gave an error of 6.1120% and NMF gave a 1% reduction in error, giving a net reconstruction error of 5.272% which consolidates the superiority of NMF mitigation technique.

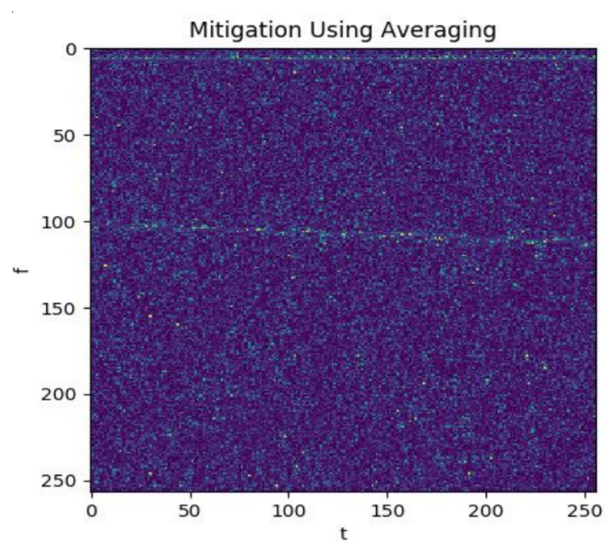

Fig. 8. Spectrogram after Mitigation using Averaging

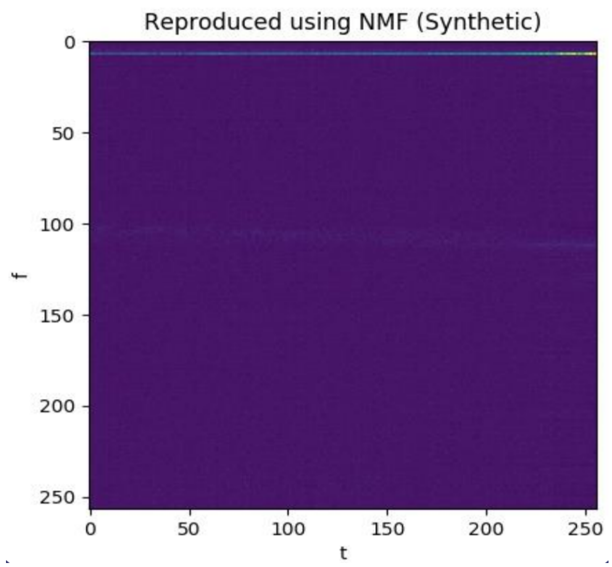

Fig. 9. Spectrogram after Mitigation using NMF

V. Acknowledgements

The authors would like to thank Yash Bhusare,a student in the department of physics IIT Madras , Prof. Uday Khankhoje and Prof. Balaji Srinivasan for aiding us in acquiring and accessing the required data.

References

[1] Olorato Mosiane, Nadeem Oozeer, Arun Aniyan,and Bruce A. Bassett , *Radio Frequency Interference Detection usingMachine Learning* , South Africa, 2012.

[2] Rob V. van Nieuwpoort , *Towards Exascale Real-Time RFI Mitigation* , eScience center and University of Amsterdam, The Netherlands, 2017.

[3] Jose Miguel Tarongi, Adriano Camps, *Radio Frequency Interference Detection and Mitigation Algorithms Based on Spectrogram Analysis*