Characterizing Interference in Radio Astronomy Observations through Active and Unsupervised Learning

Gary Doran
Mentor: Kiri Wagstaff

Jet Propulsion Laboratory, California Institute of Technology

August 16, 2012
Outline

Introduction

Data and Methods

Results

Conclusions
Radio Astronomy

- Observation of astronomical signals at frequencies of 30 MHz to 300 GHz

- **Problem:** man-made radio sources such as cell phones, satellites, spark plugs, and computers interfere with observations

- *Radio Frequency Interference* (RFI) must be reduced by physically removing or controlling sources, or via detection and removal from observed signal
Goals

**Machine Learning:**
applying algorithms that learn to perform a task (such as classification) from data

- RFI detection and removal are well-studied problems
- Little work has been done to apply machine learning techniques to classify and characterize RFI
- Help astronomers understand properties of RFI in their data to distinguish it from astronomical signals
- This work helps close the loop between detection/removal and reduction through action in the field
Parkes Multibeam Survey

- Observations made over a period of 4.5 years

- Survey contains over 1000 hours of observations (about 3.5 TB of data, in 12,000 sets of files)

- Originally a survey to search for pulsars

- Instead of searching for pulsars, we’re interested in looking for RFI
What does RFI look like?

(a) Dispersed pulsar signal

(b) Channelized near 1425 MHz

(c) Broadband, short-duration

(d) Several broadband bursts
Strategy

1. Detect RFI
2. Extract Features
3. Learn Categories
4. Interpret Results

[Images of various data visualizations and diagrams related to the strategy steps.]
Assume data is drawn from a mixture of $k$ Gaussian distributions.

Infer the means of the distributions given the data ($k$-means).

Only use a small “batch” of data at each iteration of the algorithm to speed up convergence (mini-batch $k$-means).
Active Learning

- Query human expert for labels of a few events in the most efficient way possible
- Strategy: ask for the label of event about which the current classifier is most uncertain
- Goal: classify events as accurately while minimizing human effort
General Observations

- Over 5.3 million events detected
- On average 1.7 events per second
- Most (3.4 million) events only a single time sample (125 µs) in duration
- Average duration: 2 ms (some very long events caused by highly contaminated observations)
Pointing Direction

- **Up**
- **Zenith**
- **North**
- **Azimuth**
Azimuth Angle
Clustering Results

Cluster 66 (0.90 Percent)

Cluster Average

Average Intensity

Frequency (MHz)

Modified Julian Date

Hour of Day

Zenith

0° (North)

Azimuth

0°

15°

30°

45°

60°

75°

90°
Active Learning I

Choose a label:

Submit
Conclusions

- Applying machine learning techniques to detected RFI events has the potential to improve the understanding of various RFI phenomena.

- Understanding characteristics of RFI provides valuable information that can be used to formulate and prioritize mitigation strategies.

- The techniques described above can generalize to other detection techniques (e.g. those that look for dispersed signals) to explore other kinds of RFI.

- A similar characterization can be performed for other instruments and observatories (e.g. the Deep Space Network and Greenbank Telescopes).
Acknowledgements

**Mentor:** Kiri Wagstaff

**Co-mentor:** David Thompson

The Machine Learning and Instrument Autonomy Group

**Radio Astronomers at JPL:**
Sarah Burke-Spolaor
Jake Hartman
Dayton Jones
Joseph Lazio
Walid Majid
Robert Preston

**Collaborators at NRAO:**
Paul Demorest
Mike McCarty
Mark Whitehead

Thank you!

Gary Doran
gary.doran@case.edu