RFI Novelty Detection Using Machine Learning Techniques

Stephen Harrison¹, Rory Coles², Timothy Robishaw¹, David Del Rizzo¹
¹Dominion Radio Astrophysical Observatory
²University of Ontario Institute of Technology

RFI2019 – Toulouse, France
September 25, 2019
Overview

• The Dominion Radio Astrophysical Observatory

• What are we trying to do?

• Hardware

• Signal Detection

• Novelty Detection

• Future Work
The Dominion Radio Astrophysical Observatory (DRAO)

Canadian Hydrogen Intensity Mapping Experiment (CHIME) 400-800 MHz

Dish Verification Antenna (DVA) 350 MHz – 50 GHz

John A. Galt 26m 900-1800 MHz

CHIME Pathfinder 400-800 MHz

DRAO Synthesis Telescope 408 and 1420 MHz

- Solar Flux Monitoring
- Technology Development
- Antenna Test Facilities
- Composite Dish Fabrication
- ...and more.
Telescopes and RFI Sources at DRAO

- St. Andrews Golf Course (2 km), Penticton (20 km)
- Twin Lakes Golf Course (8 km)
- Okanagan Falls, Cell Towers (5 km)
- Oliver (16 km)
- Us (0 km!!!)
- Synthesis Telescope D3A, Solar Flux
- CHIME Pathfinder
- John A. Galt 26m Solar Flux
- DVA
What Are We Trying To Do?

• Build a site RFI monitor to proactively protect our spectrum:
  • to the extent possible from 350 MHz to 2 GHz.
  • using a minimum of staff resources.
  • in a way that benefits all telescopes on the site.
Find and Analyze Signals, Make Notifications

Detect what is novel in the RF scene by first characterizing what is normal.

NOT

Capturing and storing integrated spectra for later reference.

Find a frequency/time range that corresponds to a transmission.

NOT

Flag some cells in time/frequency grid which contain RFI.

Cluster transmissions in order to study populations of events.

NOT

Studying single events or mean/min/max spectra.
CHIME ICEBoard with Custom Firmware Design
- 16/15 Oversampled Coarse Channelizer
- 10b Sampling
- 8 Dual-Pol or 16 Single-Pol Output Channels
- 125 MHz Bandwidth/Channel
- 16b+16b Complex Data Format (VITA49)
- Spectrometer Process in GPU Server
Neural Net-Based RFI Detector

<table>
<thead>
<tr>
<th>Context</th>
<th>Focus</th>
<th>f = 2</th>
<th>c = 1</th>
</tr>
</thead>
</table>

- Focus region from Luchak et al., *Extensive Deep Neural Networks for transferring small scale learning to large scale systems*, 2018.
- Github: [https://github.com/rfcoles1/RFI_Monitor](https://github.com/rfcoles1/RFI_Monitor)
Neural Net on Real Data
Versus Spectral Kurtosis

- Five 2-Level GFSK Signals Generated in GNURadio
- Channel SNR ~5, 4, 3, 2, 1 dB.
- $1 \pm 3\sigma$ SK detections using $M=1600$. 

![Neural Net and Kurtosis graphs](image)
Dimensions of RFI Events

• From bounding boxes we can predict:
  • Center Frequency
  • Bandwidth
  • Received Power / SNR
  • Transmission Duration
  • Start Date/Time
  • Time Between Transmissions

• Exploring Multidimensional Data
  • Parallel Coordinates: N parallel axes.
  • Each line represents a point in N-space.
  • Groups of lines through similar points indicate clusters in N-space.
  • Ask questions by constraining axes.
**What Are My Highest Power Interferers?**

<table>
<thead>
<tr>
<th>Frequency</th>
<th>Bandwidth</th>
<th>C(4,2)</th>
<th>C(6,3)</th>
<th>tx(s)</th>
<th>P(dB)</th>
</tr>
</thead>
<tbody>
<tr>
<td>484973278.600</td>
<td>96100.000</td>
<td>7.992</td>
<td>88.824</td>
<td>119.607</td>
<td>67.758</td>
</tr>
<tr>
<td>479977573.600</td>
<td>86743.038</td>
<td>7.110</td>
<td>79.351</td>
<td>107.655</td>
<td>63.672</td>
</tr>
<tr>
<td>474981971.200</td>
<td>77386.075</td>
<td>6.227</td>
<td>69.879</td>
<td>95.702</td>
<td>59.587</td>
</tr>
<tr>
<td>469986268.800</td>
<td>68029.113</td>
<td>5.345</td>
<td>60.406</td>
<td>83.749</td>
<td>55.502</td>
</tr>
<tr>
<td>464990566.400</td>
<td>58672.150</td>
<td>4.463</td>
<td>50.934</td>
<td>71.796</td>
<td>51.417</td>
</tr>
<tr>
<td>459980854.400</td>
<td>49315.188</td>
<td>3.581</td>
<td>41.461</td>
<td>59.843</td>
<td>47.332</td>
</tr>
<tr>
<td>454990161.200</td>
<td>39058.225</td>
<td>2.690</td>
<td>31.989</td>
<td>47.890</td>
<td>43.247</td>
</tr>
<tr>
<td>450003459.200</td>
<td>29800.263</td>
<td>1.817</td>
<td>22.516</td>
<td>35.933</td>
<td>39.162</td>
</tr>
<tr>
<td>445077568.800</td>
<td>21260.290</td>
<td>0.935</td>
<td>13.044</td>
<td>23.984</td>
<td>35.076</td>
</tr>
<tr>
<td>440012054.400</td>
<td>11887.338</td>
<td>0.033</td>
<td>3.571</td>
<td>2.031</td>
<td>30.991</td>
</tr>
<tr>
<td>435016352.600</td>
<td>2530.375</td>
<td>0.829</td>
<td>5.901</td>
<td>0.078</td>
<td>26.906</td>
</tr>
</tbody>
</table>

Freq. Axis: Just one signal.

BW Axis: ~30kHz

Transmission Duration Axis: a series of consistent short bursts.
What Are My Highest Bandwidth Interferers?

Freq. Axis: Just one signal.

BW Axis: ~90 kHz.

Power Axis: Power is lower but..

Transmission Duration Axis: It’s on all the time.
## What Else Is On All The Time?

<table>
<thead>
<tr>
<th>Frequency</th>
<th>Bandwidth</th>
<th>C(4,2)</th>
<th>C(6,3)</th>
<th>tx(s)</th>
<th>P(dB)</th>
</tr>
</thead>
<tbody>
<tr>
<td>4859713376.000</td>
<td>96100.000</td>
<td>7.992</td>
<td>86.824</td>
<td>119.607</td>
<td>67.758</td>
</tr>
<tr>
<td>476971363.600</td>
<td>86743.038</td>
<td>7.110</td>
<td>79.351</td>
<td>107.655</td>
<td>63.672</td>
</tr>
<tr>
<td>474988972.100</td>
<td>77396.075</td>
<td>6.227</td>
<td>69.879</td>
<td>95.769</td>
<td>59.587</td>
</tr>
<tr>
<td>469986260.300</td>
<td>68024.113</td>
<td>5.345</td>
<td>60.406</td>
<td>81.749</td>
<td>55.502</td>
</tr>
<tr>
<td>44990566.400</td>
<td>58672.350</td>
<td>4.463</td>
<td>50.934</td>
<td>71.796</td>
<td>51.417</td>
</tr>
<tr>
<td>437994864.000</td>
<td>49335.188</td>
<td>3.581</td>
<td>41.461</td>
<td>59.843</td>
<td>47.332</td>
</tr>
<tr>
<td>44550016.600</td>
<td>39505.225</td>
<td>2.690</td>
<td>31.080</td>
<td>47.896</td>
<td>43.247</td>
</tr>
<tr>
<td>45009858.400</td>
<td>30697.169</td>
<td>1.817</td>
<td>22.596</td>
<td>35.937</td>
<td>39.162</td>
</tr>
<tr>
<td>44502758.600</td>
<td>21254.109</td>
<td>0.935</td>
<td>14.044</td>
<td>23.084</td>
<td>35.076</td>
</tr>
<tr>
<td>440012854.400</td>
<td>11667.348</td>
<td>0.033</td>
<td>5.571</td>
<td>12.031</td>
<td>30.991</td>
</tr>
<tr>
<td>435016352.000</td>
<td>2530.375</td>
<td>0.829</td>
<td>5.901</td>
<td>0.078</td>
<td>26.906</td>
</tr>
</tbody>
</table>
Cyclic Time of Day Axis Reveals Diurnal Trends in RFI
What Else Can We Learn?

- This signal is cyclic on a one-second timescale.
- Each transmission appears to be a fixed-length.
- This is a 2-level FSK signal.
- The symbol rate is likely 4800bps.
- The frequency shift between symbols is +/- 3.6kHz.
- There is structure: Preamble, Sync Word, Data, End of Frame Marker
Detecting Novelty

• Obvious case is by frequency:
  • A signal appearing on a quiet frequency is definitely novel.

• Other changes in behavior:
  • An always-on signal has turned off.
  • Changed transmission pattern.
  • Increased power level.
  • An additional signal has appeared.

• Modify the distance metric in the clustering algorithm:
  • Frequency ranges overlap → Don’t use frequency in computing distance.
Future Work: Position/Direction Finding Monitor

- For faster RFI crushing we need position or direction of the source.
  - No telescope on site has directional capability and full coverage of the 350 MHz – 2 GHz frequency range.
  - Expand to a four-element coherent RFI monitor.

- TDoA Algorithms: Can we just ML this?
  - Coherently collect complex signals from all 4 antennas.
  - Feed as 8xN input image.

- Simulated model of our site using 4 antennas, delay phase shift and path loss, independent AWGN at each station.
- Simple regression network to estimate position and direction.
Future Work: Position Finding Using ML
Future Work: Direction Finding Using ML

- Show histogram of direction
Summary

• What are we trying to do?
  • Build a site RFI monitor to proactively protect our spectrum using minimal staff resources.

• Hardware: CHIME ICEboards, USRP for prototyping.

• Signal Detection: Neural net RFI detector based on CV techniques.

• Novelty Detection: Modified distance metric in clustering algorithm.

• Future Work: 4-Element Direction/Position Finding.
Questions?

Stephen Harrison
Digital Systems Engineer
Tel: 250-497-2328
Stephen.Harrison@nrc-cnrc.gc.ca
www.nrc-cnrc.gc.ca
• Signal type is the dimension that divides:
  • general dimensions that an observer may care about.
  • signal-specific dimensions that an RFI investigator may care about.

• How do we learn what we don’t already know?
  • Shortage of people to label data.
  • Ideal techniques will be unsupervised.

• Modulation Recognition (ModRec) approaches:
  • Traditional Approach: Cumulants (High-Order Statistics).
  • Closed-world modulation classifiers e.g. O’Shea et al., Over the Air Deep Learning Based Radio Signal Classification, 2018.
  • Clustering in autoencoder latent space e.g. O’Shea et al., Semi-Supervised Radio Signal Identification, 2017.
Pairwise Classification

• Shu et al., *Unseen Class Discovery in Open-world Classification*, 2018.
  • Are these images *the same* or *different*?
  • We can use this technique to infer new classes.

• Try test signal against all known classes.
  • High number of *different* predictions indicates a new class.
  • High number of *same* predictions indicates belonging to existing class.
Pairwise Classification Results

- Started with four examples from unlabelled 61 example set:
Pairwise Classification Results

• Incremental Discovery Of Several New Classes: