Real time machine learning to find fast transient radio anomalies
A semi-supervised approach combining detection and RFI excision

Colorado J. Reed\textsuperscript{1}, David R. Thompson\textsuperscript{2}, Walid A. Majid\textsuperscript{2}, Kiri L. Wagstaff\textsuperscript{2}

\textsuperscript{1}University of Iowa \textsuperscript{2}Jet Propulsion Laboratory, California Institute of Technology

Overview
Future optical and radio instruments with enormous data volumes can benefit from automated pattern recognition to identify events of interest for follow-up observations or subsequent processing.

Many common approaches to event detection in high time resolution data involve matched filtering to a specific anticipated profile. A dedispersion search to find fast radio transients is one example. Here we consider a complementary strategy that could operate in tandem with dedispersion: Statistical anomaly detection techniques can find events that do not match anticipated profiles, and might reveal a broader range of hitherto unknown phenomena.

Approach
We expand work by [Thompson et al. 2011] on adaptive semi-supervised anomaly detection in milliseconds data streams. This work uses adaptive eigenbases to combine (1) prior knowledge about uninteresting signals (2) online estimation of the current data properties to enable highly sensitive and precise detection of novel signals.

Here we also consider a "sparse" representation of the training examples. Sparse codebooks for signal representation can often improve generalization of the resulting dictionary. We find that performance is comparable for non-sparse and sparse methods, though sparse representations offer improved physical interpretability.

Method
At each time step, our novelty detection score measures the distance from the signal to a low-dimensional manifold that was learned to describe the recent data segment.

We hypothesize that the "regular" data lies on a linear subspace in $\mathbb{R}^d$ with $d' < d$. The novelty score for a datapoint $x$ is the reconstruction error $f(x)$ relative to this subspace, learned using the online Principal Component Analysis method from Lim et al. [2004].

$$f(x) = \|x - \hat{x}\|^2 = \|(x - \bar{x}) - AA^T(x - \bar{x})\|^2$$

Interesting signals do not compress well
They are statistically anomalous with respect to the previous data.

This work compares dense representations of $A_u$ and $A_s$ from classical PCA with sparse representations that try to limit the number of nonzero coefficients.

**Classic PCA**
Maximize $(a^TXX^Ta)$

**Sparse PCA**
Maximize $(a^TXX^Ta)^2 - \lambda \|a\|_1$

Dataset
We compare dense and sparse representations to traditional anomaly detection strategies for detecting anomalous “Peryton” events in Parkes Pulsar survey data. These do not match a perfect dispersion profile, and their proper interpretation is still not understood. Our tests focus on approximately five minutes of observation time in each of the 13 receivers. This span includes several tens of thousands of timesteps recorded at a cadence of 0.125 milliseconds in each of 96 frequency channels near 1450 MHz. Data points were formed by vectorizing width-6 windows of time/frequency samples (found to be optimal over all the methods considered).

Results
We characterized different anomaly detection methods’ abilities to retrieve hand-labeled examples of peryton events without detecting RFI. These tests suggest that adaptive semi-supervised approaches outperform traditional alternatives, with sparse and dense representations offering comparable performance.

We find that sparse learning effectively recovers the independent components of the RFI signal and as such may have independent value as an interpretive tool.

**Figure 1:** Novel “Peryton” events [Burke-Spolaor et al. 2011]

**Figure 2:** Radio frequency interference should not be detected. Note characteristic horizontal and vertical stripes, which may appear in arbitrary combinations

**Figure 3:** “Eigensignals” – principal components forming online (unsupervised) and false positive (supervised) bases. Left: dense online components; Center: dense supervised components; Right: sparse supervised components are more intuitively meaningful. This version penalizes nonzero coefficients during the learning procedure.

**Figure 4:** Performance scores for all methods.

References

