

Semi-Supervised Eigenbasis Novelty Detection

NASA's Jet Propulsion Laboratory, Pasadena, California

Recent discoveries in high-time-resolution radio astronomy data have focused attention on a new class of events. Fast transients are rare pulses of radio frequency energy lasting from microseconds to seconds that might be produced by a variety of exotic astrophysical phenomena. For example, X-ray bursts, neutron stars, and active galactic nuclei are all possible sources of short-duration, transient radio signals. It is difficult to anticipate where such signals might appear, and they are most commonly discovered through analysis of high-time-resolution data that had been collected for other purposes. Transients are often faint and difficult to detect, so improved detection algorithms can directly benefit the science yield of all such comensal monitoring.

A new detection algorithm learns a low-dimensional linear manifold for

describing the “normal” data. High reconstruction error indicates a novel signal that does not match the patterns of normal data. One unsupervised portion of the manifold model adapts its representation in response to recent data. A second supervised portion of the model is made of a basis trained in advance using labeled examples of RFI; this prevents false positives due to these events. For a linear model, an orthonormalization operation is used to combine these bases prior to the anomaly detection decision.

Another novel aspect of the approach lies in combining basis vectors learned in an unsupervised, online fashion from the data stream with supervised basis vectors learned in advance from known examples of false alarms. Adaptive, data-driven detection is achieved that is also

informed by existing domain knowledge about signals that may be statistically anomalous, but are not interesting and should therefore be ignored.

The method was evaluated using data from the Parkes Multibeam Survey. This data set was originally collected to search for pulsars, which are astronomical sources that emit radio pulses at regular periods. However, several non-pulsar anomalies have recently been discovered in this dataset, making it a compelling test case. By explicitly filtering known false alarm patterns, the approach yields significantly better performance than current transient detection methods.

This work was done by Kiri L. Wagstaff and David R. Thompson of Caltech for NASA's Jet Propulsion Laboratory. For more information, contact iaoffice@jpl.nasa.gov. NPO-48239