Radio Frequency Interference Identification and Mitigation in Pulsar Observations Using Machine Learning Techniques

Mike McCarty[1], Gary Doran[2], T. Joseph W. Lazio[2], David R. Thompson[2], John Ford[1], and Richard Prestage[1]

[1] National Radio Astronomy Observatory, Green Bank, WV, 24944, http://www.nrao.edu

[2] Jet Propulsion Laboratory, California Institute of Technology, Pasadena, CA 91109, http://ml.jpl.nasa.gov/

Pulsar observations with the Robert C. Byrd Green Bank Telescope (GBT), located in Green Bank, WV, aid researchers in understanding the basic building blocks of our existence – matter, energy, space, and time – and how they behave under extreme physical conditions. Pulsars, rapidly rotating neutron stars with clock-like timing precision, can provide insights into a rich variety of physics and astrophysics.

Advanced radio frequency interference (RFI) mitigation techniques are needed to enable sensitive pulsar searches and timing, thus allowing researchers to investigate key science questions such as testing the validity of Einstein's theory of General Relativity; understanding the formation and growth of supermassive black holes; and exploring the behavior of matter and energy in their most extreme environments. The broader bandwidths made possible by newly developed pulsar instrumentation present the opportunity to dramatically increase pulsar search sensitivity and timing precision. These will lead to dramatic advances in all of these areas. However, taking advantage of broad-band observations requires the development of improved techniques to remove RFI, which becomes a larger problem as bandwidths increase.

Machine learning algorithms, especially neural networks, have shown initial promise for automating astronomical data processing, particularly pulsar search data processing. These techniques use labeled examples (i.e. training data) to build a statistical model of RFI and pulsar patterns and extrapolate this to classify new cases. The models can exploit arbitrary combinations of linear or nonlinear relationships, finding distinguishing numerical features and exploiting patterns that the human user need not notice or articulate. This holds particular promise for automating the subtle pattern recognition problems of RFI labeling. The system can still defer ambiguous events for human review to minimize the risk of missing any real pulsars.

To date we have began to study and catalogue the different kinds of RFI in pulsar and spectral line observations collected by the GBT. A solid understanding of the RFI population will ensure that the classifier design incorporates all the relevant attributes and training data. By applying machine learning techniques to the problem we will at minimum reduce the volume of RFI candidates that must be inspected, leading to more efficient searches.

Part of this research was carried out at the Jet Propulsion Laboratory, California Institute of Technology, under a contract with the National Aeronautics and Space Administration. Government sponsorship acknowledged.