The Department of Physics at The Ohio State University

Genetic Programming: a Novel Analysis Method for Neutrino Astronomy

Kaeli Hughes, with Professor Amy Connolly

INTRODUCTION

It is theorized that cosmic rays interact with the cosmic microwave background and produce ultra high energy neutrinos. Our goal is to detect these neutrinos and utilize where they come from to further understand our universe. One way to improve the analysis is by implementing machine learning techniques.

What are neutrinos?

Neutrinos are subatomic particles with no charge and hardly any mass. Because they have no charge and interact weakly, they require an immense detection volume (like the Antarctic ice).

Askaryan Effect: When neutrinos interact in a medium, they produce a particle shower. When the negative charge of the shower exceeds the speed of light in the medium, Cherenkov radiation is produced, which is coherent for wavelengths greater than 10 cm (the transverse size of the shower)

Figure 1: A picture of the ANITA-3 payload. From kicp.uchicago.edu

The Antarctic Impulsive Transient Antenna (ANITA)

- Radio detector suspended by a balloon in Antarctica
- 4 flights, in 2006, 2008, 2014, and 2016, each about one month long
- Surveys a large area of ice and records potential neutrino events

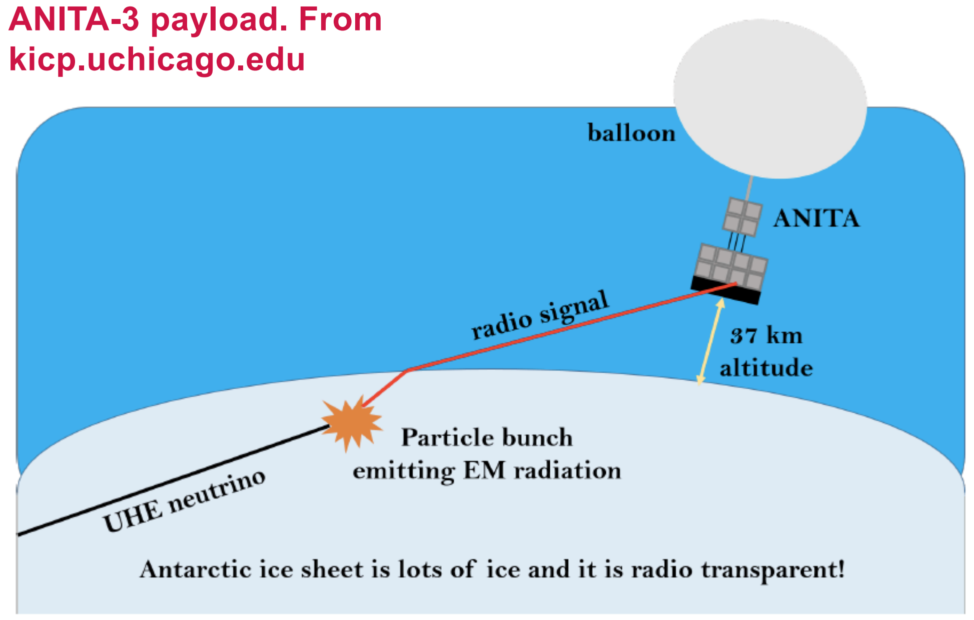


Figure 2: A diagram showing the setup of the ANITA experiment. From OSU graduate student Oindree Banerjee

WHY MACHINE LEARNING

- Anthropogenic backgrounds are hard to model
- Other experiments use Monte Carlo algorithms to predict what all their backgrounds look like: this won't work for ANITA
- Goal: to find out if machine learning can find a model for the background using two of our most important variables: the ratio of signal to noise (SNR) and the cross correlation value

ABOUT KAROO GP

It can solve three types of problems:

- Classification
  - Ex: signal vs. background
- Regression
  - Ex: finding a best fit line
- Matching
  - Ex: finding an exact function to fit data

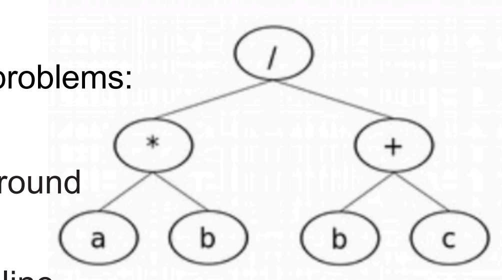


Figure 3: Karoo Functional Tree: S=a*b/(b+c)

For each problem above, functions are randomly selected in groups called "generations". The functions are tested against the data and, through tournament selection, the strongest trees are selected to parent the next generation.

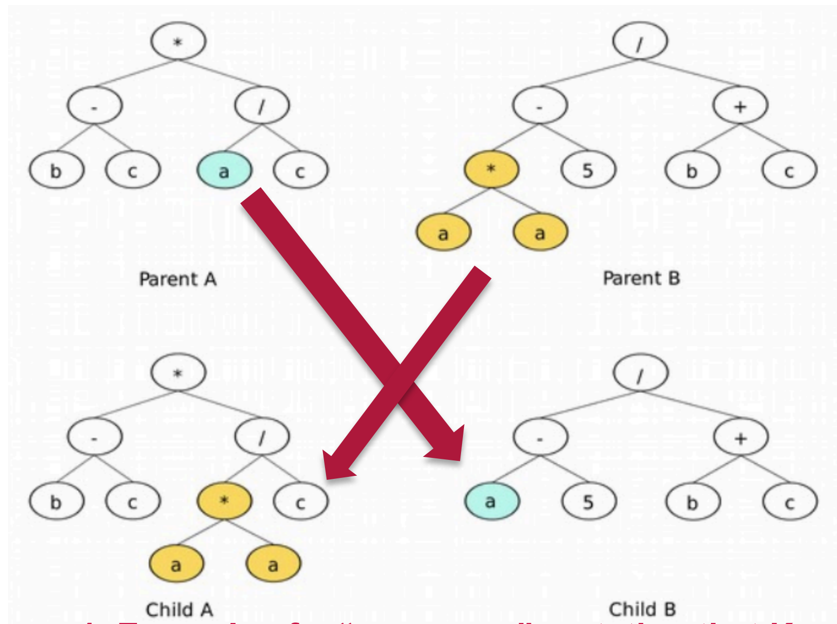


Figure 4: Example of a "crossover" mutation that Karoo does

Why Karoo was chosen:

- Functional output (not a black box)
- Easy to start using
- Powerful and quick

MODELING THE BACKGROUND

Can we find a function that describes the distribution of background events?

Expressions available:

- Arithmetic (+,-,*,/)
- Logarithms
- Exponentials
- Powers
- Trigonometric functions

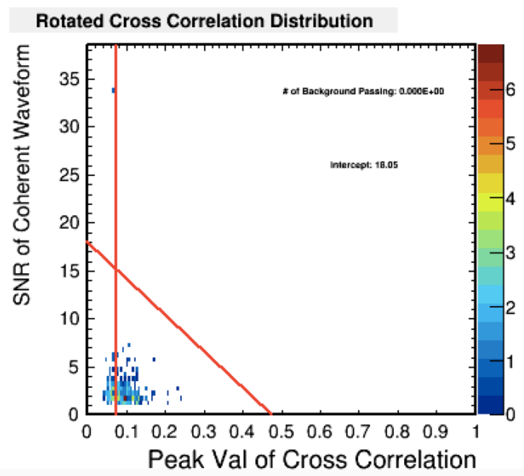


Figure 5: Plot of background events for a region of ice. From OSU PhD Brian Dailey.

Process of finding a function:

1. Zoom in on area of interest and bin data
2. Normalize the variables and center on origin
3. Create spreadsheet with variables, including constants and "features" from variables
4. Run Karoo for 100 generations with 100 trees each

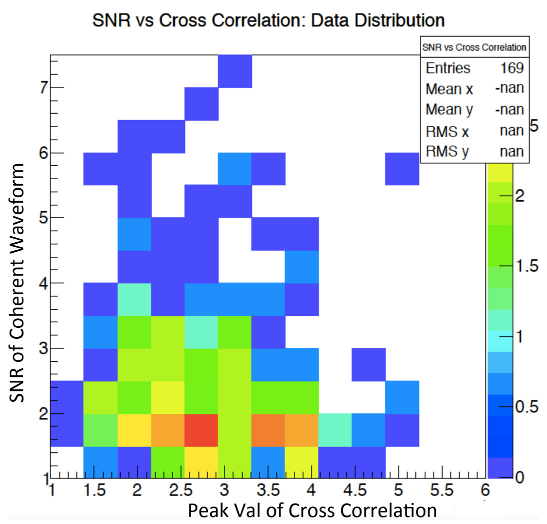


Figure 6: Same background data as above, normalized and binned

Best function so far:

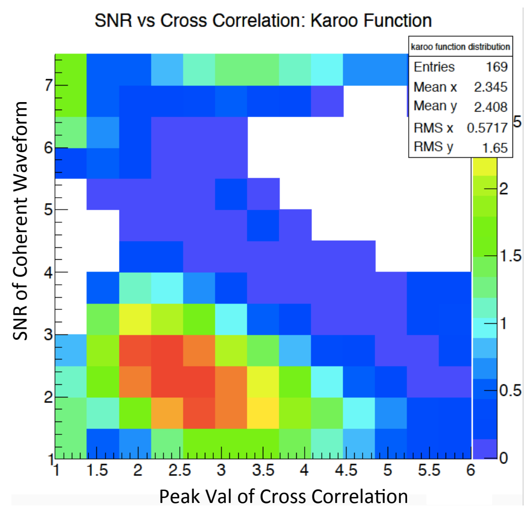
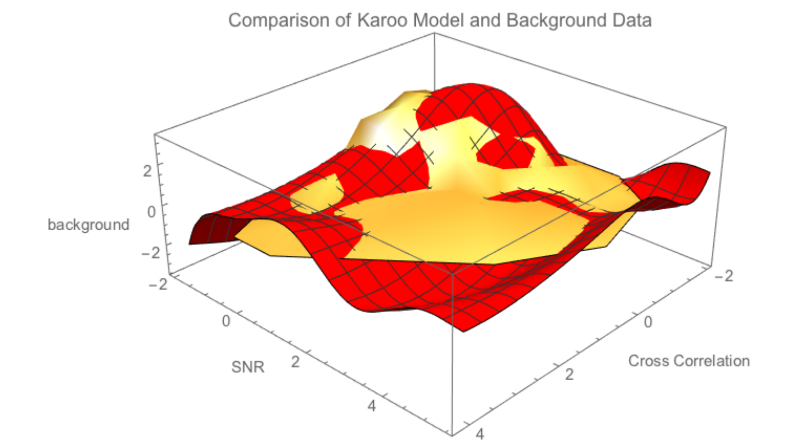


Figure 7: Plot of Karoo Function

A really messy function:
Background= 0.29*Sin[pval + 3] - Cos[snr + 3]*Cos[pval + snr] - 0.819*Cos[snr + 3] + 0.40 - 0.71*Exp[-snr - 3]*Cos[pval + 3] - 0.88*Exp[-snr]*Exp[-snr - 3] - Exp[-pval - 3]*Cos[pval + 3]*Cos[snr + 3] - 2.0*Exp[-pval - 3]*Cos[pval + 3] - 0.59*Exp[-pval - 3] - 0.61*Exp[-pval]*Cos[pval - 3]*Cos[pval + snr]
pval= Peak Val of Cross Correlation
snr = SNR of Coherent Waveform

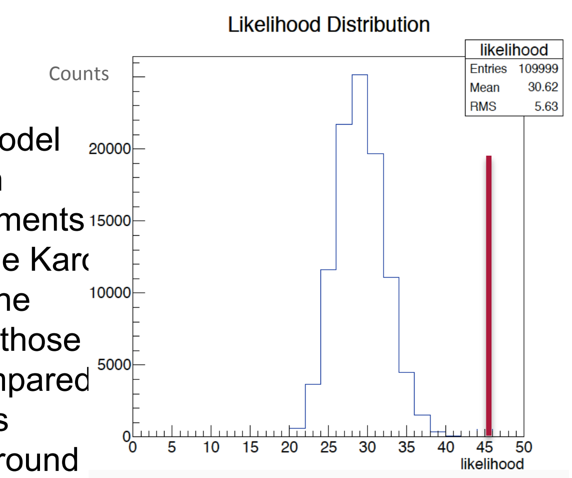
TESTING THE MODEL



A distribution of the Background data (yellow) and the Karoo model (red)

Generating Psuedo Experiments

To test how good of a model Karoo predicted, we can generate pseudo experiments that are derived using the Karoo function as truth. Then the likelihood distribution of those experiments can be compared to the likelihood that was measured for the background data.



A distribution of the data (yellow) and the Karoo model (red)

NEXT STEPS

- The likelihood of the data would fall within the distribution if it was a good fit
- More work needs to be done to find a function that fits the data accurately

NEXT STEPS

- Improve Karoo model by adding functions with the form sin(x+a) instead of just sin(x)
- Use functions selected by Karoo to re-optimize cuts for past analysis to see if there is improvement
- Work on other applications of Karoo for other problems: index of refraction, analysis for the Askaryan Radio Array (ARA), etc.

ACKNOWLEDGEMENTS

Many thanks to Kai Staats who wrote the Karoo GP code and was kind enough to share it with our group at OSU. Thank you to my advisor, Amy Connolly, who has taught me almost everything I know about particle astrophysics. Special thanks to the NSF CAREER Award 125557 and ANITA Grant NNX15AC20G.