Machine Learning Techniques to Search for 2νββ decay of 136Xe to the excited state of 136Ba in EXO-200

George Halal

1georgech@stanford.edu, Gratta Group, Physics Department, Stanford University

Introduction

Measuring the 2νββ decay of 136Xe to the excited state of 136Ba is exciting because it’s expected to introduce constraints to the Nuclear Matrix Elements, which would allow us to determine the effective Majorana neutrino mass more precisely. Therefore, in this project, we use BDT, MLP, and LSTM models with simulations of the EXO-200 detector, to increase the sensitivity of this search and set a lower limit on the half-life of this decay. We find that the LSTM performs the best.

Data

The EXO-200 detector, pictured below, is a cylindrical barrel filled with enriched 136Xe, which undergoes decays into lighter particles that are detected by sensors encompassing the walls of the detector.

We run Monte Carlo simulations of these decays that take the geometry of the detector into account. Each event is described by a number of clusters spanning the 3-dimensional volume of the detector. Each cluster measures the energy deposited in the volume of its position in the detector. We use 3,803,128 events for training, 950,782 for validation, and 1,188,477 for testing.

Models & Features

Boosted Decision Tree (BDT) & Multilayer Perceptron (MLP) features:
- Sum of cluster energies / event (normalized)
- Number of clusters / event (normalized)
- Standoff distance (normalized): minimum distance between a cluster with non-zero energy and the walls of the detector.
- $\nu_2, \nu_3, \nu_{sum}$ (normalized); $\nu_i \equiv \min_{j \neq i} |E_j - E_i|$, where $E_j$ is the energy at cluster $j$, $E_i$ is the theoretical de-excitation energy of photon / produced during the decay, and $S$ is the set of clusters.
- $A_{r_{1,2}}$ (normalized): The radial distance between the positions of the clusters with energies closest to the energies of the de-excitation photons.

Long-Short Term Memory (LSTM) features:
We order the clusters in each event based on decreasing energy deposited in them. At each time-step, we feed the network a vector containing a cluster’s energy and $x$, $y$, and $z$ positions after normalization. The number of clusters per event varies, so we pad the sequence and mask the padded values while training.

Binary output for all models:
Signal (2νββ excited state decay) or background (all other decays)

Results & Discussion

The LSTM performs the best. This is interesting, because it implies that the higher-level features used in the BDT and MLP disregard important information from the raw data used in the LSTM. Also, it implies that the order of the clusters based on decreasing energy may contain sequential information.

Future Work

Moving forward, we will apply the LSTM model to real data collected from EXO-200 to increase the sensitivity of this search and set a lower limit on the half-life of this decay.

References

3. https://kandisernate.wordpress.com/2013/01/05/dérivation-error-backpropagation-gradient-descent-for-neural-net/ (Kai Zuber)