Data Science Projects in Fundamental Physics at Yale Wright Lab

Spring/Summer 2019

Identifying neutrinos from nuclear reactors in PROSPECT

Contacts: Prof. Karsten Heeger (karsten.heeger@yale.edu), Dr. Thomas Langford (thomas.langford@yale.edu)
URL: heegerlab.yale.edu
Experiment: PROSPECT is a novel experiment designed to study the emission of neutrinos from a nuclear reactor, probe their fundamental properties, and search for new forms of matter. PROSPECT operates at the High Flux Isotope Reactor at Oak Ridge National Laboratory. The accurate identification of neutrinos from a nuclear reactor allows tests of fundamental physics and enables the monitoring of nuclear reactors for safeguard applications.
Project: We see the possibility of applying modern machine learning algorithms to our extensive simulation and calibration datasets to improve the efficiency and fidelity of our neutrino event selection.
Statistics/computing/data science challenge: One of the main challenges is separating the ~1000 neutrino events per day from the ~900k background events. So far we have utilized traditional event selection criteria based on input from simulations. However, we would like to explore alternative methods to isolate these interactions to improve our ability to study neutrinos and the nuclear reactor itself.

Developing instrument beam basis functions in measurements of Dark Energy

Contact: Prof. Laura Newburgh (laura.newburgh@yale.edu)
URL: campuspress.yale.edu/newburgh
Experiment: CHIME is a radio interferometer and takes peta-bytes of data per year.
Project: We have collected a data set to measure the CHIME beams, which is absolutely critical for achieving our science goal of understanding Dark Energy better. However, the data coverage of the full 2D sky is sparse. We will train spherical harmonic basis functions on simulated full 2D beam data and build a machine learning algorithm to use the data to fit the real beam.
Statistics/computing/data science challenge: There are a variety of challenges: The reduced beam data is smaller, but any code must be extremely efficient. In addition, no one has attempted this sort of analysis in radio astronomy, and machine learning seems like an ideal tool, but it is also a challenge with the sparsity of the data, and estimating errors will be crucial.

Energy reconstruction for the liquid xenon scintillation detector using deep learning

Contacts: Shilo Xia (shilo.xia@yale.edu), Prof. David Moore (david.c.moore@yale.edu)
URL: campuspress.yale.edu/moorelab
Experiment: The EXO-200 experiment is searching for neutrinoless double beta decay in liquefied Xe-136. If observed, this decay would determine whether there is any fundamental difference between neutrinos and antineutrinos, which has implications for why we observe matter in the universe instead of antimatter. To search for the decay, the detector collects waveforms of the light emitted by a particle interaction.
Project: The goal of this project is to use a deep neural network (DNN) to reconstruct the energy of photons produced by an incident particle in the liquid xenon detector from the raw detector waveforms.
Statistics/computing/data science challenge: The major challenges are that DNN has to learn to reconstruct the position and time dependence of the detector response, using a large dataset of ~100M calibration events collected from four radioactive sources over the course of 5 years. Despite the challenge, DNNs may substantially outperform the traditional method and reconstruct energy due to its potential power to accurately reconstruct the detector response and to optimally remove noise components from the waveforms.

Event identification in the nEXO double beta decay experiment using a deep neural network

Contacts: Zepeng Li (zepeng.li@yale.edu), Prof. David Moore (david.c.moore@yale.edu)
URL: campuspress.yale.edu/moorelab

Experiment: The next-generation Enriched Xenon Observatory (nEXO) is a proposed experiment to search for neutrino-less double beta decay (0υββ) using 5 tonnes of isotopically enriched xenon in a time projection chamber (TPC). If observed, this decay would determine whether there is any fundamental difference between neutrinos and antineutrinos, which has implications for why we observe matter in the universe instead of antimatter.

Project: 0υββ signal events could be separated from background events by performing a three-dimensional event reconstruction, in which signal events have more localized energy deposits. Deep neural networks are a promising tool for event identification in nEXO due to their ability to optimally separate events based on subtle features in the charge signals. A neural network using an existing network architecture has outperformed traditional event identification algorithm in preliminary work.

Statistics/computing/data science challenge: The network used downsampled waveforms due to GPU memory limits. In order to use the raw waveforms, the network needs to be optimized to solve the memory problem. We hope that an optimized network for event identification in nEXO would further improve the signal identification efficiency, leading to higher discovery potential for this future detector.

Interdisciplinary applications of modern analysis techniques: Classification of wood species used in cultural heritage objects

Contacts: Dr. Katherine Schilling (katherine.schilling@yale.edu), Dr. Thomas Langford (thomas.langford@yale.edu)
URL: ipch.yale.edu, research.computing.yale.edu/research-collaboration/wright-lab

Project: New techniques of chemical analysis are being applied to help understand the origin of wood used in fabrication of items of early American furniture. The Yale Institute for Preservation of Cultural Heritage has been working with the Yale Center for Research Computing and Wright Lab to apply modern data analysis techniques to better classify species of wood. This will better enable conservators and researchers to understand 17th-19th century trade dynamics and migration between Western powers and their colonies.

Statistics/computing/data science challenge: We see the possibility of applying modern machine learning algorithms to this classification problem to improve the discrimination of genus/species of different samples from historical artifacts.