Final Project Idea: Hunting Awesome Variable Stars in Sloan Digital Sky Survey Data

Brandon Bell

***Abstract*—I would like to hunt for a rare type of variable star called a Luminous Blue Variable (LBV) in Sloan Digital Sky Survey (SDSS) data with an Artificial Neural Network.**

I. Luminous Blue Variables

LBVs are are very large, about 60 times the mass of the sun, in the late stages of their journey towards exploding as a super novae. LBVs exhibit large variability in brightness and color accompanied with drastic spectral changes; this variability is thought to be connected with large stellar pulsations the can result in the ejection of entire layers of the star into space. It is through these eruptions that the stars loose much of their mass to the surrounding space creating large infared Nebulae around them. The exact nature and physics of these eruptions is not well understood at this time. One of the principal causes of our current ignorance is the extreme rarity of these objects; the LBV phase is a short lived phase of evolution of only the largest stars which are themselves, very rare. There only about 16 confirmed LBVs known with a slew of some 30 possible LBVs. Locating more LBVs in either our galaxy or, in nearby galaxies, is an import issue in the study of massive stellar evolution and import to our understanding of the physics driving very large stars in general. SDSS is a large imaging and specrographic survey of thousands of stars in both the visible and infrared with data on hundreds of stellar parameters of each star and I hope to train an artificial nueral network to find features unique to LBVs and hopeful find new LBVs hidden in the mountain of SDSS data.

II. The Code

For the code, Ive found two C libraries implementing various forms of Neural networks (Darknet, FANN) but, I would like to try and scale, or implement a new, Scikit-Learn python routine for training a back-propagating neural network in a parallel, perhaps with openMP or GPGPU support. Scikit-learns Multi-layer neural network is explicitly not intended for large scale applications as it offers no GPU support (scikit-learn.org). I could implement large scale parallel support for training with pythonMPI or, write c code to train on the data and simply produce a weights file for use Scikit-Learn, or Write the c code and a cython wrapper to integrate directly with the python code.