Shakthi Visagan
Professor Aaron Meyer
BE 188
16 February 2018

Project Proposal

I plan to implement a classifier that will identify patients as either those who have Alzheimer's/mild Dementia (AD) or those who have no condition (NC) based on a series of MRI coronal scans of their gray matter. This classifier will be based on the LeNet convolutional neural network created by Yann LeCun and Yoshua Bengio in the late 1980s and early 1990s. I will also try more modern architectures of classifier convolutional neural networks to see if I obtain better models and better accuracy. I will implement my models using the well-documented Tensorflow based Python implementation Keras, using the `Sequential()` API.

The data comes from the OASIS brain imaging data set which is easily accessible online. I used the cross-sectional data set since my goal is to classify or detect Alzheimer's, as opposed to predicting it using Longitudinal data. I have spent the last two weeks cleaning up the data from 416 patients down to 216 (I kept only the ones that had Clinical Dementia Rating scores, scores based on a scale of a patient's level of dementia). I also cleared the `NaN`s in the dataset as well as cleanly associated demographic data into an X-matrix (a patient's age, sex, handedness, socioeconomic status, education, brain volume, etc) and a Y-matrix (a one-hot encoded vector description of a patient's conditions [0,1] for AD and [1,0] for NC). Furthermore, twenty patients had two MRI scans done instead of just one scan session, which I have excluded from my training directory of images and I may consider using them as a final testing set. For each of the 216 patients, I was able to acquire coronal scans of their brain from slice 90 to 99, slice numbers I have seen in several papers due to their high variance between healthy and ill patients and are often used to diagnose Alzheimer's in clinical environments. I will try to expand my selection of slices from coronal slice 50 to slice 150 since that will also boost the power of my model.

This topic of neural networks and deep learning is interesting because several leaders in the field of machine learning have made claims that the field of healthcare and medicine will be radically changed for the better by the introduction of powerful artificial intelligence. If a computer based diagnostic model can make better detections/diagnoses/predictions on a set of scans better than a radiologist can, then the future of healthcare will be markedly different from the current systems in place.

I expected the data pre-processing step to take the longest to finish but most of it is done. Some steps left will be deciding whether or not to focus on a certain age group (only modeling patients older than 60 years old), accessing more slices and appending the paths to the correct directories, and deciding what to do with the nonimaging data (the matrix of patient demographics). I should also choose how large each of my training/validating/testing datasets are since cross-validating is not a common procedure in evaluating neural network performance. I have also prepared the images using OpenCV, such that the images are all in grayscale and normalized, which is a common practice for images being fed into LeNet. It will be relatively simple to implement the LeNet model in Keras, since that is well-documented online. The

challenge will be implementing more complicated models like those created within the past five years, especially since those are less documented.

https://github.com/shak360/be188project

My progress and the current state of my project can be tracked at the above linked Github repository. Though the project will be challenging, I am confident that I can finish creating a model in less than a month, especially since most of the data has been cleaned up, and Keras is a well documented package for creating convolutional neural networks.