Journal Report 19
2/24/20-2/28/20
Sohom Paul
Computer Systems Research Lab
Period 1, White

Daily Log

Tuesday, February 25

Installed simdna dependency and resolved dependency conflicts.
Installed CUDA tools for GPU accelerations.
Successfully changed Anaconda kernel.

Thursday, February 27

Finished install GPU-acceleration dependencies for Windows.
Found library incompatibility with Windows.
Started porting MPNN example; successfully ported 1/2 files.

Friday, February 28

Researched literature accuracies for deep learning models. Found bounds or chemical accuracy, to which we can later compare our model.

Timeline

Date	Goal	Met
Feb 17	Understand code from MPNN tutorial. Start building neural fingerprinting.	No; this goal has been expanded to fill the next 2 weeks.
Feb 24	Build first example MPNN.	No; still in process of setting up testbed to use sample code.
Mar 2	Build first example MPNN (possibly in Python 3).	Partial; need to finish building and testing.
Mar 9	Finish building example MPNN. Start implementing neural fingerprinting.	
Mar 16	Continue implementing neural fingerprinting	

Final Goal Contract	
A	Our integrated multitask-GCN model produces results superior to single task or baseline convolutional models. Our project is well-documented and works out-of-the-box with clear installation instructions available on our project GitHub. The user is given a visual interface to interactively draw molecules and predict chemical properties.
B	Our integrated multitask-GCN model produces results that are on-par with other models. Our project documentation exists and the project works with some minor hassles when installing from GitHub. The user is capable of viewing and selecting molecules from a pre-set list in order to predict chemical properties.
C	Our integrated multitask-GCN model produces results that are significantly worse than other methods. Our project is poorly documented and installation from GitHub is very difficult. The user can select molecules to test from a list, without any visuals showing molecular structure.

Reflection

Oh, the woes of working on a Windows system... for my own sanity, I will eventually have to switch to dual-booting with Ubuntu. This past week, I spent substantial amounts of time testing out hardware acceleration for Tensorflow, but it was only on Thursday that I realized that the `deepchem` library could not be installed for Windows. (I had not noticed that line hidden at the bottom of the README file. Granted, I should have anticipated that this would be an issue.) My work will likely not entirely go to waste, however, because later in the year I anticipate some of the models that I train working directly on the molecular fingerprints will not require any of the chemistry ML packages.

I have begun, but have not yet finished, porting the example MPNN to Python3. So far, the process is going relatively painlessly, as most of the Python is immediately cross-compatible. I hope to have this process finished by next week. This will allow me to quickly start implementing the neural fingerprinting that would be necessary to allow arbitrary user input for our final user interface.

I have also spent some time looking up values for chemical accuracy for the different properties we are studying. Thus far, I have been mostly looking at losses and percent errors; however,

the ultimate goal of a property prediction program is to eliminate the need for synthesis and experiment by producing results that are reasonably close to the experimental error bounds. Of course, these values are quite small and hard to achieve computationally, but they will serve as a good standard of comparison for when we start writing our final paper.