Journal Report 2
9/9/19-9/13/19
Sohom Paul
Computer Systems Research Lab
Period 1, White

Daily Log

Monday September 9

Spent 25 min of class trying to connect to the printers.
Tried running energy calculation on water molecule; test failed with a segmentation fault.
Tried updating Psi4. Update did not fix segmentation fault.

Tuesday September 10

Installed Psi4 on home computer. Ran tests scripts without segmentation faults.
Uninstalled Psi4 conda package and downloaded program binary. Successfully ran test scripts without segmentation faults.
Ran SCF and B3LYP calculations on water with cc-pVDZ and cc-pVQZ basis sets.

Thursday September 12

Learned to extract orbital energies from Psi4 to get HOMO and LUMO energies.
Experimented with using geometry optimization as a preprocessing step.
Learned to extract wavefunction from energy calculation.
My Anaconda installation broke. Reinstalled conda.

Timeline

Date	Goal	Met
Sep 2	N/A	N/A
Sep 9	Build first graph convolutional network and install relevant libraries	Partially; No backpropagation.
Sep 16	Run relevant DFT calculations; find best parameters	Yes.
Sep 23	Write Python script to run DFT calculations and pull relevant features from our dataset.	
Aug 26	Build toy networks with multitask learning	

Reflection

My primary goal for this week was to locate the commands in Psi4 for extracting the features we plan on studying (dipole moment, polarizability, HOMO, LUMO, zero-point vibrational energy, etc.). Right now, I've run tests to extract dipole moment, HOMO, and LUMO, and matched them to real data and got reasonable accuracy. I have not yet checked if the calculated thermochemical properties match with experiment, which is something I will do this coming week. My other goal for this week was to determine which sets of parameters we should use. Below I've reproduced some of the results from testing dipole moment calculations on water.

Functional	Basis	Dipole moment (D)	Error	Time (s)
SCF (HF)	cc-pVDZ	2.0595	11.05%	0.33
SCF (HF)	cc-pVQZ	2.0096	8.36%	0.64
B3LYP	cc-pVDZ	1.9113	3.06%	0.91
B3LYP	cc-pVQZ	1.9024	2.58%	2.53
CCSD	cc-pVDZ	2.0592	11.03%	1.47
CCSD	cc-pVQZ	2.0096	8.35%	12.89

There are actually several dozens of other functionals and computation methods avaiable in Psi4, but these three seemed to be the most common in the literature, so I focused my investigations on these. I found that the B3LYP functional was best able to predict dipole moments and molecular orbital energies of water. This much was consistent with the research I've conducted; the book that I'm learning DFT from cites B3LYP as the most accurate functional for small, isolated molecules. I also unnecessarily tested the cc-pVDZ, cc-pVTZ, and cc-pVQZ basis sets against each other before I realized that cc-pVQZ contained the other two as subsets, so there really is no reason not to use anything other than cc-pVQZ. I also tried running a CCSD calculation, which produced much worse results at a much slower speed.

I had not listed it as a goal on my last journal, but I was hoping that by the end of this past week I would have a parser that could take files from our dataset, convert it into a from that Psi4 can process, and programmatically save the outputs of Psi4 into a CSV. I have not yet been able to complete this task because of my unfamiliarity with Psi4 functions (and the confusing organization of its API), so I've pushed it to be my goal for the coming week. Once I have this parser written, I will be able to go back and do more rigorous testing of other functionals and evaluate accuracies over multiple molecules.

I spent an unfortunate amount of time this week fighting with the technology I'm using. On Monday, I spent much of class trying to debug segmentation faults and on Thursday, I wrested with ModuleNotFoundErrors from all conda commands after I tried to install pylint for VS Code. Both of these issues were eventually resolved by reinstalling the respective software, but not before I wasted much time reading source code/editing environment variables and .bashrc. That being said, the number of these kinds of problems ought to decrease as the year progresses.