Journal Report 18

2/24/20 - 3/1/20

Emily Ye
Computer Systems Research Lab
Period 2, White

Daily Log

Monday February 24

Continued trying to increase accuracy for LUMO GCN model

Tuesday February 25

Realized I was using the GCN model from my winter goal instead of the multitask GCN model that our final project will be based on

Continued trying to increase accuracy for LUMO GCN model

Thursday February 27

Calculated percent errors (based on 10 randomly selected molecules) for the multitask GCN model

Continued trying to increase accuracy for LUMO GCN model

Final Goal

For an 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.

For a 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.

For a 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.

Timeline

February 16	Finish combining multitask and GCN networks, test and tweak multitask GCN to improve accuracy	Finished combining multitask and GCN networks and began tweaking multitask GCN to improve accuracy
February 23	Improve accuracy for HOMO and LUMO models	Some progress, but I want to continue trying to improve accuracy
March 1	Continue improving accuracy for HOMO and LUMO models	Realized I was using the wrong model, so I want to keep trying to improve accuracy for LUMO
March 8	Continue improving accuracy for LUMO model and other underperforming properties' models	
March 15	Continue improving accuracy for underperforming properties' models	

Reflection

My main goal for this week was originally to continue improving accuracy for the HOMO and LUMO models. On Tuesday, I realized that for the past two weeks, I was using the GCN model I had made for the winter goal instead of the more recent multitask GCN model that our final project will be based on. While this means some of my progress was not as useful, I did try transferring over some of the layer changes I made with the GCN model to the multitask GCN model for LUMO. I did not do the same for HOMO since HOMO was already performing much better on the multitask GCN model (average percent error of 2.8% over 10 randomly selected molecules, compared to LUMO's average percent error of 55% over 10 randomly selected molecules). For the remainder of the week, I continued trying to improve LUMO's accuracy.

When I calculated percent errors for the multitask GCN's accuracy on each property, LUMO was by far the worst. Some other properties that I plan to look at in the upcoming weeks include A (percent error of 14.3%), B (percent error of 22.4%), C (percent error of 14.1%), and mu (percent error of 16.8%). All of the other errors were less than 5%, with the exception of gap because it is partially dependent on LUMO (which currently is the most inaccurate property). Previously, I had planned to test different convolutions to increase the model's accuracy, but since I realized I was looking at the wrong model, I think I no longer need to use different convolutions since the multitask GCN model has higher accuracy than the GCN model. Instead, I plan to recalculate average percent errors after trying to improve the specific models (by using a larger sample size of molecules rather than the 10 randomly selected molecules I am currently using) to figure out the next step for which properties I should continue improving.