Mentor Visit Assessment #12: Original Work Part III

Mentor: Hao Tian

Profession: Graduate Research Assistant

Location: Heroy Hall at SMU

Date: November 18, 2022

Assessment:

During this mentor visit, I discussed my original work and received feedback on how to refine it for feasibility and usefulness. I had already begun working on interpreting machine learning (ML) models with my mentor prior to the meeting, and I was thinking about ways to improve what we had. Through my ISM research assessments, I had learned about several new ways that computational chemists can explain the predictions of models in order to increase their trustworthiness, so I was curious to know if they could be combined. Since different ML model architectures can be combined in ensemble methods, I wondered if ML explanations could be made more effective in the same way. However, I found out that this was probably not the case because each explanation technique works differently. I also learned that it's important that the tool is model-agnostic so that it can interpret models universally, though, in theory, its performance could be improved by tailoring it to model type.

In order to trust such models, it's important that their prediction rules match the chemistry literature. For universal approaches, it's not possible to derive rules from the model structure, but we can characterize the contributions of atoms or functional groups. One of my ideas was to add a feature allowing users to compare the explanation of model predictions with textbook knowledge for certain molecules in certain datasets. For example, if I were predicting

solubility, then the tool could verify that a hydroxyl group in a given molecule has a positive contribution and a long carbon chain a negative one. Originally I was thinking about doing this for toxicity tasks using toxicophores (substructures that are associated with toxicity), but we discussed that toxicity is not nearly as well understood as solubility, and it makes sense to start with a simpler idea. The math for automatically checking contributions is also difficult because there must be a relative comparison between each atom or group. Therefore, I will start with displaying the contributions on a molecular heatmap for users, who can use their own expertise to judge whether the model is trustworthy or not. To better illustrate the process, my mentor suggested that I create a Jupyter notebook explaining each step to take. Previously, Dr. Nielsen said in an interview that having gained enough experience, he could just look at a molecule and determine its solubility, but I will have to first do some research to better understand this property. Two questions I have are whether functional group position matters for solubility and if it would be more effective to identify functional groups within molecules than going atom by atom. Of course, as I write more code, I will have a better idea of what is relevant and what can be avoided.

First in the process of creating my original work, I will experiment with the tool and determine what features could be added next. After writing more code, I plan to accompany the tool's code with a background and explanation of its functionality, which will make it accessible and useful to more people. I will also include appealing figures to present the work, which my mentor reminded me is especially important. I look forward to working on this project and hopefully building on it later in the year.