Literature Review

Aodong Liu

February 8th, 2022

Article: The potential for machine learning in hybrid QM/MM calculations
Zhang, Y.; Khorshidi, A.; Kastlunger, G.; Peterson, A. *J. Chem. Phys.* **2018**, 148(24):241740. doi: 10.1063/1.5029879.

Summary:

This paper shows the potential of incorporating machine learning into hybrid quantum mechanical/molecular mechanical (QM/MM) multi-scale simulations. With the recent advances in the atomistic machine-learning (ML) tools, the authors are able to create QM/ML hybrid models where they train the ML models using the QM calculations, and apply such models as MM calculators for the MM region. Moreover, the authors propose a scheme to re-train the ML model as more QM data calculated at different geometries become available during the algorithm, which can great improve the accuracy of the model. The authors prove the validity of their model with two illustrative examples, showing that their QM/ML algorithm can achieve chemical accuracy at a much cheaper cost compared to full QM simulations.

Critical Evaluation:

- The authors claim that one advantage that their new QM/ML algorithm possesses is that it can avoid distortion caused by mismatch between QM and MM descriptions. This is because the ML potentials are directly trained/re-trained from the QM potentials, and therefore they should share a common potential energy surface (PES) and should naturally predict identical mechanical properties. While I consider this is a great strength of the QM/ML algorithm, I don't think they've shown an example of this in the paper. I would be very interested to see if they are able to calculate a lattice constant for their Cu(100) structure from their ML potential, and see if it matches the lattice constant calculated from a pure QM calculation within the training tolerance threshold.

- I think the authors convincingly and thoroughly show that re-training steps in their proposed scheme are important by showing their algorithm side by side with an algorithm with the ML potential but in a pure QM/MM scheme (without re-training). From their figures, the readers can easily observe the difference in accuracy between the two algorithms. Moreover, by re-training the ML potential every time the geometry is updated and the QM region is re-calculated, they eliminate the need for large training-data before the algorithm starts. This feature could be significant for the field because the users now do not need to worry too much about the pre-generation of the

training data and the associated cost.

I also agree with their argument that the cost of the re-training step is correlated to how familiar the new 'QM' image is to the ML model. In other words, whether the new geometry in the QM region has been learned by the ML model before will greatly impact the cost of learning. This is well-supported in Fig. 7 of the article, since the first 'QM image' is the most expensive to learn, and the cost slowly decreases to 0 for later images as the ML model learns and improves with experience. Overall, I think their scheme of re-training is designed to be 'blackbox', and users who are not experts of machine learning can find their scheme friendly to use.

- This paper proposes a solid theoretical approach to study the ML-based potentials. Since this paper represents one of the first attempts towards incorporating ML into QM/MM models, the authors admit the many deficiencies in their algorithms at the moment, such as not being able to include long-range electronic effects, not being able to dynamically shift the QM/MM boundary, and the need for expertise to generate training data before the algorithm. However, I do not see any jumps or stretches in their logic, and their QM/ML algorithm is definitely a promising directions that future improvements can be expected.

- The quality and the presentation of the data in this paper is great. The authors realize that fitting their neural network based ML model to a set of data does not lead to a unique set of optimal parameters, therefore they independently trained and presented ten ML models using the same pre-training data set with random initial parameters. From their benchmark results, all the different models exhibit almost the exact same pattern, giving credence to their algorithm.

- This paper has adequate number of literature references.

Future Directions:

- In this paper, the way the authors update the geometry is by performing a structure optimizations in the BFGS (Broyden-Fletcher-Goldfarb-Shanno) minimization scheme. It seems to work pretty. What I would propose is to try using a transition state search scheme to update the geometry. This might allow us to explore the properties at the transition state.

- In this paper, the electrostatic QM-MM interaction has been ignored. In the future, it would be interesting to see what happens if the polarized embedding scheme is used, where the solutes are put in the QM region, and the explicit solvents are placed in the MM region.