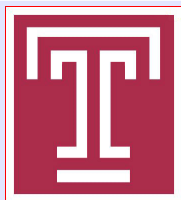


Overview

Machine learning provides a new analytical tool to interpret and extrapolate computed quantum mechanical properties. In one application (I) vibrational frequencies of a library of molecules are estimated using density functional theory with a modest model chemistry and these estimates are improved by using a neural network with molecular properties as input and training on experimental frequencies. In a second application (II), molecular potential energy surfaces are interpolated from high level *ab initio* calculations using machine learning to provide a surface for molecular dynamics in which forces and gradients can be determined 4 orders of magnitude faster than high level *ab initio*. These surfaces are robust to using molecular dynamics to simulate anharmonic vibrational spectra.



Acknowledgements

Hylleraas HPC (Notur NRC NOTUR NN4654K)
Temple HPC (NSF 1625061)
Midwest Undergraduate Computational Chemistry Consortium

I. Vibrational frequency prediction neural network

A straightforward application of machine learning is presented which shows improved prediction of vibrational frequencies following training a neural network. The steps in constructing this neural network demonstrate the common tasks associated with applying machine learning to prediction.

1. Construction of a dataset to train the model. In this case experimental vibrational frequencies of hundreds of molecules are collated along with corresponding vibrational frequencies computed using the B3LYP/6-31G(d) model chemistry.
2. A neural network is designed and coded using Python with the PyTorch (<https://pytorch.org>) module. In this design an input layer consisting of 11 inputs is connected to three hidden layers feeding into a single output reflecting the vibrational frequency. This design is reflected in Figure 1. Each input is connected to each node of the first hidden layer using a weight between 0-1. Each node between hidden layers is activated based on what is known as a RELU activation function which is the $\max(0, x)$.
3. The network is trained on experimental vibrational frequencies and the connections between the nodes reflected in weights are optimized through a technique known as back propagation using a loss function. In this case a smoothL1loss is used. This process is repeated through several epochs until a tolerance in the loss function is achieved.
4. Following training, the model is validated on the quality of prediction of experimental data typically as a function of parameters (hyper parameters).
5. The model is tested on a second set of different data to estimate the accuracy and error bars. This is shown in Figure 2.

This neural network achieves a mean squared error of about an order of magnitude better than standard scaling of computed vibrational frequencies (0.96 in this case, see Radom [1]).

References

- 1) Merrick, Jeffrey P.; Moran, Damian; Radom, Leo. "An Evaluation of Harmonic Vibrational Frequency Scale Factors". *J. Phys. Chem. A* 2007, 111 (Copyright (C) 2010 American Chemical Society (ACS). All Rights Reserved.), 11683-11700.
- 2) Smith, Jonathan M.; Nikow, Matthew; Ma, Jianqiang; Wilhelm, Michael J.; Han, Yong-Chang; Sharma, Amit R.; Bowman, Joel M.; Dai, Hai-Lung. "Chemical Activation through Super Energy Transfer Collisions". *Journal of the American Chemical Society* 2014, 136 (5), 1682-1685.
- 3) Kapil, Venkat; Rossi, Mariana, *et al.* "i-PI 2.0: A universal force engine for advanced molecular simulations". *Computer Physics Communications* 2018, 236, 214-223.
- 4) Chmiela, Stefan; Sauceda, Huziel E.; *et al.*, "sGDML: Constructing accurate and data efficient molecular force fields using machine learning". *Computer Physics Communications* 2019, 240, 38-45.

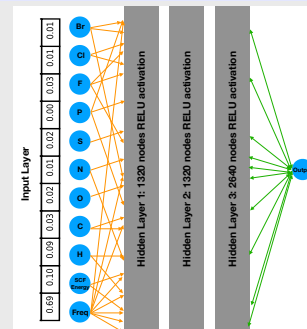


Figure 1. Schematic diagram of neural network for vibrational frequency prediction. Orange and green arrows are representative of strongly weighted connections. Table to left of inputs reflect the square of the sum of coefficients connected to that node representing the importance of each descriptor in predicting frequency.

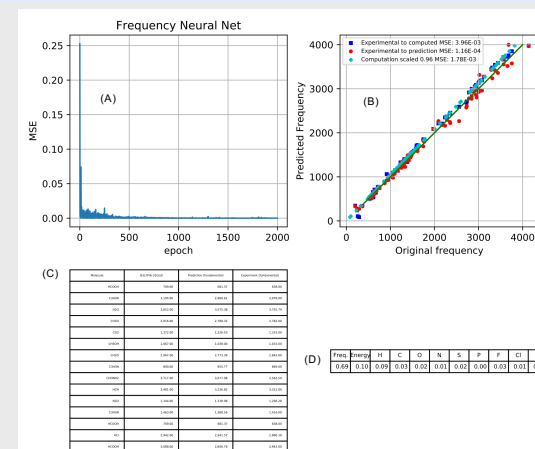


Figure 2. Selected results following training of neural network diagrammed in Figure 1. A) Evolution of mean squared error as a function of iterative refinement of coefficients. B) Plot of predicted frequencies versus experimental frequencies for B3LYP/6-31G(d) (blue), neural network prediction (red), and scaled, 0.96, B3LYP/6-31G(d) (cyan). C) Select frequency comparisons. D) Normalized coefficients on inputs including B3LYP/6-31G(d) harmonic frequency, SCF energy, # of H, C, O, N, S, P, F, Cl, Br in molecule.

II. Machine learning potential energy surfaces

Ab initio molecular dynamics is used to study chemical reaction dynamics [2] and predict other properties including vibrational spectra of energized or floppy molecules. Accurate model chemistries can require evaluation of energy and gradient for each step of a trajectory amounting to ~10 cpu sec/step with >10⁶ steps for a single trajectory at DFT accuracy for modest sized systems of 10-20 atoms. To achieve coupled cluster (CCSD(T)/cc-pVTZ) accuracy trajectories quickly become costly.

Here we demonstrate the use of accurate potential energy and gradient points to inform construction of a machine learning potential energy surface upon which we run molecular dynamics at millisecond cpu sec/step at high accuracy. The scheme is outlined in Figure 3. The calculations are carried out using i-PI [3] for path integral molecular dynamics. Infrared spectra are simulated by using the Fourier transform of the dipole autocorrelation function and averaged over several trajectories. Machine learning potential energy surface is refined with the sGDML package [4]. Figure 4 shows results for formic acid.

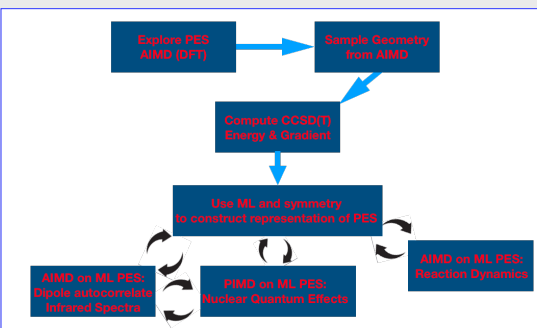


Figure 3. Schematic diagram showing steps of machine learning optimized potential energy surface creation and application

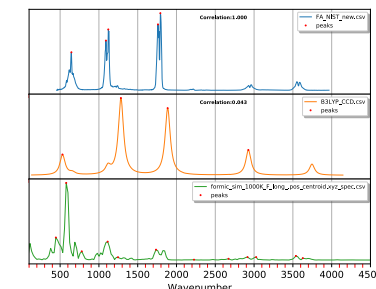


Figure 4. Formic acid gas phase experimental spectra, B3LYP/cc-pVDZ harmonic and PIMD (24 beads, 9.0 psec NVE, 0.1 fsec/step) on machine learning PES simulated infrared spectrum.