Charis Liao
Due Date: 4/6/2023

Modern chemistry aims to find and create practical molecules, which raises questions about their identification and synthesis. Computational modeling is a useful tool for chemists and biologists in this regard. Computer simulations can help create different scenarios to aid in molecule discovery, but they require a balance between cost and accuracy. Therefore, machine learning is crucial in achieving the best results.

General-purpose neural networks can efficiently predict nonlinear patterns, such as reaction thermochemistry, isomerization, and molecular torsions in drug-like molecules. In recent years, the success of machine learning techniques has increased due to better algorithms and data availability. As a result, the use of machine learning in experimental and computational chemistry has become more prevalent.

For our final project, we were given an ANI dataset containing quantum chemical calculations for training machine learning models in the prediction of molecular properties, such as energy, geometry, and vibrations. Our goal is to use a neural network to solve the Schrodinger equation for large systems while also satisfying Boltzmann Sampling via statistical mechanics.

Each molecule in the dataset is described by a SMILES string representing the chemical structure, a list of atoms in the molecule, and a tuple containing the 3D coordinates of each atom in the molecule. The dataset also includes a tuple of corresponding energy values for each molecule.

Initially, I planned to use self-organizing maps to predict energies for our deep learning exploration. However, further research showed that this method might not be suitable since it is an unsupervised learning model, unable to learn the pattern of our data. Therefore, I would choose to explore Recurrent Neural Networks (RNNs) as an alternative. RNNs can take a sequence of atomic symbols and coordinates as input and output the corresponding energy value.

As we have the option to choose homework instead of implementing a second method in the final project, I would like to choose homework for gaining more knowledge about the course content.