Training Machine Learning Models of Ionic Liquids using Ab Initio Molecular Dynamics

Azhar B. Moihdeen[1,2], Kun Yue[3], Orlando Acevedo[3]
[1]Department of Computer Science, Computing for Structure REU
[2]Department of Chemical Engineering, Georgia Institute of Technology, Atlanta, Georgia, 30332, United States
[3]Department of Chemistry, University of Miami, Coral Gables, Florida 33146, United States

Introduction

In recent years, environmental consciousness has been of utmost importance in various disciplines, including chemistry, where the wasteful practice of using hazardous solvents persists. Although a class of environmentally friendly solvents, ionic liquids, have been discovered, properties of these substances are not widely understood. By running *ab initio* Molecular Dynamics (AIMD) simulations, the energetic properties of these liquids can be understood; however, these simulations are computationally expensive in terms of time and processing power. Using the concept of machine learning, Artificial Neural Networks (ANNs) can be developed to minimize these costs while maintaining a sufficient level of accuracy.

What is Machine Learning?

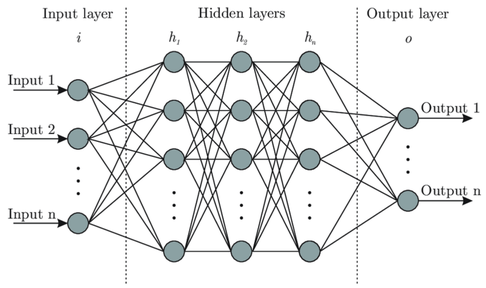
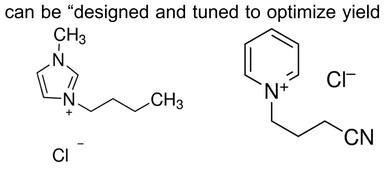


Figure 1: The structure of an Artificial Neural Network[2]

In broad terms, machine learning is a division of artificial intelligence that relies on the principle that "systems can learn from data, identify patterns, and make decisions with minimal human intervention."[1] ANNs serve as a major component of machine learning and consist of a series of layers each containing nodes that take inputs from previous layers and output results to subsequent layers, as shown in Figure 1. These structures improve upon themselves by changing the weights of certain nodes in the overall computations during each iteration until they show a sufficient level of accuracy as compared to a training set of data.

What is an Ionic Liquid?

An ionic liquid is an ionic compound, which occurs as a result of the attraction of opposite charges between a cation and an anion, that exists as a liquid at a wide range of temperatures, including room temperature. A typical characteristic of these substances is a relatively large, organic cation and a small anion, which causes the molecule to be asymmetric and thus unable to crystallize at normal temperatures. Ionic liquids are considered environmentally friendly as they are non-volatile compounds and can be "designed and tuned to optimize yield [and] selectivity."[3]



(a)

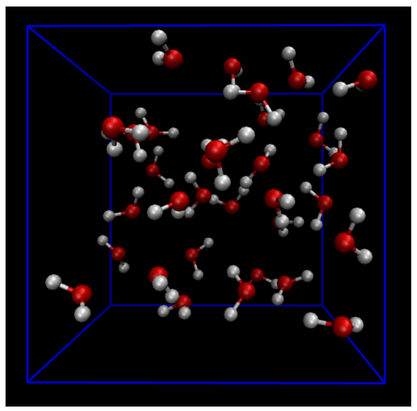
(b)

Figure 2: Examples of ionic liquids, (a) 1-Butyl-3-methylimidazolium chloride (BMIM-Cl) and (b) 1-(3-Cyanopropyl)pyridinium chloride

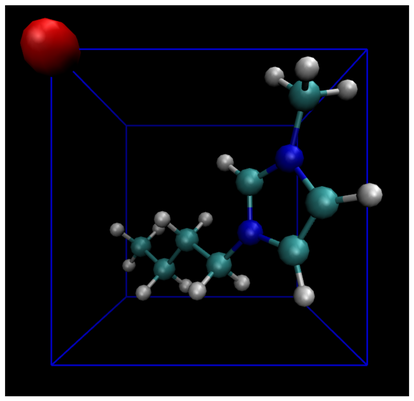
Objective

The purpose of this work was to optimize the process of obtaining data regarding the energy associated with a liquid's position in three dimensional Cartesian coordinates by using ANNs instead of AIMD simulations. By doing so, the computational costs of predicting the energy of a fixed number of molecules, known as a solvent box, of an ionic liquid are hoped to be minimized while remaining sufficiently accurate.

Methods



(a)



(b)

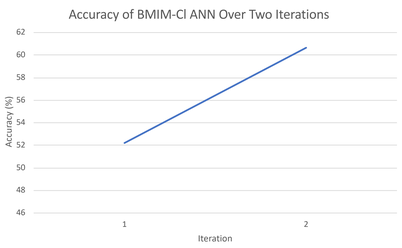
Figure 3: Images created using VMD[8] to represent the two substances studied, (a) a solvent box of 32 water molecules and (b) a single BMIM-Cl molecule

The following procedure was used to run AIMD simulations and train ANNs for the two substances studied, water and BMIM-Cl:

1. The chemical drawing software Marvin Sketch[4] was used to develop a representation of the molecule to be studied.
2. The molecule was then used to generate an input file for a solvent box (consisting, separately, of 32 water molecules and 1 molecule of BMIM-Cl) using the Packmol[5] software, which creates initial coordinates for AIMD simulations.
3. Separate simulations were run using CP2K[6], an AIMD simulation program, to find the cutoff radius for each atom regarding electron interactions with other atoms to be used in the geometry optimization in the next step.
4. Geometry optimization simulations were then run in CP2K[6] to find the lowest energy state of the solvent box and the final coordinates of the simulation were then used as the starting coordinates of the equilibration step, which finds a stable positions for the system given the temperature and pressure constraints: 298 K and 1 atm.
5. After the equilibration, AIMD simulations were run using CP2K[6] for 200,000 iterations for each substance.
6. Since periodic boundary conditions were used in all of the CP2K[6] simulations, the coordinates of each iteration were normalized such that the first atom was at the origin before training the ANN.
7. Input files for the training set of the ANN were then generated using the coordinates and energy for each iteration. A portion of the input files were then used to train the ANN using AENET[7], a software package that allows for the construction of ANNs related to atomic and molecular energies.
8. Once the training was completed, the remaining set of input files were used to test the ANN for its accuracy, comparing the results to those found from the simulations.

Results

Due to systematic errors, the entire procedure was not completed for both experiments and ideal results were not obtained in this work. The ANN for the BMIM-Cl experiment was able to complete two iterations of its training: in the first iteration, 7,000 orientations, or 3.5% of the initial 200,000 orientations, were used to train the ANN; in the second, 10,228 orientations, or 5.1% of the initial orientations were used for training, with the number increasing as compared to the previous iteration to better train the ANN in cases where it failed. The accuracy of the first iteration was found to be 52.2%, having passed 100,759 sets of coordinates out of the remaining 193,000 orientations, whereas that of the second was 60.7%, having passed a total of 115,107 orientations out of the remaining 189,772 orientations; this information is displayed in Figure 4. At the time of writing, the ANN for the experiment of the 32 water molecules has not completed an iteration of its training, meaning that data regarding its training pass percentage and the number of test cases passed could not be collected at this time.



Iteration	Number of Orientations Used for Training	Number of Orientations Used for Testing	Number of Orientations Passed	Percentage of Orientations Passed (%)
1	7,000	193,000	100,759	52.2
2	10,228	189,772	115,107	60.7

Figure 4: A summary of all the data collected from the training of the ANN of the BMIM-Cl experiment

Discussion and Future Directions

Although, due to time constraints, the ANNs for each experiment were not able to be trained completely, the data collected in this work thus far shows the potential that these constructs have in understanding the energetic properties of ionic liquids. The AIMD simulations, though consistently accurate as they use quantum mechanical calculations to determine the energy of a given orientation of molecules, were run for approximately a week, indicating an excessive amount of computational resources, whereas the data collected from the training of the ANNs was performed after a single day. Although the accuracy of the ANNs may not seem astounding, the amount of time that would be saved by using these instead of running simulations is significant and should be further studied. Future research on this subject should involve running separate AIMD simulations, extracting the coordinates, and comparing the obtained energies to those predicted by a fully trained ANN using the same set of coordinates. By doing so, the advantages of using ANNs to understand the properties of ionic liquids instead of only running simulations can be better understood.

Conclusion

- Ionic liquids serve as an environmentally friendly alternative to commonly used solvents for organic reactions.
- Several chemical properties of ionic liquids, including energetic properties, are currently not understood.
- The widely used method of running AIMD simulations to understand the energetics of these substances, though highly accurate, is computationally expensive.
- Using the concept of machine learning, ANNs can be trained to predict the energies associated with molecular orientations in a fraction of the time used to run AIMD simulations.
- Although the accuracy of the ANNs was shown to be only 60.7% in this work, they show great potential to serve as viable alternatives to running costly simulations by limiting computational expenses, which will allow research regarding ionic liquids to be conducted at a faster rate.

Acknowledgements

This material is based upon work supported by the National Science Foundation under Grant No. CNS-1659144.

Citations

1. Machine Learning: What it is and why it matters. https://www.sas.com/en_us/insights/analytics/machine-learning.html (accessed Jul 19, 2019).
2. Bre, F.; Gimenez, J. M.; Fachinotti, V. D. Prediction of Wind Pressure Coefficients on Building Surfaces Using Artificial Neural Networks. Energy and Buildings 2018, 158, 1429–1441.
3. Rogers, R. D. CHEMISTRY: Ionic Liquids--Solvents of the Future? Science 2003, 302 (5646), 792–793.
4. MarvinSketch (version 6.2.2), calculation module developed by ChemAxon, http://www.chemaxon.com/products/marvin/marvinsketch/, 2014.
5. L. Martínez, R. Andrade, E. G. Birgin, J. M. Martínez. Packmol: A package for building initial configurations for molecular dynamics simulations. *Journal of Computational Chemistry*, 30(13):2157-2164, 2009.
6. CP2K Open Source Molecular Dynamics, available at https://www.cp2k.org/.
7. N. Artrith, A. Urban, The Atomic Energy Network Package, http://ann.atomistic.net, 2015.
8. Humphrey, W., Dalke, A. and Schulten, K., "VMD - Visual Molecular Dynamics", J. Molec. Graphics, 1996, vol. 14, pp. 33-38.