A deep neural network for generation of functional organic molecules

Rhyan Barrett, Julia Westermayr and Reinhard Maurer
rhyanbarrett10@gmail.com

Sponsored by:
optibrium
dotmatics knowledge solutions

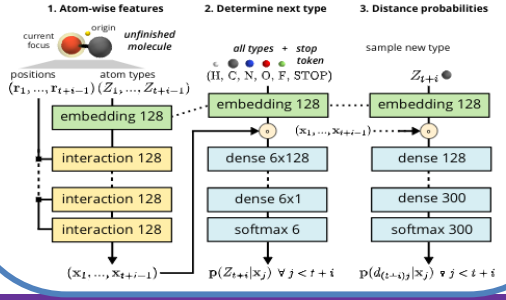
Background: The main goal of our project is to generate new molecules that have tuned optoelectronic properties using machine learning methods. This is done by adapting a generative model that was developed for small organic molecules with little chemical or structural diversity. The dataset we will be using is a spectroscopy data set of optoelectronic molecules. Additionally, we apply a deep learning model that is trained on quantum chemical properties of these systems to help the search for optimal candidates for future electronic devices.

Generative Model

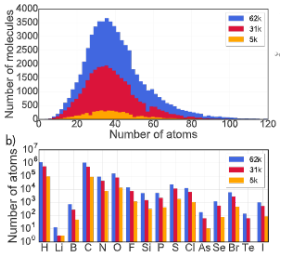
- To generate the new molecules we used an autoregressive generative model called G-SchNet[1]
- Initially G-SchNet generates a reference atom

$$p(\mathbf{r}_{t+i}|\mathbf{R}^t_{\leq i-1}, \mathbf{Z}^t_{\leq i}) = \frac{1}{\alpha} \prod_{j=1}^{t+i-1} p(d_{(t+i)j}|\mathbf{R}^t_{\leq i-1}, \mathbf{Z}^t_{\leq i}).$$

- Atoms are conditionally added to the molecule given the charges and positions of the atoms already on the molecule.
- Distance is used in the conditional probability to allow for rotational invariance of the molecule

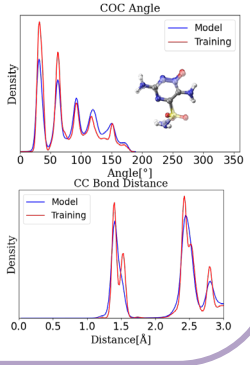


Dataset and Results

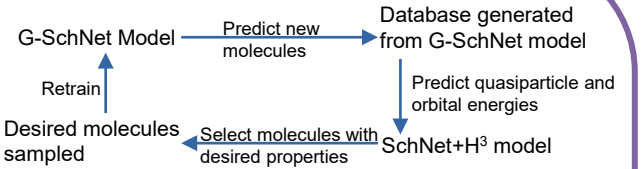


The dataset, OE62[2], consists of 62000 optoelectronic molecules which are relaxed to their lowest energy level. To the left shows the distributions of elements and sizes of the molecules in the training set.

To measure the performance of the generative model we compared bond lengths and angles for different elements. We also compared ring sizes and their frequency.

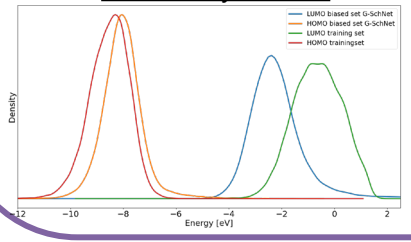


Improvement Loop



- Many of the molecules in our training set may not have the desired properties so here we attempt to build a loop that will generate molecules with useful properties, in this example a small HOMO-LUMO gap.
- We use a model (Schnet+H) to predict the HOMO and LUMO energies of all the systems in the database and to select molecules with the smallest HOMO-LUMO energy gap.
- Use this portion of the database to retrain G-SchNet which will generate molecules with a smaller HOMO-LUMO gap.

Preliminary Results



References:

- 1 Niklas Gebauer et al. Advances in Neural Information Processing Systems 32, 7566-7578, 2019.
- 2 Annika Stuke et al. Scientific Data 7(58), 2020.
- 3 Julia Westermayr et al. Chem. Sci. 12, 10755-10764, 2021.

AI³ Science Network+

PSDS PHYSICAL SCIENCES DATA-SCIENCE SERVICE

RSC INTEREST GROUP CHEMICAL INFORMATION AND COMPUTER APPLICATIONS GROUP