**1.1** **Learning a Continuous Representation of 3D Molecular Structures with Deep Generative Models**

Ragoza, Matthew & Masuda, Tomohide & Koes, David. (2020). Learning a Continuous representation of 3D Molecular Structures with Deep Generative

Models.

In this paper a fitting algorithm is presented that using deep generative models that utilizes atomic density grids and converts into discrete molecular structures. Deep learning for generative modeling can be categorized into encoder and decoder networks and Generative adversarial networks (GANs). Initially, deep generation model used the SMILES format, but problems with the same molecule can have different SMILES strings. SMILES is only capable to represent connection between molecules and not the 3D confirmations. Generative models have overcome the limitations of the other machine learning models in terms of generation and representation of 3D structures.

* + 1. **Working principle:**

Molecules are transformed into latent space using density grid. For this molecule are first converted into atom toes and their coordinates. Latent space is a compressed representation of the data where similar datapoints can be grouped together. An atom fitting algorithm is utilized for transforming density grid to atom types and their coordinates. The experiment in this paper has been performed on MolPort dataset. Implementation code can be accessed here: <https://github.com/mattragoza/liGAN>

1.2. **Machine learning based energy-free structure predictions of molecules, transition states, and solid**  
1. The work presented in this paper uses supervised learning to solve 3D molecular structures. The presented model is called G2S (Graph To Structure). There are also some modern approaches such as ETKDG6 and Gen3D7 that does the same with better efficiency, but they are not very general.

**1.2.1.** **Working principle:**

In the G2S model, kernel ridge regression is used for prediction of all the elements of the pairwise distance matrix. Which is then utilized to reconstruct the atomic coordinates of a molecule. Kernel Ridge Regression (KRR) combines ridge regression with kernel trick.**[2]**. The model only accepts stoichiometric data as input.

**1.2.2 Methodology**

* The first step is the separation of hydrogen atoms and heavy molecules.
* Then Conversion occurs of molecular bond models to graphical representations of   
  fixed size.
* One model is used for each pair of distances to calculate pairwise distances to predict the interatomic distances of molecules outside the sample.
* RDkit is used to create an adjacency matrix to use later for representation.
* The problem with distance geometry arises when there is conversion of interatomic distances into 3D coordinates. DGSOL15 is used for heavy atoms. The distance geometry problem is not model dependent.
* The experiment is performed using the QM9 dataset.
* Here is the code for this model: <https://zenodo.org/record/4792292#.YZvsq5HMLRY>

1.4. **Dataset’s chemical diversity limits the generalizability of machine learning predictions**

The chemical diversity of the dataset can limit the generalizability of machine learning predictions. This paper presents a comparative analysis using QM9 and a relatively new dataset KK9. The problem it tries to solve is the generalization of datasets.

**Test on different ML models.**

The models used in this article are Kernel Ridge Regression (KRR), Elasticity net Elastic Net and SchNet Neural Network. The models are evaluated for several energetic properties:

* Total molecular energy
* Energy of HOMO
* Energy of LUMO

There are many methods of molecular imaging are Chart view, the coulomb matrix, Bond Bag (BoB), Link and Angular ML descriptors (BAML). Coulomb matrix (CM) is used for molecular imaging in this paper. A CM is a square atom in an atomic matrix consisting of atomic nuclei charge (Z) and cartesian coordinates (R) of each atom **[3]**. Coulomb matrix eigenvalues (CME) are global three-dimensional representations. Previously Molecular structure was determination using representation of atomization energies, prioritization and interpretation of geometric searches the tactical spectrum. Attributes of CME representations and their relationships with molecular structures are determined using Hershgorin circles shelf. **[4].** EN, KNN and SchNet were first tested at QM9 and then at PM9. All the model achieves lower accuracy on the PC9 data than on the QM9 data.

**1.5.** Applying machine learning techniques to predict the properties of energetic materials

In this research work a comparison of molecular featurization methods and ML models is demonstrated with the goal to create a network model that predicts energy properties.Featurization  is the process of digital transformation of data in numeric forms which can be used by basic ML algorithms**[5].** Some featurization techniques are given below:

* Custom descriptor set
* Sum over bonds
* Coulomb matrices
* Bag of bonds
* Finger printing
* Atom-Pair
* Topologoical Torsion
* Extended Connectivity Fingerprints (ECFPs)
* E-state fingerprints
* Avalon fingerprints
* RDKit graph fingerprints
* ErG fingerprints
* physiochemical property
* fingerprints.
* Fourier series atomic radial distribution functions