UW DIRECT Capstone Project: Use Cases

Group 1 (Christine Chang, Chih-Wei Hsu, Liang Xu)

April 8, 2019

Project Summary

The goal of our project is to explore the potential use of a PNNL-developed software package, DarkChem, to relate chemical transformations to observed physical properties. DarkChem uses a variational autoencoder (VAE) to encode molecular structure and properties in latent (numerical vector) space; the reverse transformation (decoding) allows for the retrieval of structures that match desired chemical properties. In our capstone project, we will explore how two molecules in this latent space, related through chemical reactions, are related. Understanding and identifying this relationship enables the application of DarkChem towards the prediction of unknown products in a chemical reaction, as well as their molecular properties.

Use Cases and Components

Use Case #1: Predict unknown product(s) in a given reaction, as well as the properties of the product(s).

- *Input(s):* Molecular structure of reactant(s) (SMILES), type or class of reaction (e.g. oxidation, reduction, etc.)
- *Output(s):* Molecular structure of product(s) (SMILES), properties of product(s) (e.g. m/z, CCS)
- *Components:* Input structure, convert structure to latent space representation, perform reaction vector transformation on structure in latent space, use DarkChem to transform latent space representation into structure and property values

Use Case #2: Predict the structure of a small molecule given desired molecular properties.

- *Input(s):* Desired properties (e.g. m/z, CCS)
- *Output(s):* Molecular structure of product(s) matching the specified properties (SMILES, or visual representation)
- *Components:* Input properties, convert properties into latent space representation using DarkChem, decode latent space representation to retrieve structures, convert latent space structure representation to SMILES or other visualization

Use Case #3: Predict properties for a new or unknown molecule given its structure.

- *Input(s):* Molecular structure of the product (as a SMILES string)
- *Output(s):* Molecular properties predicted for the structure (e.g. m/z, CCS)
- *Components:* Input structure, convert structure to latent space representation, use DarkChem to relate latent space representation to properties, output properties

Use Case #4: Predict a possible reaction pathway given reactants and properties of desired products; in other words, predict potential products given reactants without a specified reaction. (stretch/"would be cool" goal)

- *Input(s):* Molecular structures for reactants (as SMILES strings)
- *Output(s):* Molecular properties predicted for the structure (e.g. m/z, CCS)
- *Components:* Input structure(s) and/or properties, convert structure(s) to latent space representations, use DarkChem to search latent space for potential transformations matching inputs, output possible product structures