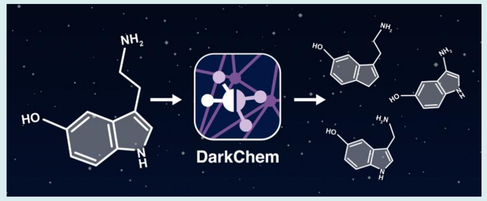
Generating Novel Molecules Using Latent Dimension Reaction Manifolds from Deep Learning

Christine Chang[1], Chih-Wei Hsu[2], Liang Xu[1], Ryan Renslow[3], Sean Colby[3]

[1]Department of Materials Science and Engineering, [2]Department of Chemical Engineering; University of Washington, Seattle, WA
[3]Pacific Northwest National Laboratory, Richland, WA

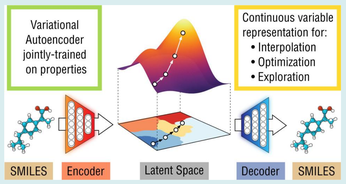
Background



DarkChem is a variational autoencoder (VAE) built at the Pacific Northwest National Laboratory (PNNL).

FEATURES

- Chemical properties prediction, e.g. mass-charge ratio (m/z) and collision cross-section (CCS)
- High reconstruction accuracy (~99%)



VAE Image: Gómez-Bombarelli, R.; Wei, J.N.; Duvenaud, D.; Hernández-Lobato, J.M.; Sheberla, D.; Aspuru-Guzik, A. et al. ACS Cent. Sci. 2018

DarKnight

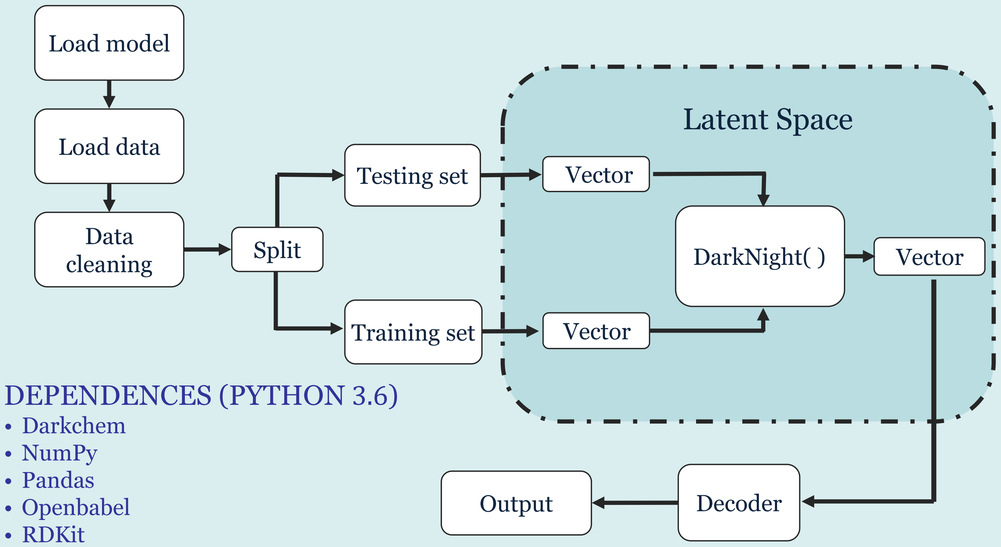
GOALS

- Determine how chemical transformations are represented in latent space
- Use latent space to predict products and product properties of chemical reactions

USE CASES

1. Predict unknown product(s) in a reaction, as well as the properties of the product(s).
   - Input(s): Reactant , Reaction
   - Output(s): Product, Product Properties
2. Predict the structure of a small molecule given desired molecular properties.
   - Input(s): Desired Properties
   - Output(s): Product matching specified properties
3. Predict properties for a new or unknown molecule given its structure.
   - Input(s): Product
   - Output(s): Predicted Properties

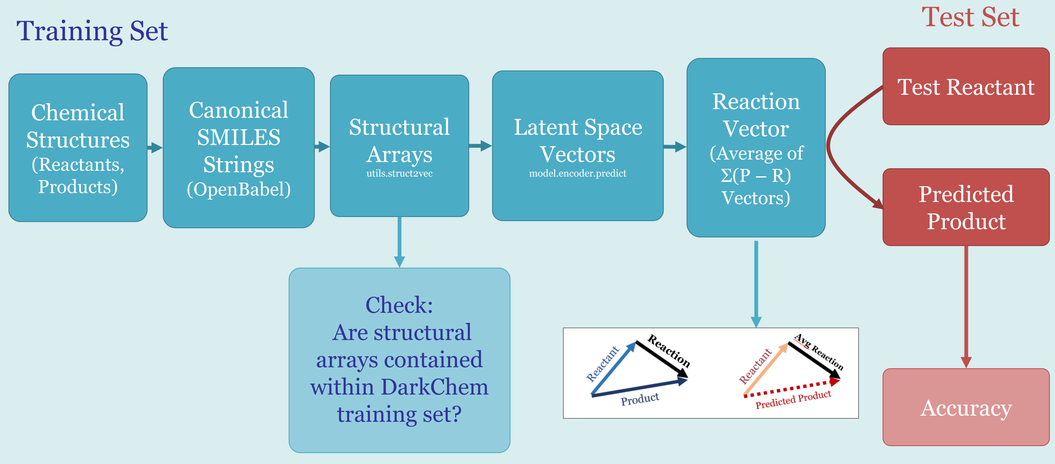
Package Flowchart



DEPENDENCES (PYTHON 3.6)

- Darkchem
- NumPy
- Pandas
- Openbabel
- RDKit
- Keras (TensorFlow backend)

Assessment of DarKnight Chemical Reactions Prediction



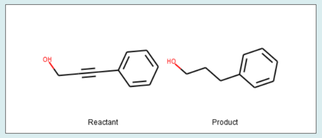
Acknowledgments

The authors acknowledge the Renslow group at the Pacific Northwest National Laboratory, including Jamie Nuñez, as well as University of Washington DIRECT program, especially Dr. David Beck and Theodore Cohen, for their support. Only open-source packages were used in this work.

Results

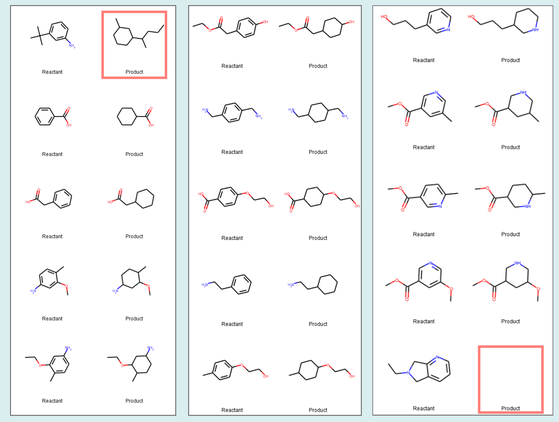
Triple Bond Reduction

- Training set: 10 reactions
- Testing set: 2 reactions



Aromatic Hydrocarbon Reduction

- Training set: 46 reactions
- Testing set: 10 reactions



Conclusions and Future Work

- Successful prediction for certain reactions (up to 90% accuracy)
- Some errors persist in predictions (e.g. extra C atom issue)

FUTURE WORK

- Larger testing sets (roadblock: must be manually generated)
- Classification of reactants to improve accuracy
- Explore and predict more complicated reactions

References

S.M. Colby, J.R. Nuñez, N. O. Hodas, C. D. Corley, R. R. Renslow. Deep learning to generate in silico chemical property libraries and candidate molecules for small molecule identification in complex samples.