ChemSpacE: Toward Steerable and Interpretable Chemical Space Exploration

Yuanqi Du[1] Xian Liu[2] Shengchao Liu[3] Jieyu Zhang[4] Bolei Zhou[2,✉]

[1]George Mason University [2]CUHK [3]MILA [4]University of Washington [✉]Correspondence

Overview

In this work, we present **ChemSpacE**, a model-agnostic method that explores the learned latent space of molecule generative models. More specicially, **ChemSpacE** aligns **chemical space** and **molecular property space** with learned **latent space** of molecule generative models and observing the manipulation path of molecules demonstrate the relationship between molecular structures and properties.

Background

Molecule Manipulation

- **Chemical Space**
  - Chemical space refers to a space where all possible molecules and chemical compounds span and are distinguished via different molecular structures.
- **Latent Space**
  - Latent space refers to a learned meaningful representation by generative models.
- **Molecular Property Space**
  - Molecular property space refers to a space where all possible molecules and chemical compounds span and distinguished via different molecular properties.

Current Issues

- **Lack of interpretability**
  - Less effort has been paid to interpret the learned reprepresentations of the molecule generative models.
- **Lack of controllability/steerability**
  - Existing methods only tackle molecule optimization/searching problems which cannot smoothly and monotonically manipulate the properties of molecules.

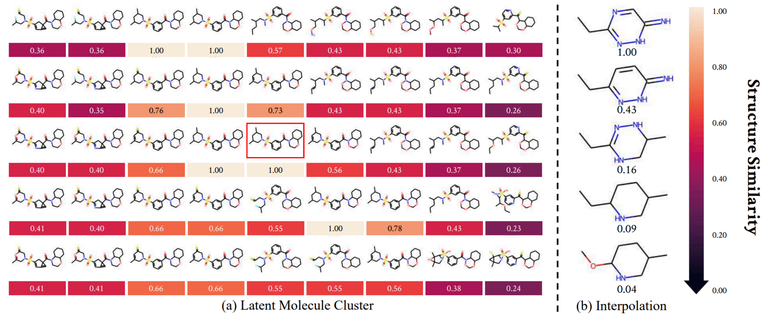
Observations/Assumptions

- **Latent Cluster Assumption**
  - Molecules with similar structures tend to cluster in the latent space.
  - Interpolating two molecules in the latent space lead to a list of smoothly-changing molecular structures.
- **Latent Separation Boundary**
  - There exists a separation boundary which separates groups of molecules (e.g. drug-like and drug-unlike) and the normal vector of the separation boundary defines a latent direction that controls the molecular property.

Chemical Space Explorer (ChemSpacE)

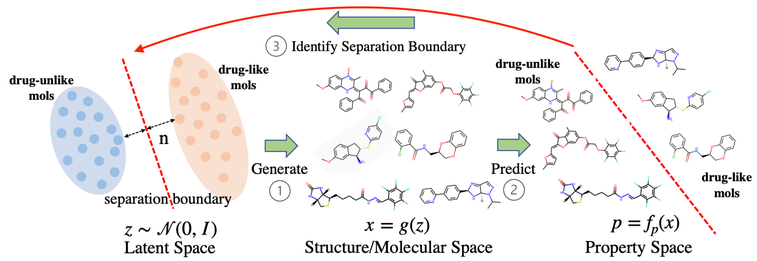
- **A new molecule manipulation task which measures the interpretability and steerability of moleucle generative models.**
- **A simple yet effective model-agnostic method ChemSpace Explorer for efficient molecule manipulation.**
- **Comprehensive experiments for quantifying interpretability and steerability in molecule generative models are conducted and an interactive demo is released.**

Observations/Assumptions



(a) Latent Molecule Cluster (b) Interpolation

ChemSpacE Framework



Quantitative and Qualitative Evaluations of Molecule Manipulation.

Quantitative Evaluation w/ Manipulation Success Rate.

- Datasets: QM9, ZINC, ChEMBL.
- Baselines: Random Manipulation, Largest Manipulation.
- Molecule Generative Models: MoFlow, HierVAE.
- Evaluation Metrics: Strict Success Rate, Relaxed Success Rate.

Table 1: Quantitative Evaluation of Molecule Manipulation over a variety of molecular properties (numbers reported are *strict success rate* in %, -R denotes model with random manipulation, -L denotes model with the largest range manipulation, -O denotes optimization-based manipulation, -C denotes model with ChemSpacE. The best performances are bold.

Datasets	Models	Avg.	QED	LogP	SA	DRD2	JNK3	GSK3B	MolWt
QM9	MoFlow-R	1.65	1.50	0.00	0.50	0.00	0.00	0.00	0.50
	MoFlow-L	3.43	1.50	1.00	0.50	0.00	1.50	0.00	0.50
	MoFlow-O	N/A	3.50	6.00	6.50	2.00	8.00	8.50	7.50
	MoFlow-C	**37.52**	**12.50**	**9.00**	**10.00**	**11.00**	**45.50**	**16.50**	**10.50**
	HierVAE-R	29.29	1.00	1.50	0.50	0.50	1.00	1.00	0.50
	HierVAE-L	30.69	0.50	0.00	0.00	0.50	2.00	0.00	0.50
	HierVAE-C	**66.23**	**27.00**	**32.00**	**35.00**	**41.50**	**51.50**	**30.00**	**39.50**
ZINC	MoFlow-R	4.25	1.50	1.50	2.50	3.00	3.50	1.50	2.00
	MoFlow-L	5.61	1.50	6.50	2.00	6.00	2.50	4.00	1.50
	MoFlow-O	N/A	1.50	9.50	0.50	2.00	15.50	23.00	0.00
	MoFlow-C	**58.08**	**52.00**	**53.50**	**51.50**	**55.00**	**56.50**	**55.50**	**53.50**
ChEMBL	HierVAE-R	25.59	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	HierVAE-L	22.98	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	HierVAE-C	**47.70**	**0.50**	**3.00**	**3.00**	**6.00**	**7.50**	**5.50**	**4.50**

Qualitative Evaluation w/ Manipulation Path Visualization.

- Molecular Properties: QED, Molecular Weight, LogP.
- Molecule Generative Models: MoFlow, HierVAE.