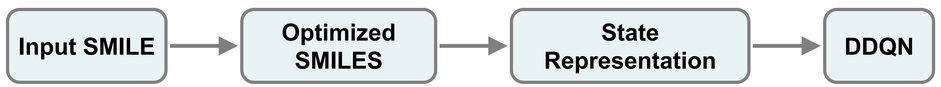
Enhancing Molecular Design through Graph-based Topological Reinforcement Learning

Xiangyu Zhang
Department of Applied Mathematics and Statistics, Johns Hopkins University

Overview

The generation of drug-like molecules is essential for drug design, requiring a balance between exploring novel chemical structures and optimizing known ones to meet **specific desirable properties** (Olivecrona et al., 2017). This work,

- Leverages the sequential decision-making capabilities (Popova et al., 2018) of **Reinforcement Learning (RL)** by formulating the modification of a molecule as a Markov Decision Process.
- Integrates both chemical and structural data of molecules
  - Multiscale Weighted Colored Graphs (MWCG)
  - Persistent Homology



Reward Function

The reward functions are tailored to specific optimization tasks, incorporating various molecular properties and constraints.

Constrained Reward Function

The constrained reward function aims to improve the Penalized $logP$ while maintaining the similarity within a specified range.

$$R_1(m) = Penalized\ logP(m) - \lambda \cdot [\mathbb{I}\{S(m,m_0) < \delta\}(\delta - S(m,m_0)) + \mathbb{I}\{\mathcal{B}(m,m_0) < \mathcal{E}\}(\mathcal{E} - \mathcal{B}(m,m_0)]$$

Target Reward Function

For tasks involving the design of new molecules with specific molecular weight or Betti number, we define the reward function as:

$$R_2(m) = (1-\omega) \cdot \mathcal{B}(m) + \omega \cdot QED(m)$$

Experiment and Result

We constructed the initial state of the model by combining one of the 800 randomly selected molecules with the lowest Penalized $logP$ value and its corresponding MWCG graph feature and persistent image vector.

Method	Penalized log P				QED			
	1st	2nd	3rd	Validity	1st	2nd	3rd	Validity
random walk	-3.99	-4.31	-4.37	100%	0.64	0.56	0.56	100%
ε-greedy, $\varepsilon = 0.1$	11.64	11.40	11.40	100%	0.914	0.910	0.906	100%
JT-VAE	5.30	4.93	4.49	100%	0.925	0.911	0.910	100%
ORGAN	3.63	3.49	3.44	0.4%	0.896	0.824	0.820	2.2%
GCPN	7.98	7.85	7.80	100%	0.948	0.947	0.946	100%
MolDQN-naïve	11.51	11.51	11.50	100%	0.934	0.931	0.930	100%
MolDQN-bootstrap	11.84	11.84	11.82	100%	0.948	0.944	0.930	100%
MolDQN-twosteps	-	-	-	-	0.948	0.948	0.948	100%
GraphTRL	11.89	11.87	11.86	100%	0.951	0.949	0.943	100%

Materials and Methods

Multiscale Weighted Colored Subgraphs (MWGCs)

Graph theory, with its focus on the connectivity between vertices and edges, provides an ideal framework for representing non-covalent interactions between atoms in molecules. The choice of MWCG has two key advantages:

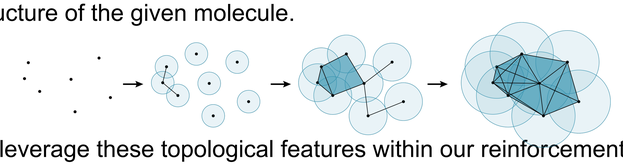
- MWCG employs color-labeled atoms to create distinct subgraphs and utilizes colored edges to represent element-specific interactions, which allows for a detailed capture of interactions at the atomic level.
- MWCG incorporates a radial basis function to scale Euclidean distances, assigning the strongest weights to edges between nearest neighbors (Rana, M. M. et al., 2023). This weighting scheme enhances the representation of local atomic environments.

$$\mu^G(\eta_{kk'}) = \sum_i \mu_i^G(\eta_{kk'}) = \sum_i \sum_j \Phi_E(\| r_i - r_j \|; \eta_{kk'})$$

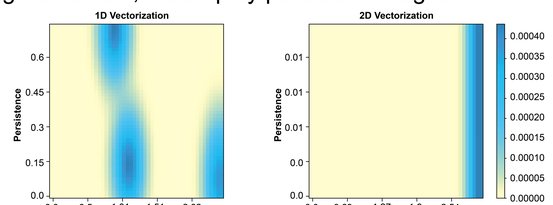
$$\varepsilon_{kk'} = \{\Phi(\| r_i - r_j \|; \eta_{kk'}) \mid \alpha_i = \tau_i, \alpha_j = \tau_j, \| r_i - r_j \| \le c\}$$

Persistent Homology

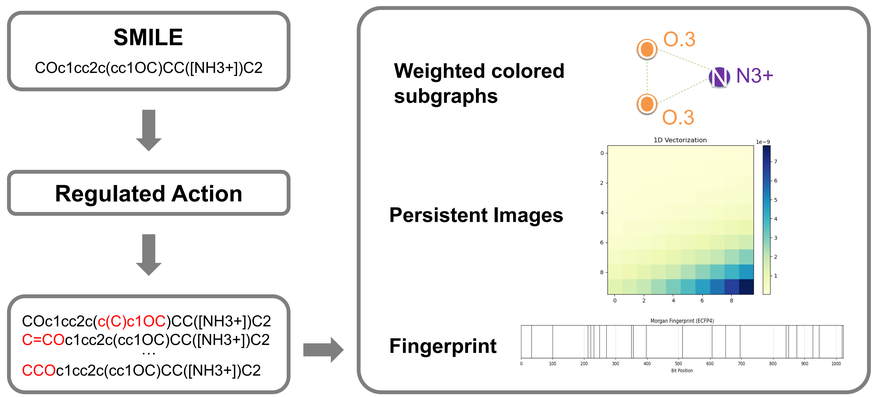
Persistent homology is a method that utilizes the filtration process to systematically reset the connectivity of a dataset according to a scale parameter, generating a series of topological spaces in various scales. Through filtration process, we can identify the intrinsic structure of the given molecule.



To leverage these topological features within our reinforcement learning framework, we employ persistent images.

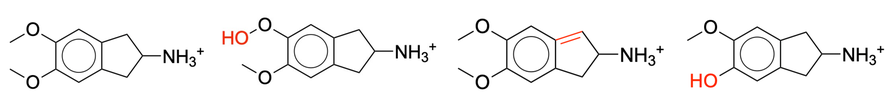


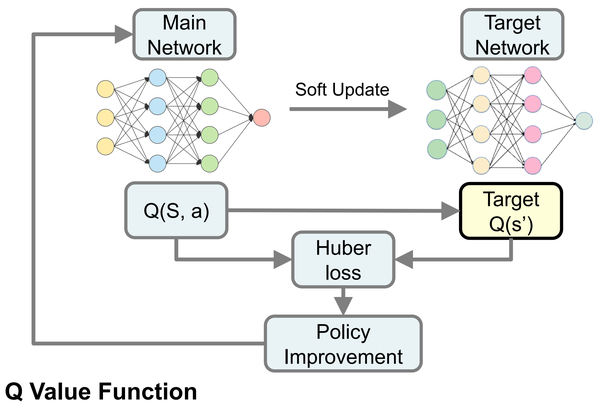
State Construction and Model Structure



For each SMILE, the Markov decision process (MDP) is denoted as $MDP(\mathcal{S}, \mathcal{A}, P_{sa}, \mathcal{R})$.

Actions:





Q Value Function

The dueling network structure has two streams to separate the contribution of state value V and the advantage A for each action.

$$Q(S_i, a_i) = V(s_i) + A(s_i, a_i)$$

Reference:

- Popova, M., Isayev, O., & Tropsha, A. (2018). Deep reinforcement learning for de novo drug design. *Science advances*, *4*(7), eaap7885.
- Olivecrona, M., Blaschke, T., Engkvist, O., & Chen, H. (2017). Molecular de-novo design through deep reinforcement learning. Journal of cheminformatics, 9, 1-14
- Rana, M. M., & Nguyen, D. D. (2023). Geometric graph learning with extended atom-types features for protein-ligand binding affinity prediction. Computers in Biology and Medicine, 164, 107250.