

Optimization of molecules via deep reinforcement learning

Björn Bebensee (2019–21343)
Topics in Artificial Intelligence

November 28, 2019

Zhou et al. [1] apply deep reinforcement learning techniques to the space of molecule optimization. In molecule design and optimization, the main objective is to discover new compounds with desired properties. This is especially relevant to drug design. As this process is typically very expensive and takes a long time and an automation of it would be desirable, there have been several approaches to e.g. build a generative model which maps molecules to a high-dimensional space in which it performs its optimization or search operations. There has also been prior work using reinforcement learning but these models either struggled with chemical validity or required pre-training on a dataset which may lead to a bias in the resulting model.

The authors introduce a model called Molecule Deep Q -Networks (MolDQN). They can ensure that their model has 100% chemical validity by modeling the modification of a molecule as a Markov decision process (MDP) with $\text{MDP}(\mathcal{S}, \mathcal{A}, \{P_{sa}\}, \mathcal{R})$ where \mathcal{S} is the state space of states (m, t) for a molecule m with t steps taken (t is limited as the search space is vast), \mathcal{A} is the action space (only actions are atom addition, bond addition, bond removal), $\{P_{sa}\}$ describes the transition probabilities where for a given action the result is deterministic (i.e. state is reached with probability 1) and \mathcal{R} gives the reward of a given state. MolDQN is then used to solve this MDP by fitting a function $Q(s, a)$ that learns to predict the future reward by taking action a from state s . They implement multi-objective rewards, that is the total reward is the cumulative reward of single-objective rewards. This is useful as a new drug may have more than one desirable property. The authors decide to use a value function learning instead of policy gradient methods as they find it to be more stable and to have a lower variance. They implement the deep Q -learning model and approximate the Q -function using a neural network.

Zhou et al. evaluate their MolDQN model on logP and QED measures against several established baseline approaches and perform equivalently or better than all of them. However, they note that they believe these metrics to be broken and not representative of actual performance. They stress that there is a strong need for a better metric in this field that is able to compare models on a meaningful task.

References

- [1] Zhou, Zhenpeng, et al. "Optimization of molecules via deep reinforcement learning." *Scientific reports* 9.1 (2019): 1-10.