Modular soft actor-critic framework for synthetic pathway generation as an approach to goal-driven de novo drug design

Aaron Tian[1] - aztian@wpi.edu

Instructor: Dr. Kevin Crowthers, Ph.D.

In the midst of a global pandemic, it is imperative that we develop robust computational models to assist in drug design in order to mitigate the inherently slow turnaround time of the traditional discovery process for small molecule drugs. The current state of computational drug design may be put in a wider context if generative models are designed to account for chemical synthesis, allowing for a streamlined transition between the discovery and clinical testing stages.

Objectives

The objective of this project is to create a goal-driven reinforcement learning model for synthesizable optimization of drug compounds via synthetic tree modification. The decomposition of sub-criteria is shown below:

Table 1. engineering design matrix

Criteria	Quantification	Weight
synthesizability	implicit	10
property optimization	t-test	10
convergence speed	computation time (s)	5
computational cost	# model parameters	6
scalability with train data	empirical	7
scalability with predictive model*	regression analysis	8
pharmacokinetic property optimization*	t-test	10

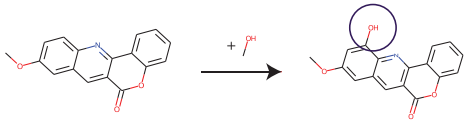
Background

Previous deep learning-based models were capable of optimizing molecular properties but failed to generate synthesizable compounds.

30% Synthesizable

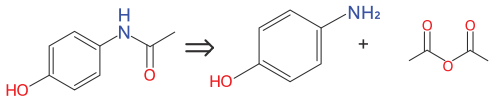
Only 30% of molecules generated by current predictive models are synthetically accessible (Gao and Coley, 2020).

This drawback can be explained by the underlying process used to generate molecules.

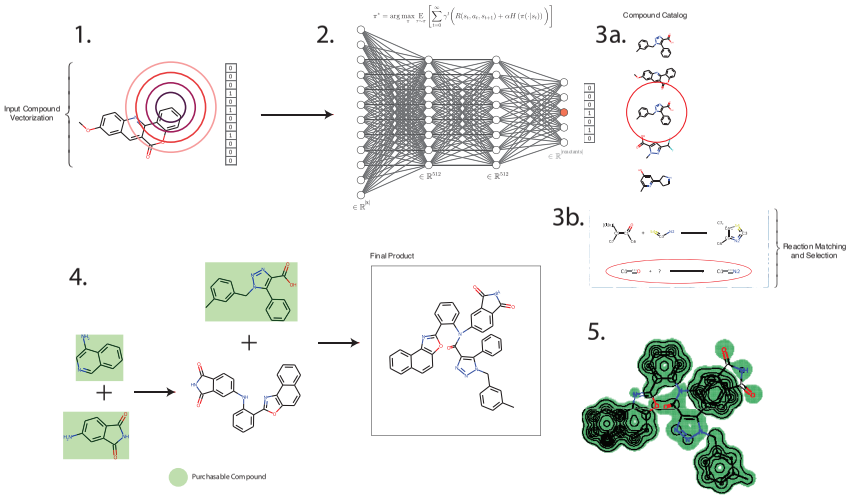


Zhou et al. (2018) utilize this approach alongsize a Deep Q-Network (DQN) to achieve state-of-the-art results in molecular property optimization. However, it was soon discovered that the generated molecules were largely unusable in a practical setting; after modifying the lead compound on the atomic scale, the product became unsynthesizable.

For a molecule to proceed to clinical trial, it must be *synthesizable*, meaning it can be built using commercially available starting compounds. We can verify this using retrosynthetic analysis, or chemical 'deconstruction.'



Graphical Abstract



Methods

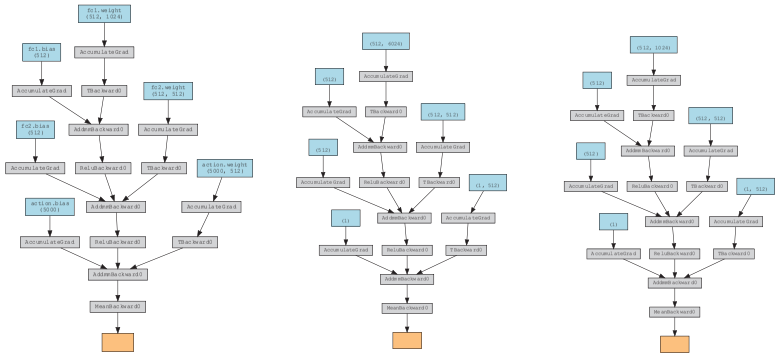


Figure 3. (a) Execution graph of the actor network. (b) Execution graph of the critic networks. (c) Execution graph of the value network.

A custom environment was built to simulate the process of reacting molecules.
A soft actor-critic model was optimized as the central generation algorithm. SAC was used for its optimization of an entropy parameter, allowing for fast Convergence alongside its off-policy nature.
The original SAC model was modified to be compatible with a discrete action space for molecule selection.

Use of Gumbel Softmax for differentiable discrete action space.

$$\frac{e^{\left(\frac{x_i}{\lambda}\right)}}{\sum_{i=1}^{n} e^{\left(\frac{x_i}{\lambda}\right)}}$$

Results

Several druglike compounds were generated by the model. The molecule and corresponding synthesis shown below has a computed QED of 0.934, a highly competitive result.

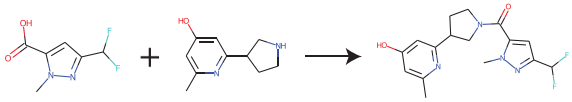


Figure 4. Best product proposed by model (QED = 0.934). The product can be traced to purchasable starting compounds.

Analysis

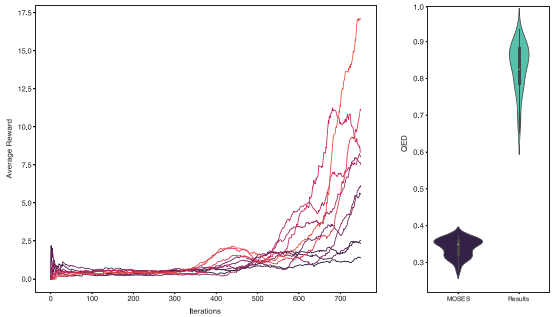


Figure 2. (a) Model performance on 10 runs with randomized reaction order. The reactant set and all model hyperparameters were kept constant. Initial model weights were loaded from a random initialization. (b) Distribution of QED across MOSES (initial lead compound set) and corresponding top-1 molecules produced after optimization.

Discussion

Figure 2a shows that the performance of the model is dependent on the order in which the reactions are passed. This suggests high sensitivity to input data.

Figure 2b demonstrates that the proposed netowork is highly capable of optimizing the properties of lead compounds for a single target. An unpaired t-test was performed on the data and yielded a p-value < 0.001.

Conclusion

The model proposed in this project shows promizing scalability with repsect to the quality of the input data, suggesting that its performance can be further improved by engineering curated datasets for the purpose of lead optimization. The following outline promizing areas of future research:

- Multi-discrete action space for simultaneous prediction of reactants and reactions
- Thorough tuning of reward signal weights and model hyperparameters
- Incorporation of predictive models in the reward signal
- Development of predictive models for prediction of pharmacokinetic properties