MANDREL: Modular Reinforcement Learning Pipelines for Material Discovery

Clyde Fare[1]*, George K. Holt[2], Lamogha Chiazor[1], Michalis Smyrnakis[2], Robert Tracey[1], Lan Hoang[1]*

[1] IBM Research UKI, United Kingdom
[2] STFC Hartree Centre, United Kingdom
clyde.fare@ibm.com, george.holt@stfc.ac.uk, lamogha.chiazor@ibm.com, michail.smyrnakis@stfc.ac.uk, robert.tracey@ibm.com, lan.hoang@ibm.com

Abstract

AI-driven materials discovery is evolving rapidly with new approaches and pipelines for experimentation and design. However, the pipelines are often designed in isolation. We introduce a modular reinforcement learning framework for inter-operable experimentation and design of tailored, novel molecular species. The framework unifies reinforcement learning (RL) pipelines and allows the mixing and matching of choices for the underlying chemical action space, molecular representation, desired molecular properties, and RL algorithm. Our demo showcases the framework's capabilities applied to benchmark problems like quantitative estimate of drug-likeness and PLogP, as well as the design of novel small molecule solvents for carbon capture.

Introduction

The advancement of chemical research for applications such as drug and materials discovery hinges upon the generation of novel, performant molecular species. Traditionally this process is challenging since it requires significant domain knowledge and extensive experiments. This makes the process expensive and time consuming. Deep learning can reduce the effective cost of evaluating a potential material via providing accurate experiment proxies; however it requires large amounts of annotated data (Hu 2021). Reinforcement learning (RL) is a sequential, adaptive solution to this discovery problem. It offers a targeted, flexible approach to molecular design that is more general and widely applicable than other optimisation techniques.

Unfortunately, the field suffers from fragmented pipelines and frameworks. Research usually relies on integrated, monolithic pipelines that combine a specific implementation of a molecular action space, representation and reward function. It is therefore difficult to compare the performance across different pipelines or leverage the advances present within one pipeline for another.

To address the problem of fragmented pipelines for benchmarking RL for material discovery, we present MANDREL (MAterial aNd Discovery using REinforcement Learning): an integrative, modular Python framework that allows researchers to explore the different training features of an RL pipeline. It allows switching across (i) different chemical representations and formulation of the discovery tasks, (ii) RL algorithms, and (iii) target properties for rewards, ranging from proxies to values calculated from simulation models, thus accelerating the discovery of small, generic but novel molecular species with desired properties.

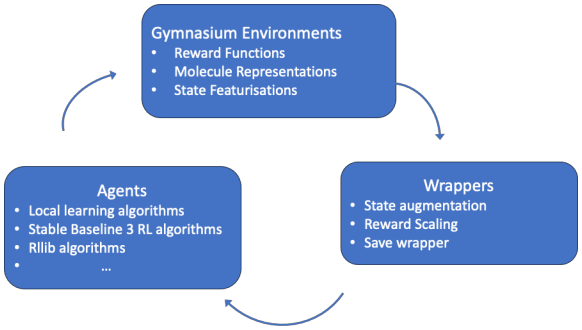

Figure 1: Schematic representation of the MANDREL molecular RL toolkit

Framework Components

Within MANDREL, the generation of molecules possessing particular properties is described by a Markov decision process (MDP). As shown in Figure 1, the reward function, along with functions to featurise the state and/or full molecular graph, are treated as inputs to a Gymnasium environment (Towers et al. 2023), which is defined for a particular choice of action space. The combination of these three components (reward function, action space and featuriser) allow for the composition of an MDP. The final component is the choice of behavioral policy, which governs how the agent interacts with the molecular design environment, with the goal of learning how to design performant molecules.

Our contribution can be summarised as follows: MANDREL presents a standardised form of the material discovery problem. These environments span several different popular means for traversing chemical space: SMILES (Weininger 1988), SELFIES (Krenn et al. 2022), STONED-SELFIES (Nigam et al. 2021) and the action space introduced by Zhou et al. (2019). The modular character of the framework means reward functions representing different

*corresponding authors

experiments or simulation proxies can be easily plugged into and combined with constraints on the molecules generated based on e.g., synthesisability heuristics (Gao and Coley 2020). Different choices for featurising the growing molecules can be swapped in and out. Finally, as our framework is based on the latest Gymnasium version, it is compatible out-of-the-box with current and future state-of-the-art RL algorithms present in popular libraries such as Stable-Baselines3 (SB3) and Ray RLlib, without the need to create a tailored implementation of each algorithm.

Environment Module

Four types of environment with distinct action spaces are currently implemented:

1. SMILES Gym: Uses SMILES characters to construct the molecule, with each action being the addition of a SMILES character.
2. SELFIES Gym: Uses SELFIES characters to construct the molecule, with each action being the addition of a SELFIES character.
3. STONED-SELFIES Gym: STONED-SELFIES is an approach based on SELFIES, but rather than each action consisting of an additional character, string mutation operators are applied instead to generate new possible SELFIE strings. These strings are filtered based on a distance measure to the previous step, and the action is then the choice of which of the screened mutant strings to accept.
4. The environment introduced in Zhou et al. (2019) can be viewed as a precursor to the STONED-SELFIES action space. This environment uses a set of heuristic rules to define valid mutations of a SMILES string, and the action again selects which of these mutations to accept.

Each environment allows the user to traverse different versions of the chemical space they define by adjustment of suitable hyperparameters. For example, the choice of character primitives for SMILES/SELFIES environments, the rules for adding bonds to systems containing rings for the environment based on Zhou et al. (2019), and the degree of locality imposed on a single step for the STONED-SELFIES environment.

Agent Component

Our framework is compatible with the popular libraries SB3 and Ray RLlib. We also provide an implementation of the Masked Deep Q-Network (DQN) algorithm used in Zhou et al. (2019).

Reward Component

MANDREL currently includes five different illustrative reward functions:

- LogP: A measure of solubility (i.e., water-octanol partition coefficient) (Guimaraes et al. 2017).
- PLogP: Penalized LogP objective (Nigam et al. 2019).
- QED: Quantitative estimate of drug-likeness (Bickerton et al. 2012).
- CC-Capacity: Carbon capture absorption capacity metric for solvent-based capture of carbon dioxide (Van Kessel et al. 2023). This uses a surrogate model based on AdaBoost with 75 estimators (Hastie et al. 2009).
- CC-Rate: Carbon capture rate metric for solvent-based capture of carbon dioxide (Van Kessel et al. 2023). This uses a surrogate model based on a Gaussian process (Williams and Rasmussen 2006).

Featurisation Component

The featurisation of a growing molecule is divided into two categories:

A) Molecular featurisation: A universal routine across all environments, this takes the form of a function that transforms a SMILES string into a vector representation. Currently, this is implemented as a Morgan fingerprint (Morgan 1965) but any molecular graph featurisation algorithm can be plugged in.

B) State-specific featurisation: Optional and specific to the active environment, this one-hot chartecter encoding featurisation is used to provide additional featurisation where either a complete molecule is not present at each step (e.g., this is the case in the SMILES environment) or the complete molecule is not sufficient to fully characterise the state at each step (e.g., as occurs in the SELFIES environment).

Interface

The framework is showcased within a Dash application that allows users to mix and match choices of the molecular design task, action space and algorithm; observe the resulting training curves (along with the hyperparameters that were used); and interact with the discovered molecules. An overview of this interface can be found in the video submission.

Discussion

MANDREL serves as a platform for direct molecular RL over explicit graph representations of molecular space. It provides a modular, flexible, and user-friendly framework for researchers in the field. The platform offers insights across the experiments, as illustrated by the examples in this demo. For example, we recover the QED results presented in Zhou et al. (2019) within MANDREL and find that using their masked DQN variant shows marked performance improvements over both the vanilla DQN and masked PPO algorithms implemented within SB3 (matching hyperparameters as closely as possible). Exploring the effect of switching to use the SELFIES environment yields a stable agent with reduced reward variance but that takes a greater number of steps to converge. Meanwhile the STONED-SELFIES environment exhibits similar behaviour to that in Zhou et al. (2019).

Acknowledgments

This work was supported by the Hartree National Centre for Digital Innovation, a collaboration between STFC and IBM.

References

Bickerton, G. R.; Paolini, G. V.; Besnard, J.; Muresan, S.; and Hopkins, A. L. 2012. Quantifying the chemical beauty of drugs. *Nature chemistry*, 4(2): 90–98.

Gao, W.; and Coley, C. W. 2020. The Synthesizability of Molecules Proposed by Generative Models. *Journal of Chemical Information and Modeling*, 60(12): 5714–5723. PMID: 32250616.

Guimaraes, G. L.; Sanchez-Lengeling, B.; Outeiral, C.; Farias, P. L. C.; and Aspuru-Guzik, A. 2017. Objective-reinforced generative adversarial networks (organ) for sequence generation models. *arXiv preprint arXiv:1705.10843*.

Hastie, T.; Rosset, S.; Zhu, J.; and Zou, H. 2009. Multi-class adaboost. *Statistics and its Interface*, 2(3): 349–360.

Hu, W. 2021. Reinforcement learning of molecule optimization with bayesian neural networks. *Computational Molecular Bioscience*, 11(4): 69–83.

Krenn, M.; Ai, Q.; Barthel, S.; Carson, N.; Frei, A.; Frey, N. C.; Friederich, P.; Gaudin, T.; Gayle, A. A.; Jablonka, K. M.; Lameiro, R. F.; Lemm, D.; Lo, A.; Moosavi, S. M.; Nápoles-Duarte, J. M.; Nigam, A.; Pollice, R.; Rajan, K.; Schatzschneider, U.; Schwaller, P.; Skreta, M.; Smit, B.; Strieth-Kalthoff, F.; Sun, C.; Tom, G.; Falk von Rudorff, G.; Wang, A.; White, A. D.; Young, A.; Yu, R.; and Aspuru-Guzik, A. 2022. SELFIES and the future of molecular string representations. *Patterns*, 3(10): 100588.

Morgan, H. L. 1965. The Generation of a Unique Machine Description for Chemical Structures-A Technique Developed at Chemical Abstracts Service. *Journal of Chemical Documentation*, 5(2): 107–113.

Nigam, A.; Friederich, P.; Krenn, M.; and Aspuru-Guzik, A. 2019. Augmenting genetic algorithms with deep neural networks for exploring the chemical space. *arXiv preprint arXiv:1909.11655*.

Nigam, A.; Pollice, R.; Krenn, M.; dos Passos Gomes, G.; and Aspuru-Guzik, A. 2021. Beyond generative models: superfast traversal, optimization, novelty, exploration and discovery (STONED) algorithm for molecules using SELFIES. *Chemical science*, 12(20): 7079–7090.

Towers, M.; Terry, J. K.; Kwiatkowski, A.; Balis, J. U.; Cola, G. d.; Deleu, T.; Goulão, M.; Kallinteris, A.; KG, A.; Krimmel, M.; Perez-Vicente, R.; Pierré, A.; Schulhoff, S.; Tai, J. J.; Shen, A. T. J.; and Younis, O. G. 2023. Gymnasium. *Zenodo*.

Van Kessel, T.; Cipcigan, F.; Mcdonagh, J.; Wunsch, B.; Elmegreen, B.; Gifford, S.; and Zavitsanou, S. 2023. Machine Guided Discovery of Novel Carbon Capture Solvents. In *American Chemical Society (ACS) Fall Meeting*.

Weininger, D. 1988. SMILES, a chemical language and information system. 1. Introduction to methodology and encoding rules. *Journal of Chemical Information and Computer Sciences*, 28(1): 31–36.

Williams, C. K.; and Rasmussen, C. E. 2006. *Gaussian processes for machine learning*, volume 2. MIT press Cambridge, MA.

Zhou, Z.; Kearnes, S.; Li, L.; Zare, R. N.; and Riley, P. 2019. Optimization of molecules via deep reinforcement learning. *Scientific reports*, 9(1): 10752.