

非専門家でも手軽に使えるデータ駆動型深層強化学習ライブラリの開発

— d3rlpy : 世界で最も実用的な深層強化学習ライブラリ —

1. 背景

深層強化学習は深層学習と強化学習を組み合わせた機械学習手法であり、対象のシステムと試行錯誤的にインタラクションを行いながらオンラインで最適化を行うことが可能である。特に、囲碁AIのAlphaGoに代表されるような、人手でアルゴリズムを設計することが不可能であったり、データで入出力を定義することができないような問題を扱うことができるため、機械学習分野で非常に研究が活発である。

しかし、深層強化学習はアカデミアや産業界での注目度の高さに対して、実際のプロダクトや応用事例が非常に少なく、原因として実問題への応用の難しさと実装難易度の高さが考えられる。

まず、実問題への応用を困難にしている要因として、オンラインで学習できる問題設定が非常に限られることが挙げられる。深層強化学習では大量の試行錯誤が必要のため、強化学習エージェントが環境と長時間インタラクションを行う。一方で、応用先として注目されているロボット制御などの分野ではバッテリーやアクチュエータの制約があるため、前述したような長時間の試行錯誤を行いながら学習を行うことができない。

次に、深層強化学習アルゴリズムの実装が困難である要因として、非研究者向けに設計されたライブラリが存在しないことが挙げられる。深層強化学習をスクラッチで実装する場合、実装者には深層学習と強化学習の知識に加えて、論文の数式を解釈して実装に落とし込む能力が要求される。また、すでに多くの深層強化学習ライブラリが公開されているが、全てが機械学習の専門家・研究者向けに設計されているため、機械学習に精通していない利用者が使いこなすのが非常に難しい。

2. 目的

本プロジェクトでは、強化学習の応用の難しさと実装難易度の高さの両方を解決するようなライブラリ・ソフトウェアを開発することで、より多くの人が強化学習を使いこなせる世界を実現することを目的とした。

具体的には、実問題への応用の難しさに対して、近年注目が高まっているデータ駆動型強化学習のサポートを行うこととした。データ駆動型強化学習とは事前に集められたデータセットのみを用いてオフラインで学習を行うアプローチで、今まで強化学習を適用できなかった問題を扱うことができる。また、実装難易度の高さに対しては、ユーザが強化学習の非専門家であっても手軽に使えるAPIを提供することで解決を行うこととした。

3. 開発の内容

① d3rlpy : 深層強化学習ライブラリ

d3rlpyはPython向けの深層強化学習ライブラリである。最大の特徴としてデータセットのみからオフラインで学習ができるデータ駆動型強化学習アルゴリズムを多数実装している。図1にオフライン学習の利用例を示す。非常に短いコードで最先端のアルゴリズムが利用できる。また、従来のオンライン学習もサポートしているため、あらゆる強化学習問題に利用することができる。図2にオンライン学習での利用例を示す。

```
import d3rlpy # import package

# prepare dataset and environment for evaluation
dataset, env = d3rlpy.datasets.get_pendulum()

# prepare algorithm
cql = d3rlpy.algos.CQL()

# prepare metrics
env_scorer = d3rlpy.metrics.scorer.evaluate_on_environment(env)

# start training
cql.fit(
    dataset, # dataset to train
    eval_episodes=dataset, # dataset to evaluate
    scorers={ # evaluation metrics
        'environment': scorer,
        'td_error': d3rlpy.metrics.scorer.td_error_scorer
    }
)
```

図1. オフライン学習の例

```
import d3rlpy # import package
import gym # import package

# prepare environment
env = gym.make('Pendulum-v0')
eval_env = gym.make('Pendulum-v0')

# prepare algorithm
sac = d3rlpy.algos.SAC()

# start training
sac.fit_online(env, eval_env=eval_env)
```

図2. オンライン学習の例

さらに、d3rlpyでは強化学習アルゴリズムの原論文で示されている性能を超えるための機能が多数実装されており、本ライブラリ単独で専門家が実装したアルゴリズムよりも遥かに高い性能を得ることができる。

- ・ 分布型価値関数
- ・ 価値関数アンサンブル
- ・ N-step TD誤差計算
- ・ データオーギュメンテーション

他にも実際の製品に利用されることを想定したONNXエクスポート機能や、データセットをCythonによって高速に処理するなど、利用者の利便性を最大化するための機能が盛り込まれている。

② d4rl-pybullet / d4rl-atari : データ駆動型強化学習向けデータセット

本プロジェクトではd3rlpyの評価と強化学習コミュニティへの貢献を目的として、データ駆動型強化学習向けのデータセットを2種類公開した。

- ・ **d4rl-pybullet** : PyBulletシミュレータを利用した完全に無料のデータセット
- ・ **d4rl-atari** : Googleが公開しているAtari 2600のデータセットを手軽に使えるようにしたラッパーライブラリ

③ MINERVA : 深層強化学習GUIソフトウェア

さらに、本プロジェクトではコードを全く書かずに最先端のデータ駆動型深層強化学習の利用を可能にしたGUIソフトウェアMINERVAを開発した。MINERVAはWebベースのソフトウェアであり、フロントエンドはReactによって構築し、バックエンドの強化学習アルゴリズムにd3rlpyを利用している。図3にMINERVAの実際のUIを示す。

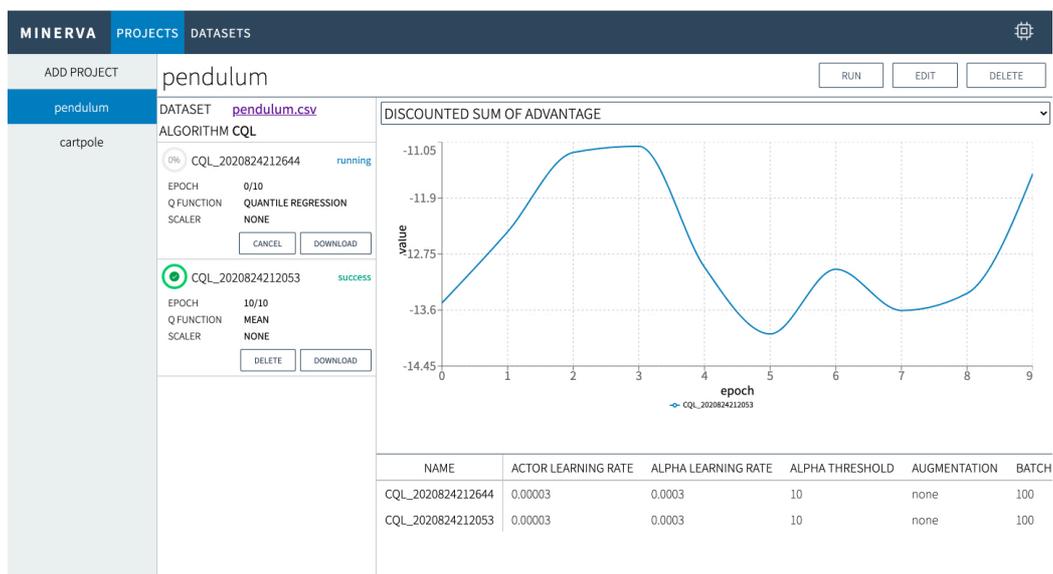


図3. MINERVAのUI

4. 従来の技術（または機能）との相違

① データ駆動型強化学習をサポートした初めてのライブラリ

深層強化学習は非常に活発な研究分野であるため、現在までに多くのライブラリが公開されてきたが、データ駆動型アルゴリズムをサポートしたライブラリはd3rlpyが初めてである。さらに、従来のオンラインアルゴリズムもサポートしているため、あらゆる強化学習問題に対して利用することができる。

② 使いやすいAPI

従来の深層学習ライブラリは研究者向けに設計されているため、拡張性を追求している反面、ユーザに多くのコードを書かせることが多かったが、d3rlpyでは拡張性を犠牲にせず、最小限のコードで利用できるように設計した。また、特定の深層学習ライブラリの知識なしに利用できる強化ライブラリはd3rlpy以外に存在しない。

③ 高い実用性

既存の強化学習ライブラリは論文の性能を再現することに念頭に開発されているため、実際の製品で利用するには性能が足りないことが多い。そこで、d3rlpyでは多くの性能を向上させるアルゴリズムが実装されており、このような試みを行っているライブラリは全くなく、既存のライブラリで同じことをやろうとしても大きな設計変更が必要なため、簡単に真似することはできない。

④ 世界初のGUIツール

GUIのみで機械学習を行うツールやサービスは非常にたくさん提案されてきたが、GUIのみで深層強化学習を行うツールはMINERVAが初めてである。実装としてはd3rlpyのGUIインタフェースであるため、今後のd3rlpyの発展に伴ってMINERVAも機能が増えていく。論文を超える性能が得られる深層強化学習ライブラリは現状ではd3rlpyだけであるため、第三者が強力な性能を備えたGUIツールの開発を行おうとしても簡単にはできないはずである。また、UIのデザインも非常に洗練されており、研究者が開発したGUIツールにありがちな野暮ったいものとは一線を画している。

5. 期待される効果

本プロジェクトのテーマである深層強化学習は主な応用先がロボティクスであることから、日本の強みであるロボット産業や自動運転と非常に親和性が高い。また、強化学習の非専門家の人々が今回開発したソフトウェアを利用することで、アカデミックの分野では想定されてこなかったような独創的な応用例などを発掘されることを期待している。さらに、d3rlpyでは人類がまだ挑戦したことのないようなタスクであっても、非専門家の人々がアルゴリズムの性能を限界まで引き出せるような機能を詰め込んでいるので、継続的に新たな応用“実例”が生まれてくるはずである。

6. 普及（または活用）の見通し

本プロジェクトの開発物はGitHub上で公開しながら開発を行っており、すでに多数のGitHubスターやissue、プルリクエストという形で反響をいただいている。本報告書執筆時点でd3rlpyは130を超えるスター、38件のissue報告、13件のプルリクエストが届いている。本プロジェクトで開発したリポジトリ全体でのGitHubスターは240を超えている。d3rlpyは現在GitHub経由で徐々にスターを伸ばしており、今後の記事執筆などの広報活動でさらに多くのスターを獲得できると期待している。

7. クリエータ名（所属）

妹尾 卓磨（慶應義塾大学）

(参考) 関連URL

- d3rlpy : <https://github.com/takuseno/d3rlpy>
- d4rl-pybullet : <https://github.com/takuseno/d4rl-pybullet>
- d4rl-atari : <https://github.com/takuseno/d4rl-atari>
- MINERVA : <https://github.com/takuseno/minerva>