Instructions and Template for Assignment 3

Manyou Ma
School of Artificial Intelligence, Shenzhen Technology University, Shenzhen, China
email: mamanyou@sztu.edu.cn

Assignment 3: Performance Comparison

Assignment 3 builds upon the work completed in Assignment 1, where you proposed an initial project plan and implemented a preliminary version of the simulator. In this assignment, you are expected to further refine and complete your simulator, apply multiple deep reinforcement learning (DRL) algorithms, and conduct a comprehensive simulation-based evaluation.

Specifically, you must finalize your simulator implementation and ensure that it supports end-to-end training and evaluation. Based on this simulator, you are required to solve the environment using *at least three* DRL algorithms. Each DRL algorithm must be implemented in from scratch, including network architectures, loss functions, optimization procedures, and training loops. You are required to implement all DRL algorithms *from scratch*. The use of high-level DRL training libraries, such as Stable-Baselines3, is *not allowed*. In addition, you must include the non-DRL baseline derived in Assignment 1 as a reference method for performance comparison.

You are required to report the experimental results through a structured simulation study. You must clearly describe the experimental setup, including the selected baseline algorithms and all relevant simulation and training parameters. The simulation results must include comparisons of episode reward, episode length, and training loss during the training process, with at least three corresponding plots.

All implementation code must be provided in a public GitHub repository. You must continue working with the provided LaTeX template and update the report accordingly. In particular, you are required to include the complete simulator code in the appendix and provide a separate attachment listing all hyperparameters used for each DRL algorithm.

I. Motivation

Nam dui ligula, fringilla a, euismod sodales, sollicitudin vel, wisi. Morbi auctor lorem non justo. Nam lacus libero, pretium at, lobortis vitae, ultricies et, tellus. Donec aliquet, tortor sed accumsan bibendum, erat ligula aliquet magna, vitae ornare odio metus a mi. Morbi ac orci et nisl hendrerit mollis. Suspendisse ut massa. Cras nec ante. Pellentesque a nulla. Cum sociis natoque penatibus et magnis dis parturient montes, nascetur ridiculus mus. Aliquam tincidunt urna. Nulla ullamcorper vestibulum turpis. Pellentesque cursus luctus mauris.

Nulla malesuada porttitor diam. Donec felis erat, congue non, volutpat at, tincidunt tristique, libero. Vivamus viverra fermentum felis. Donec nonummy pellentesque ante. Phasellus adipiscing semper elit. Proin fermentum massa ac quam. Sed diam turpis, molestie vitae, placerat a, molestie nec, leo. Maecenas lacinia. Nam ipsum ligula, eleifend at, accumsan nec, suscipit a, ipsum. Morbi blandit ligula feugiat magna. Nunc eleifend consequat lorem. Sed lacinia nulla vitae enim. Pellentesque tincidunt purus vel magna. Integer non enim. Praesent euismod nunc eu purus. Donec bibendum quam in tellus. Nullam cursus pulvinar lectus. Donec et mi. Nam vulputate metus eu enim. Vestibulum pellentesque felis eu massa.

Quisque ullamcorper placerat ipsum. Cras nibh. Morbi vel justo vitae lacus tincidunt ultrices. Lorem ipsum dolor sit amet, consectetuer adipiscing elit. In hac habitasse platea dictumst. Integer tempus convallis augue. Etiam facilisis. Nunc elementum fermentum wisi. Aenean placerat. Ut imperdiet, enim sed gravida sollicitudin, felis odio placerat quam, ac pulvinar elit purus eget enim. Nunc vitae tortor. Proin tempus nibh sit amet nisl. Vivamus quis tortor vitae risus porta vehicula.

II. Related Work

Reinforcement Learning [1] and DQN [2] Nam dui ligula, fringilla a, euismod sodales, sollicitudin vel, wisi. Morbi auctor lorem non justo. Nam lacus libero, pretium at, lobortis vitae, ultricies et, tellus. Donec aliquet, tortor sed accumsan bibendum, erat ligula aliquet magna, vitae ornare odio metus a mi. Morbi ac orci et nisl hendrerit mollis. Suspendisse ut massa. Cras nec ante. Pellentesque a nulla. Cum sociis natoque penatibus et magnis dis parturient montes, nascetur ridiculus mus. Aliquam tincidunt urna. Nulla ullamcorper vestibulum turpis. Pellentesque cursus luctus mauris.

Nulla malesuada porttitor diam. Donec felis erat, congue non, volutpat at, tincidunt tristique, libero. Vivamus viverra fermentum felis. Donec nonummy pellentesque ante. Phasellus adipiscing semper elit. Proin fermentum massa ac quam. Sed diam turpis, molestie vitae, placerat a, molestie nec, leo. Maecenas lacinia. Nam ipsum ligula, eleifend at, accumsan nec, suscipit a, ipsum. Morbi blandit ligula feugiat magna. Nunc eleifend consequat lorem. Sed lacinia nulla vitae enim. Pellentesque tincidunt purus vel magna. Integer non enim. Praesent euismod nunc eu purus. Donec bibendum quam in tellus. Nullam cursus pulvinar lectus. Donec et mi. Nam vulputate metus eu enim. Vestibulum pellentesque felis eu massa.

Quisque ullamcorper placerat ipsum. Cras nibh. Morbi vel justo vitae lacus tincidunt ultrices. Lorem ipsum dolor sit amet, consectetuer adipiscing elit. In hac habitasse platea dictumst.

Integer tempus convallis augue. Etiam facilisis. Nunc elementum fermentum wisi. Aenean placerat. Ut imperdiet, enim sed gravida sollicitudin, felis odio placerat quam, ac pulvinar elit purus eget enim. Nunc vitae tortor. Proin tempus nibh sit amet nisl. Vivamus quis tortor vitae risus porta vehicula.

Fusce mauris. Vestibulum luctus nibh at lectus. Sed bibendum, nulla a faucibus semper, leo velit ultricies tellus, ac venenatis arcu wisi vel nisl. Vestibulum diam. Aliquam pellentesque, augue quis sagittis posuere, turpis lacus congue quam, in hendrerit risus eros eget felis. Maecenas eget erat in sapien mattis porttitor. Vestibulum porttitor. Nulla facilisi. Sed a turpis eu lacus commodo facilisis. Morbi fringilla, wisi in dignissim interdum, justo lectus sagittis dui, et vehicula libero dui cursus dui. Mauris tempor ligula sed lacus. Duis cursus enim ut augue. Cras ac magna. Cras nulla. Nulla egestas. Curabitur a leo. Quisque egestas wisi eget nunc. Nam feugiat lacus vel est. Curabitur consectetuer.

III. System Model and Problem Formulation

Nam dui ligula, fringilla a, euismod sodales, sollicitudin vel, wisi. Morbi auctor lorem non justo. Nam lacus libero, pretium at, lobortis vitae, ultricies et, tellus. Donec aliquet, tortor sed accumsan bibendum, erat ligula aliquet magna, vitae ornare odio metus a mi. Morbi ac orci et nisl hendrerit mollis. Suspendisse ut massa. Cras nec ante. Pellentesque a nulla. Cum sociis natoque penatibus et magnis dis parturient montes, nascetur ridiculus mus. Aliquam tincidunt urna. Nulla ullamcorper vestibulum turpis. Pellentesque cursus luctus mauris.

Nulla malesuada porttitor diam. Donec felis erat, congue non, volutpat at, tincidunt tristique, libero. Vivamus viverra fermentum felis. Donec nonummy pellentesque ante. Phasellus adipiscing semper elit. Proin fermentum massa ac quam. Sed diam turpis, molestie vitae, placerat a, molestie nec, leo. Maecenas lacinia. Nam ipsum ligula, eleifend at, accumsan nec, suscipit a, ipsum. Morbi blandit ligula feugiat magna. Nunc eleifend consequat lorem. Sed lacinia nulla vitae enim. Pellentesque tincidunt purus vel magna. Integer non enim. Praesent euismod nunc eu purus. Donec bibendum quam in tellus. Nullam cursus pulvinar lectus. Donec et mi. Nam vulputate metus eu enim. Vestibulum pellentesque felis eu massa.

Quisque ullamcorper placerat ipsum. Cras nibh. Morbi vel justo vitae lacus tincidunt ultrices. Lorem ipsum dolor sit amet, consectetuer adipiscing elit. In hac habitasse platea dictumst. Integer tempus convallis augue. Etiam facilisis. Nunc elementum fermentum wisi. Aenean placerat. Ut imperdiet, enim sed gravida sollicitudin, felis odio placerat quam, ac pulvinar elit purus eget enim. Nunc vitae tortor. Proin tempus nibh sit amet nisl. Vivamus quis tortor vitae risus porta vehicula.

Fusce mauris. Vestibulum luctus nibh at lectus. Sed bibendum, nulla a faucibus semper, leo velit ultricies tellus, ac venenatis arcu wisi vel nisl. Vestibulum diam. Aliquam pellentesque, augue quis sagittis posuere, turpis lacus congue quam, in hendrerit risus eros eget felis. Maecenas eget erat in sapien mattis porttitor. Vestibulum porttitor. Nulla facilisi. Sed a turpis eu lacus commodo facilisis. Morbi fringilla, wisi in dignissim interdum, justo lectus sagittis dui, et vehicula libero dui cursus dui. Mauris tempor ligula sed lacus. Duis cursus enim ut augue. Cras ac magna. Cras nulla. Nulla egestas. Curabitur a leo. Quisque egestas wisi eget nunc. Nam feugiat lacus vel est. Curabitur consectetuer.

Suspendisse vel felis. Ut lorem lorem, interdum eu, tincidunt sit amet, laoreet vitae, arcu. Aenean faucibus pede eu ante. Praesent enim elit, rutrum at, molestie non, nonummy vel, nisl. Ut lectus eros, malesuada sit amet, fermentum eu, sodales cursus, magna. Donec eu purus. Quisque vehicula, urna sed ultricies auctor, pede lorem egestas dui, et convallis elit erat sed nulla. Donec luctus. Curabitur et nunc. Aliquam dolor odio, commodo pretium, ultricies non, pharetra in, velit. Integer arcu est, nonummy in, fermentum faucibus, egestas vel, odio.

Sed commodo posuere pede. Mauris ut est. Ut quis purus. Sed ac odio. Sed vehicula hendrerit sem. Duis non odio. Morbi ut dui. Sed accumsan risus eget odio. In hac habitasse platea dictumst. Pellentesque non elit. Fusce sed justo eu urna porta tincidunt. Mauris felis odio, sollicitudin sed, volutpat a, ornare ac, erat. Morbi quis dolor. Donec pellentesque, erat ac sagittis semper, nunc dui lobortis purus, quis congue purus metus ultricies tellus. Proin et quam. Class aptent taciti sociosqu ad litora torquent per conubia nostra, per inceptos hymenaeos. Praesent sapien turpis, fermentum vel, eleifend faucibus, vehicula eu, lacus.

Pellentesque habitant morbi tristique senectus et netus et malesuada fames ac turpis egestas. Donec odio elit, dictum in, hendrerit sit amet, egestas sed, leo. Praesent feugiat sapien aliquet odio. Integer vitae justo. Aliquam vestibulum fringilla lorem. Sed neque lectus, consectetuer at, consectetuer sed, eleifend ac, lectus. Nulla facilisi. Pellentesque eget lectus. Proin eu metus. Sed porttitor. In hac habitasse platea dictumst. Suspendisse eu lectus. Ut mi mi, lacinia sit amet, placerat et, mollis vitae, dui. Sed ante tellus, tristique ut, iaculis eu, malesuada ac, dui. Mauris nibh leo, facilisis non, adipiscing quis, ultrices a, dui.

IV. Simulation Study

This section presents the simulation-based evaluation of the proposed environment and learning algorithms. We first describe the experimental setup, including the baseline methods and simulation parameters. We then report and compare the performance of different algorithms through multiple training metrics.

A. Experimental Setup

1) Baseline Algorithms: In addition to the DRL-based approaches, we include the non-DRL baseline algorithm developed in Part 1 of the project. This baseline follows a handcrafted decision rule derived from domain knowledge and does not rely on learning. The baseline serves as a reference point to evaluate the effectiveness of reinforcement learning in the proposed environment.

Students must clearly describe the baseline policy and its decision logic here.

2) DRL Algorithms: We evaluate at least three DRL algorithms on the same simulator. Each algorithm is implemented from scratch and trained under identical environment settings to ensure a fair comparison. You should list and properly cite the selected algorithms and briefly describe their key characteristics.

3) Simulation Parameters: All algorithms are trained and evaluated using the same simulation parameters. Table I summarizes the key parameters used in the experiments.

TABLE I: Simulation Parameters

Parameter	Value
Maximum episode length	
Discount factor (γ)	
Learning rate	
Batch size	
Training episodes / steps	
Random seed(s)	

B. Simulation Results

We evaluate the training performance of all algorithms using three key metrics: episode reward, episode length, and training loss. All results are averaged over training iterations, and smoothing is applied where appropriate for visualization.

1) Episode Reward: Figure 1 shows the evolution of the episode reward during training for different algorithms. Higher episode reward indicates better long-term performance in the proposed environment.

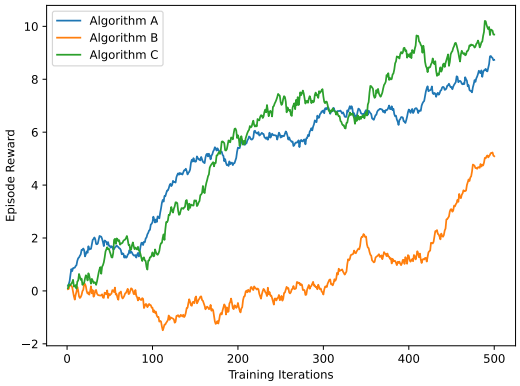

Fig. 1: Episode reward versus training iterations.

2) Episode Length: Figure 2 compares the episode length observed during training. This metric reflects the agent's ability to reach terminal conditions efficiently.

3) Training Loss: Figure 3 illustrates the training loss for the learning-based methods. The specific definition of the loss depends on the algorithm (e.g., TD loss for value-based methods or critic loss for actor–critic methods).

APPENDIX

A. Environment Implementation

In this appendix, students must include the *complete implementation* of the simulator used in their experiments. The code provided here must contain all *core environment functions*, including at minimum the constructor (`__init__`), `reset`, `step`, and observation generation. The simulator code included in this appendix must be identical to the version used for training and evaluation, and must be consistent with the implementation in the accompanying GitHub repository.

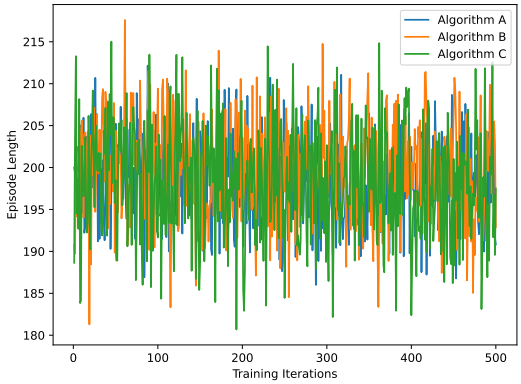

Fig. 2: Episode length versus training iterations.

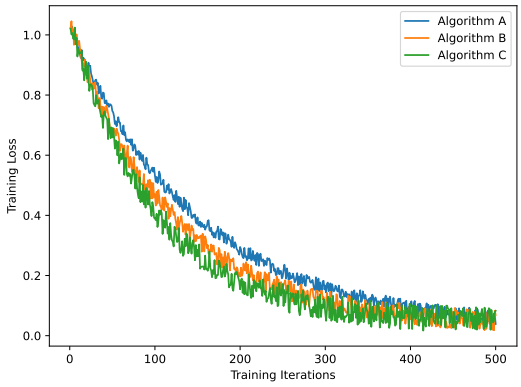

Fig. 3: Training loss versus training iterations.

```
import numpy as np


class ResourceEnv:
    def __init__(self, max_q=20, horizon=200,
        seed=0):
        self.max_q = max_q
        self.horizon = horizon
        self.rng = np.random.default_rng(seed)
        self.state_dim = 3
        self.action_dim = 2
        self.reset()

    def reset(self):
        self.q = self.rng.integers(0, self.
            max_q // 2)
        self.h = self.rng.uniform(0.2, 1.0)
        self.t = self.horizon
        return self._obs()

    def step(self, a):
        r = 0.0
        if self.rng.random() < 0.25:
            self.q += 1
```

```
        if a == 1 and self.q > 0:
            if self.rng.random() < self.h:
                self.q -= 1
                r += 1.0
            else:
                r -= 0.2
        else:
            r -= 0.05
        if self.q > self.max_q:
            r -= 1.0
            self.q = self.max_q
        self.h = self.rng.uniform(0.2, 1.0)
        self.t -= 1
        done = self.t <= 0
        return self._obs(), r, done, {}

    def _obs(self):
        return np.array(
            [self.q / self.max_q, self.h, self
                .t / self.horizon],
            dtype=np.float32,
        )
```

REFERENCES

[1] R. S. Sutton and A. G. Barto, *Reinforcement Learning: An Introduction, 2nd Edition*. MIT Press, 2018.

[2] V. Mnih, K. Kavukcuoglu, D. Silver, A. A. Rusu, J. Veness, M. G. Bellemare, A. Graves, M. Riedmiller, A. K. Fidjeland, G. Ostrovski *et al.*, "Human-level control through deep reinforcement learning," *Nature*, vol. 518, no. 7540, pp. 529–533, Feb. 2015.