
Assignment 1 - Defining & Solving RL Environments

Shaurya Mathur
University at Buffalo
Buffalo, NY
smathur4@buffalo.edu

Abstract

1 The report presents the code and results for the checkpoint for first assignment
2 for CSE 546 - Reinforcement Learning. The goal of the assignment is to acquire
3 experience in defining and solving RL environments, following Gymnasium stan-
4 dards.

5 1 Defining RL Environments

6 1.1 B.2 Traffic Light Control

7 **Scenario:** A traffic light controller operates at a 4-way intersection. The goal is to minimize the
8 average wait time of cars by optimizing the traffic light switching strategy.

9 1.2 Environment Setup

- 10 • **Grid Size:** 4x4 grid representing the intersection.
- 11 • **Cars:** Cars arrive at the intersection and must wait until they can move forward.
- 12 • **Goal:** Minimize the average wait time of cars at the intersection.
- 13 • **Actions:** Switch to Red, Green, or Yellow for each of the four directions.
- 14 • **Rewards:**
 - 15 – -1 for each second a car waits.
 - 16 – +5 for each car that successfully passes through the intersection.
- 17 • **Terminal State:** Defined by a maximum steps reached or a certain number of cars pro-
18 cessed.

19 1.3 Deterministic and Stochastic Environments

20 Deterministic Environment:

- 21 • The traffic flow is fixed, meaning cars arrive at fixed regular intervals from each direction.
- 22 • The timing and number of cars arriving at the intersection are predictable.
- 23 • Rewards: -1 per second a car waits, +5 for each car passing through the intersection.

24 Stochastic Environment:

- 25 • The traffic flow is random, with cars arriving at irregular intervals.
- 26 • The reward function remains the same as in the deterministic setting.

27 • To simulate the arrival of cars on the intersection, the environment uses a probabilities for
28 each directino totaling to 1. Each probability denotes the likelihood of a car reaching the
29 intersection in that direction.

30 Other

31 • The environment has the capability to simulate different traffic conditions, such as heavy
32 traffic during rush hour and light traffic during off-peak times.

33 1.4 Environment Constraints

34 • **Legal Light Switching:** Traffic lights can't perform illegal action sequences such as
35 switching to the same color twice or 1. Green 2. Yellow 3. Green. Environment termi-
36 nates on an illegal action.

37 • **Light Timings:** Green traffic light stays and allows a single car to cross the intersection
38 for 3 seconds and yellow light stays for 2 second.

39 • **Direction Constraints:** At a time step, only 1 direction can be green/yellow, all others will
40 be red.

41 2 Visualizations of Environment

42 2.1 Initial Environment State

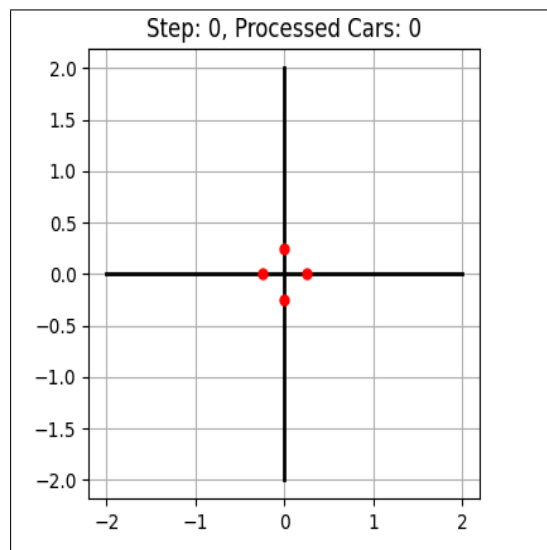


Figure 1: Initial Environment State.

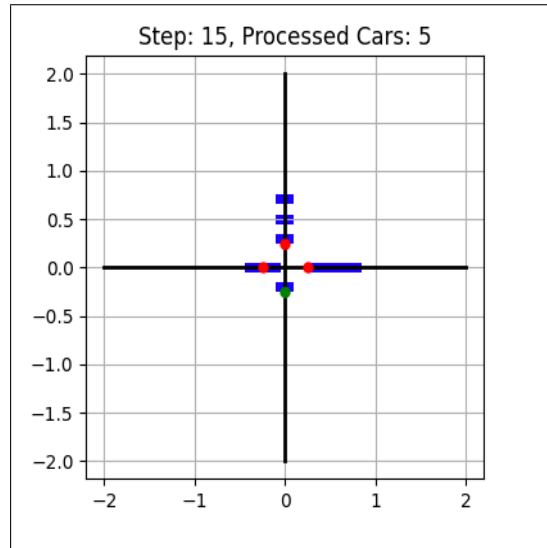


Figure 2: Intermediate Environment State 1

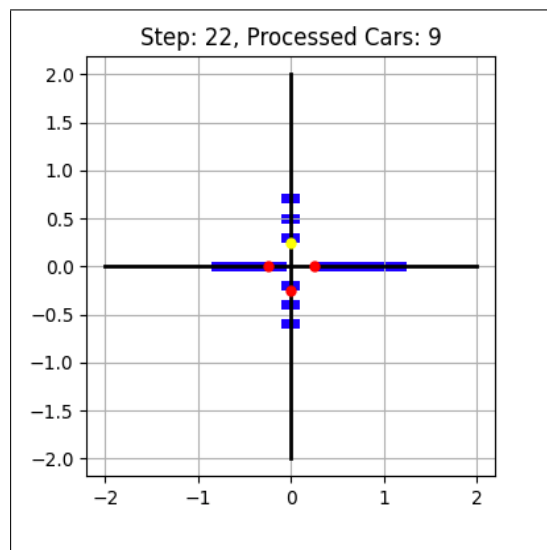


Figure 3: Intermediate Environment State 2

2.3 Terminal Environment State

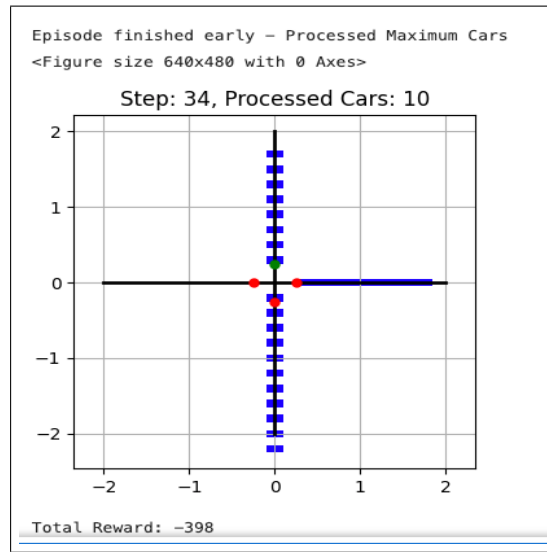


Figure 4: Terminal Environment State

3 Safety in AI

Ensuring safety in the traffic light control environment is crucial to prevent unrealistic or unsafe actions by the reinforcement learning (RL) agent.

- We enforce constraints on the agent's action space, ensuring that it follows legal traffic light sequences (e.g., preventing a transition from Green directly back to Green without a Yellow phase). I have written a `_is_legal_action` function which adds relevant constraints on the environment.
- The state space is well-defined, ensuring the agent only operates within the valid 4x4 grid representing the intersection and the traffic light is changed only in one of the four directions.
- The reward policy discourages unsafe behaviorspenalizing excessive wait times while rewarding efficient traffic flow.
- In the stochastic environment, randomness in car arrivals is carefully controlled to avoid unrealistic congestion or deadlock scenarios.
- The environment is capable of handling different traffic conditions, such as heavy traffic during rush hour and light traffic during off-peak times.
- Lastly, extensive testing and validation are performed to ensure that the trained agent generalizes well to various traffic conditions while maintaining safety constraints.

4 Link to github

<https://github.com/ShauryaMathur/CSE546-RL-Assignment1>

67 **4.1 Github Commit History**

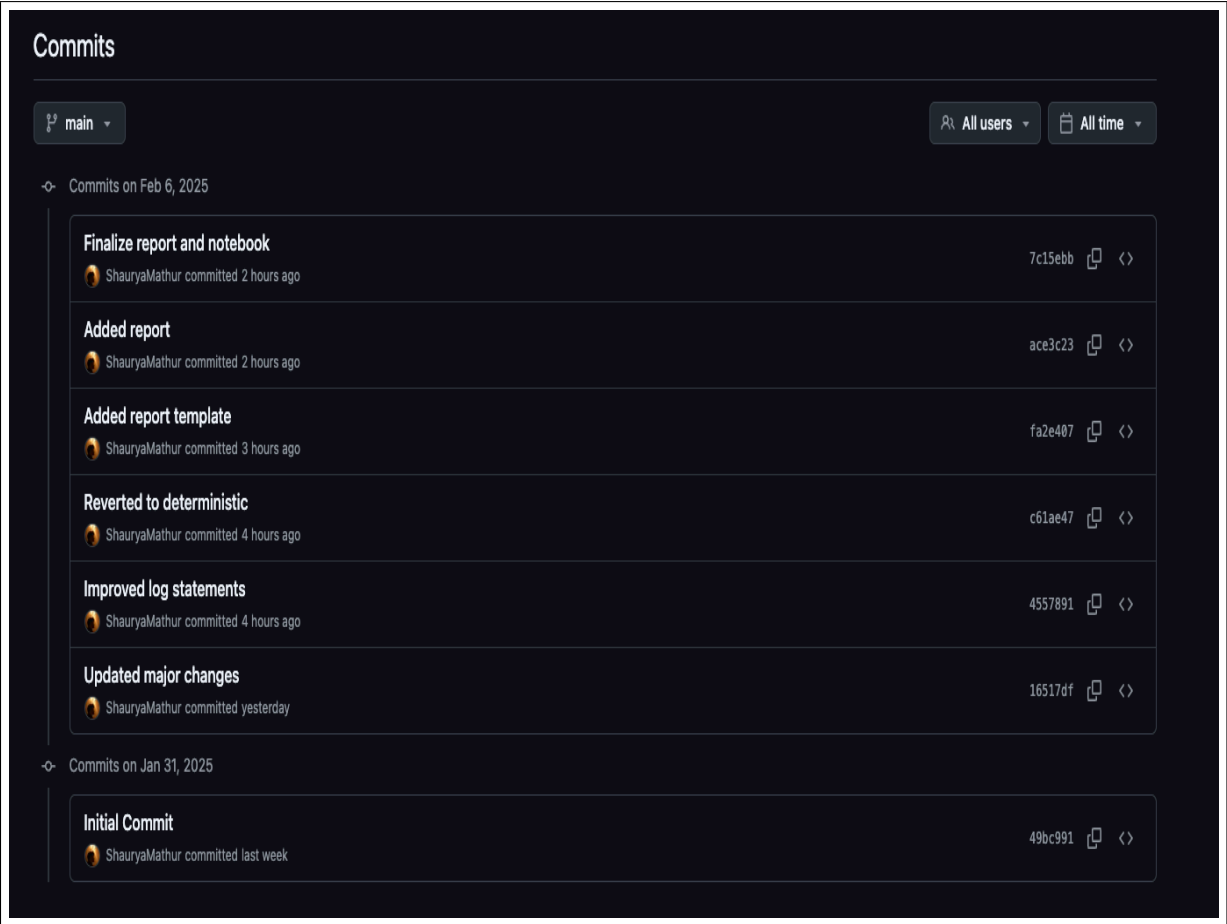


Figure 5: Github Commit History

68 **References**

69 [1] <https://gymnasium.farama.org/api/env/>.
70 [2] Lecture slides.
71 [3] <https://matplotlib.org/stable/index.html>.
72 [4] <https://docs.python.org/3/library/dataclasses.html>