Assignment 1 - Defining & Solving RL Environments

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

Abstract

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

1 Defining RL Environments

1.1 B.2 Traffic Light Control

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

1.2 Environment Setup

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

1.3 Deterministic and Stochastic Environments

Deterministic Environment:

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

Stochastic Environment:

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

Submitted to Prof. Alina Vereshchaka for CSE546 Assignment 1.

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

Other

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

1.4 Environment Constraints

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

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

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

2 Visualizations of Environment

2.1 Initial Environment State

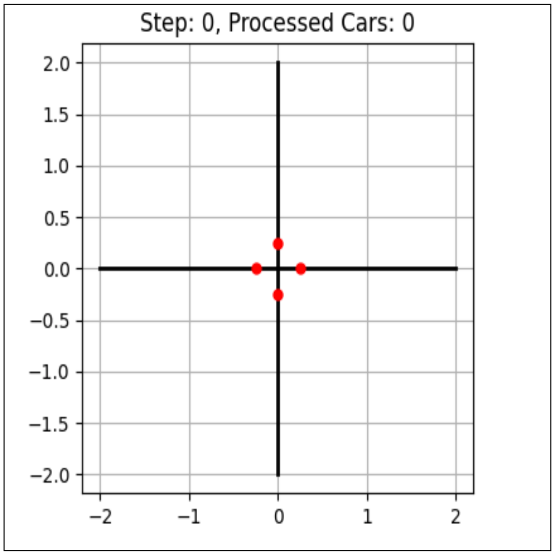


Figure 1: Initial Environment State.

2.2 Intermediate Environment States

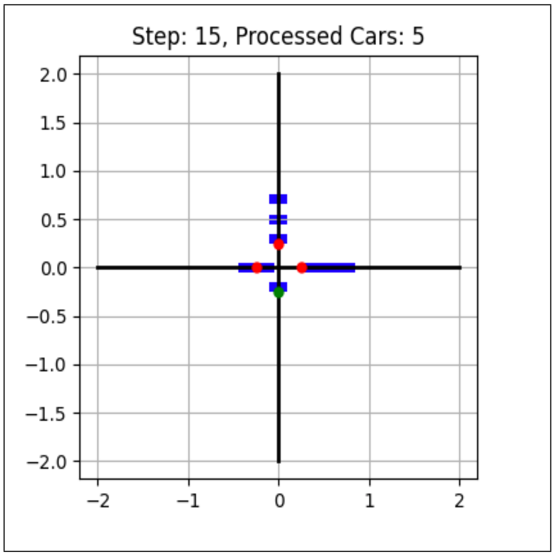


Figure 2: Intermediate Environment State 1

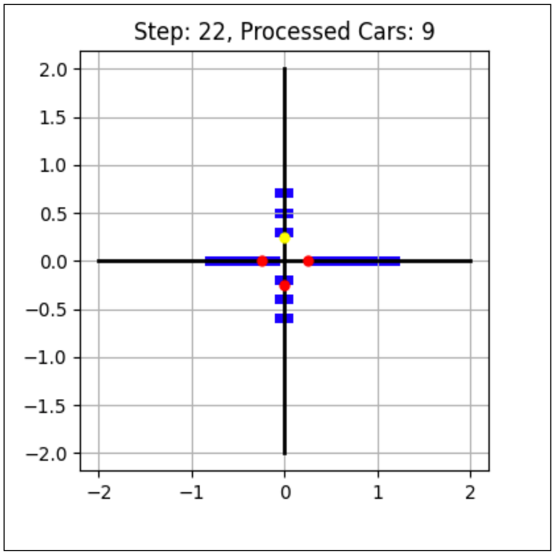
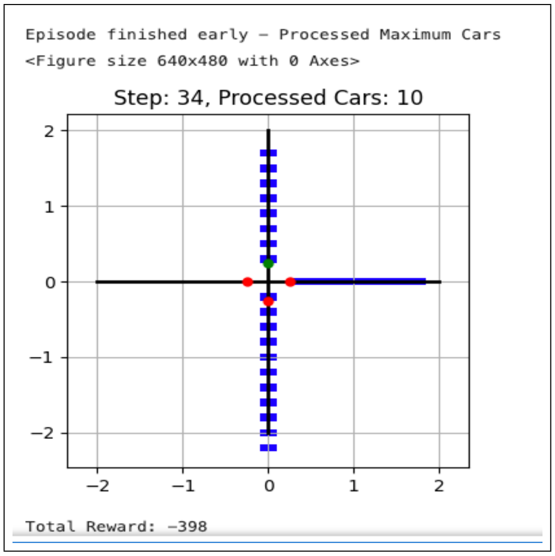


Figure 3: Intermediate Environment State 2

2.3 Terminal Environment State



```
Episode finished early – Processed Maximum Cars
<Figure size 640x480 with 0 Axes>
```

```
Total Reward: −398
```

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.

References

[1] https://gymnasium.farama.org/api/env/.

[2] Lecture slides.

[3] https://matplotlib.org/stable/index.html.

[4] https://docs.python.org/3/library/dataclasses.html