Comparing Exploration Approaches in Deep Reinforcement Learning for Traffic Light Control

Yaniv Oren | Supervisors: Frans Oliehoek, Rolf Sterre | June 2020 | CSE3000

[1] THE SETTING

- **Traffic** is a global **pandemic**. Traffic flow can be improved by **better traffic light control.**
- Specifically, a **dynamic**, **optimized** traffic light **policy** per intersection.
- We apply reinforcement learning (RL) ***(1)*** to this setting, by optimizing traffic light control.

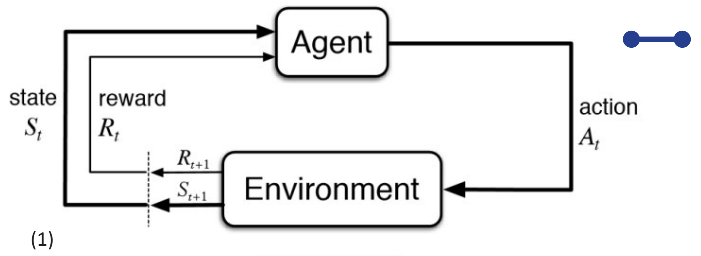
[2] THE QUESTION

- **Exploration** is a fundamental principle of RL. To find an optimal policy through experience alone, an agent must explore its environment.
- There are many different exploration approaches, with different achievements and computational costs.

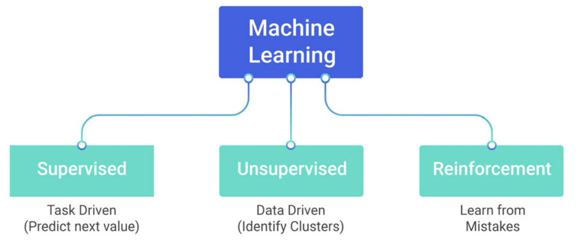
We investigate a comparison between different exploration approaches in deep RL for traffic light control, to identify the value of different exploration approaches in this setting.

[3] THE METHOD

- A baseline **DQN** agent ***(2)*** using epsilon-greedy is compared to two state of the art exploration approaches - **Bootstrapped DQN *(3)*** and **Randomized Prior Functions *(4)***.
- This is done with **8 different agents** – six learners employing different combination of the approaches, and two static agents to evaluate the final policies learned.
- The traffic simulator SUMO ***(5)*** is used to simulate different traffic profiles against different road-maps: a **basic grid** map, a simulation of **real traffic in Manhattan, New York** . and custom **heavy traffic** in the same road-map.



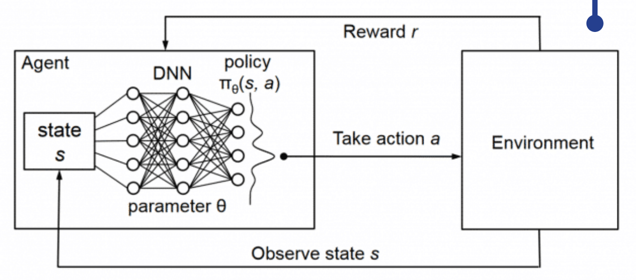
(1)



(6)

(1) REINFORCEMENT LEARNING

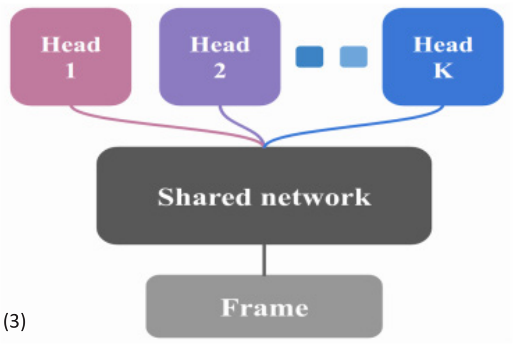
Reinforcement learning **(RL)** is an area of **machine learning** where, **an agent operates in an environment**, and attempts to learn an **optimal policy**, such that **the reward over time** is **maximizied.**



(2)

(2) DQN

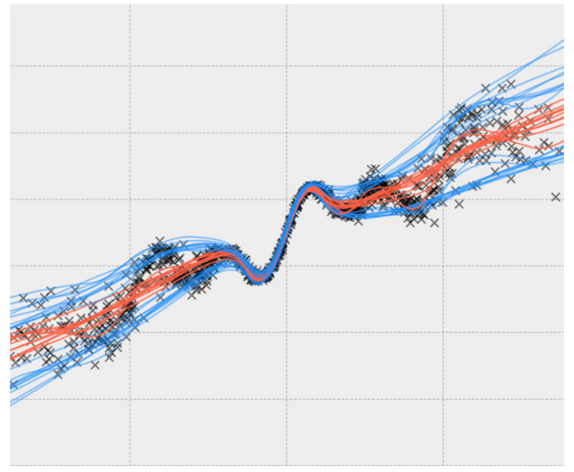
DQN is a **deep reinforcement learning algorithm**, where deep neural networks are used as function estimators.



(3)

(3) BOOTSTRAPPED DQN

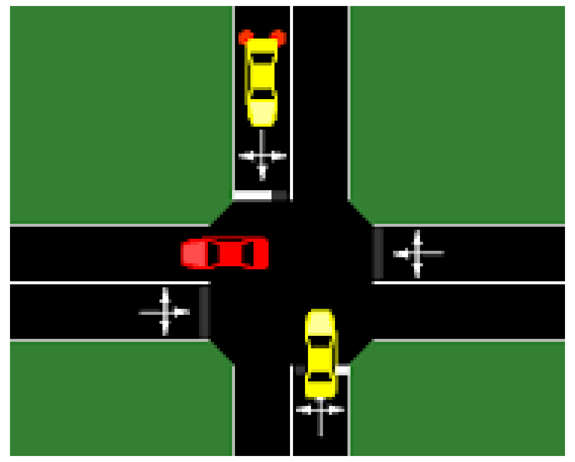
Bootstrapped DQN attempts to achieve **deep exploration**, by keeping several estimations of the Q value.



(4)

(4) RANDOMIZED PRIOR FUNCTIONS

Randomized prior functions add an untrainable, network to the Q-value, to give each Q estimation an inherent "tendency" to go in some direction, to **improve the uncertainty** mechanism.

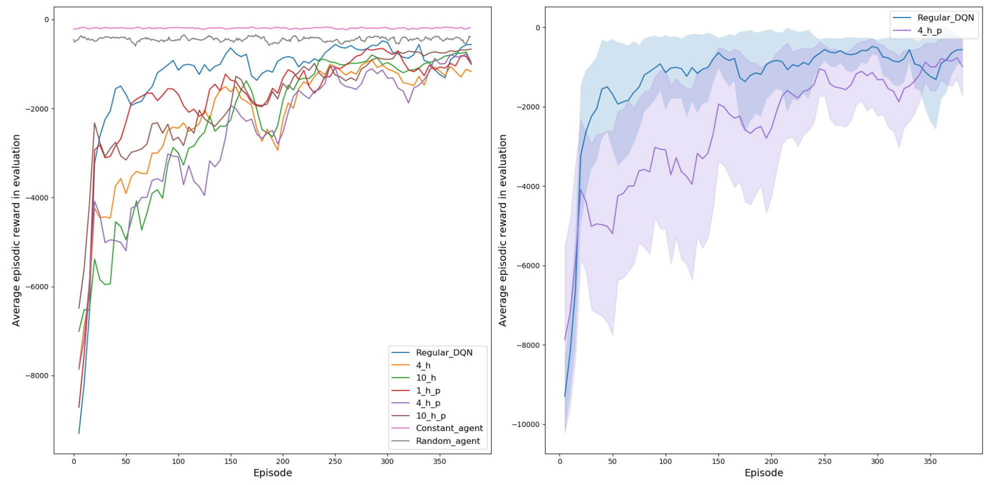


(5)

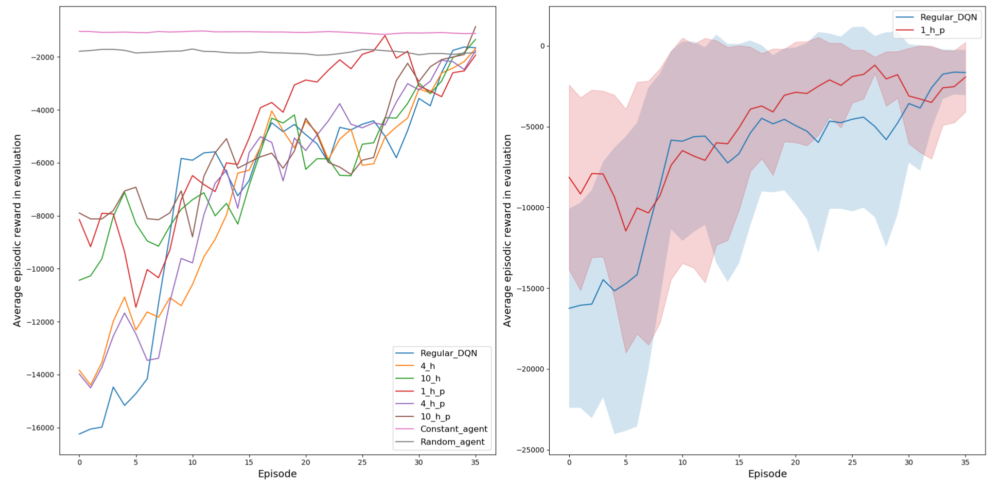
(5) SUMO

SUMO is an **open source traffic simulator** used to test the agents' ability to learn effective policies.

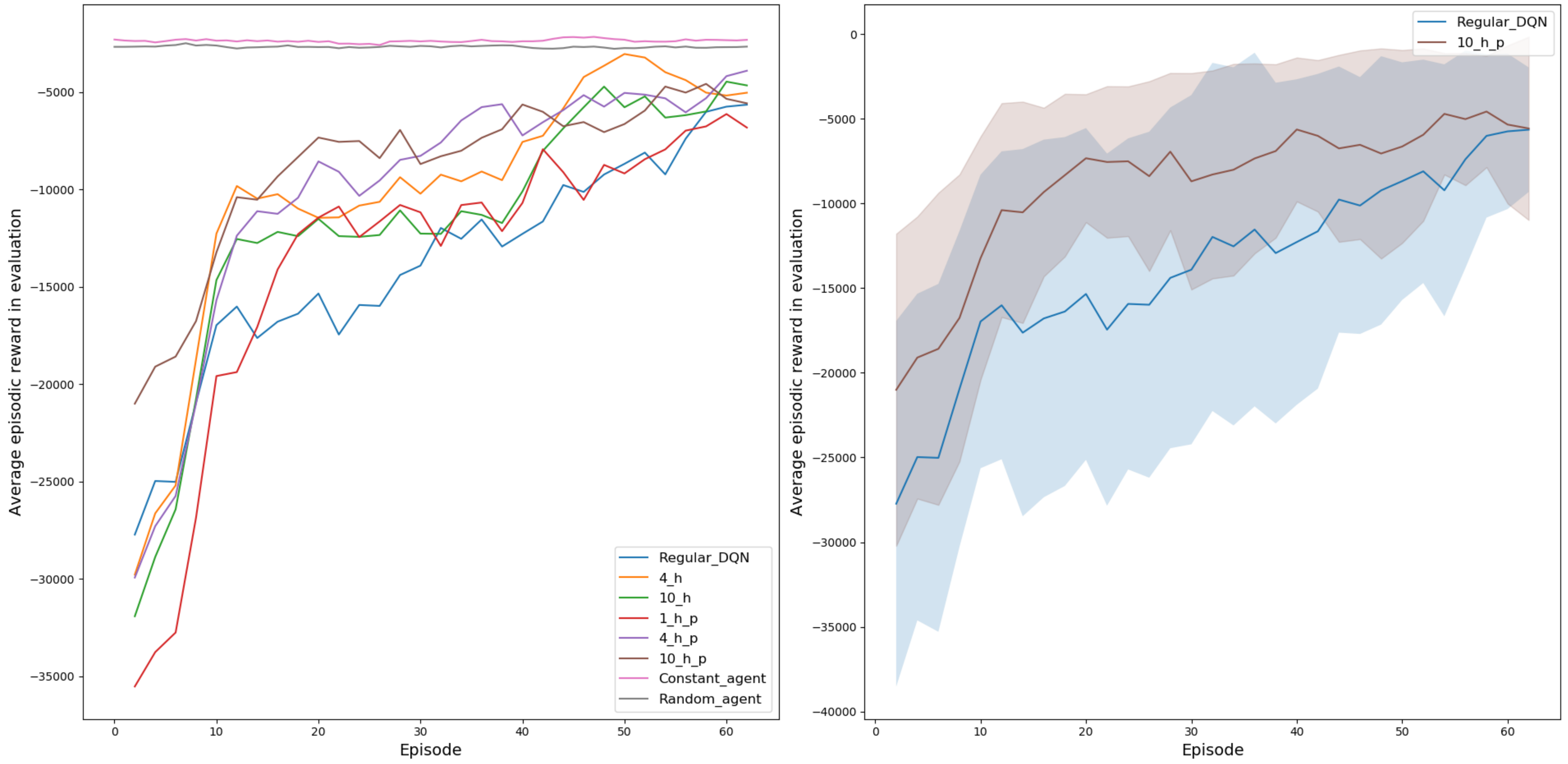
Results



Evaluating the agents against the **basic grid** scenario.
The **regular DQN** achieves the **best** average evaluation score.



Evaluating the agents against the **real** traffic scenario of **high** average speed.
The **different agents achieve mostly similar evaluation scores.**



Evaluating the agents against the custom, **heavy** traffic scenario of **low** average speed. The **agents appear to perform according to their level of complexity.**

(1): Sutton, Richard S. "Reinforcement Learning." (1999).
(2): Mao, Hongzi et al., "Resource Management with Deep Reinforcement Learning." (2016).
(3): Osband, Ian, et al. "Deep exploration via bootstrapped DQN." Advances in neural information processing systems. 2016.
(5): Marmerola, Guilherme D., "Risk and Uncertainty in Deep Learning", https://gdmarmerola.github.io/risk-and-uncertainty-deep-learning/ (2019)
(6) Demush, Rostyslav "Reinforcement Learning Applications: A Brief Guide on How to Get Business Value from RL", https://perfectial.com/blog/reinforcement-learning-applications/ (2018)