Exploration In Deep Reinforcement Learning for Traffic Light Control

Yaniv Oren
Supervisors: Frans Oliehoek, Rolf Sterre
June 2020

[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

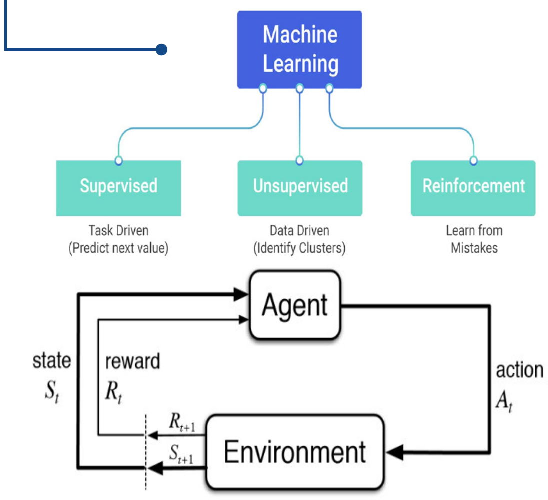
One of the fundamental principles of RL is exploration. In order to find an optimal policy through experience alone, the agent must explore its environment. There are many different exploration approaches, with different achievements and computational costs. We investigate Do state-of-the-art exploration approaches improve a deep learning agent's ability to learn optimal traffic light control policies?

[3] THE METHOD

We compare the performance of a baseline DQN agent (2) to two different state of the art exploration approaches - Bootstrapped DQN (3) and Randomized Prior Functions (4). We use the traffic simulator SUMO (5) to simulate a small section of the road-map of Manhattan, New York, and real traffic data to generate random traffic flows over the simulation road-map.

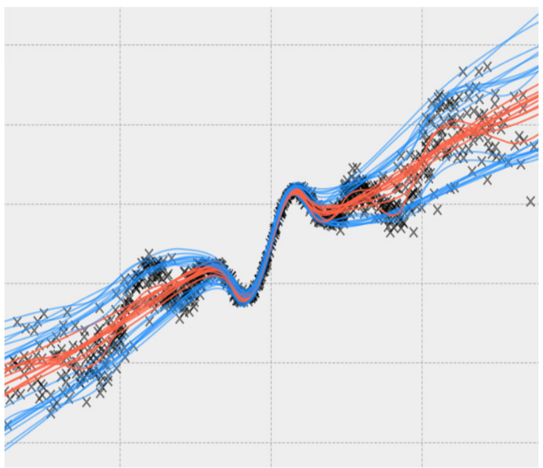
(1) REINFORCEMENT LEARNING

Reinforcement learning (RL) is an area of machine learning. In reinforcement learning (RL), an agent operates in an environment, executes actions and receives rewards. The environment provides information regarding the state the agent is in, and what actions it can execute. The agent is attempting to learn an optimal policy, such that the reward over time is maximizied.



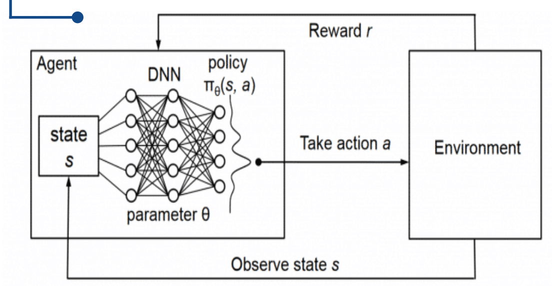
(4) RANDOMIZED PRIOR FUNCTIONS

Prior functions add an additional, untrainable, neural network to the Q-value estimation, to give each Q estimation an inherent "tendency" to go in some direction. As a result, the agent is pushed to explore even in states it doesn't have any prior experiences of.



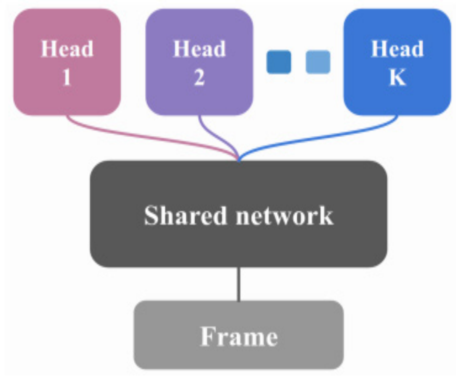
(2) DQN

DQN is a reinforcement learning algorithm. In DQN, the agent uses deep neural networks to estimate the function describing the value of each action in each state.



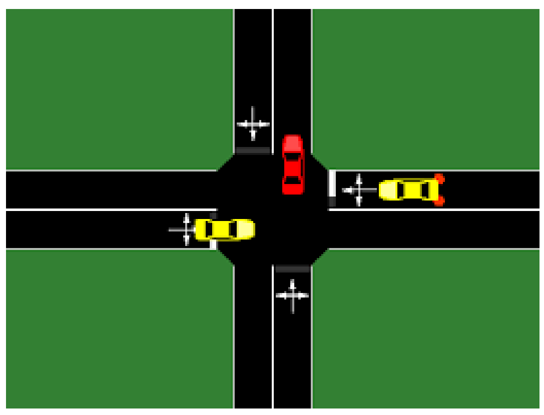
(3) BOOTSTRAPPED DQN

Bootstrapped DQN is an evolved exploration approach that tries to capture uncertainty and deep exploration, by keeping several estimation of the Q value instead of just one, and sampling from them randomly.



(5) SUMO

SUMO is an open source traffic simulator we use to to test our agents' ability to explore learn effective policies.



RESULTS