Traffic Signal Control: Q-Learning with Different Rewards

Kyle Masters

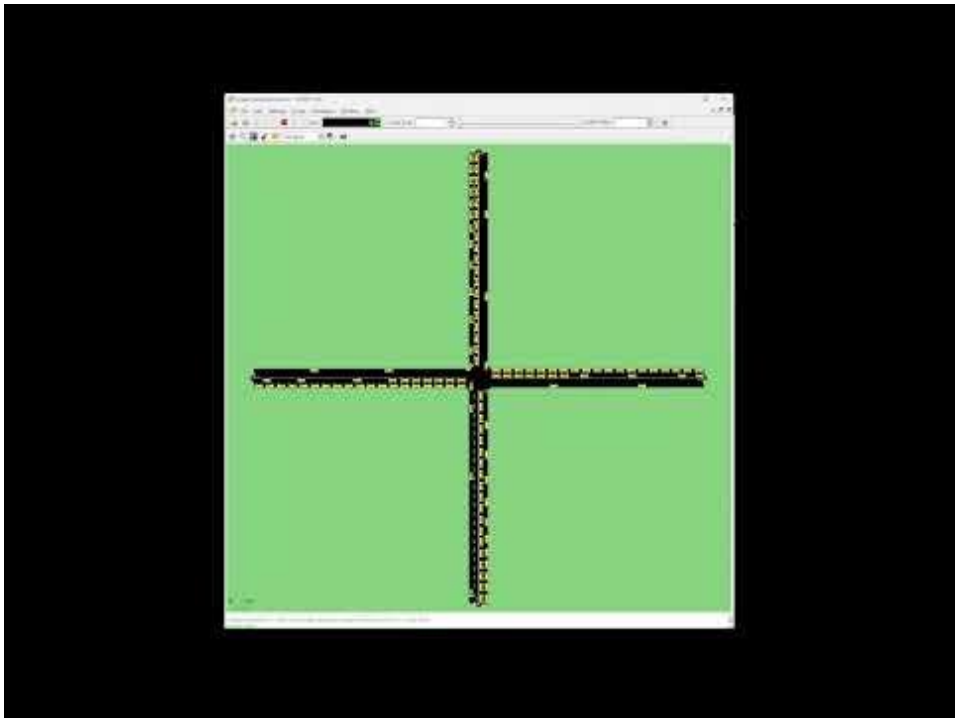
Traffic Signal Control: Q-Learning with Different Rewards

Mini-World

- 4 traffic lights
- 8 lanes (2 each direction)
- Vehicles driving through intersection

Traffic issues cause

- Carbon emissions
- Travel time



Environment

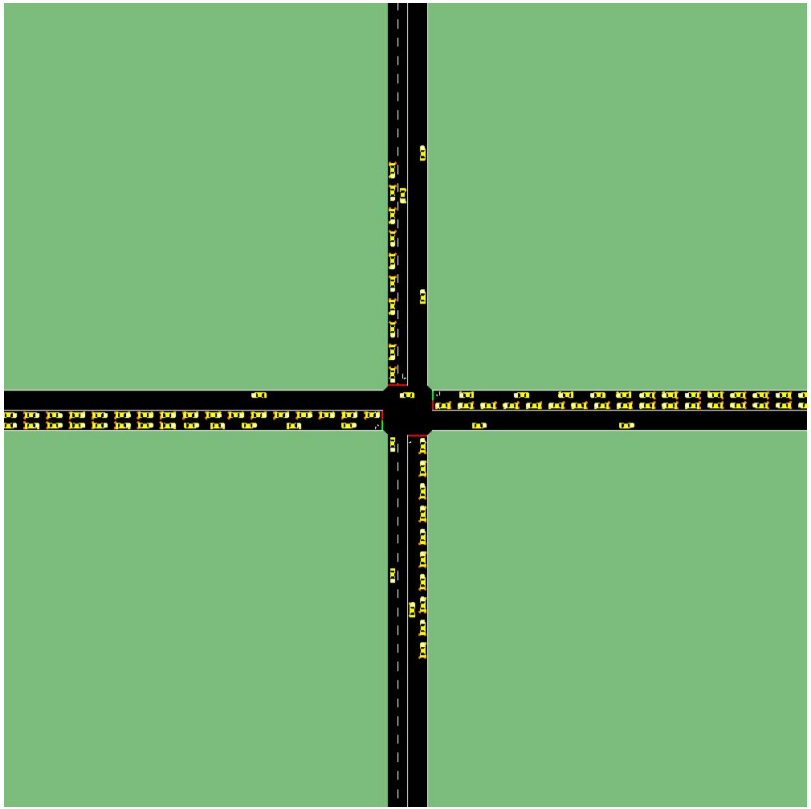
Environment Used: SUMO-RL

Input (State Space)

- Current Phase Vector (4)
- Minimum Time Boolean (1)
- Lane density (8)
- Lane queue (8)

Output (Action Space)

- NS Straight
- NS Left Turn
- EW Straight
- EW Left Turn



Method

DQN (Deep Q Learning)

2 Dense Neural Networks:

- Policy - Often Updated
- Target - Rarely Updated

Replay Buffer

4 different reward functions

- Wait Time
- Speed
- Queue Length
- Vehicle Count

1 evaluation metric

- Total vehicle throughput

Results

Successful Rewards

- Wait Time
- Speed

Unsuccessful

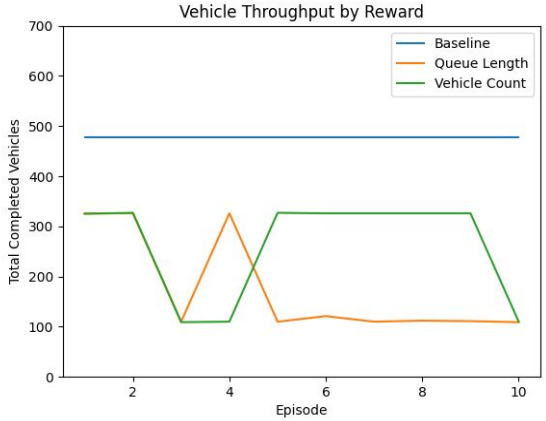
- Queue Length
- Vehicle Count

In general:

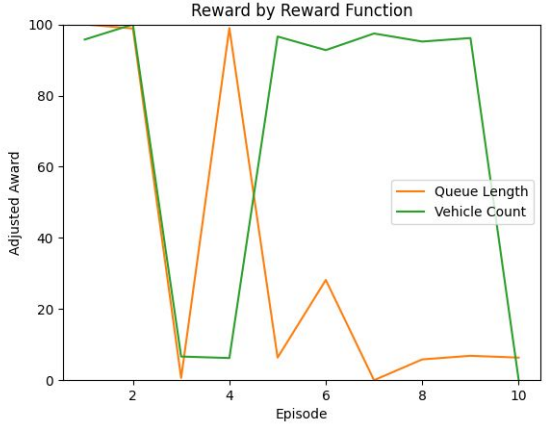
- Susceptible to local maxima

	Total Vehicles Through
Baseline	476
Wait Time	686
Speed	628
Queue Length	326
Vehicle Count	327


Vehicle Throughput by Reward
Baseline
Wait Time
Speed
Total Completed Vehicles
Episode
0
100
200
300
400
500
600
700
2
4
6
8
10


Vehicle Throughput by Reward
Baseline
Queue Length
Vehicle Count
Total Completed Vehicles
Episode
0
100
200
300
400
500
600
700
2
4
6
8
10


Reward by Reward Function
Wait Time
Speed
Adjusted Award
Episode
0
20
40
60
80
100
2
4
6
8
10


Reward by Reward Function
Queue Length
Vehicle Count
Adjusted Award
Episode
0
20
40
60
80
100
2
4
6
8
10

Summary

Successful agents

- Made intelligent decisions

Unsuccessful agents

- Fell into local maxima
- Solutions focused on single lanes

Reward heavily impacted results