COMPSCI 687, Reinforcement Learning
Course Project

Dilip Chakravarthy Kavarthapu

December 21, 2017

Introduction

Traffic congestion is one of the most common issues around the world. Heavy concentration of vehicles in any major city and inadequate road network infrastructure contributes to heavy traffic jams that have become a part and parcel of our lives. One solution that can easen this burden is by controlling traffic lights at every signal in a coordinated fashion to ensure free flow of traffic.

Reinforcement Learning can be one of the methods that can be applied to optimize this road traffic network. By formulating a traffic signal at any junction as an agent that can take an action based on the traffic situation around, it can ensure a better traffic condition than an uncontrolled traffic signal. In this project, I have formulated the traffic movement control as a Markov Decision Process and attempted to optimize the flow of traffic using Q Learning and SARSA Learning algorithms. In the rest of the report, I give details on problem formulation, behaviour as an MDP, solution and results.

Problem Formulation

In this section, I describe more about the environment that has been developed.
Environment :-

The traffic junction has 4 roads leading to it - north, south, east and west, having 2 lanes each. Traffic signal has a total of 4 actions - North-South roads in green, East-West roads in green, North-South roads(only left turn) in green and East-West(only left turn) in green. There are 2 types of vehicles that can be spawned and simulated - one resembling a car and another resembling a heavy duty vehicle like a truck/bus(size, acceleration and max. speed have been adjusted accordingly. Certain random behavior like skipping signals after long wait is also included.

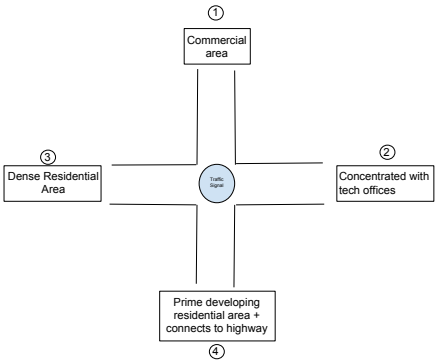

Figure 1: Pictorial Representation of the imaginary traffic junction

Traffic Simulation:-

The traffic has been simulated according to an imaginary junction with the layout above. The traffic simulation was done with the help of the SUMO - free traffic simulation suite(1).

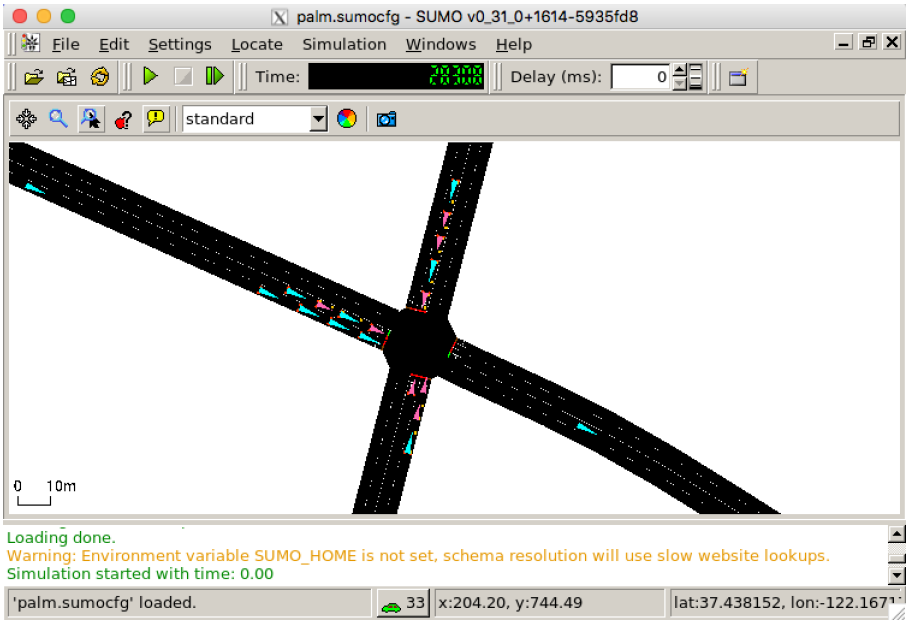

Figure 2: Pictorial Representation of the imaginary traffic junction

Traffic density, as it happens in daily lives has been modelled with a bivariate normal distribution as shown below. Peak timings are chosen from a uniform random distribution between 10-11 hrs(Variable x1) in the morning and between 18.5-19.5 hrs(Variable x2) in the evening. Standard deviation has been chosen to be 3.5 and 2.5 respectively. The following is the pictorial representation of the traffic density w.r.t time.

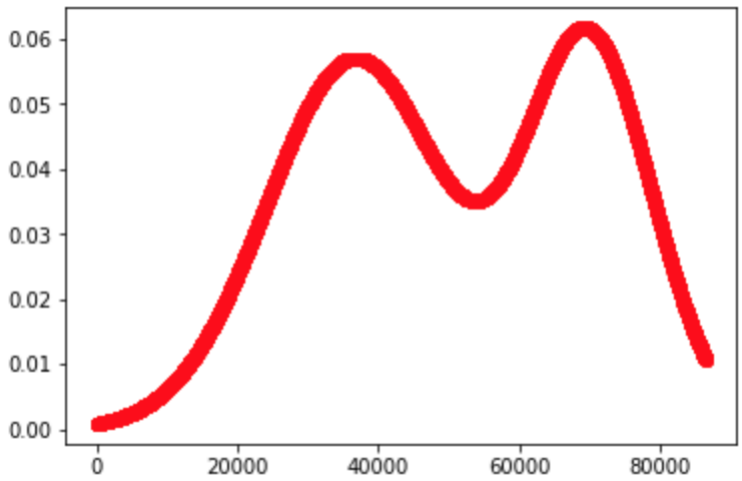

Figure 3: Scaled version of the traffic density at the junction. X - axis is time in seconds.

Given the nature of the areas in the layout above, the inflow and outflow of traffic in various lanes have been set as follows. Closeness to reality was verified with results from(2).

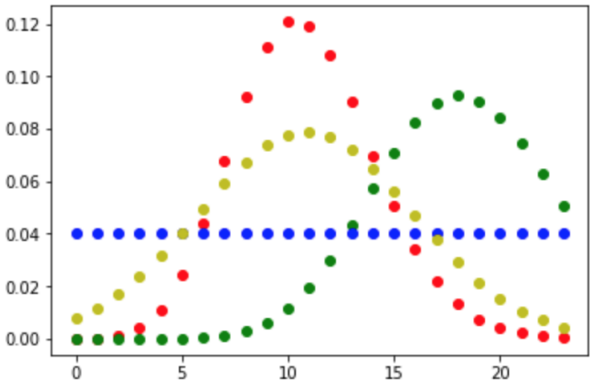

Figure 4: This plot shows the probability pattern of inflow of vehicles from various roads. Area 1 : Blue, Area 2 : Green, Area 3 : Red, Area 4 : Yellow. X-axis is Hours.

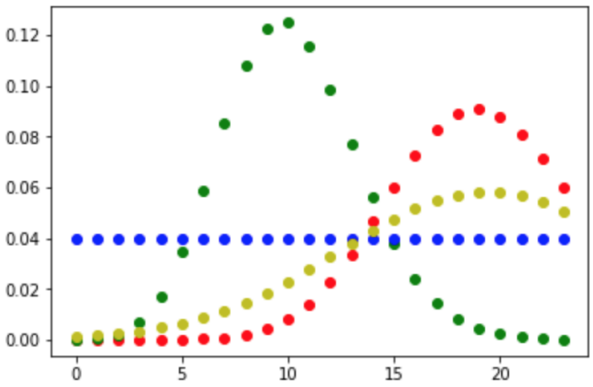

Figure 5: This plot shows the probability pattern of outflow of vehicles from various roads. Area 1 : Blue, Area 2 : Green, Area 3 : Red, Area 4 : Yellow. X-axis is Hours.

We can see that the inflow of vehicles from area 3 is higher in the morning since it is a residential area. Similarly higher inflow of vehicles from area 4 in the morning, while the inflow of vehicles is higher for area 2 in the evening since people leave offices to home. Inflow into area 1 remains kind of constant since it is a commercial area. Similarly, the outflows have been modelled which are more or less inverted versions of the inflow graphs. For Area 3 and Area 2, Poisson distributions was used, for Area 4, Gaussian distribution with standard deviation(v3) picked from uniform distribution between 5-7.

This problem can be formulated as a **Markov Decision Process** as follows:-

State S : At any time instant, the total number of vehicles, halted number of vehicles, cumulative waiting time of the vehicles in each of the 8 lanes at the signal form the state which means a total of 24 feature vector state. Each of this is scaled by dividing by possible maximums which are found by simulating the episode with a suboptimal baseline policy(20 seconds for each traffic signal in turns).

Action A : There are a total of 4*4 actions that are possible. North-South roads in green, East-West roads in green, North-South roads(only left turn) in green and East-West(only left turn) in green. Each of these actions have varying time settings : 8, 16, 24, 32 and followed by 4 seconds of yellow light.

Transition Function P : This will be stochastic given the large number of states possible. It'll be hard to compute though because of large number of states possible, different variety of vehicles existing

on roads and random behavior of drivers on adverse situations like skipping traffic signal/long wait/etc.

Reward function R : Reward here is the negative of the sum of the cumulative waiting time in each of the 8 lanes. It is also scaled down by the time duration of the signal in action to be comparable between actions.

Initial state distribution : It is kind of already determined because there are no vehicles at the start at 00:00 hrs.

Reward discount parameter : Can be set manually between 0-1. However, intuitively it seems like should be close to 1 and experimentally too proves the same.

Q Learning and SARSA

I attempted to solve this problem by using standard Q Learning(3)(4) and SARSA(4) using ϵ-greedy action selection. The state size is 24 and each parameter has been scaled down by it's maximum value which is found by simulating a whole episode in a suboptimal way(20 seconds for each traffic signal in turns) before learning.

The state representation used was in scaled down linear basis. ϵ was initially set to 0.1 and was scaled down until 0.05 by multiplying with 0.99 at every learning step. This helped in learning quicker as observed in initial experiments.

The following are the results from Q and SARSA learning respectively. The parameters that were used are $\gamma = 0.95$, $\alpha = 0.01$, ϵ as described above.

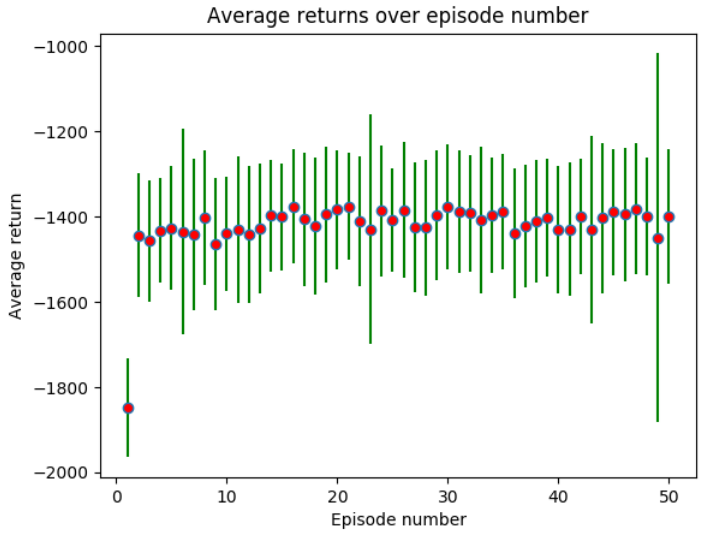

Figure 6: Average returns vs episode for Q Learning

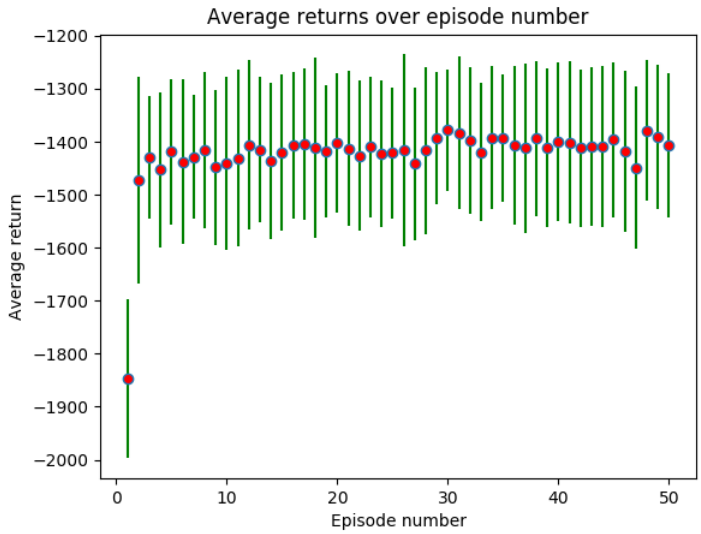

Figure 7: Average returns vs episode for SARSA

As we can see here, the algorithm learns fastly due to the very long length of an episode, we can see that on average, the agent learns the optimal policy by the end of the first episode.

A good improvement is seen with respect to the standard traffic lights rotation at equal time intervals. The average returns using the baseline version of traffic light control is around -1650. This shows the success and possibility of using RL algorithms to solve traffic issues in major cities.

However, even though there is improvement over the baseline, it would have been better to have been able to compare this method of solution with the optimal case. Estimating the optimal reward is tough and presents a tough and good opportunity to develop relevant methods that can help in optimizing it further.

Conclusion

Traffic at an imaginary junction was simulated and Reinforcement Learning algorithms were to optimize the traffic signal control to ensure free flow of traffic. Initial results like certify the usage of Q Learning and SARSA on solving traffic light control problems.

A further study that is really the need of the hour and could be really interesting is to study the behavior of entire traffic in an area/city and using multiple agents to control each traffic light by working together in unison and ensure free traffic flow.

References

[1] Simulation of Urban MObility. *http://sumo.dlr.de/wiki/Simulation_of_Urban_MObility_-_Wiki*

[2] Jeffrey Glick *Reinforcement Learning For Adaptive Traffic Signal Control.*

[3] Watkins, C. J. and Dayan *Q-learning. Machine learning.*

[4] Sutton, R. S. and Barton, A. G. *Reinforcement Learning: An Introduction.*

[5] Prashanth LA, Shalabh Bhatnagar *Reinforcement Learning With Function Approximation for Traffic Signal Control*