Reinforcement Learning based Urban Traffic Control System

Authors: Alvin Dey(MT18066) ,Pranay Raj Anand(MT18079), Swagatam Chakraborti(MT18146)

I. PROPOSED ALGORITHM

We are using the Temporal Difference Q-Learning algorithm which uses the following Q-function to update the Q-Matrix,

$$Q(s,a)^{new} = (1-\alpha)\,Q(s,a)^{old} + \alpha(R(s,a) + \gamma \max_i Q(s',a_i))$$

In Figure 1, q_1 is sum of queue length at stopline 1 and queue length at stopline 2, q_2 is sum of queue length at stopline 3 and queue length at stopline 4, T_1 is the green phase time for stopline 1 and stopline 2, T_2 is the green phase time for stopline 3 and stopline 4

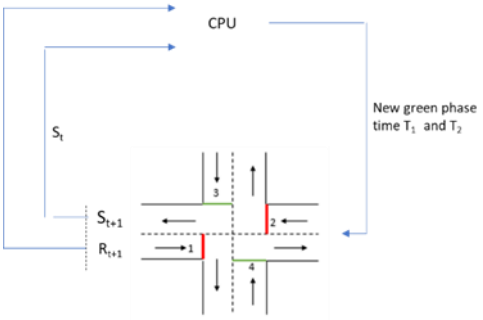

Figure 1 Reinforcement Learning Agent for Traffic Signal Control

II. RULE BASE

A. Calculation of green phase time

$$T_{phase}^{new} = T_{phase}^{old} + \rho\Delta \quad (2)$$

where ρ=i/12 where i=1,2,3….12

Δ is the factor that we consider while changing the phase time and it depicts the green phase utilization time i.e the percent of green phase time used by cars to cross the intersection.

$$\Delta = q - 0.2667 \times T_{phase}^{old} \quad (3)$$

B. Lane occupancy

It is defined as the ratio of number of cars occupying the lanes to the number of cars that can occupy a lane.

$$L_{occupancy} = \frac{(q_1+q_2)}{lane\ capacity} \quad (4)$$

C. Traffic influx

It defined as change in total cars in queue from previous state(5) to current state.

$$C_{influx} = (queue_{previous} - queue_{current})/queue_{previous} \quad (4)$$

where *queue*=q_1+q_2

III. PROBLEM FORMULATION

A. Traffic state formulation: A traffic state is defined as the combination of lane occupancy(3) and traffic influx(4).On basis of lane occupancy: low($L_{occupancy}$<30%) , medium(30%<=$L_{occupancy}$ <70%), high($L_{occupancy}$>70%).On basis of traffic influx :decreasing(C_{influx}<0), stable(30%<=C_{influx}<60%), increasing(C_{influx}>=60%).So together they constitute 9 traffic states. (5)

B. Action: The ρ we mentioned in (2) can have 12 discrete values which will constitute our action space.We are using *Epsilon-Greedy strategy* for selecting an action. It selects an action considering the exploration and exploitation trade-off. (7)

C. Reward: The reward *R* given to agent for taking an action *a* on state *s* is defined below,

$$R(s,a) = (queue_{previous} - queue_{current}) \times 100 \quad (6)$$

where *queue* is as defined in (4)

IV. ALGORITHM

1. Set α=0.1, γ=0.8 and ϵ=0.6 for trade-off between convergence and precision.
2. For episodes=1 to 6
   - 2.1. Create a random initial traffic condition in simulator.
   - 2.2. On the basis of <q_1,q_2,T_1,T_2> given by simulator and the previous state traffic tuple <q'_1,q'_2,T_1',T_2'> the state *s* of current traffic is set using conditions specified in (5)
   - 2.3. Select an action *a* randomly or optimally on basis of Epsilon-Greedy strategy (7)
   - 2.4. Calculate the new phase times T_1 and T_2 using equation (2)
   - 2.5. Set a reward *R(s,a)* according to equation (6)
   - 2.6. Update the Q-value *Q(s,a)* in the Q-Matrix using the equation (1)
   - 2.7. Give the new phase times to the simulator.
   - 2.8. Set new traffic state as the current traffic state.
   - 2.9. Repeat from 2.2 until we have transited through 10 traffic states

V. MODEL SPECIFICATION & DATA

Car: length(5m), max acceleration(1.6m/s^2), max deacceleration(4.4m/s^2),inflow rate(variable). *Intersection:* roads(4), lanes(1 forward and 1 backward), traffic signals(4). edit has been completed, the paper is ready for the template.

VI. RESULTS & ANALYSIS

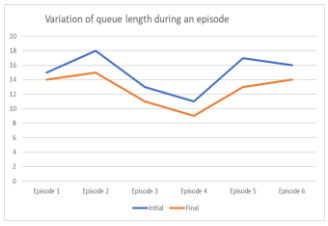

Figure 2 Average queue length vs Episodes

As per observation there is a decrease in the average queue length from initial to final state during an episode. This decrement increases with each episode as the agent learns.

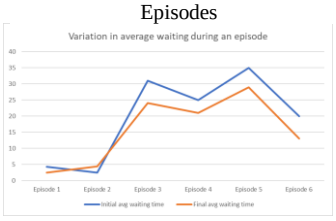

Figure 3 Average waiting time vs Episodes

As per observation there is a decrease in the mean waiting time from initial to final state during an episode. This decrement increases as the agent learns.

VII. FUTURE WORK

1. Integrating one more intersection in the model and using the two intersection to exchange the information of their traffic state with each other and optimize the traffic flow.

2. Comparing and analysing our algorithm results with the results when no such algorithm is employed at the intersection and static green phase times are used.

VIII. REFERENCE

[1] Namrata S.Jadhao,Ashishs S Jadhao. A,"Traffic Signal Control Using Reinforcement Learning", 2014 Fourth International Conference on Communication Systems and Network Technologies,2014.

[2] P.G. Balaji, X. German, D. Srinivasan . An,"Urban traffic signal control using reinforcement learning agents",2009,IET Intelligent Transport Systems,2009.