MSc Thesis Summary: Multi Agent Deep Recurrent Q-Learning for Different Traffic Demands

Azlaan Mustafa Samad

December 2020

1 Introduction

In today's world due to rapid urbanisation there has been a shift of population from rural to urban areas especially in developing countries in search of better opportunities. This has lead to unplanned urbanisation leading to a particularly important issue of increased traffic congestion. This has in turn led to environmental degradation and health issues among people. With the current advancement in Artificial Intelligence, especially in the field of Deep Neural Networks various attempts have been made to apply it in the field of Traffic Light Control. This thesis was an attempt to take forward the problem of solving traffic congestion thereby reducing the total travel time. One of the contributions of the thesis was to study the performance of Deep Recurrent Q-network models in different traffic demands for Multi-Agent systems. Another contribution was to apply different coordination algorithms along with Transfer Learning in Multi-Agent Systems or multiple traffic intersections and study their behaviour. Lastly, the performance of these algorithms were also studied when the number of intersections and the demand increase.

2 Mathematical Framework

The traffic flow problem can be treated as a sequential decision making problem (mathematical framework for Reinforcement Learning), in which an agent (traffic light intersection), first observes the traffic environment. The environment conveys the agent relevant information about the traffic. Then the agent based on its observation takes an action, that is it changes the traffic signals thereby receiving rewards. The goal of the agent is to maximise the reward(or reduce the average travel time). There are two types of RL: Model-based and Model-free. Due to the enormous size of the state space, model free method is viable to use. In model free approach, the agent relies on experience while in model based approach the agent can predict how the environment reacts to its action. The RL algorithm used in the thesis was to predict Q-values using Deep Convolutional Recurrent Neural Network for a given state s when an action a was taken. A Q-value basically tells how good a particular state is. Formally, it is the expected value of rewards when in a state s and taking an particular action a. Thus, a large Q-value means that the expected reward is higher for that state-action pair. The RL agent learns to assign a Q-value to every state-action pair from experience. This means that it tries different actions for different state and thus learn a policy to recommend action for a given state.

3 Implementation and Conclusion

The state was defined in terms of a binary matrix representing a traffic intersection, where 1 corresponds to a car at that position. This is inputted into the neural network which outputs the Q-value. For inferential purpose, the action corresponding to the maximum Q-value is implemented. An open source software Simulation of Urban MObility (SUMO) was used for simulating traffic scenarios. Reward functions were modelled in terms of average delay and average waiting time at each time step where delay accounts for slow moving cars and waiting time for stationary cars. In real life scenario, each traffic intersection is dependent on its nearby traffic intersection and thus it is important for them to coordinate in order to reduce the total travel time of the cars. This was done using three different algorithms namely Max-Plus (MP), Brute Force (BF) and Individual Coordination (IC). In MP, the different intersections communicate with each other to share information and take actions while in BF, the action combination which yields the maximum Q-value is selected. In IC, each intersection acts independently and takes action corresponding to its maximum Q-value. The variant of Transfer learning used was to train agents on single, two or three intersection and then reuse them on bigger intersections.

This was implemented for different traffic demands and for different number of traffic light intersections. The Transfer Learning approach saves a lot of computational time and energy. The algorithm performance for IC was unpredictable while its good for MP and BC. With increase in demand and number of intersections, fluctuations are introduced in the average travel time and performances are hard to predict.