**Traffic Light Signal Control using Deep Reinforcement Learning**

An efficient transportation system is the need of the hour in countries like India where incessant traffic congestions have become increasing costly and pose an obstacle towards unceasing mobility. Traffic conditions need to be refined in order to improve the economy and ease the daily life of people and hence an intelligent traffic light control is indispensable for alleviating increasing traffic pressure. However, existing systems and methodologies for controlling traffic signals are insufficient for addressing the problem as they are not devised by observing real traffic data. These traditional traffic controllers such as Adaptive Traffic Signal Control (ATSC) based on sensors placed around traffic intersections and pre-defined fixed-time controllers optimized by traffic engineers are unable to sufficiently resolve or ease the traffic congestions by accommodating changing traffic patterns. Though they are capable of regulating nominal traffic flow, they are vulnerable to unexpected changes in traffic patterns and abnormal circumstances such as accidents, construction, and other events. Deep reinforcement learning (DRL) has been shown to have the potential to dynamically adjust traffic lights according to real-time traffic, in recent studies. Recent discoveries in the deep learning community can be quite valuable in designing learning-based controllers that meet the required goals in this setting.

H. Wei et al [1] proposed an effective deep reinforcement learning model, IntelliLight for traffic light control and tested their method on a large-scale real traffic dataset obtained from surveillance cameras after training their model on synthetic data using the simulation platform SUMO (Simulation of Urban MObility). The developed model considered a four-way road intersection as the environment with two phases of traffic light as green light on west-east direction (Green-WE) and red light on west-east direction (Red-WE). They designed an intelligent traffic light agent which would achieve the goal of giving the optimal **action** of changing the light to the next phase or keeping the current phase, in order to maximize the reward in the long run. The **reward** **function** is defined as the weighted sum of the factors: sum of queue length, sum of average waiting time for lanes, number of light switches, sum of delay for vehicles, total time (in minutes) that vehicles spent on approaching lanes and total number of vehicles that passed the intersection during the time interval after an action. The traffic light agent takes an appropriate action according to ϵ-greedy strategy combining exploration and exploitation after observing the environment state at equal time intervals. The **state representation** includes queue length on a lane, number of vehicles on that lane, average waiting time of vehicles on that lane, an image representation of vehicles’ position, and current and next phase of traffic light. The agent learned a **Deep Q-network** for estimating the reward. To resolve the difficulties in distinguishing the decision process for different phases, the study proposed a special sub-structure known as Phase Gate and also utilized the Memory Palace theory to improve the fitting capability of the network to predict the reward accurately. Their method IntelliLight is able to achieve the best reward, queue length, delay and duration over the best performance among baseline methods on real world data.

K. Tan et al [2] proposed a DRL-based adaptive traffic signal control framework that explicitly considers realistic traffic scenarios, sensors, and physical constraints. In this framework, they also proposed a novel reward function that shows significantly improved traffic performance compared to the typical baseline pretimed and fully-actuated traffic signals controllers. The framework utilized readily available real-world data sensor streams to learn the optimal policy for the agent in the simulation platform VISSIM. The environment considered is a simple four-way intersection having one lane in each direction. The **state** is formulated as the average travel time and upstream queue length for each lane. The possible **actions** are changing the current phase or keeping it. The reward function comprises of the weighted sum of Queue length for all upstream lanes, Total delay for vehicles in all upstream lanes, Total number of vehicles crossed the intersection, and Total number of residual queues. The framework was tested using **Deep-Q network (DQN)** algorithm and the DRL agent’s performance was tested on real traffic data during high traffic demand periods which outperformed both the baseline controllers.

A. Vidali et al [3] presented a Reinforcement Learning approach to traffic lights control, coupled with a microscopic agent-based simulator (Simulation of Urban MObility - SUMO). The environment taken into account is a simple four-way intersection having 4 lanes per arm. The **state** design includes only spatial information about the vehicles hosted inside the environment to let the agent know the position of vehicles at each step. The agent obtains the set of cells that describe the presence or absence of vehicles in every area of the incoming roads upon observing the state of the environment. The **action** set consists of four actions: North-South Advance, North-South Left Advance, East-West Advance, and East-West Left Advance. Two agents were trained using two different **reward** functions: total waiting time and accumulated total waiting time. The **Deep Q-learning** mechanism which is a value based RL method was involved in the study conducted.

D. Li et al [4] built a truly adaptive traffic signal control model in a traffic microsimulator, i.e., “Simulation of Urban Mobility” (SUMO), using the technology of modern deep reinforcement learning. The model proposed was based on a **deep Q-network algorithm** and both **single-agent** (for single 4-way intersection) **and multiagent** (for multiple intersections) cases were demonstrated. The possible **actions** are changing the current phase or keeping it with the decision being made on fixed time intervals. Four traffic signal **phases** (NS-G, NSL-G, EW-G, EWL-G) are taken into consideration. The **reward** function utilized is made inversely proportional to the average waiting time of each vehicle. They tested with data sets pertaining to three different traffic conditions, and proved that the proposed model is better than other methods (e.g., Q-learning method, longest queue first method, and Webster fixed timing control method) for all cases. The proposed model reduces both the average waiting time and travel time, and it becomes more advantageous as the traffic environment becomes more complex.

Sajad et al [5] proposed two kinds of reinforcement learning algorithms: deep policy-gradient and value-function based agents to predict the best possible traffic signal. According to their working, at each time step, the agent receives snapshot of the current traffic state from their choice of traffic simulator which was again SUMO. In their policy-gradient agent, there is mapping of observations to the control signal whereas in the value-functioned agent, first the values are estimated for all legal signals, and the optimal one is selected. The action set in their work contains only 2 actions – North south green and East-West green, one of them can be true at a time. The reward function their experimentation is the total cumulative delay between two consecutive actions, i.e., the positive reward values imply the taken actions actually leads to a decrease in the total cumulative delay, and a negative value indicated increase in delay. They performed gradient descent on the policy parameters. In the value-functioned based approach, they used deep neural networks to estimate the action-value function, which maps the input state to the action values (i.e., the future reward that can be achieved for that given state and action). They evaluated their performance based on average reward, average cumulative delay, and average queue length.

Liang et al [6] used data collected from different sensors and vehicular networks which are then converted to small grids. They then experiment with a variable traffic signal green light duration and propose a model with double Q-Learning networks, dueling networks and prioritized experience replay. Again, they also perform all their simulations in the Simulation of Urban MObility (SUMO). The general idea is that they extend the duration for the phase that has more vehicles in that direction, and hence this adaptive phase duration optimization is their key problem. Their states contain two pieces of information about the vehicle – their position and their speed. Their action space contains four phases – north-south green, east-north & west-south green, east-west green, and east-south & west-north green, thus considering all types of turns in a 4-way crossroad. Their choice of reward function is the cumulative waiting time between two neighboring cycles. It means the reward is equal to the increment in cumulative waiting time between before taking the action and after the action. They use state of the art techniques along with CNN and propose a network termed Double Dueling Deep Q Network, and its specifications are explained. Their evaluation method is to maximize the defined reward – cumulative delay of all vehicles, and reducing the average waiting time of vehicles, and also perform an evaluation under rush hour, which means that the traffic flow from all lanes is not homogenous, and is indicative of the real-life conditions where the traffic flow rate from a particular direction is 2-3 times the flow from the other direction.

**Possible Problem definitions**

We can model our environment as a four-way intersection with one lane in each direction and train and validate the Deep Reinforcement Learning (DRL) agent on the microsimulator SUMO across a number of episodes (each episode simulating fixed hours of traffic simulation time in SUMO). The central intersection controller for this intersection will be the DRL agent. Our agent design problem will consist of three fundamental components of DRL: state, action, and reward.

STATE

Some notable RL approaches towards defining states are grid [7] and sensor-based representations [8,9,10]. However, these works are limited by scalability as the state and action spaces grow exponentially for larger traffic networks. With neural networks, various research papers have utilized convolutional neural networks (CNN) [11] to capture state representations of the network via visual top-down images [1], discretized matrix representation [12,13,14,15], etc.

We can formulate the state as average travel time, queue length, number of vehicles, average waiting time of vehicles for each lane and an image representation of vehicles’ position, and current and next phase of traffic light.

ACTION

Since our problem only involves a simple intersection, the agent can only choose between two legal actions— change current phase conditions or keep it, otherwise we can even include more than two specific actions such as described in the literature above [2],[4],[6]. When an action is selected, a fixed amount of time will last before the agent can select a new configuration [12]. Some works gave the agent more flexibility by defining phase duration with variable length [16]. Hence, we can have either fixed or variable phase durations for the agent.

REWARD

Researchers experimented with a multitude of reward signals over the past few years to properly define a solid indicator for the DRL agent to learn the optimal policy. There are some which focus on typical traffic congestion metrics like change in cumulative delay [10, 17], average delay [12, 15, 9], and multiple attributes [18, 1, 14, 19]. Most of the works include the calculation of the change between cumulative vehicle delay between actions, where the vehicle delay is defined as the number of seconds the vehicles is steady [20], [21]. Similarly, the cumulative vehicle staying time can be used, which is the number of seconds the vehicle has been steady since his entrance in the environment [12]. Moreover, some works combine multiples indicators in a weighted sum [1], and we can devise a reward function based on the same or investigate different reward functions for different agents as mentioned in the above literature [4].

METHOD/MODEL/LEARNING ALGORITHM

There are three types of reinforcement learning methods: value-based, policy-based and actor-critic methods [22]. Value-based methods are generally known as Q-learning. In Q-learning, a Q-table is used to store each state and a corresponding Q-value owned by each action in this state, but maintaining a Q-table is quite expensive when the environment becomes very complex. Another algorithm Deep Q Network (DQN), which combines the benefits of Q-learning and convolutional neural networks (CNNs) (to compute the environment state and learn features from an image [21] or a spatial representation [20], [12]) can overcome this problem very well. We can choose Deep-Q Network algorithm (DQN) to test our framework because of the simplicity and sample efficiency of DQN algorithms (value-based) compared to policy-based algorithms.

**References**

1. H. Wei et al, “IntelliLight: A Reinforcement Learning Approach for Intelligent Traffic Light Control,” Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining, pp. 2496–2505, July 2018.
2. K. Tan et al, “Deep Reinforcement Learning for Adaptive Traffic Signal Control”, Proceedings of the ASME 2019, Dynamic Systems and Control Conference DSCC2019, November 2019.
3. A. Vidali, L. Crociani, G. Vizzari, and S. Bandini, ‘‘A deep reinforcement learning approach to adaptive traffic lights management,’’ in Proc. Workshop ‘From Objects Agents’, pp. 42–50, 2019.
4. Li, D.; Wu, J.; Xu, M.; Wang, Z.; Hu, K. “Adaptive Traffic Signal Control Model on Intersections Based on Deep Reinforcement Learning,” J. Adv. Transp. 2020, 2020.
5. Sajad Mousavi, S., Schukat, M. and Howley, E., “Traffic Light Control Using Deep Policy-Gradient and Value-Function Based Reinforcement Learning,” arXiv e-prints, pp. arXiv-1704, 2017.
6. Liang, X., Du, X., Wang, G. and Han, Z.,“Deep reinforcement learning for traffic light control in vehicular networks,” arXiv preprint, arXiv:1803.11115, 2018.
7. Wiering, M.,“Multi-agent reinforcement learning for traffic light control”. In Machine Learning: Proceedings of the Seventeenth International Conference (ICML’2000), pp. 1151–1158, 2000.
8. Medina, J. C., Hajbabaie, A., and Benekohal, R. F.,“Arterial traffic control using reinforcement learning agents and information from adjacent intersections in the state and reward structure”. In 13th International IEEE Conference on Intelligent Transportation Systems, pp. 525–530, 2010.
9. Arel, I., Liu, C., Urbanik, T., and Kohls, A.,“Reinforcement learning-based multi-agent system for network traffic signal control”. IET Intelligent Transport Systems, 4(2), pp. 128–135,2010
10. El-Tantawy, S., Abdulhai, B., and Abdelgawad, H.“Multiagent reinforcement learning for integrated network of adaptive traffic signal controllers (marlin-atsc): methodology and large-scale application on downtown toronto”. IEEE Transactions on Intelligent Transportation Systems, 14(3), pp. 1140–1150,2013.
11. Le Cun, Y., Bengio, Y., et al.,“Convolutional networks for images, speech, and time series”. The handbook of brain theory and neural networks, 3361(10), p. 1995.
12. Gao, J., Shen, Y., Liu, J., Ito, M., and Shiratori, N.,“Adaptive traffic signal control: Deep reinforcement learn learning algorithm with experience replay and target network”. arXiv preprint arXiv:1705.02755, 2017.
13. Genders, W., and Razavi, S.,“Using a deep reinforcement learning agent for traffic signal control”. arXiv preprint arXiv:1611.01142, 2016.
14. Muresan, M., Fu, L., and Pan, G.,“Adaptive traffic signal control with deep reinforcement learning an exploratory investigation”. arXiv preprint arXiv:1901.00960, 2019.
15. LIU, M., DENG, J., XU, M., ZHANG, X., andWANG,W., “Cooperative deep reinforcement learning for traffic signal control”, 2017.
16. L. Li, Y. Lv, and F.-Y. Wang, “Traffic signal timing via deep reinforcement learning,” IEEE/CAA Journal of Automatica Sinica, vol. 3, no. 3, pp. 247–254, 2016.
17. Liang, X., Du, X., Wang, G., and Han, Z., “A deep reinforcement learning network for traffic light cycle control”. IEEE Transactions on Vehicular Technology, 68, pp. 1243–1253, 2019.
18. Van der Pol, E., and Oliehoek, F. A. “Coordinated deep reinforcement learners for traffic light control”.
19. Chu, T., Wang, J., Codec`a, L., and Li, Z., “Multiagent deep reinforcement learning for large-scale traffic signal control”. IEEE Transactions on Intelligent Transportation Systems, 2019.
20. W. Genders and S. Razavi, “Using a deep reinforcement learning agent for traffic signal control,” arXiv preprint arXiv:1611.01142, 2016.
21. S. S. Mousavi, M. Schukat, and E. Howley, “Traffic light control using deep policy-gradient and value-function-based reinforcement learning,” IET Intelligent Transport Systems, vol. 11, no. 7, pp. 417–423, 2017.
22. Sutton, R. S., and Barto, A. G., 2018. Reinforcement learning: An introduction.