Adaptive Traffic Control System using Reinforcement Learning

Kranti Shingate
Department of Computer Engineering
FCRIT, Vashi, Navi Mumbai,
India

Komal Jagdale
Department of Computer Engineering
FCRIT, Vashi, Navi Mumbai,
India

Yohann Dias
Department of Computer
Engineering FCRIT, Vashi, Navi
Mumbai, India

Abstract— **The advent of the automobile revolution has led to various traffic congestion problems. People can't arrive at their destination on time because of gigantic traffic. The framework utilized for coordinating traffic isn't reliant on the ongoing situation of an intersection. Traffic Light Control System with pre-set clocks are broadly used to invigilate and control the traffic generated at the intersections of numerous streets. However, the synchronization of multiple traffic light systems at adjacent intersections is a complicated problem given the various parameters involved. To handle such traffic either expansion of road networks or adaptive traffic control system which handles such traffic intelligently. This paper presents a system which handles traffic using Artificial Intelligence technique for adapting signal according to the density of traffic thereby automatically increasing or decreasing traffic signal time using Experience Replay mechanism. In this system, the Reinforcement Learning algorithm was used to determine optimal traffic light configuration and using deep Neural Networks the obtained results were used to extract the features required to make a decision.**

Keywords— Reinforcement Learning (RL), Traffic Light Control System (TLCS), Experience Replay mechanism, Artificial Intelligence, Deep Neural Networks

I. INTRODUCTION

Traffic congestion is ceaselessly developing everywhere throughout the world and it has become a hindrance for commuters. As an outcome of increasing population and urbanization, the transportation request is consistently rising in the cities around the world. The broad routine traffic volumes carry pressures to existing urban traffic foundation, bringing about ordinary traffic clogs. One more issue in automobile overload is a deferral of red light. Traffic blockage can likewise be advanced by huge red light delay. This delay issue is incited because lights in the rush hour gridlock control are systematized and it isn't subject to actual traffic.

The existing system widely used are traffic signals with pre-set timers which operate under fixed time operation and display green light to each approach for the same time every cycle regardless of the traffic conditions. This may be best suited for heavily congested areas but for low traffic density, the sequence is not as beneficial as no vehicles are waiting. With advancements in technology, the Adaptive Traffic Signal Control System has been developed in Bhubaneswar city. The system gets input from sensors embedded in the road and synchronizes the group of traffic signals accordingly. This signalling system is run on solar power.

The system is infeasible and costly since it requires a system embedded in roads. LQF (longest queue first) scheduling algorithm minimizes the queue sizes at each approach to the intersection. The goal is to lower vehicle delay as compared to a current state signal control method. A focus is given by giving preference vehicles (such as emergency vehicles or large trucks). As the system concentrates on reducing the queue length, the stability of the system is a major concern. Therefore way out for this issue is a Traffic Control System using reinforcement learning (RL) – an AI structure that endeavours to estimate an ideal basic leadership policy.

The framework gives a solution for diminishing traffic in metropolitan urban areas by contemplating constant traffic situations and the reinforcement learning algorithm to improve after some time. Since the traditional traffic control framework utilizes basic convention that alternate green and red light for a fixed interval. Such traffic control frameworks work admirably when there is a limited amount of traffic network. To build up a traffic control framework that handles and directs the traffic shrewdly using reinforcement learning algorithm and to accomplish smooth transportation of vehicles and to curb natural issues like raised air pollution, wastage of fuel and danger of mishap.

In contrast, the adaptive traffic control system offers a response to reduce traffic in metropolitan urban networks by considering continuous traffic circumstances and reinforced learning computation to improve over time. The system will without a doubt examine a 4-way intersection for the incoming traffic density to take an optimized step towards reducing it. The estimation used learns over a while so the fundamental periods of system most likely won't give perfect results for the recognized traffic.

As a result, a reinforcement learning algorithm that naturally extracts all highlights (machine-created highlights) helpful for versatile traffic signal control from raw real-time traffic data furthermore and learns the ideal traffic signal control arrangement is needed.

II. RELATED WORK

The reinforcement learning-based framework uses the present traffic situation for creating an improved traffic light configuration. The recognition of vehicles in a partially observable condition utilizes DSRC (Dedicated short-range communication) [1]. If an intersection consists of an enormous distinction of detected and undetected vehicles then the agent makes a one-sided move in support of detected vehicles .Hence, this system gives better outcomes for vehicles empowered with remote correspondence than for ones that stay undetected. Another significant component that improves algorithm stability is the experience replay and target network [2] used during the training phase of the agent (traffic signal).

The experience replay mechanism contains the information needed for learning in the form of a randomized group of samples is called 'batch'. This information is submitted to the agent, but instead of immediately submitting the information that the agent gathers during the simulation, the batches are stored in a data structure called memory. In this memory, every sample is stored which is collected during the training .Also this framework can give precise outcomes as it utilizes machine crafted features instead of human-crafted features (e.g. vehicle queue length, position and speed of vehicles) for analysing real-time traffic and developing an optimal policy for adaptive traffic signal control.

To make the system increasingly receptive to the actual traffic it is necessary to emphasize the feasibility and value of applying a model-less temporal difference reinforcement learning algorithm [3] for traffic light control. The main drawback of such a system is environment involves four-way intersections but allows traffic stream in either horizontal or vertical not both [3].

The multi-agent system for network traffic signal control introduces the use of a multi-agent system and reinforcement learning algorithm to obtain an efficient traffic signal control [4].In this, two types of agents are used i.e. central agent and an outbound agent. The outbound agents schedule traffic signals using the Linear Queue First (LQF) algorithm and the central agent learns a value function (Q-learning) driven by its local and neighbour's traffic conditions. At low arrival rates, the LQF scheduling algorithm performs slightly better than the multi-agent Q Learning system.

Adaptive traffic signal control, which adjusts traffic signal timing according to real-time traffic, is an effective method to reduce traffic congestion. Another set of multi-agent model-based Reinforcement Learning systems was formulated under the Markov Decision process model for traffic light control [5]. The system does not rely on heuristics equations but learns the optimal control by improving its experience by interacting with the environment. Such systems can be improved by adding public transport which should give priority for public transport since they carry more passengers [5].

TABLE I.

COMPARATIVE STUDY OF TRAFFIC CONTROL SYSTEM USING REINFORCEMENT LEARNING

No.	Title	Methodology	Limitations
1.	Intelligent Traffic Signal Control: Using Reinforcement Learning with Partial Detection	RL algorithm for partially observable ITS based on DSRC. (Dedicated Short Range Communications).	Better performance for the detected vehicles than undetected vehicles.
2.	Adaptive Traffic Signal Control: Deep reinforcement learning algorithm with experience replay and target network	Reinforcement learning algorithm using experience replay and target network	Gives accurate results for machine - crafted features only
3.	Reinforcement Learning for Intelligent Traffic Light Control	To determine the feasibility and value of applying a model-less temporal difference Reinforcement learning algorithm to traffic light control.	Environment involves four-way intersections but allows traffic to flow in either horizontal or vertical not both.
4.	Reinforcement learning-based multi-agent system for network traffic signal control	The outbound agents schedule traffic signals (LQF algorithm) and central agent learns a value function (Q-learning) driven by its local and neighbours traffic conditions.	At low arrival rates, LQF scheduling algorithm performs slightly better than the multi-agent Q Learning system. Require lane system.
5.	Multi-Agent Reinforcement Learning for Intelligent Traffic Light Control	A set of multi-agent model-based Reinforcement Learning system for traffic light control	Should give priority to public transport, since they carry more passengers

III. PROBLEM FORMULTION AND DESIGN

In modern cities, we have lots of traffic on the road and most of the time traffic management systems will not be able to handle such traffic congestion problems. Traffic problems may occur due to emergencies, construction on road sites, or tourist vacations etc. Traditional traffic light control system or traffic signal with the pre-set timers have fixed cycles of changing phase or alternatives signals which is not suitable for real world traffic congestions and results in inefficient traffic flow.

Current traffic controllers are either pre timed control system or actuated control systems. Pre timed control systems have pre-set of timings on green signal light. A longer green light duration during peak hours and a shorter duration during afternoons and late nights. Another is actuated control systems which is responsive to dynamic traffic but does not really serve well in long term traffic scenarios. To avoid or overcome this issues, adaptive or intelligent traffic light control system have been designed to cope up with real time traffic congestion problems.

So, chances of the modification in traffic flow through an intersection which is managed by traffic light controller. The analysis will be conducted with simulation where an agent is used to make a choice of which traffic light should be activated in order to reduce the traffic congestion problem and optimizing the traffic efficiency. To choose the best action in every situation, some learning mechanisms are used by learning agent and those learning techniques are related to reinforcement and deep learning algorithms.

Each reinforcement learning framework has two primary parts - Agent and Environment (Fig. 1.). The agent is the traffic light framework that is liable for taking actions. The environment is the present state of distribution of vehicles in an intersection.
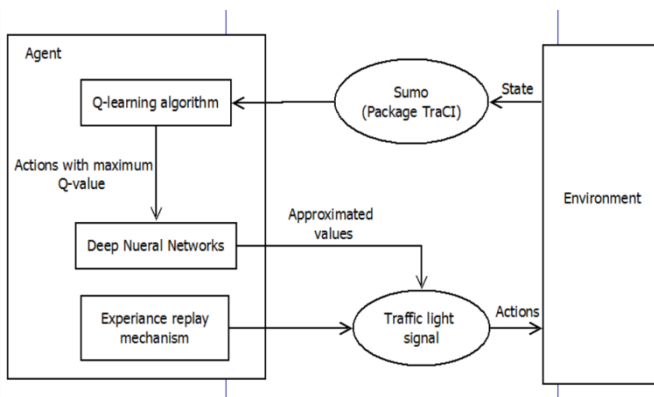

Fig. 1. Workflow of the Traffic Control System

In the proposed framework, the Sumo (Simulation of Urban Mobility) test system produces an arbitrary number of vehicles picking irregular source and distribution, which in turn gives a contribution as a state to the agent. This state is utilized by the Q-learning algorithm and the action with the highest Q-value is picked. Deep neural systems are utilized to get the approximated values to improve results and traffic light signal performs actions that influence the environment (vehicle distribution).

In the design of this system which is based on the reinforcement learning algorithm, it is important to characterize the environment, states, actions and rewards and learning mechanisms involved. In the simulation, the environment is represented by 4-way intersection which contain 4 incoming lanes and 4 outgoing lanes (Fig. 2.).Each incoming lane defines the possible direction that vehicles can follows: left most lane is used by left turn only, right most lane is used for right turn and for going straight and two middle lanes are dedicated to only going straight.

In the environment, there are 8 traffic lights traffic lights which are indicated by a colour on the stop line of every incoming lane that represents the status of traffic light for that particular lane. For example, whenever cars are coming from south direction and if that vehicle want to go straight or turn right then as Fig. 2. shows green for that particular lane and red for remaining lanes.
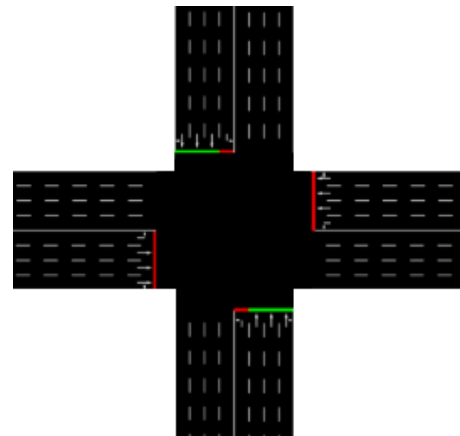

Fig. 2. Four-way intersection without vehicles

Rules for traffic lights:

- The colour phase transition for every traffic light is always red-green-yellow-red.

- Duration for traffic light is fixed.10 seconds for green traffic light and 4 seconds for yellow traffic light and duration for red traffic light is defined as the amount of time since the last phase change.

- At least one traffic light is yellow or green phase

- Every traffic light is not in the red phase simultaneously.

In each lane of the crossing point, incoming lanes are discretized (Fig. 3.) in the cells that can recognize the appearance or non-appearance of vehicle inside them. In each arm there are 20 cells.10 of them are set along the left most path whereas 10 are set within others three lanes. So, in entire system there are total 80 cells.
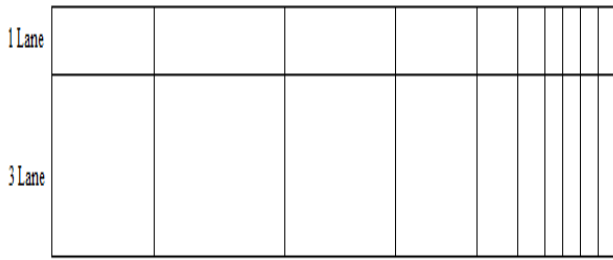
Fig. 3. Design of state representation

IV. LEARNING METHODOLOGY

A. Reinforcement learning Algorithm

In real life, we play out various undertakings to seek after our dreams. After performing tasks, we get a few prizes which is either positive or negative. Along with this rewards, we continue investigating various ways and attempt to make sense of which activity may prompt better rewards. For reasons unknown, the entire thought of reinforcement learning is truly observational in nature. Reinforcement learning is a branch of artificial intelligence which lets machine learns on your own in a way different from traditional machine learning. Reinforcement learning is nothing but taking suitable action to maximize reward in particular situation.

In reinforcement learning, input should be an initial state from which the model start and there are many possible outputs as there are a variety of solutions to a particular problem. The training is based upon the input, the model will restore a state and the user will decide to reward or rebuff the model based on its output.

Rewards in Reinforcement is either positive or negative based on the decision taken by agent. Positive reward is characterized as when an event occurs due to a specific behaviour, builds the quality and the frequency of the behaviour. In other words it has a positive effect on the behaviour. After getting positive reward agent maximizes the performances and sustain change for a long period of time. Negative Reinforcement is defined as the strengthening of a behaviour because a negative condition is stopped or avoided. At the point when agent gets negative rewards it increases behaviour and provide defiance to minimum standard of performance.

Adaptive traffic signal control system is a system that can be implemented using RL techniques. Reinforcement learning offers numerous feasible solutions to address the traffic flow problem. To solve these problems it emerges different algorithm and neural network. In this, one or more independent agents have the objective of increasing the proficiency of traffic flow that drives through one or more crossing point controlled by traffic light controller. To describe the context of traffic light signal controller RL components that is state, action, reward are widely used.

Several reasons to use RL for traffic light control system:

● Agent can make decision without supervision prior knowledge of the environment

● Agent can adopt different situations such as accident or weather conditions

● Agent learns using the system performance i.e. rewards so there is no need to describe every variable of the environment

B. Q-learning algorithm

The agent's learning mechanism is Deep Q-learning. The q-learning function learns from activities that are outside the present policy, such as taking random activities, and hence a policy isn't required. Q-learning seeks to learn a policy that maximizes the total reward. The 'q' in q-learning represents quality. Quality for this situation speaks to how helpful a given activity is in increasing some future reward. It is a blend of Deep Neural Networks and Q-learning. Q-learning is a basic yet very powerful algorithm for our agent since this enables the agent to make sense of precisely which activity to perform. In deep Q-learning, we utilize a neural system to approximate the Q-value function. The state is given as the input and the Q-value of every single imaginable activity is created as the output. It is a model free reinforcement method which includes assigning a Q-value to an activity performed by the agent.

Steps involved in reinforcement learning using deep Q-learning networks:-

a. All the past experience is stored by the user in memory
b. The next action is determined by the maximum output of the Q-network
c. The loss function here is mean squared error of the predicted Q-value and the target Q-value – Q*. This is basically a regression problem. However, we do not know the target or actual value here as we are dealing with a reinforcement learning problem. Going back to the Q-value update equation derived from the Bellman equation.
Q-value is defined as-

$$Q(s_t, a_t) = Q(s_t, a_t) + \alpha(r_{t+1} + \gamma \cdot \max_A Q(s_{t+1}, a_t) - Q(s_t, a_t)) \quad (1)$$

Where,

$Q(s_t, a_t)$ is value of action at performed in state st
$Q(s_{t+1}, a_t)$ is the Q-value of immediate next step
r_{t+1} is reward agent gets after performing action at

γ is the discount factor determines the significance of future rewards. Discount factor ranges between 0 and 1. Discount factor 0 makes the agent opportunistic by only thinking about current rewards whereas factor 1 make it strive for long term rewards.

α is the learning rate in which factor 0 will make the agent not learning anything whereas a factor 1 defines that agent consider only the most recent information.

The Q-value for state-action is upgraded by an error, balanced by the learning rate(α).The learning rate decides to what degree recently obtained data overrides old information. Q-value speaks to the conceivable rewards gotten within another time stamp for performing an action in state S, also the

discounted future rewards obtained from the next state-action perception.

C. Reinforcement learning Models

We consider an adaptive traffic light control system, which takes reward and state perception from the environment and chooses an action. In this subsection, we introduce our design of actions, rewards, and states (Fig. 4.).

1) State: Positions of vehicles inside the environment. This system is used for controlling traffic generated in a 4-way intersection with four incoming lanes and four outgoing lanes per arm.

2) Agent Action: A configuration of the traffic light that implies the green phase for some lanes for a fixed amount of time. Car follows the possible directions defined by the incoming lanes: left-most lane (left-turn only),right-most lane (right-turn and straight) two middle lanes (only for going straight)

3) Reward: Reward is kind of compensate which is received by agent after performing some action a based on some states. Decision taken by agent might be right or wrong for particular state. Rewards can be either positive or negative. A positive reward is a consequence of a good action, while a negative reward is received after a bad action.

4) Environment: Environment is the place where agent performs action based on some decision. In the environment, there are 8 traffic lights traffic lights which are indicated by a colour on the stop line of every incoming lane that represents the status of traffic light for that particular lane.
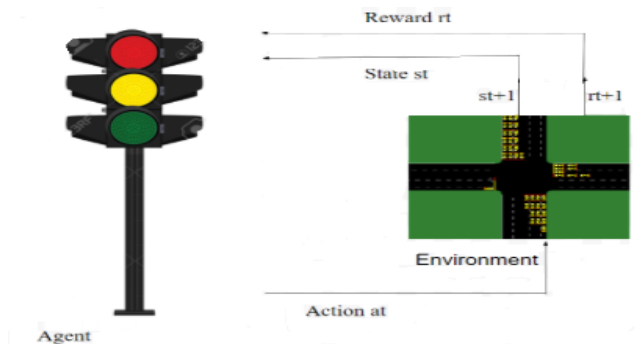

Fig. 4. Traffic light as reinforcement learning agent

V. CONCLUSION

The learning agent of the Adaptive traffic control system is designed with state representation that identifies the position of vehicle in environment and makes decisions according to real time traffic. Based on the decision agent gets reward which is further used by the agent to make appropriate decisions to reduce traffic on the basis of its rewards. The system can be designed as a multi-agent system to take decisions for more than one intersection at a time. Also public transport, emergency vehicles like fire brigade, ambulance should be given higher preference. The system requires lane system for its working so an improvisation in the system for functioning on roads without lanes can be implemented.

REFERENCES

[1] Rusheng Zhang ,Akihiro Ishikawa,Wenli Wang,Benjamin Striner ,Ozan Tonguz:"Intelligent Traffic Signal Control: Using Reinforcement Learning with Partial Detection"

[2] Juntao Gao, Yulong Shen, Jia Liu, Minoru Ito and Norio Shiratori : "Adaptive Traffic Signal Control: Deep Reinforcement Learning Algorithm with Experience Replay and Target Network"

[3] "Reinforcement Learning for Intelligent Traffic Light Control"

[4] I. Arel, C. Liu, T. Urbanik, A.G. Kohls : "Reinforcement learning-based multi-agent system for network traffic signal control"

[5] "Multi-Agent Reinforcement Learning for Intelligent Traffic Light Control"