International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 1 Issue 4, June - 2012

Reinforcement Learning Based for Traffic Signal Monitoring and Management

Ms Namrata S. Jadhao[1], Mr. .Parag A. Kulkarni[2]

[1]*G.H.Raisoni college of Engineering & management, Wagholi, Pune*
[2]*R & D Head, EKLaT Research , Pune*

ABSTRACT - ***To obtain more accurate patterns insight into traffic signal by analyzing within and between day variations in traffic volumes, using the methods of machine learning. Proposed system is based on reinforcement learning (RL) for traffic signal control. RL uses multi agent structure where vehicles and traffic signals are working as agents. Reinforcement learning is to learn the optimal policy by a trial-and-error process including observing the environment and choosing an action according to current states and receiving rewards from the environment. The policy which maximizes the expected long-term reward is considered as the optimal one System objective is to optimize traffic states using RL-algorithm. This paper describes traffic management using reinforcement learning based on paramic simulation. Expected outcomes of the algorithm will work more efficiently than other traffic system.***

Keywords—**Agent Based System, Intelligent Traffic Signal Control, Multi Objective Scheme, Optimization Objectives, Reinforcement Learning.**

I. INTRODUCTION

The research objective involves optimal control of a heavily congested traffic across a two dimensional road network. RL is a field of study in machine learning where an agent, by interacting with and receiving feedback from its environment, attempts to learn an optimal action selection policy [9]. A promising approach is to make use of machine learning techniques to control the traffic. Such methods allow the control system to automatically learn a good, or even optimal policy. Thus, intelligent algorithms have been used in attempts to build an efficient traffic control system, such as fuzzy control technology, artificial neural network and genetic algorithm, which greatly improve the efficiency of traffic control [5]. An RL problem is defined once states, actions and rewards. At each simulation time step, the local state of an intersection is based on local traffic statistics are clearly defined and according to it action is selected. A reward is provided to an intersection agent after executing a given action. The reward ranges from -1 to 1. On the other hand, the agent is subject to a penalty if an increased average delay is observed [9]. This paper uses reinforcement learning (RL) to optimize the traffic light controllers in a traffic network. Reinforcement learning is basically a method of machine learning algorithms consisting of Q learning, temporal difference ,SARSA algorithm and so on .Reinforce learning is a self-learning algorithm which doesn't need an explicit model of the environment. Hence it can be applied in traffic signal control effectively to response to the frequent change of traffic flow and outperform traditional traffic control algorithm. Reinforcement learning is to learn the optimal policy by a trial-and-error process including observing the environment and choosing an action according to current states and receiving rewards from the environment. The policy which maximizes the expected long-term reward is considered as the optimal one [5].

Q learning a form of reinforcement learning in which the agent learns to assign values to state-action pairs. We need first to make a distinction between what is true of the environment and what the agent thinks is true of the environment. First let's consider what's true of the world. If an agent is in a particular state and takes a particular action, we are interested in any immediate reinforcement that's received but also in future reinforcements that result from ending up in a new state where further actions can be taken, actions that follow a particular policy. Given a particular action in a particular state followed by behavior that follows a particular policy, the agent will receive a particular set of reinforcements. This is a fact about the world. In the simplest case, the Q-value for a state-action pair is the sum of all of these reinforcements, and the Q-value function is the function that maps from state-action pairs to values. The derivation of the class of prediction-learning techniques that are now formally known as Temporal Difference learning TD(λ) procedures. TD procedures are particularly attractive in that they allow for weight updates based just on the current state xt, and the next state xt+1 [1].

Thorpe studied reinforcement learning for traffic light control in1997. He used a neural network to predict the waiting time for all cars standing at the intersection and

selected the best control policy using Sarsa algorithm Abdulhai et al. presented a basic framework of applying Q-learning to traffic signal control and got effective results while applying it to an isolated intersection. MIKAMI el al. combined evolutionary algorithm and reinforcement learning for cooperative traffic signal control .However, the above methods used traffic-light based value functions which means a large number of states need to be handled. Therefore, these methods suffer from the "dimension curse" and result with limited success when applied to large-scale road network. Wiering et al. utilized a car-based value function to solve this problem. They made a predictor for each car to estimate the overall waiting time given possible choices of a traffic light using reinforcement learning, and selected the decision which minimized the sum of waiting time of all cars in the network. This method effectively reduced the states space and thus can be applied to large network control. Experiment in a network with 12 edge nodes and 16 junctions proved the effectiveness of this method. . In real traffic system, consider different optimization objectives in different conditions, which is called multi-objective control scheme .In this paper, in the free traffic condition, we try to minimize the overall number of vehicles stops of the network; while in the medium traffic condition, the overall waiting time is considered as the optimal goal. In congested traffic condition, queue spillovers must be avoided to keep the network from large-scale congestion, thus the queue length must be focused on. Therefore, multi-objective control scheme can adapt to various traffic conditions and make a more intelligent control system.

II *AGENT BASED MODEL OF TRAFFIC SYSTEM*

A more advanced approach to traffic simulation and optimization is the Agent based System approach in which agents interact and communicate with each other and the infrastructure. We use an agent-based model to describe the practical traffic system. Vehicles and traffic signal controllers in the road network are regarded as two types of agents. Exchanging of data can takes place between these agents. The Wiering's model is used to built the road network as shown in figure 1. There are six possible settings for each traffic controllers to prevent accidents: two traffic lights from opposing directions allow cars to go straight ahead or to turn right, two traffic lights at the same direction of the intersection allow the cars from there to go straight ahead, turn right or turn left. The capacity of each road lane is defined according to its practical length. At each time step, new cars are generated with a particular destination and enter the network from outside. After new cars have been entered, traffic light decisions are made and each car moves to the subsequent lane if it is not occupied or the car's predecessor is moved forward. Thus, each car is at a specific traffic node, a direction at the node i.e. dir, a position in the queue i.e. place and has a particular destination des. Thus we can use [node, dir, place, des] to denote the state of each vehicle The optimization objectives include waiting time, stops and queue length, which will be selected according to the traffic situation. We use Q([node, dir, place, des],action) to denote the total expected value of optimized indices for all traffic lights for each car until it arrives at the destination given its current node, direction, place and the decision of the light. It should be noticed Q([node, dir, place, des],action) doesn't only refer to the waiting time but also stops and queue lengths. This is the most import difference between our model and Wiering's model.

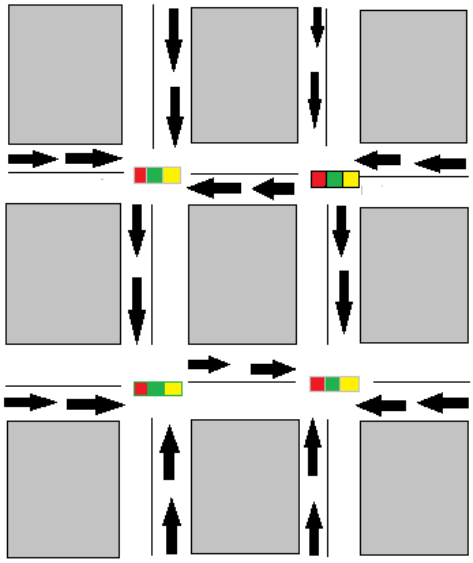

Fig 1 Agent Based Model

III *MULTI RL ALGORITHM*

The control algorithm is extended to a multi-objective scheme by choosing optimization objective according to real-time traffic condition. The multi-objective control algorithm considers three types of traffic situations as follows less traffic situation, medium traffic situation and congested traffic situation.

A. *Less traffic condition*

In this condition, our goal is to minimize the number of stops.

$$Q([n,d,p,des],Green) = \sum_{(n',d',p')} P(Red|[n',d',p',des])$$

$$(R[n,d,p,des],[n',d',p',des] + \gamma Q([n',d',p',des],Green)$$

The probability that a traffic light turns red is calculated as follows

$$P(Red|[n,d,p,des]) = \frac{C([n,d,p,des],Red)}{C([n,d,p,des])}$$

$$IF(vechicle\ speed == 0)\{$$

$$Q([n,d,p,des],Red) = current\ time;$$

$$
\begin{aligned}
&\}\\
&IF(vechicle\ speed\ !=0)\{\\
&Q([n,d,p,des]),Green)=current\ time;\\
&\}\\
&Now\\
&Waiting\ time=Q([n,d,p,des],green)-Q([n,d,p,des],red)
\end{aligned}
$$

Here waiting time of each vehicle at each signal is culating. The number of stops will increase when a vehicle moving at a green light in current time step meet a red light in the next time step.

B. *Medium Traffic condition*

In this condition, our goal is to minimize the overall waiting time of vehicles.

$$V([n,d,p,des])=\sum_{L}P(L|[n,d,p,des],L)Q([n,d,p,des],L)\quad(3)$$

$$
\begin{aligned}
&Q([n.d,p.des]L)=\sum_{(n',d',p')}P([n,d,p,des],L[n',d',p',des])\\
&(R[n,d,p,des],[n',d',p',des])+\gamma V([n',d',p',des]))
\end{aligned}
$$

$$
\begin{aligned}
&IF(vechicle\ speed\ ==0)\{\\
&Q([n,d,p,des],Red)=current\ time;\\
&\}\\
&IF(vechicle\ speed\ !=0)\{\\
&Q([n,d,p,des]),Green)=current\ time;\\
&\}\\
&Now\\
&Waiting\ time=Q([n,d,p,des],green)-Q([n,d,p,des],red)
\end{aligned}
$$

C. *Congested traffic condition*

In this condition, spillovers of queue must be avoided which will minimize the traffic control effect and probably cause large-scale traffic congestion.

$$
\begin{aligned}
&Q([n,d,p,des],Green)=\sum_{(n',d',p')}P([n,d,p,des],Green(n',d',p',des])\\
&(R|[n,d,p,des],[n',d',p',des]+\alpha R'([n,d,p,des],[n',d',p',des])\\
&+\gamma Q([n',d',p',des]))
\end{aligned}
$$

$$
\begin{aligned}
&Q([n,d,p,des],Red)=\sum_{(n',d',p')}P([n,d,p,des],Red,[n',d',p',des])\\
&(R([n,d,p,des],[n',d',p',des])+\gamma V([n',d',p',des]))
\end{aligned}
$$

$$
\begin{aligned}
&IF(vechicle\ speed\ ==0)\{\\
&Q([n,d,p,des],Red)=current\ time;\\
&\}\\
&IF(vechicle\ speed\ !=0)\{\\
&Q([n,d,p,des]),Green)=current\ time;\\
&\}\\
&Now\\
&Waiting\ time=Q([n,d,p,des],green)-Q([n,d,p,des],red)
\end{aligned}
$$

The queue length is taken into consideration when design the Q learning procedure. Denote the maximum queue length at the next traffic light as tl' , can be written as K. The capacity of the lane of next traffic light, *L is given* , then the adjusting factor α is determined by the queue length *K*.

$$\alpha=\left\{\begin{matrix}0\\10\left(\frac{K-0.8}{2}\right)\\L\end{matrix}\middle|\begin{matrix}if\ K\leq 0.8L\\if\ 0.8L<K\leq L\\ifK>L\end{matrix}\right\}$$

D. *Priority Control For Buses And Emergent Vehicles*

Emergent vehicles such as ambulances enter the road network, they should have priority to pass through. To realize the priority control of these special vehicles without or least disturbance to the regular traffic order is very essential. So that a priority factor is added to describe the emergent degree of these special vehicles.

$$
\begin{aligned}
&Q([N,D,P,DES],L)=\sum_{(n',d',p')}P([n,d,p,des],L[n',d',p',des])\\
&(\beta R([n,d,p,des],[n'.d'.p'.des])+\gamma V([n',d',p',des]))
\end{aligned}
$$

$$
\begin{aligned}
&IF(vechicle\ speed\ ==0)\{\\
&Q([n,d,p,des],Red)=current\ time;\\
&\}\\
&IF(vechicle\ speed\ !=0)\{\\
&Q([n,d,p,des]),Green)=current\ time;\\
&\}\\
&Now\\
&Waiting\ time=Q([n,d,p,des],green)-Q([n,d,p,des],red)
\end{aligned}
$$

IV RESULTS

Since it is very hard to apply a signal control model to real traffic system management, traffic simulation was chosen to do the case studies. Paramics V6.3 was selected as the simulation platform because it is a professional traffic simulation tool . A practical road network was modeled in Paramics containing 7 intersections (N1-N7)and 8 OD zones(Zone1-Zone8). The simulation ran for 10000 time steps, the former 4000 steps was the learning process, and the latter

6000 steps was used to collected the simulation results. Factor γ is set to be 0.9 and β is set to be 3. The lanes in the network are divided into cells with length of 7.5 m. The capacity of the lanes equals to the number of the cells. We compared our method with fixed control and variable control. In our model, when the traffic volume entering the network in a minute is less than 90, it is regarded as free traffic; when the volume is larger than 90 but less than 180, it is regarded as medium traffic; when the traffic volume is larger than 180, it is regarded as congested traffic condition.

TABLE I
COMPARISION OF FIXED CONTROL , VARIABLE CONTROL AND RL.

Time Slot	Fixed Time (sec.)	Variable Time (sec.)	Time from Learning (Reinforcement Learning) (sec.)
9-12	90	60	30
12-3	90	30	10
3-6	90	30	10
6-9	90	60	40

V CONCLUSION

In this paper, we have presented the multi-objective control algorithm-I based on reinforcement learning. The simulation indicated that the multi-RL-I got the minimum stops under free traffic, although not the minimum waiting time; the multi-RL had the similar performance with the RL method under medium traffic, which was better than fixed control and variable control; under congested condition, multi-RL-I could effectively prevent the queue spillovers to avoid large scale traffic jams. There are still some system parameters that should carefully be determined by hand. For, example, the adjusting factor α indicating the influence of the queue at the next traffic light to the waiting time of vehicles at current light under congested traffic condition. This is a very important parameter, which we should further research its determining way based on traffic flow theory. In addition, some phenomenon in real traffic system such as the lane changing of cars will influence their travel time. We should further take these into consideration and build a model more close to the real traffic system.

REFERENCES

[1] Temporal Difference Learning: A Critique. ESWAR SIVARAMAN Submitted In Partial Fulfillment Of The Course Requirements For "Neural Networks" ECEN 5733 May 2000

[2] Intelligent Traffic Light Control,Marco Wiering, Jelle Van Veenen, Jilles Vreeken, And Arne Koopman Intelligent Systems Group ,Institute Of Information And Computing Sciences Utrecht University Padualaan 14, 3508TB Utrecht, The Netherlands Email: Marco@Cs.Uu.Nl July 9, 2004

[3] Sutton, R. S., And Barto, A. G. ~1998!. Reinforcement Learning—An Introduction,MIT Press, Cambridge, Mass.

[4] Thorpe, T. L. ~1997!. ''Vehicle Traffic Light Control Using SARSA".

[5] Multi-Objective Reinforcement Learning For Traffic Signal Coordinate Control YIN Shcengchao; DUAN Houli; LI Zhiheng: ZHANG Yi.

[6] Sutton, R. S., And Barto, A. G. 1998!. Reinforcement Learning—An Introduction,MIT Press, Cambridge, Mass.

[7] Reinforcement Learning For True Adaptive Traffic Signal,Control Baher Abdulhai; Rob Pringle; And Grigoris J. Karakoulas,In May/June 2003

[8] SELF-ORGANIZING URBAN TRAFFIC CONTROL ARCHITECTURE WITH SWARM-SELF ORGANIZING MAP IN JAKARTA: SIGNAL CONTROL SYSTEM AND SIMULATOR ,W. Jatmiko, A.Azurat ,Herry, A. Wibowo,H. Marihot,M. Wicaksana, . Takagawa,K. Sekiyama, And T. Fukuda.-2010

[9] Reinforcement Learning-Based Multi-Agent System For Network Traffic Signal Control I. Arel, C Liu, T. Urbanik, A.G. Kohls In 2010