Improved Multi-Agent Reinforcement Learning for Minimizing Traffic Waiting Time

Vijay Kumar
M.T.U
India

B. Kaushik
K.E.C., M.T.U.,
India

H. Banka
ISM, India

ABSTRACT
This paper depict using multi-agent reinforcement learning (MARL) algorithm for learning traffic pattern to minimize the traveling time or maximizing safety and optimizing traffic pattern (OTP). This model provides a description and solution to optimize traffic pattern that use multi-agent based reinforcement learning algorithms. MARL uses multi agent structure where vehicles and traffic signals are working as agents. In this model traffic area divide in different-different traffic ZONE. Each zone have own distributed agent and these agent will pass the information one zone to other threw the network. The Optimization objectives include the number of vehicle stops, the average waiting time and maximum queue length of the next (node) intersection. In addition, This research also introduce the priority control of buses and emergent vehicles into this model. Expected outcome of the algorithm is comparable to the performance of Q-Learning and Temporal difference learning. The results show significant reduction in waiting time comparable to those algorithms for the work more efficiently than other traffic system.

General Terms
Learning Algorithm, Artificial Intelligence, Agent based learning.

Keywords
Agent Based System, Intelligent Traffic Signal Control, Multi Objective Scheme, Optimization Objectives, RL, Multi-Agent System (MAS).

1. INTRODUCTION
Manage the traffic in high traffic areas is a big problem. Increasing population size requires more efficient transportation systems and hence better traffic control system. Even developed countries are suffering high costs because of increasing road congestion levels. In the European Union (EU) alone, congestion costs 0.5% of the member countries' Gross Domestic Product (GDP) [11], [8], and this is expected to increase to roughly 1% of the EU's GDP by 2009 if the problem is not dealt with properly. In 2002, the number of vehicles per thousand persons had reached 460 which is nearly double the number (232) in 1974.In high traffic situations and bad driving in the EU (European Union) accounts for up to 50% of fuel consumption on road networks resulting in deadly emissions that could otherwise be diminished. High traffic transport contributes 41% of carbon dioxide to give out from road traffic in the EU thus resulting in serious health and safety problems. In these cases to avoid the high costs that give by these threats, UTC has to provide some solutions to the problem of traffic management [11], [8]. To achieve the global goal UTC optimization, increasing global such threats and vehicles infrastructure communicating between some systems may be provide some extra detail .These detail may provide help for local view of the traffic conditions.

In case medium traffic conditions the Wiering's method reduce the overall waiting time for vehicles. This method reduce the waiting time for vehicles and optimize the goal. In real traffic system, this model should consider different-different optimization objectives in different traffic situation, which is called multi-agent control scheme in this paper. In the free traffic situation, presented model try to minimize the overall number of stops of vehicles in the traffic network. In case medium traffic situation this research tries to minimize the waiting time on behalf optimal goal. In congested traffic condition main focused on queue length. So multi-agent control scheme can adapt to different traffic conditions and make a more intelligent traffic control system. Therefore, this model, propose a multi-agent control strategy using MARL. Multi-objective control and paramic simulation model both have some problems .First node traffic situation pass to the all next nodes. If first node has a free traffic, this condition will passes all the next nodes, this is not good way for real traffic so this model will calculate traffic situation individually for each node.

In congested traffic situation, queue spillovers must be avoided to keep the network from large-scale congestion, thus the queue length must be focused on [6]. In this model cycle is prevented. The value of β is not fix (3) it depends on traffic control admin in this model .This may be 4, 5 etc. On behalf the value of β this model will manage green light for emergent vehicles in traffic network. In this model data exchange between vehicles and roadside traffic equipments is necessary, thus vehicular ad hoc network is utilized to build a wireless traffic information system.

Therefore distributed network helpful for utilized to develop a wireless traffic information system. Different researchers have chosen variant types of artificial intelligence algorithms and methods for the optimization of the traffic flow in real traffic conditions. Genetic algorithm or evolutionary algorithm is one of the most common methods introduced into the traffic control system. Routing of traffic flow using genetic algorithm has shown some improvement in the traffic control. Fuzzy logic control is also useful into the traffic light systems for better control of traffic flow. Increase performance of real traffic light system is build with some idea such that increases green light time period for vehicles. Another approach to improve the traffic control is using wireless network communications between vehicles and traffic control systems to get traffic information for traffic flow. This information can use for optimization in traffic system in medium and high traffic conditions. Reinforcement learning technique is used in certain research studies for the traffic flow control and

optimizations. So reinforcement learning technique can be applied in traffic signal control effectively to response to the frequent change of traffic flow and outperform traditional traffic control algorithm that helpful for optimality, reducing traffic delay and build a better traffic light system.

This model are minimizing travel time or maximizing safety, Minimizing vehicle travel time, reducing traffic delay, increasing vehicle velocity, and prioritizing emergency traffic Since OTP controllers by hand is a complex and tedious task ,this research study how multi-agent reinforcement learning(MARL) algorithms can be used for this goal.

2. AGENT–BASED MODEL OF TRAFFIC SYSTEM

In this model use an agent-based model to describe the practical traffic system. In the road, there are two types of agent one is vehicles and another is traffic signal controllers called as distributed agents. Traffic information will be exchange between these agents. There are some possibility for each traffic controllers that prevent traffic threats and accidents. Two traffic lights from opposing directions allow vehicle to go straight ahead to turn right, two traffic lights at the same direction of the intersection allow the vehicle from there to go straight ahead, turn right or turn Left. When new vehicle have been added the traffic light decisions are made and each vehicle moves to cell if cell is not occupied .This decision control by the traffic system according to traffic conditions. There for, each vehicle is at a traffic node, a direction at the node (dir), a position in the queue (place) and has a particular destination (des). This model use [node, dir, place, des], in sort ([n, d, p, des] to denote the state of each vehicle [7].The main object is optimization with reduce waiting time ,number of stops and traffic queue length.

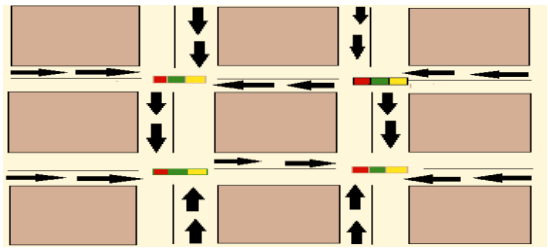

Fig 1: Agent Based Model.

In this model Q([n,d,p,des], action) to represent the total expected value of optimized indices for all traffic lights for each vehicle. This process will be continue until vehicles arrive at the destination goal. In Wiering's model, consider first node traffic situation pass to the all next nodes. If first node has a free traffic, this condition passes all the next nodes but this model will calculate traffic situation individually for each node. This is the most import difference between this model and Wiering's model.

3. REINFORCEMENT LEARNING FOR TRAFFIC CONTROL

Previously several methods for learn traffic have been developed like Sarsa and Q-learning .These all techniques suffered with same problem in high traffic conditions. In urban or congested traffic, these technique are not scale to multi-agent Reinforcement Learning. In urban traffic may be possible that traffic grows dynamically. So need a dynamic method for handle urban traffic that grow dynamically. Q-learning and Sarsa they are applied only to small network. One name is Reinforcement Learning that support dynamic environment using dynamic programming. A more popular approach is to use model-based reinforcement learning, in which the transition and reward functions are estimated from experience and then used to find a policy via planning methods like dynamic programming.

3.1 Simple model

Figure 2 shows the learning process of an agent. At each time step, the agent receives a reinforcement feedback from the environment along with the current state. The goal for the agent is to create an optimal action selection policy p to maximize the reward. In many cases, not only the immediate reward but also the subsequent rewards Delayed rewards? should be considered when actions are taken.

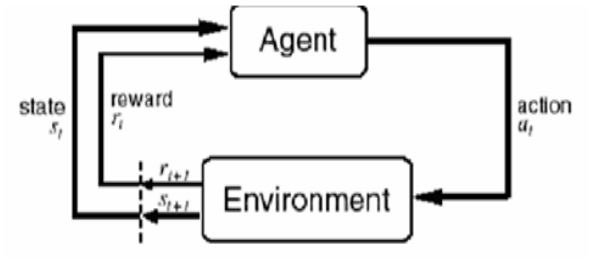

Fig 2: Agent with state and action

Agent and environment interact at discrete time steps:
$t = 0,1,2,k$

Agent observes state at step: $t : s_t \in S$

Produces action at step: $a_t \in A(s_t)$

Gets resulting reward: $r_{t+1} \in R$

And resulting next state: s_{t+1}

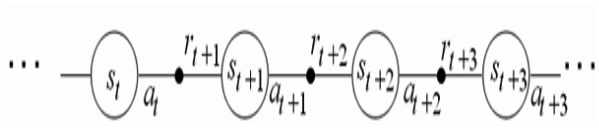

Fig 3: A general process model of RL [8]

3.2 This Basic Elements of Reinforcement Learning

1. Model of the process
2. Reward functions.
3. Learning objective.
4. Controllers.
5. Exploration.

3.3 Multi-agent Frame work

The multi-agent framework is based on the same idea of Figure 2 but, this Time, there are several agents deciding on actions over the environment. The big difference resides in the fact that all each agent probably has some effect on the environment and, so, actions can have different outcomes depending on what the other agents are doing. Next Fig. shows the multi-agent model or framework.

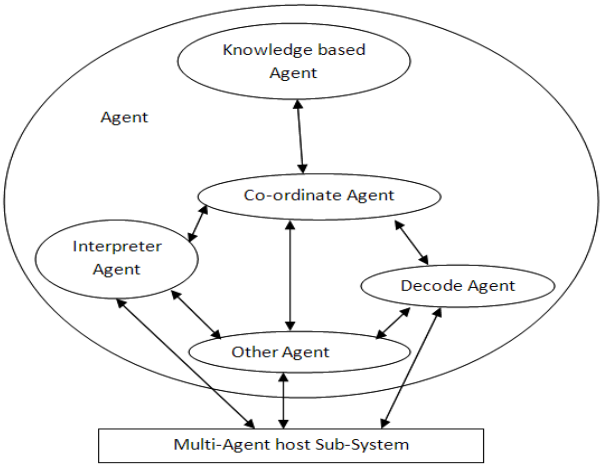

Fig 4: ***Multi-Agent Model***

In addition to benefits owing to the distributed nature of the multi-agent solution, such as the speedup made possible by parallel computation, multiple RL agents can harness new benefits from sharing experience, e.g., by communication, teaching. Conversely, besides challenges inherited from single-agent RL, including the curse of dimensionality.

4. MULTI-AGENT CONTROL ALGORITHM BASED ON REINFORCEMENT LEARNING

The multi-agent control algorithm considers three types of traffic situations as follows less traffic (low traffic or free traffic) situation, medium traffic situation and congested traffic situation.

4.1 Free traffic condition

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. In free traffic condition the main goal is to minimize the number of stops. So use Q ([node, dir, pos, des], Green) as the expected cumulative number of stops. The formulation of Q ([node, dir, pos, des], Green) is shown as follows.

$$Q([node, dir, pos, des], Green) = \sum_{(node', dir', pos')} P(\mathrm{Re}d \mid [node', dir', pos', des])(R([node, dir pos, des], [node', dir', pos', des]) + \gamma Q([node', dir' pos', des], Green)) \quad (1)$$

Where $[node', dir' pos', des])$ means the state of a vehicle in next time step; $P(\mathrm{Re}d \mid [node', dir', pos', des])$ gives the probability that the traffic light turns red in next time step; $R([node, dir, pos, des], [node', dir' pos', des])$ is a reward function as follows: if a vehicle stays at the same traffic light, then R=1, otherwise R=0, (the vehicle gets through this intersection and enters the next one); γ is the discount factor ($0 < \gamma < 1$) which ensure the Q-values are bounded. The probability that a traffic light turns red is calculated as follows.

$$P(\mathrm{Re}d \mid [node', dir', pos', des]) = C([node, dir, pos, des], \mathrm{Re}d) / C([node, dir, pos, des]) \quad (2)$$

Where $C([node, dir, pos, des]$ is the number of times a vehicle in the state of $[node, dir, pos, des]$, $C([node, dir, pos, des], \mathrm{Re}d)$ is the number of times the light turns red in such state.

4.2 Medium traffic condition

In medium traffic condition main goal of this model is to minimize the overall waiting time of vehicles. If number of vehicles are larger 100 but less than 150, it is consider as medium traffic.

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

$$Q([n, d, p, des]L = \sum_{(node', dir' pos)} P([n, d, p, des], L[n', d', p', des])(R[n, d, p, des], [n', d', p', des]) + \gamma V(n', d', p', des]\beta)) \quad (4)$$

Where is L the traffic light state (red or green), $P(L \mid [n, d, p, des]$, is calculated in the same way as equation 2, $(R[n, d, p, des], [n', d', p', des])$ is defined as follows as: if a vehicle stays at the same traffic light, then R=1, otherwise R=0 and β use for force to be green light $\beta = 10 - \beta$.

4.3 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.

$$Q([node, dir, pos, des], Green) = \sum_{(node', dir' pos)} P([node, dir, pos, des], Green, [node', dir', pos' des]) (R([node, dir, pos, des], [node', dir' pos', des]) + \alpha R'([node, dir, pos, des], [node', dir' pos', des]) + \gamma V(node', dir', pos', des'])) \quad (5)$$

$$Q([node, dir, pos, des], \mathrm{Re}d) = \sum_{(node', dir', pos')} P([node, dir, pos, des], \mathrm{Re}d, [node', dir', pos, 'des]) (R([node, dir, pos, des], [node', dir', pos' des]) + \gamma V([node', dir', pos', des]\beta)) \quad (6)$$

Where $Q([node, dir, pos, des], L)$ and $V([node, dir, pos, des]$ have the same meanings as

under medium traffic condition. Compared equation 5 with equation 4, another reward function $R'([node, dir, pos, des],[node', dir' pos', des])$ is added to indicate the influence from traffic condition at the next node, and β use for force to be green light, $\beta = 10 - \beta$ $R([node, dir, pos, des],[node', dir' pos', des])$ Is the reward of vehicles waiting time while $R'([node, dir, pos, des],[node', dir' pos', des])$ indicates the reward from the change of the queue length at the next traffic node. Consider queue length when design Q learning procedure, $t_{l'}$ denote the max queue length at next traffic light so $t_{l'}$ can written as K. L is the capacity of the lane of next traffic light and α is the adjusting factor that determine queue length $K_{tl'}$ as follows:

$$\alpha = 0 \begin{cases} 0 & IF \quad K_{tl'} < 0.8L \\ (k_{\underline{Tl-0.8}}) \quad 1.0 & IF \quad 0.8L < K_{tl'} \\ L \quad .2 & IF \quad K_{tl>L} \end{cases} \quad (7)$$

The largest α value is set to .2 in this model.

4.4 Priority Control for Emergent Vehicles

In case emergency vehicles like Fire Truck ambulances, Prime Minister Vehicles etc. so need to manage traffic light when these conditions were arise. For these situations give high priority for these types of vehicle. The traffic administrator can manage traffic light according to traffic conditions. If emergency condition arises the admin of traffic control can reduce time of the green light that is set priority according to type of vehicles for green light. In priority condition the main focus manage green light on behalf this, present model can reduce waiting time for emergency vehicles.

$$Q([node, dir, pos, des], L) = P([node, dir, pos, des], Green([node', des', pos'])(\beta R([node, dir, pos, des],[node', des' pos' des]) + \gamma V[node', dir', pos' des])) \quad (8)$$

5. RESULT

In this research 1000 time steps use for simulation. For learning process 2000 steps use, and 2000 steps were also used for simulation result. The value 0.9 set to factor γ in this model. β is set to be according to emergent Vehicles situation that is for Fire Truck and ambulance the priority of green light may be differ, not 3(fix). If in a minute number of vehicles are 100 entering in traffic network, it is consider as free traffic. If number of vehicles are larger 100 but less than 150, it is consider as medium traffic, and number of vehicles are larger than 150 it is consider as congested (high traffic) traffic condition.

5.1 Comparison of average waiting time

Comparison of average waiting time regard to the increasing of traffic volume rapidly is shown in figure 5 .TD means temporal difference, QL means Q-learning algorithm, MARL means Multi-agent reinforcement learning algorithm the model proposed in this paper. The next table shows a data set used in TD, QL, and MARL.

Table 1 Visiting Points with Q-Capacity and Q-Length

visiting Points	q-capacity	q-length
Lambeth	1000	50
Watford	500	150
WestDrayton	800	100
Leatherhead	900	200
Otford	800	700
Dartford	950	200
Loughton	600	105
Aylesford	800	600

Table 2 Visiters distance

Visitors	Lambeth	Watford	WestDrayton	Leatherhead	Otford	Dartford	Loughton	Aylesford
Lambeth	0	25	30	28	-1	27	22	-1
Watford	25	0	40	-1	-1	-1	52	-1
WestDrayton	30	40	0	45	-1	-1	-1	-1
Leatherhead	28	-1	45	0	47	-1	-1	-1
Otford	-1	-1	-1	47	0	22	-1	35
Dartford	27	-1	-1	-1	22	0	32	33
Loughton	22	52	-1	-1	-1	32	0	-1
Aylesford	-1	-1	-1	-1	-1	33	-1	0

In Table 2 visitors distance,-1 show there is no any path between two visitor nodes.

Number of stops under the multi-agent RL control will be less than those under other control strategies like TD and Q-learning. Reinforcement learning who minimize number of stops comparable to TD and Q-learning technique in case medium traffic and congested traffic conditions.

6. CONCLUSION

This paper presented the multi-agent RL control algorithm based on reinforcement learning. The simulation indicated that the MARL got the minimum waiting time under free traffic, comparable QL, TD. MARL 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 node 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 fuzzy logic approach such as crisp to fuzzy conversion such as Lambda cuts for minimizing traffic pattern. Neural network as a tool can also be used for detecting trends in traffic patterns and to predict minimal waiting time for traffic.

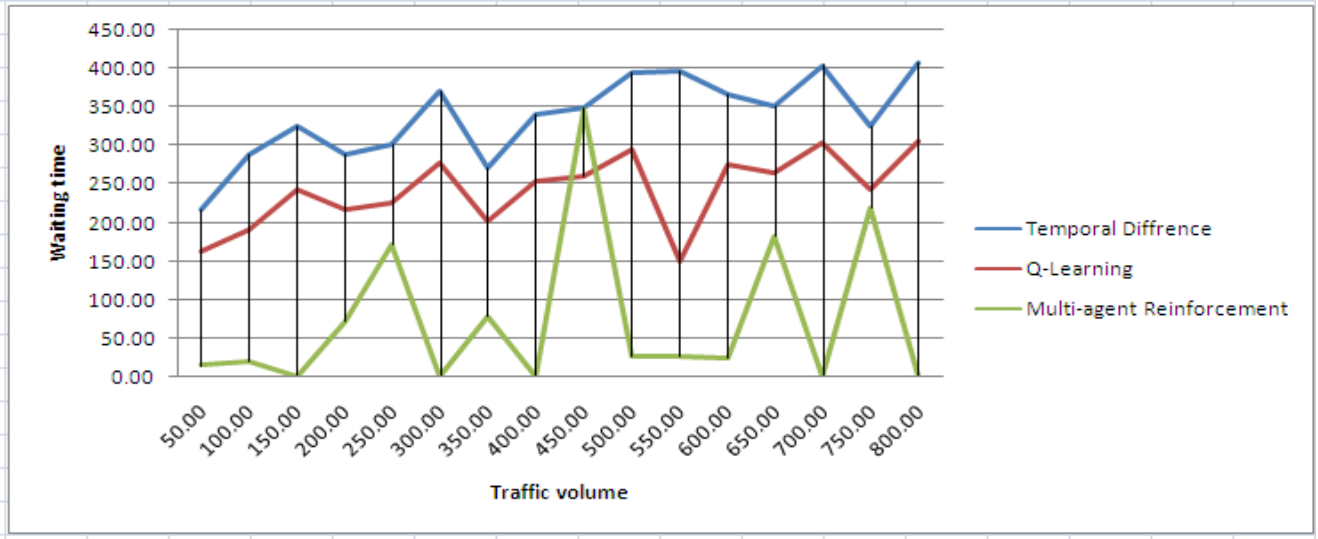

Fig 5: Simulation between TD, QL and MARL by increasing the opposite traffic length.

7. ACKNOWLEDGMENTS

First and foremost, I would like to express my sincere thanks to my paper advisor Associative Prof. Baijnath Kaushik for providing me their precious advices and suggestions. This model wouldn't have been a success for me without their cooperation and valuable comments and suggestions.

I also want to express my gratitude to Prof. P. S. Gill (H.O.D.) and Associative Prof. Sunita Tiwari (M.Tech. Coordinator) for their support, kind help, continued interest and inspiration during this work.

8. REFERENCES

[1] Bowling,M.: Convergence and no-regret in multiagent learning. In: L.K.Saul, Y.Weiss, L. Bottou (eds.) Advances in Neural Information Processing Systems 17, pp. 209–216. MIT Press (2005).

[2] Bus¸oniu, L., De Schutter, B., Babuˇska, R.: Multiagent reinforcement learning with adaptive state focus. In: Proceedings 17th Belgian-Dutch Conference on Artificial Intelligence (BNAIC-05), pp. 35–42. Brussels, Belgium (2005).

[3] Chalkiadakis, G.: Multiagent reinforcement learning: Stochastic games with multiple learning players. Tech. rep., Dept. of Computer Science, University of Toronto, Canada (2003).

[4] Guestrin, C., Lagoudakis, M.G., Parr, R.: Coordinated reinforcement learning. In: Proceedings 19th International Conference on Machine Learning (ICML-02), pp. 227–234. Sydney, Australia (2002)

[5] Hu, J., Wellman, M.P.: Nash Q-learning for general-sum stochastic games. Journal of Machine Learning Research 4, 1039–1069 (2003)

[6] M.Wiering, et al (2004). Intelligent Traffic Light Control. Technical Report UU-CS-2004-029, University Utrecht.

[7] M.Wiering (2000). Multi-Agent Reinforcement Learning for Traffic Light Control. Machine Learning: Proceedings of the 17th International Conference (ICML' 2000), 1151-1158.

[8] Mitchell, T. M. (1995) the Book of Machine Learning: McGraw-HILL INTERNATIONAL EDITIONS.

[9] Nunes L., and Oliveira, E. C. Learning from multiple sources. In Proceedings of the 3rd International Joint Conference on Autonomous Agents and Multi Agent Systems, AAMAS (New York, USA, July 2004), vol. 3, New York, IEEE Computer Society, pp. 1106–1113.

[10] Oliveira, D., Bazzan, A. L. C., and Lesser, V. using cooperative mediation to coordinate traffic lights: a case study. In Proceedings of the 4th International Joint Conference on Autonomous Agents and Multi Agent Systems (AAMAS) (July 2005), New York, IEEE Computer Society, pp. 463–470.

[11] Price, B., Boutilier, C.: Accelerating reinforcement learning through implicit imitation Journal of Artificial Intelligence Research 19, 569–629 (2003).

[12] Tan, M.: Multi-agent reinforcement learning: Independent vs. cooperative agents. In: Proceedings 10th International Conference on Machine Learning (ICML-93), pp. 330– 337. Amherst, US (1993).