Urban traffic control based on pre-adapted reinforcement learning

Mahshid Helali Moghadam, Nasser Mozayani
Computer Engineering Department
Iran University of Science and Technology
Tehran, Iran
mahshidhelali@comp.iust.ac.ir, Mozayani@iust.ac.ir

Abstract— **Urban traffic control is one of the appropriate research grounds in various artificial intelligence fields such as multiagent systems and learning methods. Dynamism, continuous changes of states, and the necessity to give prompt response are among the specific characteristics of the environment in a traffic control system. Proposing an appropriate and flexible strategy to meet the existing requirements is always a major challenge in the traffic control field. In this article, we have proposed an efficient method to control urban traffic using multiagent systems and a kind of reinforcement learning augmented by an auxiliary pre-learning stage. In this technique, the agent primarily uses some statistical traffic data and applies traffic theories to compute appropriate values for the traffic system parameters. Having these primary values, then it enters into a reinforcement learning phase and starts the learning based on the obtained basic knowledge. Experimentally, the agents in the proposed method showed better learning performance than that of pure RL; and the results show that this method gives a better control on traffic based on theoretical values and different conditions in the environment.**

Keywords-urban traffic control; multiagent system; auxiliary pre-learning stage; reinforcement learning;

I. INTRODUCTION

Traffic control is one of the foremost and challenging issues in our world. Today, considering everyday increasing growth in number of vehicles and traffic congestion in the streets and roadways, development of traffic infrastructures could hardly give a perfect response to resolve traffic problem in our networks. This requires us to have a system that controls traffic lights optimally and appropriately under different conditions. Classic and common techniques of traffic control like fixed-time control and time-of-day control may act well when there is a certain amount of traffic in the given network, but on the occasions when model of traffic volume is variable in the network within different hours, such techniques could not be effective appropriately. Thus, proposing a method that can make decision dynamically with respect to different conditions and within the least period of time, may probably meet the existing need.

In recent years, different artificial intelligence approaches such as multiagent systems have widely been favored for traffic control in the field of intelligent transportation systems. For this system, concept of intelligent agents may be adapted to different parts of the system such as traffic lights, automobiles and pedestrians. Intelligent agents can learn and adapt themselves to different conditions. Taking a step forward, these agents are able to cooperate with each other and create coordination and guide traffic flow more optimally in the network at larger scale. In many of the multiagent systems, reinforcement learning is used to train agents. This learning is very similar to human learning process.

We consider traffic network as a system which is composed of intelligent agents and use incremental reinforcement learning along a pre-learning stage in order to assist the agents for better performance. In this system, traffic lights controller which is installed on each intersection is considered as an intelligent agent. These agents receive some statistical traffic information at pre- learning stage and estimate primary traffic parameters according to the existing theories. After calculating primary parameters based on the given theory, they enter into the main phase. They use an incremental Q-learning. In this article, we briefly deal with discussion on Q-learning, and then we will describe the proposed method and compare it's outcomes by Fixed- time control.

II. RELATED WORKS

In the field of using intelligent agents and multiagent systems for traffic control, several efforts have been made during the recent years. Wiering uses a model based reinforcement learning method in a multiagent system to learn traffic control strategies. In his technique, he uses car-based value functions to approximate cars waiting time [1]. Bakker et al developed Wiering's technique by adding some coordination among the agents. In both methods, the agent computes its optimal action with respect to local mode [2]. Moriarty et al tries to utilize distributed artificial intelligence to formulate traffic control and coordinate changes of traffic lights to preserve the arbitrary speed [3]. In [4] Bingham has used reinforcement learning within the framework of Neuro-Fuzzy technique to control traffic lights; however, with respect to low sensitivity of this method, limited exploration and off-line approach for value updating, one could achieve a restricted success by using this method. In [5], Dresner uses a different technique based on reservation mechanism to prevent from accident in intersections in such a way that each intersection has a controller agent and vehicle transmits

a request to this controller agent to pass through intersection; if this request is accepted, vehicle can pass through the intersection securely; otherwise it should wait for a certain time. In [6], Dresner improves the proposed idea and examines the problem by assuming more complexities such as possibility of cars U-turns, acceleration in intersections etc. In [7], one technique is proposed which is based on multiagent systems. In this method, each intersection is modeled as an independent intelligent agent or a player taking part in a dynamic process in which not only agents own local goals but also a global one has to be taken into account. So, all agents eventually move toward achieving the global goal. In this method, concept of evolutionary game theory is mainly applied. In this technique, role of a traffic control manager is also considered which is tasked to make decision on traffic control policies and manage tactically while agents in the intersections are responsible for operational tasks. In [8], Bazzan et al. organize agents in groups of limited size. These groups are then coordinated by another agent, a tutor or supervisor. In this work, multiagent reinforcement learning for control of traffic signals will be implemented in two situations: agents act individually and agents can be ''tutored'', meaning that the tutor agent will recommend a joint action.

III. REINFORCEMENT LEARNING

Reinforcement learning can be interpreted simply and concisely in that: Learning through interaction with environment in order to achieve a certain objective. We call the decision maker and one who learns, an agent and something with which the agent interacts (which in fact includes all things out of the agent) is called the environment. This interaction takes place continually. Accordingly, the agent may make decision and based on its decision, it does an action. So in response to such an action, it will be rewarded and then it moves toward the new state. In other words, more accurately, the agent and environment interact sequentially with each other during time intervals.

Policy or strategy of the agent is a probability function which gives the possibility of selecting any action at each state. So the optimal policy is a mapping from states to actions such that the discounted future reward is maximized. Q-learning is a model-free approach to reinforcement learning. It's an online learning algorithm that can learn a controlling policy for Markov decision process based on the received rewards. Output of this algorithm is values in Q-table. Followed by this, the stages of Q- learning are given as follows:

1-Giving the primary value 0 to values of Q-table

2-Receiving the current environment state (*s*)

3-Repeating the following loop until the goal state is reached;

3-1- Selection of an action (*a*) in one of the two following modes:

3-1-1- Randomly (exploration)

3-1-2- According to Q- table which is built until now (extraction)

3-2-Being rewarded by the environment (*r)*

3-3-Receiving the new state of the environment (*s')*

3-4-Changing the value in Q-table according to following expression:

$$Q(s,a) \leftarrow \alpha (r + \gamma \max Q(s' , a')) + (1 - \alpha) Q(s,a) \quad (1)$$

3-5- Taking the next state as the current state ($s \leftarrow s'$)

This algorithm starts by one initial state and achieves the goal state by taking a series of actions and receiving reward. Each of episodes starts from one initial state and comes to one goal state.

In this algorithm, $\alpha \in [0, 1]$ is the learning rate for the agent and it characterizes that to what extent the newly obtained information is substituted by the former information. Value 1 for this rate causes the agent only to consider the latest information while value 0 may cause the agent not to learn anything.

$\gamma \in [0, 1]$ is discount factor to characterize significance of the future rewards where value 0 shows that the agent only considers the current reward. On the other hand, approaching to value 1 will make it strive for a long-term high reward. If the discount factor meets or exceeds 1, the *Q* values will diverge. [9].

IV. URBAN TRAFFIC CONTROL BASED ON AIDED REINFORCEMENT LEARNING

The approach proposed here to control traffic lights in intersections is based on the use of incremental reinforcement learning accompanied with an auxiliary pre-learning stage. In this approach, the traffic network is considered as a system composed of intelligent agents and the traffic lights controller, installed on each intersection, is considered as an intelligent agent. In this approach, before start of the learning process, agents are equipped with some useful traffic information and then they begin to learn. At first agents acquire some statistical traffic information including flows and saturation flows of the network streets. This statistical information either is given to the agents in advance, or is gathered by agents through the supervision over traffic network. Thus, a fixed-time control is initially applied to the network and for certain period of time agents observe the network and approximate passage flows through the streets. Then, according to some formulas, each agent computes the appropriate cycle length and primary green-times for the intersection. Afterwards, the calculated cycle length and green-time values are implemented in the intersection. The agent enters into reinforcement learning phase after the first cycle length and gradually learns to adaptively control the traffic in the network under different conditions. Through the reinforcement learning, the agent computes the reward at the end of every cycle length. Then it updates Q-table and selects the next action.

In this reinforcement learning, 24 states have been considered. These states indicate the traffic volume of the streets in proportionate to each other. For instance, State#1 shows the following state.

Traffic volume in West st >= North st >= East st >= South st

Accordingly, State#2 is related to a state where traffic volume in the western street is greater than the northern street and the traffic volume in the northern street is greater than the southern street, and traffic volume in the southern street is also more than the eastern street. In this technique, 19 possible actions have been considered for each agent. Accordingly, the agent can change the existing intervals by adding or subtracting 1s in such a way that the cycle length remains constant or it can use the existing intervals without any changes. For example, Action#1 does not change the existing intervals and keeps them constant for the next cycle length. Action#2 adds one second to West Green Time but subtracts one second from North Green Time. For instance, one of the other Actions does it as follows:

WestGreenTime - 1
NorthGreenTime -1
EastGreenTime +1
SouthGreenTime +1

It should be mentioned that a minimum value also is considered for each green time in order to prevent the agent from reducing it any more. It means if one of the green times has minimum value and the agent has selected an action that reduces it, The agent is not allowed to apply this action and should select action #1 (The action that doesn't change the time intervals).

In this approach, the received reward by each agent is calculated as follows:

Reward is minus ratio of total existing cars in the streets of the intersection to total capacity of these streets. (Capacity of each street is the maximum number of cars which can be placed through the given street.)

V. EXPERIMENTAL RESULTS

This method has been implemented on a network composed of six intersections and 17 double-lane main streets. Statistical traffic data of each intersection are shown in table 1, 2 and 3. Following this trend, the method for the calculation of cycle length and the related Green Time intervals is given as an example for intersection #1.

Junction #1:
Sum of ratios Y= 0.75
Start up lost time: l_1 =2 s

l_2 = Yellow +all red time – e, e =2 (2)
Yellow +all red time = 3 s

Clearance lost time: l_2 = 1 s
Total lost time per phase: $t_L = l_1 + l_2 = 3$ s (3)
Lost time per cycle: $L = N \times t_L = 4 \times 3 = 12$ s (4)

Cycle Length: $C = L / (1- V_c / [(3600/h) \times PHF \times v/c])$ (5)

$C = 12 /(1 – 1361 / [1800 \times 0.90 \times 0.95]) = 100$ s

Total Effective Green Time: G = C- L = 100 – 12 = 88 s

Effective West green time : $g_1 = G \times (y_1 / Y) = 88 \times (0.26 / 0.75) \approx 31$ s
Effective North green time: $g_2 = G \times (y_2 / Y) = 88 \times (0.14 / 0.75) \approx 17$ s
Effective East green time: $g_3 = G \times (y_3 / Y) = 88 \times (0.17 / 0.75) \approx 20$ s
Effective South green time: $g_4 = G – (g_1 + g_2 + g_3) = 20$ s

$g_i = G_i$ + (Yellow +all red time) - t_{Li} (6)

Actual West green time : $G_1 = 31 – 3 + 3 = 31$ s
Actual North green time: $G_2 = 17 – 3 + 3 = 17$ s
Actual East green time: $G_3 = 20 – 3 + 3 = 20$ s
Actual South green time: $G_4 = 20 – 3 + 3 = 20$ s

In (2), e is encroachment of vehicles into yellow and all-red time. A default value of 2.0 s is used for e. In (4), N is the number of phases in the cycle. In (5), PHF is the peak hour factor to estimate the flow rate in the worst hours and v/c is the desired volume to capacity ratio. In the computation of intersection cycle length, start up lost time l_1 = 2s and clearance lost time l_2 = 1s have been considered. Regarding (2), value of "yellow + all red time" is equal to 3 sec. In our simulation, "yellow time" and "all red time" have been considered as zero and 3, respectively. With respect to the given expressions, in each cycle length, total lost time is calculated as 12 seconds [10].

Like the above method, cycle length and green time phases of other intersections are computed. Table 4 indicates the calculated cycle length and green time phases of each intersection.

The resultant outcomes from implementing this approach show that this approach has better performance than the fixed-time control with the resulting parameters from traffic theories. In other words, this kind of learning, flexibly deals with street traffic control under various conditions. Fig.1 indicates a comparison between stop time which is resulted from the application of fixed-time control with the parameters calculated based on traffic theory and the resulting stop time from implementing the proposed method on the same network. Fig. 2 also shows travel time in both methods.

These figures signify that at the beginning of simulation, both methods have identical behavior but gradually the proposed learning indicates clearly its superiority in performance.

VI. CONCLUSION

The multiagent systems and reinforcement learning have an appropriate potential for application in the field of traffic control. The resulting outcomes from simulation show that the suggested approach may control more dynamically and flexibly traffic volume in street network than fixed-time control under different conditions. Benefitting the resultant information from traffic theories, as an information infrastructure in reinforcement learning, is considered as one of special advantages in the proposed approach. This approach causes an improvement in the performance of

reinforcement learning within the environment by injecting useful traffic information in the learning process.

REFERENCES

[1] M. Wiering, "Multi-Agent Reinforcement Learning for Traffic Light Control", Proc. Seventeeth International Conference on Machine Learning and Applications, pp. 1151–1158, 2000.

[2] B. Bakker, M. Steingrover, R. Schouten, E. Nijhuis and L. Kester, "Cooperative multi-agent reinforcement learning of traffic lights", Proc. Workshop on Cooperative Multi-Agent Learning, European Conference on Machine Learning, ECML'05, 2005.

[3] David E.Moriarty, Simon Handley, and Pat Langley, "Learning Distributed Strategies for Traffic Control", Proc. Fifth International Conference of the Society for Adaptive Behavior, Zurich, Switzerland, pp. 437-446, 1998.

[4] Ella Bingham, "Reinforcement Learning in Neurofuzzy Traffic Signal Control", European Journal of Operational Research, pp.232-241, 2001.

[5] K. Dresner and P. Stone. "Multiagent traffic management: A reservation-based intersection control mechanism", Proc. Third International Joint Conference on Autonomous Agents and Multiagent Systems, pp. 530–537, New York, USA, 2004.

[6] Dresner and P. Stone. "Multiagent traffic management: An improved intersection control mechanism", Proc. Fourth International Joint Conference on Autonomous Agents and Multiagent Systems, pp. 471–477, Utrecht, The Netherlands, 2005.

[7] Bazzan, A.L.C, A distributed approach for coordination of traffic signal agents. Autonomous Agents and Multiagent Systems 10 (1), pp. 131–164, 2005.

[8] Ana L.C.Bazzan, Denise de Oliveira a, Bruno C. da Silva, Learning in groups of traffic signals, Engineering Applications of Artificial Intelligence 23, pp.560–568, 2010.

[9] Richard S.Sutton and Andrew G.Barto, "Reinforcement Learning: An Introduction", MIT Press, 1998.

[10] Roger P.Roess, Elena S.Prassas, William R.Mcshane, Traffic Engineering, Pearson Education, Inc, 2004.

Table ١ . Traffic statistical data of junction #1 and #2

	Junction #1				**Junction #2**			
	West St	***North St***	***East St***	***South St***	***West St***	***North St***	***East St***	***South St***
Flow	475	257	308	321	454	379	384	162
Saturation flow	1800	1800	1800	1800	1800	1800	1800	1800
Ratio: y_i	0.26	0.14	0.17	0.18	0.25	0.21	0.21	0.09

Table٢ . Traffic statistical data of junction #3 and #4

	Junction #3				**Junction #4**			
	West St	***North St***	***East St***	***South St***	***West St***	***North St***	***East St***	***South St***
Flow	345	484	296	227	360	275	368	239
Saturation flow	1800	1800	1800	1800	1800	1800	1800	1800
Ratio: y_i	0.19	0.27	0.16	0.13	0.20	0.15	0.20	0.13

Table٣ . Traffic statistical data of junction #5 and #6

	Junction #5				**Junction #6**			
	West St	***North St***	***East St***	***South St***	***West St***	***North St***	***East St***	***South St***
Flow	346	294	279	322	344	333	266	227
Saturation flow	1800	1800	1800	1800	1800	1800	1800	1800
Ratio: y_i	0.19	0.16	0.16	0.18	0.19	0.19	0.15	0.13

Table٤ . cycle length and green time phases of each intersection

	Junction #1	Junction #2	Junction #3	Junction #4	Junction #5	Junction #6
Cycle Length	100 s	120 s	100 s	80 s	80 s	60 s
West green time	31 s	36 s	23 s	20 s	19 s	14 s
North green time	17 s	30 s	32 s	15 s	16 s	14 s
East green time	20 s	30 s	19 s	20 s	16 s	11 s
South green time	20 s	12 s	14 s	13 s	17 s	9 s

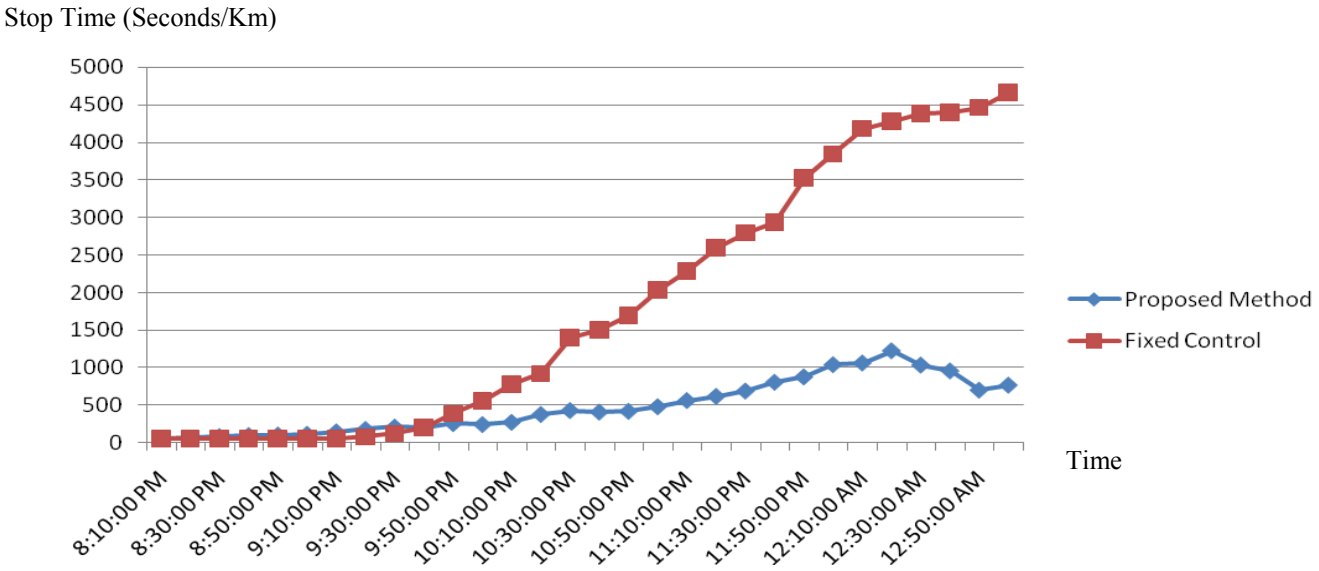

Figure 1. A comparison between stop time which is resulted from the application of fixed-time Control and the resulting stop time from implementing the proposed method

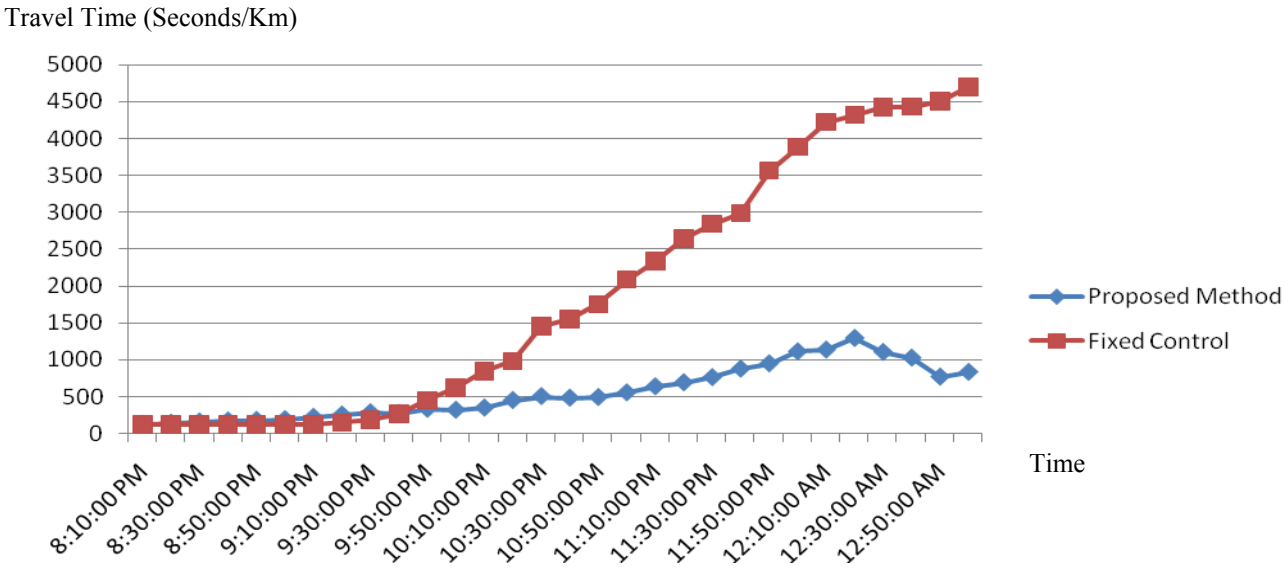

Figure 2. A comparison between the values of travel time in both methods