Arabian Journal for Science and Engineering
https://doi.org/10.1007/s13369-017-3018-9

RESEARCH ARTICLE - SYSTEMS ENGINEERING

Fuzzy Q-Learning-Based Multi-agent System for Intelligent Traffic Control by a Game Theory Approach

Abolghasem Daeichian[1] · **Amir Haghani**[2]

Received: 11 December 2015 / Accepted: 3 December 2017

Abstract
This paper introduces a multi-agent approach to adjust traffic lights based on traffic situation in order to reduce average delay time. In the traffic model, lights of each intersection are controlled by an autonomous agent. Since decision of each agent affects neighbor agents, this approach creates a classical non-stationary environment. Thus, each agent not only needs to learn from the past experience but also has to consider decision of neighbors to overcome dynamic changes of the traffic network. Fuzzy Q-learning and Game theory are employed to make policy based on previous experiences and decision of neighbor agents. Simulation results illustrate the advantage of the proposed method over fixed time, fuzzy, Q-learning and fuzzy Q-learning control methods.

Keywords Traffic control · Multi-agent system · Game theory · Fuzzy Q-learning

1 Introduction

Urbanization, increasing number of vehicles, and lack of transport infrastructures have increased travel time, fuel consumption, and air pollution. Therefore, urban life equals with waste of time, less clean air, and acoustic pollution. Conventional fixed traffic management systems are not able to fight complexity and dynamic of large traffic networks. While artificial intelligence (AI) are greatly employed to develop intelligent traffic systems (ITS) [6,7,19,24], multi-agent system is an approach to model ITS [25,30]. This framework consists of a population of intelligent and autonomous agents work together in an environment [27]. Traffic lights [20], vehicles [3], and pedestrians [29] are considered as agents in modeling of urban traffic networks. Each agent needs to learn from the past experiences which is a key point to approximate a better decision-making policy. Multi-agent model-based [32] as well as model-free [12] reinforcement learning (RL) techniques are widely used in researches on ITS [6,23].

In a multitude of researches, any agent only considers its own traffic state in order to determine the control policy. For example, single intersection with two phases is investigated in [2]. Length of vehicles queue waiting on the light is considered as state which can be measured by the agent. It decides on extend green time or change it to the next phase so that the number of vehicles waiting on the light is minimized. The results show superiority of Q-learning agent over uniform traffic flows and constant-ratio traffic flows. In [32], traffic lights are considered as agents which communicate with vehicles. The vehicles estimate their mean waiting time and transmit this time to traffic light where a popular RL algorithm, namely Q-learning, is used to provide a control for traffic signal scheduling. Results of this study show 22% reduction in waiting time compared to constant time lights. Multi-objective reinforcement learning is utilized to control several traffic lights in [17]. Optimization goals include number of stops of a vehicle, mean stopping time, and length of vehicles' queue on the next intersection. Its results indicate that multi-RL can effectively prevent the queue spillovers under congested condition to avoid large-scale traffic jams. Bull et al. [10] used learner classifiers to control light traffic including 4 intersections. In this research, traffic lights include two phases at each intersection, where one phase is for moving north–south and one is for east-west. Controller at each intersection obtains optimum phase time through extracting if-then rules. Its results show that performance of

✉ Abolghasem Daeichian
a-daeichian@araku.ac.ir; a.daeichian@gmail.com

[1] Department of Electrical Engineering, Faculty of Engineering, Arak University, Arak 38156-8-8349, Iran

[2] Department of Electrical Engineering, Payam Institute of Higher Education, Golpayegan, Isfahan, Iran

Published online: 13 December 2017

the traffic light using learner classifier system has improved significantly compared to constant time traffic light. In [28], the learning purpose is modeled in such a way that states indications are based on the summation of the cars waiting times. Obviously, the more cars information is received, the model will be more complicated and state space will be larger. This issue is one of the significant problems of large networks. Adaptive control, which is introduced in [23], uses the approximate of a function as mapping of states to scheduling. Fuzzy inference engine is exploited to decrease systematic faults of Q-algorithm in [22]. The results demonstrate that not only learning in fuzzy framework is done faster than Q-learning but also delay in intersections is decreased considerably. A multi-agent fuzzy approach is proposed in [18], where Q-learning updates the set of rule base in fussy inference engine. In [13], a new method which has the capability to estimate an incomplete model of environment is described for a given non-static environment. This method is applied in a network composed of 9 intersections. The reported results show that this method has better performance than the model-free methods and model-based methods, but could not be generalized and used in larger networks.

In other researches, agents consider other agents in determination of their own control policy. For instance, coordination among agents is desired in [21] where the agents not only consider number of waiting vehicles on its own intersection, but also they consider number of vehicles which have stopped in adjacent intersections. The RL is applied on 5 intersections within three different scenario. The overall results show improvement in delay time. In [32], RL is used to control the traffic in a grid where a type of cooperative learning simultaneously controls the traffic signals and determines the optimal routes. One of the main drawbacks of this method is the high costs of communication and information exchange, specifically when intersections of network are increased. Cooperative RL tries to extract the knowledge from neighbor agents in a scheduling learning [26]. This method is implemented in an area of Dublin including 64 intersections.

This paper introduces a hybrid fuzzy Q-learning and Game theory method for control of traffic lights in multi-agent framework. It exploits the benefits of fuzzification as well as interaction with other agents. The traffic network is modeled by considering an autonomous agent controls in which each intersection decides on duration of green phase. The number of vehicles in different inputs of the intersection are measured by the corresponding agent. Any agent interacts with neighbor agents by getting a reward from each decision. This paper proposes that each agent fuzzify the inputs and utilizes in a fuzzy inference system for fuzzy estimation of traffic model states. The agent uses a Q-learning approach modified by Game theory to learn from the past experiences and consider the interaction with neighbor agents. The agent gets a reward proportional to its own traffic state and a reward from each decision from neighbor agents to update its Q-learning algorithm. The neighbor reward and its weighting in Q-value update is proposed to be fuzzy in the proposed method. The proposed method is applied on a five-intersection traffic network. The simulation results indicate that proposed method outperforms the fixed time, fuzzy, Q-learning and fuzzy Q-learning control methods in the sense of average delay time.

This paper is unfolds as follows. After this introduction, Q-learning and its fuzzy version are described in the next section. Section 3 is devoted to application of Game theory in ITS. Sections 4 and 5 are about problem statement and proposed solution, respectively. Simulation results are given in Sect. 6. Finally, the paper is concluded in Sect. 7.

2 Q-Learning and Fuzzy Q-Learning

The objective of agents which act in dynamic environments is making optimum decisions. If the agents are not aware of rewards corresponding to various actions, selecting a proper action would be challenging. To achieve this goal, learning adjusts agents' action selection based on collected data. Each agent tries to optimize its actions with dynamic environment via trial and error in reinforcement learning (RL). The RL is actually how different situations are mapped upon actions to receive the best results or the highest reward. In many cases, actions influence the reward of next steps as well as affect the reward of its corresponding step. There are model-based [32] as well as model-free [12] RL techniques. In model-free RL, the agent does not need explicit modeling of the environment because its actions could be directly selected based on rewards. Q-learning is a model-independent approach where the agent does not access to transfer model [1,31]. Suppose that the agent is in a state s, performs an action a, from which it gets the rewards r from the environment and the environment changes to state s'. This is given by a tuple in the form of (s, a, r, s'). State-action value which represents the expected total reward resulting from taking action a in state s is denoted by Q-value $Q(s, a)$. The agent starts with random value and after each action they receive a tuple in the form of (s, a, r, s'). For each tuple, the value of state-action could be calculated according to the following equation:

$$Q(s,a) = (1-\alpha)Q(s,a) + \alpha\left[r + \gamma \max Q(s',a') - Q(s,a)\right] \quad (1)$$

where $\alpha \in [0, 1]$ is the learning rate of agent. $\alpha = 1$ means that merely new information is considered and zero means that the agent does not have any learning. $\gamma \in [0, 1]$ is discount factor which determines future rewards. Zero value for this factor makes the agent opportunist which means that the agent only considers current reward. On the other hand,

$\gamma = 1$ means that the agent will wait for a longer time to achieve a large reward. Q-learning will converge to optimum value $Q^*(s, a)$ with probability of one if all state-action pairs are experienced repetitively and learning rate decrease during the time [22]. Generally, RL is useful for solving problems with small dimension discrete state and action space. When the dimension of state and action space becomes larger, the size of search table will be so large that it makes the algorithm very slow due to computational time. On the other hand, when the states or actions are stated continuously, using search table will not be possible. To tackle this problem, fuzzy theory is employed. If the intelligent agent has a proper fuzzy set as expert knowledge about the desired area, the ambiguity could be resolved. Thus, intelligent agent can understand vague objectives and unknown environment. In practice, the action in large spaces is facilitated by eliminating Q-values table. In this method everything is based on quality values and fuzzy inference. Fuzzy inference system (FIS) deals with input and Q-learning algorithm uses the follower section and its active rules as states. Reward signal of Q-algorithm is built in accordance with fuzzy logic, environment reward signal and performance estimation of current action. It is tried to select the action which maximizes the reward signal [9,14]. Learning system is able to select one action among j actions for each rule. j-th possible action in i-th rule is denoted by $a[i, j]$ and its value is shown by $q[i, j]$ consider the following rules [9]:

$$\begin{aligned} &\text{If } x \text{ is } s_i \quad \text{then} \quad a[i,1] \text{ with } q[i,1] \\ &\qquad\qquad \text{or} \quad a[i,2] \text{ with } q[i,2] \\ &\qquad\qquad \vdots \\ &\qquad\qquad \text{or} \quad a[i,j] \text{ with } q[i,j] \end{aligned} \tag{2}$$

Learning should find the best result for each rule. If the agent selects an action which results in high value, it may learn optimum policy. Thus, fuzzy inference system may obtain necessary action for each rule [9].

3 Game Theory in ITS

Relation between agent-oriented environments and games theory originates from the fact that each state of agent-oriented environments can be resembled to a game environment. Profit function of players would be current state of the environment and goal of players is to move toward balanced or equilibrium point (reaching the best decision-making policy). Some scholars have studied the application of Game theory to control of traffic lights [15,16]. They integrate Game theory into the multi-agent interaction approach. Some of them suit the traffic problem into a rigorous mathematical game model [5,8,11], while others modify the learning method of agents based on Game theory [33]. In [5], signalized intersections are modeled as finite controlled Markov chains and each intersection is seen as non-cooperative game where each player try to minimize its queue. The solutions are given as Nash equilibrium and Stackelberbg equilibrium and the simulation results indicate shorter queue length than adaptive control. In [8], a two-player non-cooperative game is articulated between user seeking a path to minimize the expected trip cost and choosing link performance scenarios to maximize the expected trip cost. It shows that the Nash equilibrium point measures network performance. Intelligent traffic control is expressed as a Cournot game where the traffic authority and the users choose their strategies simultaneously and as a bi-level Stackelberg game where the traffic authority is the leader which determines the signal settings in anticipation of the user reactions. In [33], Game theory is used to address coordination between agents based on traffic signal control with Q-learning. It specifies strategies ($C(m)$ ={red light time plus 4 s, red light time plus 8 s, red light time minus 4 s, red light time minus 8 s,unchangeably}) and actions ($S(n)$ ={east west straight and right turn, south north straight and right turn, east west left turn, south north left turn}). Then, an interaction mathematical model via Game theory as a four parameter group $G = \{B, A, I, U\}$ is presented. B is a group of decision-makers as players. A is a group of any possible strategies and actions, i.e. $A = C(m) * S(n)$. I represents the information which agents masters. U is the benefit function which adopts Q-value. So, the Nash equilibrium is [33]:

$$U_i(a_i^*, a_{-i}^*) \geq U_i(a_i, a_{-i}^*) \tag{3}$$

where a_i and a_{-i} denote action of i-th agent and actions of other agents, respectively. a_i^* and a_{-i}^* represent the actions at Nash equilibrium. The renewed Q-values in distributed reinforcement Q-learning are used to build the payoff values. Q-value function is updated as:

$$\begin{aligned} Q_i(s_i, a_i) = {} & (1-\alpha_i) Q_i(s_i, a_i) \\ & + \alpha_i \left[r_i(s_i, a_i) + \sum_{j=1, j\neq i}^{n} f(i,j) r_j(s_i, a_i) \right. \\ & \left. + \gamma \ \max(Q_i(s_i', a_i') - Q_i(s_i, a_i)) \right] \end{aligned} \tag{4}$$

where α and γ are learning rate and discount factor, respectively. s_i and a_i are current state of traffic environment and current action, respectively. s_i' is its next state, n is the number of traffic signal control agents surrounding i-th agent, $Q_i(s_i, a_i)$ is the Q-value function for i-th agent when selects action a_i in state s_i. $r_i(s_i, a_i)$ is reward function of i-th agent and $r_j(s_i, a_i)$ is reward function of j-th agent neighboring i-

th agent. $f(i, j) \in [0, 1]$ is a weighted function which shows the effect of $r_j(s_i, a_i)$ on i-th agent. Mathematical functions are suggested in [33] for $r(s, a)$ and $f(i, j)$. Assumption of discrete action-state space and determination of reward and weighting functions are drawbacks of that work.

4 Problem Statements

Consider a traffic network in which the lights of each intersection is controlled by an autonomous agents without any centralized management. Some sensors which are installed below the surface of surrounding streets or traffic cameras of each intersection provide information about traffic situation for the corresponding agent. An agent has to decide on duration of green light at north–south (NS) and west–east (WE) paths. Also, any agent interacts with neighbor agents. Anyway, the agent is expected to schedule traffic lights optimally, in the sense of average delay, based on the received information from its sensors and received information from neighbor agents.

The agents may have little knowledge about others' decision due to distribution of information. Even if an agent has previous known information about others' decision, it is not valid as other agents are also learning. Thus, the environment is dynamic and the behavior of other agents may change during time. Lack of prediction of other agents causes uncertainty in problem solving procedure. This paper looks for a decision-making algorithm for lights control agents which considers neighbor agents information in addition to its own information.

5 Proposed Algorithm

We consider a constant duration T for green plus red phases. So, if the agent determines the green phase duration t_g, then the red phase duration is $t_r = T - t_g$. Any typical agent i receives number of vehicles on the NS and WE streets from its own sensors and the green phase duration of neighbor agent j in order to schedule its own green phase duration. This paper proposes an autonomous agent with structure in Fig. 1 to control each intersection.

The number of vehicles in WE and NS streets which are measured by sensors are fuzzified. Then, a fuzzy inference engine with rules as Eq. 2 are employed to fire the corresponding output membership functions. Finally, defuzzification results to duration of green phase in NS path (t_g^{NS}). Thus, the duration of green phase in other path, WE, is $t_g^{WE} = T - t_g^{NS}$. We propose that, Q-value function which is updated by Eq. 4 be the value of each action in Eq. 2 which is denoted by $q[i, j]$. This update equation takes the neighbor agents' decision into account.

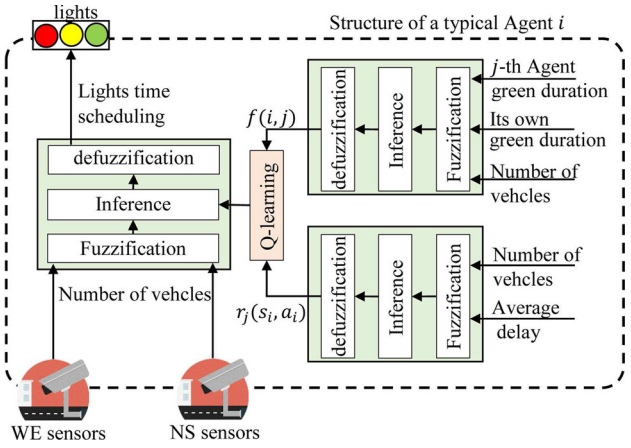

Fig. 1 The proposed structure for a typical agent

The i-th agent takes decision of neighbor agent j into account by reward $r_j(s_i, a_i)$ and a weighting function $f(i, j)$. The reward is calculated based on average delay obtained from the decision made by the agent and current traffic situation in a fuzzy manner. A fuzzy inference engine obtains these two inputs after fuzzification and gives the reward after defuzzification; see Fig. 1. weighting function $f(i, j) \in [0, 1]$ shows the effect of $r_j(s_i, a_i)$ on the decision of i-th agent. This weight is also calculated by a fuzzy inference engine. This engine takes its own t_g, the neighbor agents' t_g, and number of waited vehicles and gives $f(i, j)$. Suitable choice for reward and weighting function plays a significant role in agent learning. The agent with structure in Fig. 1 runs the following algorithm:

1. Initial value of Q_i-value for i-th traffic signal control agent is in the form of $\forall(s_i, a_i) : Q_i(s_i, a_i) = 0$.
2. Observing s_i by WE and NS sensors which is the current state of i-th intersection.
3. Selecting a proper estimation for desired state by fuzzy inference system.
4. Calculating the reward related to i-th and j-th traffic signal control agent and the weighting function for neighboring agents separately.
5. Observing new state s_i'.
6. Updating Q_i-value according to Eq. 4.
7. Returning to step 2 till the variation of Q-value becomes less than ϵ.

6 Simulation Results

Consider a traffic network with a center and four neighbor intersection. The delay in each intersection depends on physical characteristics of the intersection, traffic light scheduling and number of cars in input streets. We utilized traffic model

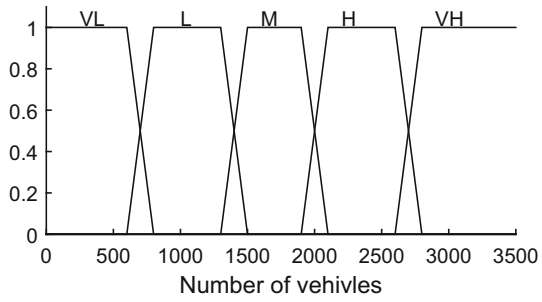

Fig. 2 Membership function of number of vehicles enter the street for reward FIS

which is given by the American Highway Capacity Manual (HCM) [4, Eq.20]:

$$d = 0.38\frac{C(1-\lambda)^2}{1-x} + 173x^2\left[(x-1)+\sqrt{(x-1)^2+\frac{16x}{C}}\right] \quad (5)$$

where d, C, λ, and x are average delay (s), cycle time (s), green ratio, and degree of saturation, respectively. $\lambda = \frac{g}{c}$ and $x = \frac{v}{c}$, where c, g, and v are capacity (vehicle per hour), green time (sec), and input volume, respectively. We use this model to calculate average delay based on the green phase duration and number of vehicles. For more details of this equation, we refer to [4].

Assume that $C = T = 100$ s and $c = 3500$ veh/h. v is volume of vehicles entering each street which varies between 0 to 3500 veh/h. g is duration of the green phase which each agent selects considering fuzzy Q-learning and interaction with adjacent agents. The traffic network simulation algorithm is as follow:

1. The volume of vehicles entering each intersection (v) are randomly generated by a discrete uniform distribution on the interval [0, 3500].
2. Average delay is calculated by Eq. 5.
3. Each agent decides on the time of green phase g.
4. Go to step 1 until end of simulation time.

Assume structure of the agents as in Fig. 1 with the Mamdani FIS with input membership function as in Fig. 2 for number of input vehicles and Fig. 3 for average delay to calculate the reward functions $r_j(s_i, a_i)$. Centroid defuzzification by the output membership function as in Fig. 4 is considered to estimate a reward value in interval $[-3, 3]$.
The weighting function FIS has number of vehicles, its own green phase duration and the neighbor agents' green phase duration as inputs. Figure 2 shows the membership function for number of vehicles, and Fig. 5 depicts the membership

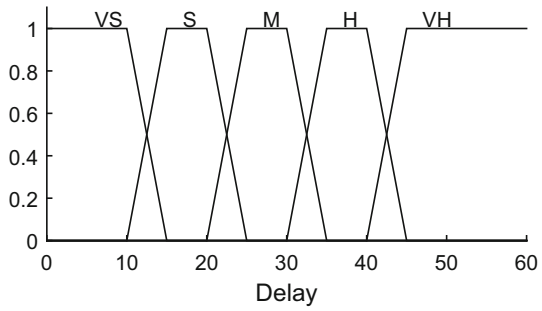

Fig. 3 Membership function of average delay for reward FIS

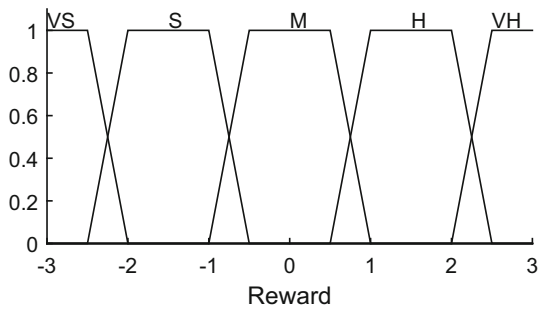

Fig. 4 Membership function of output for reward FIS

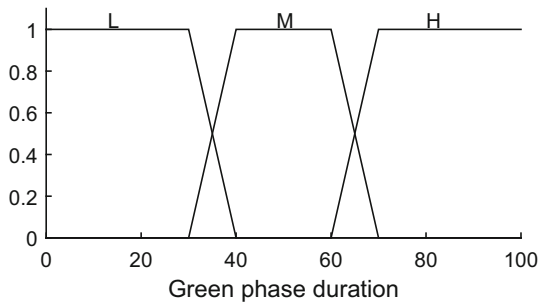

Fig. 5 Membership function of green phase duration for weighting function FIS

function for its own and neighbor green phase duration. Centroid defuzzification is applied to calculate weights on output membership function as in Fig. 6 which should be a value between 0 and 1.
Finally, the agent uses fuzzy Q-learning (Eq. 2) with Q-value update rule (Eq. 4) where learning and discount factor are selected to be 0.5 and 0.7, respectively. The membership function for each measured number of vehicles is shown in Fig. 7. The output estimates green phase duration with membership functions as in Fig. 8.
The proposed method is compared with Fuzzy Q-learning (using Eq. 2 where $q[i, j]$ is the Q-value which updates with Eq. 1), Q-learning (using Q-learning method with Q-value which updates with Eq. 1), fuzzy(using traditional fuzzy inference method) and fixed time ($t_g = 60$ s) in the sense of total average delay. Average delay in each

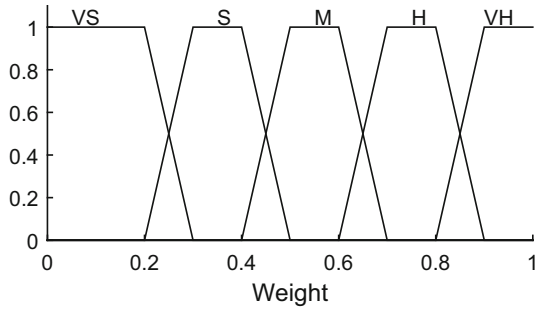

Fig. 6 Membership function of output for weighting function FIS

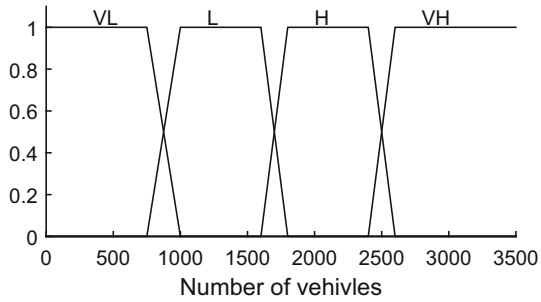

Fig. 7 Membership function of number of vehicles for fuzzy Q-learning

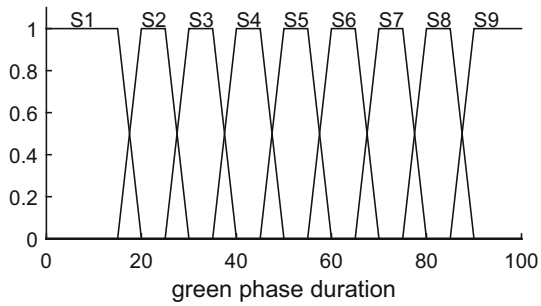

Fig. 8 Membership function of green phase duration for fuzzy Q-learning

time interval is depicted in Fig. 9, and the total average delay is illustrated in Fig. 10. The results illustrate that total average delay decrease from more than 50 s for fixed time scheduling to approximately 15 s for the proposed method.

7 Conclusion

In this study, an intelligent control method of a controlling traffic network was performed to decrease average delay time. Each traffic light is considered as a learning agent. This paper proposed a structure for the agents. Each agent learn to decide on the duration of green phase through a fuzzy Q-learning algorithm which is modified by Game theory. Each agent receives a reward from neighbor agents.

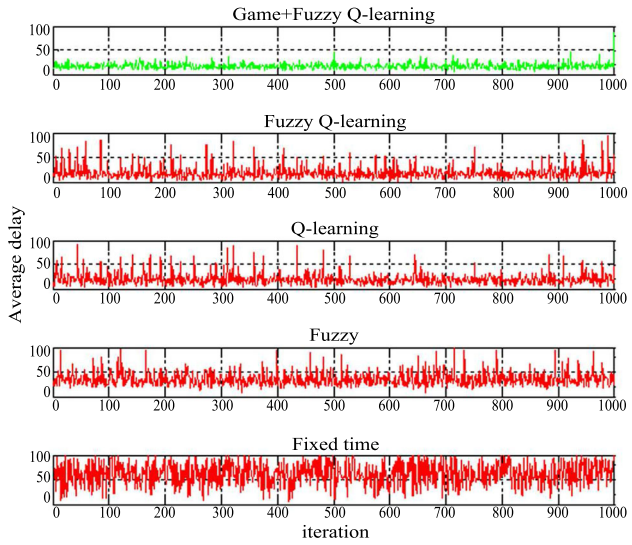

Fig. 9 Delay of the proposed method, fixed time, fuzzy Q-learning, Q-learning and fuzzy in each time step

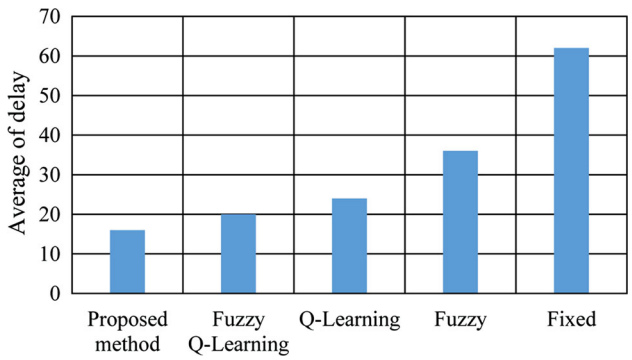

Fig. 10 Average of delay for the proposed method, fixed time, fuzzy, Q-learning, fuzzy Q-learning

The reward received from the neighbor and weighted functions of neighboring agents are factors learning algorithm. These parameters are fuzzified through a FIS. Also, the number of vehicles in each street is measured and fuzzified to be used in decision-making process. The simulation results were compared with fixed time method and other intelligent methods. The results revealed that our proposed method achieves considerable reduction of average delay in intersections.

References

1. Abdoos, M.; Mozayani, N.; Bazzan, A.L.: Traffic light control in non-stationary environments based on multi agent q-learning. In: 14th International IEEE Conference on Intelligent Transportation Systems (ITSC), pp. 1580–1585. IEEE (2011)
2. Abdulhai, B.; Pringle, R.; Karakoulas, G.J.: Reinforcement learning for true adaptive traffic signal control. J. Transp. Eng. **129**(3), 278–285 (2003)

3. Adler, J.L.; Satapathy, G.; Manikonda, V.; Bowles, B.; Blue, V.J.: A multi-agent approach to cooperative traffic management and route guidance. Transp. Res. Part B Methodol. **39**(4), 297–318 (2005)
4. Akgungor, A.P.; Bullen, A.G.R.: Analytical delay models for signalized intersections. In: 69th ITE Annual Meeting, Nevada, USA (1999)
5. Alvarez, I.; Poznyak, A.; Malo, A.: Urban traffic control problem a game theory approach. In: 47th IEEE Conference on Decision and Control, pp. 2168–2172. IEEE (2008)
6. Balaji, P.; German, X.; Srinivasan, D.: Urban traffic signal control using reinforcement learning agents. IET Intell. Transp. Syst. **4**(3), 177–188 (2010)
7. Bazzan, A.L.; Klgl, F.: A review on agent-based technology for traffic and transportation. Knowl. Eng. Rev. **29**(03), 375–403 (2014)
8. Bell, M.G.: A game theory approach to measuring the performance reliability of transport networks. Transp. Res. Part B Methodol. **34**(6), 533–545 (2000)
9. Bonarini, A.; Lazaric, A.; Montrone, F.; Restelli, M.: Reinforcement distribution in fuzzy q-learning. Fuzzy Sets Syst. **160**(10), 1420–1443 (2009)
10. Bull, L.; ShaAban, J.; Tomlinson, A.; Addison, J.D.; Heydecker, B.G.: Towards distributed adaptive control for road traffic junction signals using learning classifier systems. In: Bull, L. (ed.) Applications of Learning Classifier Systems, pp. 276–299. Springer, Berlin (2004)
11. Chen, O.; Ben-Akiva, M.: Game-theoretic formulations of interaction between dynamic traffic control and dynamic traffic assignment. Transp. Res. Rec. J. Transp. Res. Board **1617**, 179–188 (1998)
12. Chin, Y.K.; Bolong, N.; Kiring, A.; Yang, S.S.; Teo, K.T.K.: Q-learning based traffic optimization in management of signal timing plan. Int. J. Simul. Syst. Sci. Technol. **12**(3), 29–35 (2011)
13. Da Silva, B.C.; Basso, E.W.; Perotto, F.S.; C Bazzan, A.L.; Engel, P.M.: Improving reinforcement learning with context detection. In: Proceedings of the Fifth International Joint Conference on Autonomous Agents and Multiagent Systems, pp. 810–812. ACM (2006)
14. Glowaty, G.: Enhancements of fuzzy q-learning algorithm. Comput. Sci. **7**, 77–87 (2005)
15. Goyal, T.; Kaushal, S.: An intelligent scheduling scheme for real-time traffic management using cooperative game theory and ahp-topsis methods for next generation telecommunication networks. Expert Syst. Appl. **86**, 125–134 (2017)
16. Groot, N.; Zaccour, G.; De Schutter, B.: Hierarchical game theory for system-optimal control: applications of reverse stackelberg games in regulating marketing channels and traffic routing. IEEE Control Syst. **37**(2), 129–152 (2017)
17. Houli, D.; Zhiheng, L.; Yi, Z.: Multiobjective reinforcement learning for traffic signal control using vehicular ad hoc network. EURASIP J. Adv. Signal Process. **1**, 724,035 (2010)
18. Iyer, V.; Jadhav, R.; Mavchi, U.; Abraham, J.: Intelligent traffic signal synchronization using fuzzy logic and q-learning. In: International Conference on Computing, Analytics and Security Trends (CAST), pp. 156–161. IEEE (2016)
19. Kponyo, J.; Nwizege, K.; Opare, K.; Ahmed, A.; Hamdoun, H.; Akazua, L.; Alshehri, S.; Frank, H.: A distributed intelligent traffic system using ant colony optimization: a netlogo modeling approach. In: International Conference on Systems Informatics, Modelling and Simulation (SIMS), pp. 11–17. IEEE (2016)
20. Liu, Z.: A survey of intelligence methods in urban traffic signal control. IJCSNS Int. J. Comput. Sci. Netw. Secur. **7**(7), 105–112 (2007)
21. Medina, J.C.; Hajbabaie, A.; Benekohal, R.F.: Arterial traffic control using reinforcement learning agents and information from adjacent intersections in the state and reward structure. In: 2010 13th International IEEE Conference on Intelligent Transportation Systems (ITSC), pp. 525–530. IEEE (2010)
22. Pacheco, J.C.; Rossetti, R.J.: Agent-based traffic control: a fuzzy q-learning approach. In: 13th International IEEE Conference on Intelligent Transportation Systems (ITSC), pp. 1172–1177. IEEE (2010)
23. Prashanth, L.; Bhatnagar, S.: Reinforcement learning with function approximation for traffic signal control. IEEE Trans. Intell. Transp. Syst. **12**(2), 412–421 (2011)
24. Rida, M.: Modeling and optimization of decision-making process during loading and unloading operations at container port. Arab. J. Sci. Eng. **39**(11), 8395–8408 (2014)
25. Roess, R.P.; Prassas, E.S.; McShane, W.R.: Traffic Engineering. Prentice Hall, Englewood Cliffs (2004)
26. Salkham, A.; Cunningham, R.; Garg, A.; Cahill, V.: A collaborative reinforcement learning approach to urban traffic control optimization. In: Proceedings of IEEE/WIC/ACM International Conference on Web Intelligence and Intelligent Agent Technology, pp. 560–566. IEEE Computer Society (2008)
27. Schaefer, M.; Vokřínek, J.; Pinotti, D.; Tango, F.: Multi-agent traffic simulation for development and validation of autonomic car-to-car systems. In: McCluskey, Th.L., Kotsialos, A., Müller, J.P., Klügl, F., Rana, O., Schumann, R. (eds.) Autonomic Road Transport Support Systems, pp. 165–180. Springer, Berlin (2016)
28. Steingrover, M.; Schouten, R.; Peelen, S.; Nijhuis, E.; Bakker, B.: Reinforcement learning of traffic light controllers adapting to traffic congestion. In: BNAIC, pp. 216–223. Citeseer (2005)
29. Teknomo, K.: Application of microscopic pedestrian simulation model. Transp. Res. Part F Traffic Psychol. Behav. **9**(1), 15–27 (2006)
30. Vilarinho, C.; Tavares, J.P.; Rossetti, R.J.: Intelligent traffic lights: green time period negotiation. Transp. Res. Procedia **22**, 325–334 (2017)
31. Watkins, C.J.; Dayan, P.: Q-learning. Mach. Learn. **8**(3–4), 279–292 (1992)
32. Wiering, M.: Multi-agent reinforcement learning for traffic light control. In: ICML, pp. 1151–1158 (2000)
33. Xinhai, X.; Lunhui, X.: Traffic signal control agent interaction model based on game theory and reinforcement learning. In: International Forum on Computer Science-Technology and Applications, vol. 1, pp. 164–168. IEEE (2009)