Sixth International Conference on Intelligent Control and Information Processing
Wuhan, China; November 26-28, 2015

Reinforcement Learning Based Overtaking Decision-Making for Highway Autonomous Driving

Xin Li, Xin Xu, and Lei Zuo
The College of Mechatronics and Automation
National University of Defense Technology
Changsha, China

Abstract—**In this paper, we develop an intelligent overtaking decision-making method for highway autonomous driving. The key idea is to use reinforcement learning algorithms to learn an optimized policy via a series of simulated driving scenarios. A vehicle model based on data fitting of real vehicles as well as a traffic model is established to simulate driving scenarios and validation tests of obtained policies. Human driving experiences are considered in designing the reward function. A reinforcement learning method called the Q-learning algorithm is used to learn overtaking decision-making policies. Simulations show that our method can learn feasible overtaking policies in different traffic environments and the performance is comparable or even better than manually designed decision rules.**

Keywords—reinforcement learning; Q-learning; overtaking decision-making; vehicle model; highway road

I. INTRODUCTION

Autonomous vehicles are one of the most important parts of intelligent transportation systems. An autonomous vehicle is a synthetic system containing four main modules: perception, decision-making, planning and control [1], [2]. The decision-making module can be viewed as the brain of an autonomous vehicle. It receives information from the perception module and then makes corresponding decisions for the planning module and the control module. During driving, an autonomous vehicle needs to make decisions on a variety of driving behaviors. Overtaking behaviors generally take place many times during driving and are closely related to security and speediness. A basic overtaking process can be treated as a problem with three stages [3], [4]: (1) lane-changing from the origin lane, (2) overtaking the vehicle in adjacent lane, (3) driving back to the previous lane safely.

There have been a variety of research attempts that have been carried out to solve the overtaking decision-making problem. Jula et al. [3] studied the minimum longitudinal distances in some special cases of lane-changing or merging scenarios. Shamir [4] studied how one vehicle overtook a single slower-moving vehicle in front, the equations of motion were used and plenty of experiments were taken to seek for the optimal values of the unknown variables. Perez et al. [5] proposed a decision control algorithm based on fuzzy logic which achieved two consecutive lane-change actions. Karaduman et al. [6] studied the driving scene with one vehicle in front and another fall behind, and came up with a Bayesian belief network to determine the probability of collisions. Nilsson and Sjoberg [7] used a mixed logical dynamical system to solve a predictive control problem of overtaking in a highway traffic scene with four other vehicles. Despite of the above progresses, there are still two problems to be solved: 1) the values of adjustable policy parameters are difficult to be determined, 2) in more complicated driving conditions, it is difficult for previous methods to obtain optimized performance.

Reinforcement learning aims to solve sequential decision-making problems in uncertain environments [8], [9]. Reinforcement learning originates from the trial-and-error learning principle and its main purpose is to maximum the accumulated delayed rewards from the environment. Instead of an accurate model or teacher signal, reinforcement learning emphasizes the interaction with environment. There have been some recent attempts to use reinforcement learning for overtaking decision-making problems. Loiacono et al. [10] developed an overtaking behavior by using the Q-learning method [11] and integrated it into the behavior-based architecture of TORCS and reinforcement learning was combined with a behavior-based fuzzy logical architecture. Ngai and Yung [12], [13] studied the vehicle overtaking problem with a multiple-goal reinforcement learning framework, where seven sub-goals were considered and a fusion function was used to obtain a comprehensive decision.

In this paper, we mainly consider to improve the effect of intelligent decision-making of autonomous vehicles in three aspects: 1) simulating different driving scenarios, 2) learning overtaking policies which are complying with highway driving habits or rules, 3) testing the applicability of the overtaking policies. Previous work of Zheng et al. [16] proposed a vehicle model which is generated via data fitting of real vehicles and a simple reinforcement learning framework for overtaking decision-making. Based on their work, we make further use of the vehicle model and improve the reinforcement learning framework for overtaking decision-making. The vehicle model is used to simulate the autonomous vehicle and a traffic model is employed to simulate the traffic flow. The vehicle model is also used in evaluating the performance of overtaking policies. Two main indicators are chosen to evaluate the performance: the average velocity of the autonomous vehicle and the minimum distance from other vehicles through the driving. To make the learnt overtaking policy to be more accordant with driving habits of human drivers, several driving experiences are used to design the reward function. An improved reinforcement learning framework for overtaking decision-

This work is supported by the National Natural Science Foundation of China under Grant 91220301.

making is established, where the Q-learning algorithm is used to learn the overtaking policy. Simulation results show that our method can obtain feasible overtaking policies in different traffic environments.

The paper is organized as follows. In section II, we introduce the problem formulation for autonomous decision-making of intelligent vehicles. Section III proposes the reinforcement learning framework for overtaking decision making in details and Section IV provides the simulation results under different traffic conditions using obtained overtaking policies. Conclusions and discussions for further work are given in the last section.

II. Problem Formulation

A. The Vehicle Model

In order to simulate different driving scenarios, a vehicle model that can reflect the real vehicle dynamic properties is required. In this paper, based on our previous work, we used a vehicle model with 14 degrees of freedom [14]. The vehicle model is established via a data-driven method by using the real driving data from HQ3 (a vehicle style of HongQi).

As shown in Fig. 1, the autonomous vehicle steers around the instantaneous turn center and the two front wheels turn a certain angle according to the attitude angle of the vehicle body, while the rear wheels make no changes in angle. The mapping relationship from the attitude angle to the steering angles of the front wheels is defined as:

$$\begin{cases} \delta_{fl} = \arctan \dfrac{L \tan \delta_{fm}}{L - l_f \tan \delta_{fm}} \\ \delta_{fr} = \arctan \dfrac{L \tan \delta_{fm}}{L + l_f \tan \delta_{fm}} \end{cases} \tag{1}$$

$$\delta_s = \alpha \cdot \delta_{fm} \tag{2}$$

where δ_{fm} is the equivalent angle of the left front wheel while δ_{rm} stands for the right one. L is the distance between the front and rear wheels, and l_f is the vertical distance between the center of mass with the front wheels. α is an approximate constant coefficient . And the rolling model of the wheels is:

$$I_w \dot{\omega} = T_t - T_b - f_x R_l \tag{3}$$

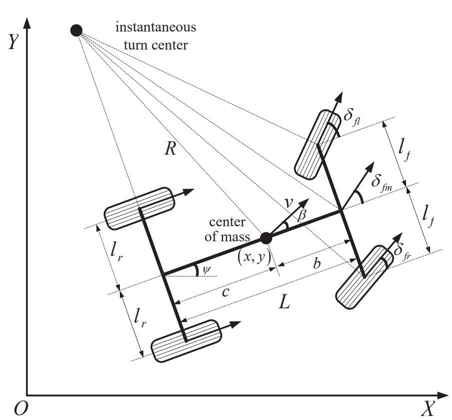

Fig. 1. The steering scenario of a vehicle [16].

where T_t is the driving torque while T_b stands for the braking torque. f_x is the longitudinal force of the wheel and R_l is the load radius. I_w is the moment of inertia which is the same to all wheels, and its value is given in advance. In this way the angular acceleration   of wheels can be obtained.

According to the above formulas, the steering wheel input is mapping to the equivalent angle of each front wheel. In addition, the overall stress in the dynamic model contains the stress from wheels, gravity and air resistance:

$$F = \mathbf{T}\begin{bmatrix} 0 \\ 0 \\ \delta_{fr} \end{bmatrix} F_{fl} + \mathbf{T}\begin{bmatrix} 0 \\ 0 \\ \delta_{fr} \end{bmatrix} F_{fr} + F_{rl} + F_{rr} + F_{wind} + G \tag{4}$$

where T stands for the coordinate transformation matrix while F_{fl}, F_{fr}, F_{rl}, F_{rr} represent the stress from each wheels, respectively. G is the gravity and F_{wind} is the resistance from air:

$$F_{wind} = \begin{bmatrix} -f_{wind} & 0 & 0 \end{bmatrix}' \tag{5}$$

$$f_{wind} = C_w A v^2 g / 16 \tag{6}$$

where C_w is a constant coefficient which is given 0.35 in this paper and g is the acceleration of gravity. A is the front cross sectional area of a vehicle and v is relative velocity between the vehicle and wind.

When using this model, the expected velocity and lateral drift is required. These two parameters will be converted to engine output and steering wheel angle with a PID controller.

B. The Relationship between Vehicles

We assume that all the vehicles drive on highway roads with a two-lane structure which contains a driving lane and an overtaking lane. In order to decide whether to overtake, only nearest vehicles surrounding the autonomous vehicle will be taken into consideration. Fig. 2 shows the relationship between the autonomous vehicle and its surroundings.

TABLE I. Longitudinal Velocity Planning

The current state	Velocity difference (m/s)	Velocity planning (m/s)
d_1-d_{front}>10& d_1>d_e	$\lvert v_1-v_a\rvert$>3.6	v_a+0.25(d_1-d_{front})+1.5(v_1-v_a)
	$\lvert v_1-v_a\rvert$≤3.6	v_a+0.25(d_1-d_{front})+1.0(v_1-v_a)
-4<d_1-d_{front}≤10& d_1>d_e	$\lvert v_1-v_a\rvert$>3.6	v_a+0. 5(d_1-d_{front})+1.5(v_1-v_a)
	$\lvert v_1-v_a\rvert$≤3.6	v_a+0.5(d_1-d_{front})+1.0(v_1-v_a)
d_1-d_{front}≤-4& d_1>d_e	$\lvert v_1-v_a\rvert$>3.6	v_a+1.5(d_1-d_{front})+1.5(v_1-v_a)
	$\lvert v_1-v_a\rvert$≤3.6	v_a+1.5(d_1-d_{front})+1.0(v_1-v_a)
d_1≤d_e	--	0

a. d_e is the minimum braking distance: de = $v_a^2/2a^2$-$v_1^2/2a^2$+10,

b. a is the emergent braking deceleration

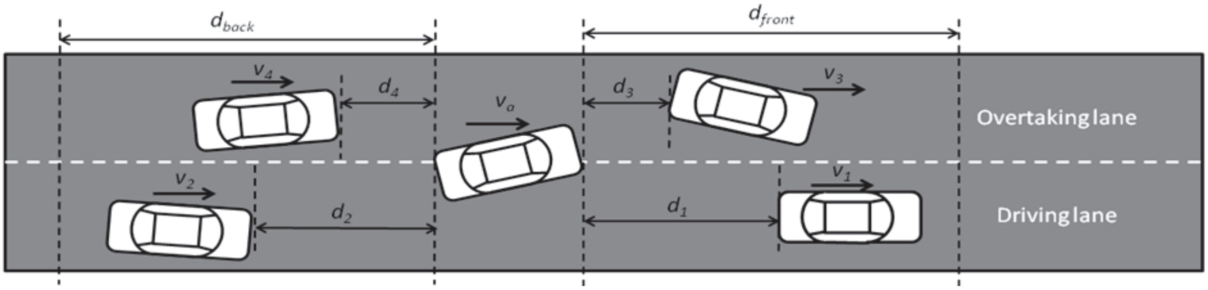

Fig. 2. Relationship between the autonomous vehicle and its surrounding vehicles.

In the two lane driving environment, we named the nearest surrounding vehicles as vehicle 1 (frontward in the driving lane), vehicle 2 (backward in the driving lane), vehicle 3 (frontward in the overtaking lane), and vehicle 4 (backward in the overtaking lane). $\{d_1, d_2, d_3, d_4\}$ are used to describe the related locations between corresponding vehicles. Only the longitudinal distances are considered. d_{front} and d_{back} stand for the maximum forward sight range and maximum backward sight range, respectively. When there are no vehicles surrounded or the other vehicles are too far away to be sensed, the corresponding relative distances should be d_{front} and d_{back}. Similarly, $\{v_1, v_2, v_3, v_4\}$ are used to describe the current longitudinal velocities of corresponding vehicles. When there are no vehicles surrounding or the other vehicles are out of the perception scope, v_i will be assigned by v_a, which is the current velocity of the autonomous vehicle. The expected longitudinal velocities of vehicles are determined from a longitudinal velocity planning module, as shown in Table I.

C. Simulating the Driving Scenarios

Using the vehicle model mentioned above, we can simulate several basic overtaking scenarios. Two cases are mainly taken into consideration: 1) the autonomous vehicle is driving in the driving lane and there is a vehicle in low velocity driving in front, then it needs to decide whether to overtake. 2) the autonomous vehicle is driving in the overtaking lane after passing the vehicle in the driving lane, then it needs to decide when to change back to the driving lane to complete a whole overtaking process. In both cases, there should be at least one other vehicle driving in the other lane. In the initialization of driving scenarios, d_i, v_i and v_a need to be determined first. d_i varied from 10(m) to 100(m) with interval of 10 (m). v_a and v_i varied from 64(km/h) to 98(km/h) with interval of 4(km/h) (i=1,2,3,4).

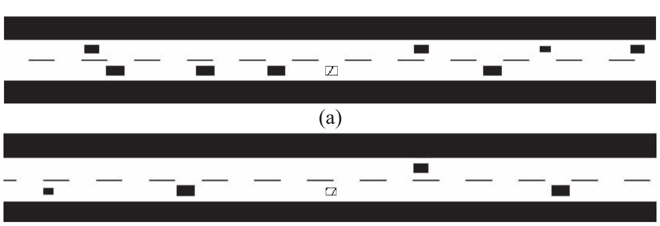

Fig. 3. Simulations of traffics with different density values. "▨" represents the autonomous vehicle and "■" stands for other vehicles. (a) μ=8. (b) μ=4 (μ is the traffic density sign which represents the number of vehicles in the current 500(m) range around the autonomous vehicles).

D. Learning the Overtaking Policy

To obtain the overtaking policy, based on the work in [16], we designed an improved reinforcement learning framework to describe the decision-making task and use the Q-learning algorithm to learn the policy. To make the decision-making method more aligned with habits of human drivers, several driving experiences are considered. Detailed description is given in the next section.

E. Evaluating the Overtaking Policy

After the overtaking policy is learnt via the driving scenarios, it will be applied in a series of different simulated traffic flows to do the evaluation.

To simulate the traffic environment that always contains other vehicles in front of the autonomous vehicle, a 500(m) range around the autonomous vehicle is chosen as the current scene of traffic. When a vehicle falls behind of this range, it will be destroyed and a new vehicle will be created in front. The traffic is initialized and updated randomly with a controllable density. Here 'density' means the current number of vehicles in the scene of traffic. The traffic density is quantified as sign μ (value defaults as an integer). Fig. 3 shows a simple example of simulated traffic with different densities.

Besides the autonomous vehicle, the other vehicles in the traffic should also have the ability to overtake or keep in lane. We propose a simple expert system to achieve the overtaking decision-making. The definition is offered in Table II. When evaluating the overtaking policy, the autonomous vehicle will drive in the simulated traffic flow with multiple other vehicles. Two main indicators are chosen to evaluate the performance: the average velocity of the autonomous vehicle and the minimum distance from other vehicles through the driving.

TABLE II. THE RULES OF A SIMPLE EXPERT SYSTEM

The current lane	Action	Condition
1	Move to lane 2	$(d_1<50 \mid v_1-v_f<-3)$ & $d_4>80$ &$(d_3-d_1>10)$ & $(v_3-v_1>1)$
	Keep in lane 1	*else*
2	Keep in lane 2	$((d_1>150$ & $v_1-v_f>1) \mid v_1>v_3)$ & $d_2>60$
	Move to lane 1	*else*

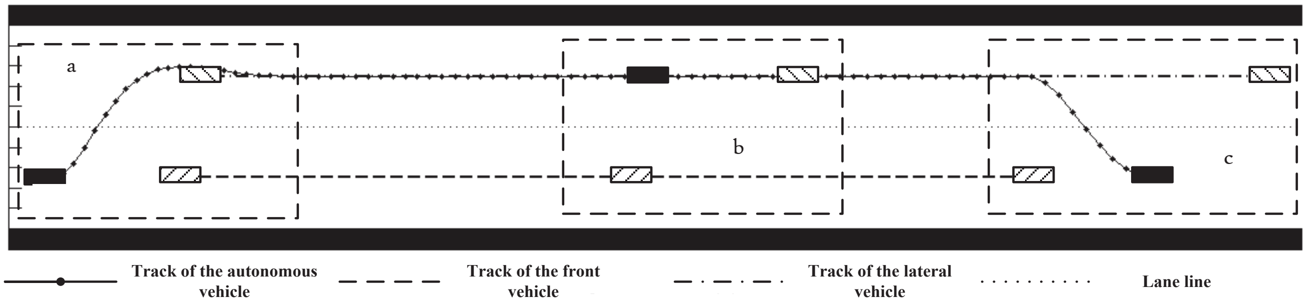

Fig. 4. Verification of the overtaking policy in constant traffic. "▨" is the autonomous vehicle while "▧"and "■" stands for other vehicle.

III. The Reinforcement Learning Framework for Overtaking Decision-Making

Reinforcement learning has been studied to solve a class of sequential decision-making problems which describes how an agent learns to achieve goals from the interactions with unknown or dynamic environment [10]. Instead of exact teacher signals, reinforcement learning receives numerical rewards by evaluating the state of agent and environment after taking a certain action. A greedy strategy is often chosen to search for the action which maximizes the reward value.

The reinforcement learning framework is an important abstraction of the problem for goal-directed learning during interaction and we normally assume it has the Markovian property. The following three basic elements are required to represent the problem of learning goal-directed behaviors:

1) *State*. States of the agent and environment describe the overall change of the system when taking a certain action. Also, a certain state is the basis on which the choices are made.
2) *Action*. Optional actions represent the choices made by the agent. When a choice is made, states of the whole system will change immediately and we can always find a series of choices to achieve the goals.
3) *Reward*. The usage of reward signals to describe the achievement of a goal is one of the most distinctive features of reinforcement learning. In particular, the reward signal is the description of what the agent is expected to achieve.

A. State

The state set of the reinforcement learning framework contains the current statue of the autonomous vehicle and its surroundings, which is defined as follows:

$$s^{(k)} = \left\{ v_a^{(k)}, v_f^{(k)}, v_i^{(k)}, d_i^{(k)}, l^{(k)} \right\}, \quad (i = 1,2,3,4) \tag{7}$$

$$S = \left\{ s^{(k)} \mid k = 0,1,... \right\} \tag{8}$$

where l is the current lane number. v_f is the expected upper limit of v_a. The superscript k represents the current step.

B. Action

In the two-lane driving environment, the autonomous vehicle needs to determine whether to do the lane change. Here we define the driving lane as lane 1 and the overtaking lane as lane 2. When the vehicle locates at the boundary between the driving lane and the overtaking lane, the lane number becomes 1.5.The action set of our reinforcement learning framework for overtaking decision-making is defined after the expected lane number, as shown in Table III. Because of the usage of the vehicle model, we only need to consider the lateral offset from the current lane to the objective lane, and the vehicle model will translate the offset into velocity and angle changes autonomously. In this way, two actions are able to distinguish six different situations.

C. Reward

In order to make the overtaking policy more compliant with human driving habits in highway roads, several driving experiences are referred to represent the rewards of the reinforcement learning framework. Here we mainly consider the following aspects:

1) Collisions must be avoided. The lowest reward will be got when a collision happens.
2) To remain security, velocity change should be controlled strictly. When it changes too fast, the autonomous vehicle will obtain an extreme low reward as half of a collision.
3) A vehicle should not occupy in the overtaking lane for a long time, so the reward value in the overtaking lane is always lower than that in the driving lane.

TABLE III. The Action Set of Reinforcement Learning Framework for Overtaking Decision-Making

The current lane	The objected lane	Action
1	1	Keep in lane 1
	2	Move to lane 2
2	1	Move to lane 1
	2	Keep in lane 2
1.5	1	Move to lane 1
	2	Move to lane 2

TABLE IV. THE Q-LEARNING ALGORITHM

Start Initialize matrix Q as zero matrix For each episode: Select random initial state Do While $(e^{(k)} > 0.1)$ // not yet converged $Q_{old} = Q(s^{(t)}, a^{(t)})$ // Inherit the old Q value Select one among all possible actions for the current state get immediate reward R_{k+1} of state $s^{(t+1)}$ list all the feasible action a for $s^{(t+1)}$ search the maximum consequent reward $\pi(\varepsilon) = \begin{cases} MAX, & P(MAX) = 1-\varepsilon \\ random, & P(random) = \varepsilon \end{cases}$ //using greedy strategy update the optimal current Q value $\hat{Q}(s^{(t)}, a^{(t)}) = (1-\alpha)Q(s^{(t)}, a^{(t)}) + \alpha\left[R_{t+1} + \gamma \max_{a^{(t+1)}}\left(Q(s^{(t+1)}, a^{(t+1)})\right)\right]$ Set the next state as the current state End End End

4) When the current velocity of the autonomous vehicle is higher than its expected upper limit, a deceleration process is encouraged.

5) When overtaking a vehicle in front in the driving lane, d_1 should be kept approximately equal to the value of v_a.

The reward function is defined in the following formula:

$$r^{(k)} = \begin{cases} -300 & c=1 \\ -150 & c=0\,\&\, v_d^{(k)} > 10 \\ -0.1 v_d^{(k)} & else\,\&\, v_a > v_f \\ v_a^{(k)} - v_f^{(k)} - 0.1 v_d^{(k)} & else\,\&\, l=2 \\ v_a^{(k)} - v_f^{(k)} - 0.1 v_d^{(k)} + (d_1 - v_1) & else\,\&\, l=1 \end{cases} \tag{9}$$

$$J = \sum_{k=0}^{T} \gamma^k r^{(k)} \tag{10}$$

where c is the sign of collision. v_d is the velocity difference between the current moment and the last moment.

D. The Learning Algorithm

The Q-learning algorithm is an iterative algorithm in which an iteration error is required to determine the convergence of the algorithm. Here we use a 2-norm form expression:

$$e^{(k)} = L^2(\hat{Q}(k+1) - \hat{Q}(k)) \tag{11}$$

The update process of Q value in Q-learning should be according to the following formula:

$$\hat{Q}(s^{(t)}, a^{(t)}) = (1-\alpha)Q(s^{(t)}, a^{(t)}) + \alpha\left[R(s^{(t)}, a^{(t)}) + \gamma \max_{a^{(t+1)}}\left(Q(s^{(t+1)}, a^{(t+1)})\right)\right] \tag{12}$$

where α is the learning rate valued between 0 and 1, and it determines to what extent the newly acquired information will override the old information. γ is the discount factor valued between 0 and 1, which determines the importance of future rewards. The algorithm flow is shown in Table IV. In our method, α = 1and γ = 0.95.

IV. SIMULATION AND DISCUSSION

A. Evaluation in Basic Traffic Conditions

Firstly, a simple overtaking scenario which contained two other vehicles drove in constant velocity and stayed in their initial lane was defined as the basic traffic condition.

As shown in Fig. 4, (a), (b) and (c) represent the three stages in an overtaking process. In (a), the autonomous vehicle caught the overtaking opportunity and changed lane from the driving lane to the overtaking lane. And in (b), the autonomous vehicle drove passed the vehicle in driving lane but the distance was not afford to make the next lane-change. In this way it went on driving till (c), in which the second lane-change action took place.

B. Policies from Learning and Non-learning Methods

In this part, the policies from Q-learning and the expert system (the same as all other vehicles) were compared. The autonomous vehicles with these two policies were firstly set in a common traffic condition with the density μ=10. As the states of the autonomous vehicle and its surroundings changed, the immediate rewards were recorded in Fig. 5. Obviously, the policy from Q-learning always obtained higher reward than the policy from the expert system.

Then these two policies were further tested with a series of traffic conditions with the density μ varied from 5 to 12. For each density, the simulation was repeated for 50 times. The terminal condition of the simulation was a driving distance greater than 3000(m) or facing a collision. Fig. 6 recorded the performance of different policies from Q-learning and the expert system in different traffic environments. Fig. 6 (a) showed the change of average velocities as the traffic density increased. In low density (μ<7), the policy from the expert system earned higher velocity than Q-learning, but when the traffic density increase, the velocities from the two policies preformed approximately the same. Fig. 6 (b) showed the great difference in the minimum relative distance between vehicles. The policy from Q-learning get a steady change with more reasonable range (30-55(m)) than the policy from the expert system (12-70(m)). We concluded that the learned policy from Q-learning performed better than that from the expert system in maintaining security and performed approximately the same as the expert system in maintaining speediness.

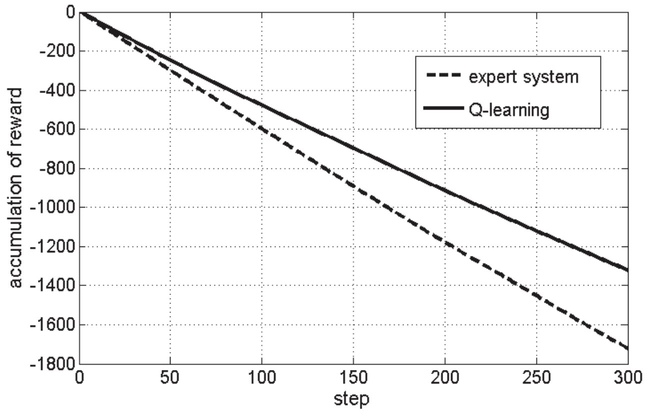

Fig. 5. Accumulation of rewards of Q-learning and the expert system.

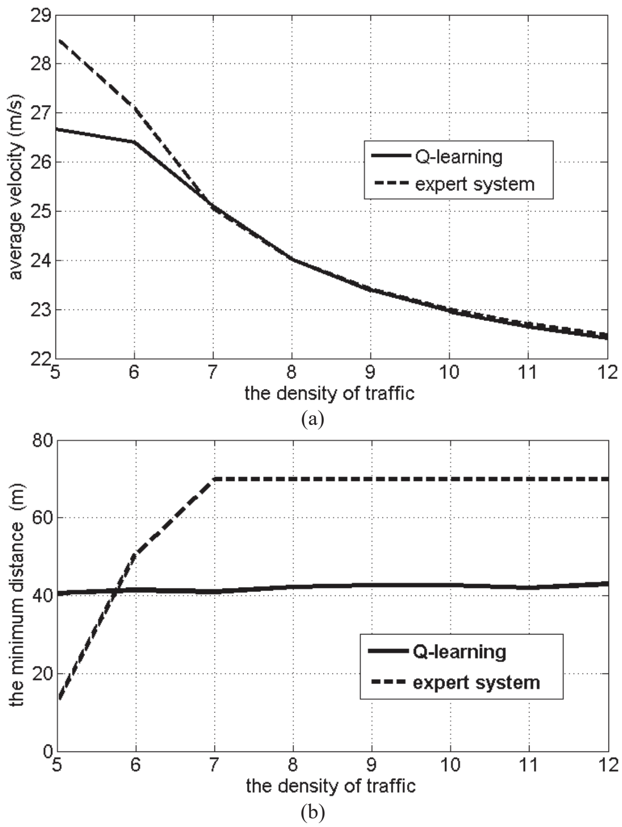

Fig. 6. Comparison of policies from Q-learning and the expert system.

C. *Policies from Q-learning in Different Iteration Stages*

Q-learning is an iterative algorithm, and in different iteration stages different overtaking policies will be obtained. In this part, we compared the policies from different iteration stages. Here we defined three stages: 1) the initial condition before the first update occurs was chosen as the state "before", from which we obtained the "initial policy". 2) When the algorithm converged, the state reached "end" and "final policy" was generated. 3) the "middle" state was determined when the iteration error firstly became no larger than the square root of the initial value, as shown in follow:

$$\min_{k}\left(e^{(k)} \le \sqrt{e^{(0)}}\right) \tag{13}$$

These three policies were tested in a series of traffic conditions with the density $\mu \in [5, 15]$. In each condition, the simulation was repeated for 50 times, the simulation stopped when the autonomous vehicle has drove 3000(m) or faced a collision. Average values of results from each policy were recorded in Fig. 7.

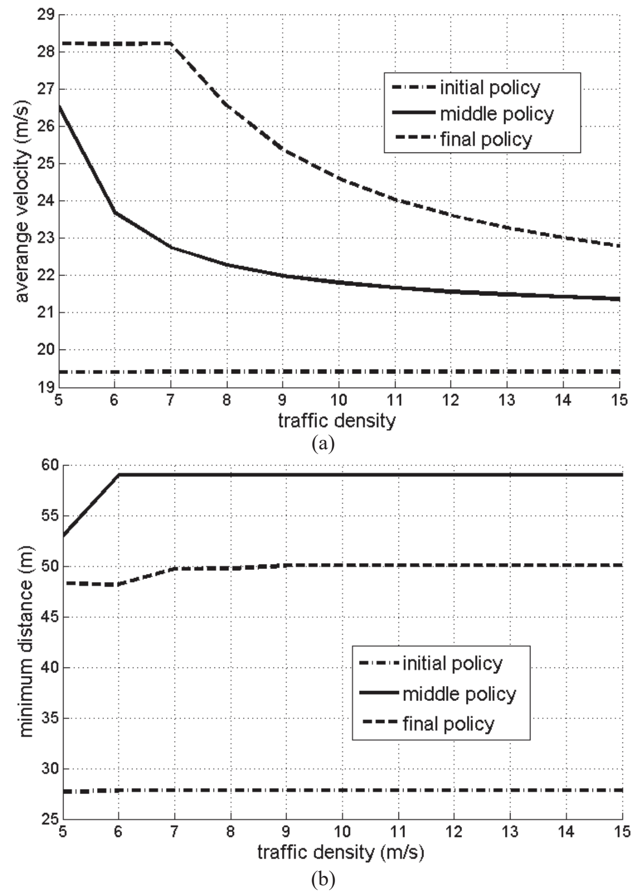

Fig. 7. Comparison of policies from Q-learning with different iteration

Fig. 7 (a) showed that the final policy achieved the highest average velocity while the initial policy obtained the lowest. The curve of middle policy dropped immediately with the increasing of traffic density, but curve of final policy approximately kept the same in low traffic density ($\mu \in [5, 7]$), and the decline appeared when μ exceeded 7. Fig. 7 (b) showed the minimum distance between the autonomous vehicle and other vehicles in the driving process. Here the final policy achieved better results while the results from the initial policy were too close to the front vehicles to ensure the safety and results from the middle policy is too far from the front vehicles to obtain a higher velocity.

V. Conclusion

In this paper, we studied a reinforcement learning based overtaking decision-making method for highway autonomous driving. We use the real data fitting vehicle model to simulate the autonomous vehicle and design a density controllable traffic model. In order to improve the reinforcement learning framework for overtaking decision-making, several highway driving experiences are considered. We use the Q-learning algorithm to obtain the overtaking policies and evaluate their validities in several traffic conditions with two main indicators named average velocity and minimum distance. Simulation results indicate that the method can obtain appropriate policies under different traffic conditions and perform better than non-learning methods. For further study, we will continuously improve the reinforcement learning framework for autonomous decision-making. In addition, we will also pay attention to multi-objective reinforcement learning methods.

REFERENCES

[1] N. Suganuma and T. Uozumi, "Development of an autonomous vehicle—system overview of test ride vehicle in the tokyo motor show 2011," 2012 Proceedings of SICE Annual Conference (SICE), pp. 215-218, 2012.

[2] N. Suganuma and Y. Hayashi, "Development of autonomous vehicle-overview of autonomous driving demonstration in ITS world congress 2013," 11th International Conference on Informatics in Control, Automation and Robotics (ICINCO), vol. 2, pp. 545-549, 2014.

[3] H. Jula, E.B. Kosmatopoulos, and P. A. Ioannou, "Collision avoidance analysis for lane changing and merging," IEEE Transactions on Vehicular Technology, vol. 49, no. 6, pp. 2295-2308, November 2000.

[4] T. Shamir, "How should an autonomous vehicle overtake a slower moving vehicle: design and analysis of an optimal trajectory," IEEE Transactions on Automatic Control, vol. 49, no. 4, pp. 607-610, April 2004.

[5] J. Perez, V. Milanes, E. Onieva, J. Godoy, and J. Alonso, "Longitudinal fuzzy control for autonomous overtaking," IEEE International Conference on Mechatronics (ICM), pp. 188-193, 2011.

[6] O. Karaduman, H. Eren, H. Kurum, and M. Celenk, "Interactive risky behavior model for 3-car overtaking scenario using joint bayesian network," 2013 IEEE Intelligent Vehicles Symposium (IV), pp. 1279-1284, 2013.

[7] J. Nilsson and J. Sjoberg, "Strategic decision making for automated driving on two-lane, one way roads using model predictive control," IEEE Intelligent Vehicles Symposium (IV), pp. 1253-1258, 2013.

[8] R.S. Sutton and A.G. Barto, "Reinforcement learning: an introduction," Cambridge: MIT press, vol. 1, no. 1, 1998.

[9] L. Busoniu, D. Ernst, B. De Schutter, and R. Babuska, "Approximate reinforcement learning: an overview," 2011 IEEE Symposium on Adaptive Dynamic Programming And Reinforcement Learning (ADPRL) Paris, pp. 1-8, 2011.

[10] D. Loiacono, A. Prete, P.L. Lanzi, and L. Cardamone, "Learning to overtake in TORCS using simple reinforcement learning," IEEE Congress on Evolutionary Computation (CEC), pp. 1-8, 2010.

[11] Q. Wang and Z.L. Zhan, "Reinforcement learning model, algorithms and its application," IEEE International Conference on Mechatronic Science, Electric Engineering and Computer (MEC), pp. 1143-1146, 2011.

[12] D.C.K. Ngai and N.H.C. Yung, "Automated vehicle overtaking based on a multiple-goal reinforcement learning framework," IEEE Intelligent Transportation Systems Conference (ITSC), pp. 818-823, 2007.

[13] T.C. Du, M.T. Cox, D. Perlis, J. Shamwell, and T. Oates, "From robots to reinforcement learning," IEEE 25th International Conference on Tools with Artificial Intelligence (ICTAI), pp. 540-545, 2013.

[14] R. Zheng, C.M. Liu, and Q. Guo, "A decision-making method for autonomous vehicles based on simulation and reinforcement learning," International Conference on. Machine Learning and Cybernetics (ICMLC), vol. 1, pp. 362-369, 2013.