Proceedings of the 2013 International Conference on Machine Learning and Cybernetics, Tianjin, 14-17 July, 2013

A DECISION-MAKING METHOD FOR AUTONOMOUS VEHICLES BASED ON SIMULATION AND REINFORCEMENT LEARNING

RUI ZHENG, CHUNMING LIU, QI GUO

College of Mechatronics and Automation, National University of Defense Technology, Changsha 410073, P. R.China

Abstract:

There are still some problems need to be solved though there are a lot of achievements in the field of automatic driving. One of those problems is the difficulty of designing a decision-making system for complex traffic conditions. In recent years, reinforcement learning (RL) shows the potential in solving sequential decision optimization problems, which can be modeled as Markov decision processes (MDPs). In this paper, we establish a 14-DOF dynamic model of an autonomous vehicle and use RL to build a decision-making system for autonomous driving based on simulation. The decision-making process of the vehicle is modeled as an MDP, and the performance of the MDP is improved using an approximate RL. At last, we show the efficiency of the proposed method by simulation in a highway environment.

Keywords:

Autonomous Vehicles; Reinforcement learning; Markov Decision Process; Autonomous driving; Decision-making

1. Introduction

With the improvement of technology and people's living standards, the safety of car driving is drawing more and more attention. With the demands of transportation and safety, intelligent vehicles have been studied. As the first intelligent machine called AGVS was made by Barrett Electrics in USA, the study on intelligent vehicle has never been stopped since 1950. After 1990, in order to ensure the traffic safety and apply this technology to national defense, the study has turned into a new phase which the intelligent vehicle has become the center in this field. Now, the existing work about intelligent vehicle includes automatic control [1], speed safety, lane departure warning [2-4], hitting prediction [5], lane keeping [6], corner safety and so on.

Vehicle decision-making is an internal part of intelligent vehicle system, including lane keeping and lane changing [7]. The main objective of lane keeping is to perform automatic steering of the vehicle in order to keep the vehicle in the middle of the road. The control system of lane keeping usually designed to detect any difference between the intelligent vehicle and environment [8]. However, the purpose of lane changing is to turn the vehicle to the right lane in time in order to takeover other vehicles. This system is usually based on the distance or the velocity of host vehicle and others. Generally, there are two methods to design the vehicle decision-making system [9]. One is to modify or configure a real vehicle for actual road testing. The United States, Europe, Japan [10], and Tsinghua University [11] and Jilin University have carried out relevant studies, and made sample vehicle. The common feature of these projects is the usage of real cars to do some safe driving studies on certain aspects [9]. The second method is to establish simulation by using virtual environment. For example, the University of Iowa of the United States developed the world's most advanced simulator-the U.S. National Advanced Driving Simulator (NADS) in 1999 [12]. In 2009, the Automobile Study Institute of Tsinghua University bought a dynamic virtual driving test-bed from Japan, which is now being used for a number of studies on driver's driving behaviors. Though there are many achievements in these years, there are still many problems existing in complex condition. The most remarkable one is the difficulty of modeling the complex vehicle in virtual environment. Another is the definition of decision-making process.

In RL, the learning agent interacts with an initially unknown environment and modifies its action policies to maximize its cumulative payoffs. Thus, RL provides a general methodology to solve complex uncertain sequential decision problems. The environment of RL is typically modeled as a Markov decision process or Markov decision problem (MDP). However, different from traditional dynamic programming methods in operations research, an RL agent is assumed to learn the optimal or near-optimal policies from its experiences without knowing the parameters of MDP. In recent years, there are many successful developments in RL.

978-1-4799-0260-6/13/$31.00 ©2013 IEEE

To solve difficult decision and control problems in uncertain dynamic systems, data-driven learning control methods, especially reinforcement learning (RL), have received much study interests in recent years. In 1988, Richard S. Sutton proposed a widely-used method called temporal-difference learning (TD) [13]. With the help of the theory of linear least-squares function approximation, Bradtke and Andrew G. Barto established two algorithms which were called Least-Squares TD (LSTD) and Recursive Least-Squares TD (RLSTD) [14] in 1996. In 2003, combining value function approximation with linear architectures and approximate policy, Michail G. Lagoudakis and Ronald Parr proposed an approach called least-squares policy iteration (LSPI). What's more, a new method called Kernel-based Least-Squares policy iteration (KLSPI) [14][15] that adopts the kernel-function as the approximator was proposed by Xu et al., so that it can solve value function approximation problems in MDPs.

In this paper, we first establish a 14-DOF dynamic simulation model of the vehicle which adapted to the highway environment; then we apply RL to the decision-making process and verify it in the simulation model. The rest of this paper is organized as follows. In Section 2, a vehicle model of 14 degrees of freedom is given. The decision-making method based on RL is presented in Section 3, where the reward function and control design are discussed in detail. At last, simulation results are provided to illustrate the effectiveness of the method. Section 4 draws conclusions.

2. Vehicle model with 14 degrees of freedom

The establishment of the vehicle dynamic model is a very complex process. As the mutual coupling effects between the various subsystems of the vehicle, the dynamic characteristics of the vehicle become very complicated. The dynamic model of the vehicle includes the steering system model, the body suspension model and the motion model. The first two models can be modeled via the analysis of the mechanical structure of the car. But as the high speed condition, motion model is difficult to model through the structural analysis, so we complete it by data driven manner.

2.1 Steering system model

The steering system is the apparatus used to change or maintain the direction when the vehicle is traveling forward or backward. In normal conditions, the angle of the steering wheel δ_s and the angle of the vehicle's cartwheel δ_{fm} is approximated as a linear relationship:

$$\delta_s = \alpha \cdot \delta_{fm} \tag{1}$$

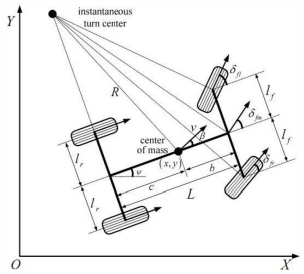

Figure 1. A dynamic model of the vehicle at low velocity

Figure 1 shows the dynamic model of the vehicle at low speed. x and y is the coordinate of vehicle's centroid in the earth reference frame OXY, v is the speed of vehicle's centroid, ψ is the vehicle's transverse swing angle, δ_{fl}, δ_{fr} and δ_{fm} are the angle of the vehicle's front revolver, the angle of the vehicle's front right wheel and the equivalent steering wheel angle, respectively. β is the sideslip angle of the vehicle's centroid, R is the instantaneous turning radius, L is the distance of the front axle to the rear axle, l_f and l_r are the half length of the front axle and the rear axle, respectively.

In the process of the vehicle's turning at slow speed, the cartwheel satisfies the nonholonomic constraints relative to the ground. The vehicle's center of instantaneous has been located in the extension line of the rear axle of the vehicle. It means that the velocity direction of each point on the rear axis is always perpendicular to the rear axle, and the speed of the entire rear axle and the rear wheel is 0 at the rear axle direction. So the speed of each point on the rear axle should satisfy the following:

$$\dot{x}_C \sin\psi - \dot{y}_C \cos\psi = 0 \tag{2}$$

The relationship between the two harmony angle of the front wheel and the equivalent rotation angle of the steering wheel is:

$$\begin{cases} \delta_{fl} = \arctan \dfrac{L \tan \delta_{fm}}{L - l_f \tan \delta_{fm}} \\ \delta_{fr} = \arctan \dfrac{L \tan \delta_{fm}}{L + l_f \tan \delta_{fm}} \end{cases} \tag{3}$$

2.2 Body and suspension model

Assuming that the road surface where the vehicle traveling on as the plane Ω, the projective point O_l which the car's centroid O_c projected on the plane Ω as the origin, the projection which the forward direction of the car projected on the plane Ω as the direction of X-axis, perpendicular to the plane Ω and toward the upper direction as the direction of Z-axis, the direction of Y-axis is

determined by the right hand rule, we can establish the Cartesian coordinate system $O_l X_l Y_l Z_l$ to the local coordinate system, and by moving the local coordinate system $O_l X_l Y_l Z_l$ to the car's centroid O_c, the body coordinate system $O_c X_c Y_c Z_c$ shown in Figure 2 is achieved.

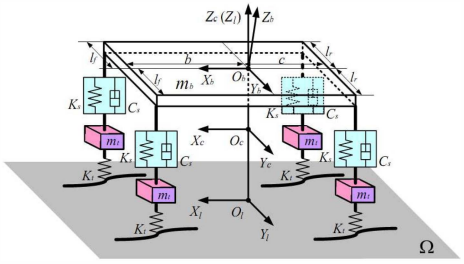

Figure 2. Dynamics model of body and suspension

Assuming that the point which the plane of the body's rectangle and the $O_l Z_l$-axis intersects is the origin, the line which the plane of the body's rectangle and the plane $O_l X_l Z_l$ intersects is the X-axis, the line which the plane of the body's rectangle and the plane $O_l Y_l Z_l$ intersects is the Y-axis, we can establish the body coordinate system $O_b X_b Y_b Z_b$.

The force acting on the vehicle in the local coordinate system $O_l X_l Y_l Z_l$ is:

$$\begin{aligned} F &= T_{n2o}(0,0,\delta_{fl})F_{wheel\_fl} + T_{n2o}(0,0,\delta_{fr})F_{wheel\_fr} \\ &+ F_{wheel\_rl} + F_{wheel\_rr} + F_{wind} + G_l \end{aligned} \tag{4}$$

where G_l is the gravity and F_{wind} is the air resistance,

$$G_l = m_c \begin{bmatrix} g_{lx} \\ g_{ly} \\ g_{lz} \end{bmatrix} = m_c T_{o2n}(0,0,\psi_{rb}) T_{o2n}(\phi_{er},\theta_{er},\psi_{er}) \begin{bmatrix} 0 \\ 0 \\ -g \end{bmatrix} \tag{5}$$

$$F_{wind} = \begin{bmatrix} -f_{wind} & 0 & 0 \end{bmatrix}^T$$

m_c Is the quality of the vehicle, $f_{wind} = C_w A v^2 g / 16$, $C_w = 0.35$ is the coefficient of the air resistance, which in the range of 0.3 to 0.6, $A = 2.7m^2$ is the cross-sectional area of the vehicle, v is the relative velocity of the vehicle and the wind.

So the kinematic equation is described as follows:

$$F = m_c a_l = m_c \begin{bmatrix} a_{lx} \\ a_{ly} \\ a_{lz} \end{bmatrix} = m_c \left(\frac{dv_l}{dt} + \dot{\psi}_{rb} \times v_l \right) \tag{6}$$

which v_l and a_l are the speed and the acceleration of the vehicle under the local instantaneous coordinate system $O_l X_l Y_l Z_l$.

2.3 Motion Model

For the front-wheel drive car, the torque named T_e come from the engine, through the clutch, transmission, assigned to the two front wheels by the differential finally. Assuming the torque which was passed to the shuck of the differential is T_t, it was distributed to the left axle and the right axle as:

$$\begin{cases} T_t = T_{t\_fl} + T_{t\_fr} \\ I_s \dot{\omega}_s = \frac{1}{2}\left[\left(T_{t\_fl} - T_{t\_fr} \right) \frac{r_s}{r_t} \right] - T_s \end{cases} \tag{7}$$

where T_{t_fl} and T_{t_fr} are the torque assigned to the front left wheel and the front right wheel respectively, I_s is the moment of inertia of the differential planetary gear, ω_s is the speed of angle of the differential planetary gear, r_s is the radius of the differential planetary gear, r_t is the radius of the input axle of differential frictional torque. T_s is the torque of friction which come from the differential.

According to the movement relationship of the differential: $\dot{\omega}_s = \frac{1}{2}\left(\dot{\omega}_{fl} - \dot{\omega}_{fr} \right) \frac{r_t}{r_s}$ (8)

Then we can get the left and right side output torque of the differential:

$$\begin{cases} T_{t\_fl} = (aT_t + c)/(a+b) \\ T_{t\_fr} = (bT_t - c)/(a+b) \end{cases} \tag{9}$$

Among them:

$$\begin{cases} a = I_{w\_fl} I_{w\_fr} r_s^2 - I_s I_{w\_fl} r_t^2 \\ b = I_{w\_fl} I_{w\_fr} r_s^2 - I_s I_{w\_fr} r_t^2 \\ c = 2I_{w\_fl} I_{w\_fr} r_s r_t T_s + I_s r_t^2 (I_{w\_fl} T_{b\_fr} + I_{w\_fl} f_{x\_fr} R_{l\_fr} \\ \quad I_{w\_fl} f_{x\_fr} R_{l\_fr} - I_{w\_fr} T_{b\_fl} - I_{w\_fr} f_{x\_fl} R_{l\_fl}) \end{cases} \tag{10}$$

For the accelerating model and the braking model of the system, we use the data-driven method. The particular is the least squares method, and the base function is selected as the polynomial function. The real vehicle model is selected as the HQ430.

For the accelerating model, the base function is selected

as the cubic polynomial. Then we obtain the relationship between the longitudinal acceleration named a, the velocity named v and the throttle opening degree named o:

$$a = \omega_1 v^3 + \omega_2 v^2 o + \omega_3 v o^2 + \omega_4 o^3 + \omega_5 v^2 + \omega_6 v o + \omega_7 o^2 + \omega_8 v + \omega_9 o + \omega_{10} \quad (11)$$

Among them:

$$\omega_1 = -5.24e-05, \omega_2 = 0.00015, \omega_3 = -7.73e-05,$$
$$\omega_4 = -4.84e-05, \omega_5 = -0.00063, \omega_6 = -0.0034$$
$$\omega_7 = 0.0072, \omega_8 = 0.065, \omega_9 = -0.15, \omega_{10} = 0.56$$

Similarly, for the braking model, the base function is selected as the quadratic polynomial. Then we obtain the relationship between the longitudinal acceleration noted as a, the velocity v and the braking pressure p:

$$a = \omega_1 v^3 + \omega_2 v^2 p + \omega_3 v p^2 + \omega_4 p^3 + \omega_5 v^2 + \omega_6 v p + \omega_7 p^2 + \omega_8 v + \omega_9 p + \omega_{10} \quad (12)$$

Among them:

$$\omega_1 = -5.24e-05, \omega_2 = 0.00015, \omega_3 = -7.73e-05,$$
$$\omega_4 = -4.84e-05, \omega_5 = -0.00063, \omega_6 = -5.04e-06,$$
$$\omega_7 = 3.82e-06, \omega_8 = -0.053, \omega_9 = -0.012, \omega_{10} = 0.21$$

2.4 Verification of the simulation model

Through the comparison of the simulation of the dynamic response and the experimental results of the real vehicle, we will validate the dynamics simulation model based on the HQ430. The simulation tests include the simulation of the accelerating system and the braking system.

The input of the simulation of the accelerating system comprises the angle of the steering wheel, the throttle opening degree and the braking pressure. Among them, the angles of the steering wheel and the braking pressure are 0.

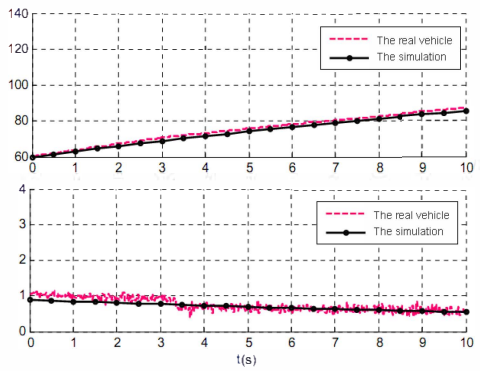

Figure 3. The comparison of the real vehicle with the simulation of the accelerating system

The input of the simulation of the braking system includes the angle of the steering wheel, the throttle opening degree and braking pressure. Among them, the angle of the steering wheel is 0 and the throttle opening degree is 17.

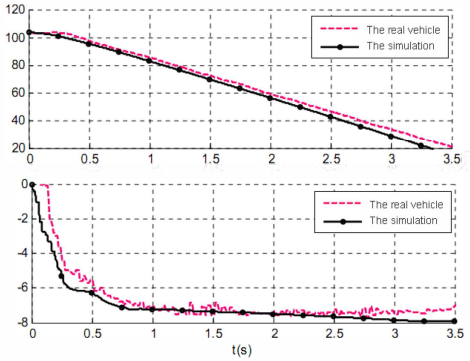

Figure 4. The comparison of the real vehicle with the simulation of the braking system

Figures 3 and Figure 4 show the experiment results of the accelerating system and the braking system respectively. The results show that the simulation model and the real vehicle is consistent, which means that this model can reflect the response performance of the real vehicle.

3. Decision-making design based on simulation model

3.1 The control system design of the autonomous vehicle

Generally, the plan of the vehicle's speed consists of two parts: the speed keeping and the vehicle following. The plan is usually associated with the specific task if there don't have any other vehicles and obstacles in front of this vehicle; and the plan is usually chosen for following the car which driving in front of it (obstacle can be seen as the vehicle which speed is 0) to maintain a safe distance.

The process of following a car is complex and it always depend on the experience of the people. The experience of following a car should be reflected by two levels: the state transition and the distance following. Generally, the following distance and the emergent braking distance can be defined as:

$$d_{follow} = \frac{v_0^2}{2a_1} + T_1 v_0 + 5$$
$$d_{emergency} = \frac{v_0^2}{2a_2} - \frac{v_1^2}{2a_2} + T_2 v_0 + 10 \quad (13)$$

Which a_1 is the braking deceleration which is set in advance, T_1 is the driver's reaction time when followed the

vehicle, a_2 is the emergent braking deceleration which is set in advance, T_2 is the driver's reaction time when emergent braking.

Combine the changing of the distance with the changing of the state in the process of following the vehicle, and the PID controller which is used to maintaining distance, we get the plan of following the vehicle, as shown in Table 1 (the unit of the distance is m and the speed is km / h)

For the highway accord with the normative structure, in which the track of the vehicle is also simple, so is the path planning. Under the local rectangular coordinate system OXY , the path planning results of the vehicle is shown in Table 2.

TABLE 1. PLAN OF FOLLOWING THE VEHICLE

	The speed	The plan
$d_1 - d_{follow} > 10$ and $d_1 > d_{emergency}$	$v_1 - v_0 > 3.6$	$v_0 + 0.25(d_1 - d_{follow}) + 1.5(v_1 - v_0)$
	$3.6 \geq v_1 - v_0 > -3.6$	$v_0 + 0.25(d_1 - d_{follow}) + 1.0(v_1 - v_0)$
	$v_1 - v_0 \leq -3.6$	$v_0 + 0.25(d_1 - d_{follow}) + 1.5(v_1 - v_0)$
$10 \geq d_1 - d_{follow} > -4$ 且 $d_1 > d_{emergency}$	$v_1 - v_0 > 3.6$	$v_0 + 0.5(d_1 - d_{follow}) + 1.5(v_1 - v_0)$
	$3.6 \geq v_1 - v_0 > -3.6$	$v_0 + 0.5(d_1 - d_{follow}) + 1.0(v_1 - v_0)$
	$v_1 - v_0 \leq -3.6$	$v_0 + 0.5(d_1 - d_{follow}) + 1.5(v_1 - v_0)$
$d_1 - d_{follow} \leq -4$ 且 $d_1 > d_{emergency}$	$v_1 - v_0 > 3.6$	$v_0 + 1.0(d_1 - d_{follow}) + 1.5(v_1 - v_0)$
	$3.6 \geq v_1 - v_0 > -3.6$	$v_0 + 1.0(d_1 - d_{follow}) + 1.0(v_1 - v_0)$
	$v_1 - v_0 \leq -3.6$	$v_0 + 1.0(d_1 - d_{follow}) + 1.5(v_1 - v_0)$
$d_1 \leq d_{emergency}$		0

For the highway accord with the normative structure, the trajectory of the vehicle is simple, so is the path planning. Under the local rectangular coordinate system OXY , the path planning results of the vehicle is shown in Table 2.

TABLE 2. PATH PLANNING OF FOLLOWING THE VEHICLE

The state of the vehicle	The desired path
Driving on the driving lane	$y = \frac{d}{2}$
Driving on the overtaking lane	$y = \frac{3d}{2}$
Changing from the driving lane to the overtaking lane	$y = \begin{cases} d - \frac{d}{2}\cos\left(\frac{\pi t}{T_{change}}\right) & T_{change} \geq t > 0 \\ \frac{3d}{2} & t > T_{change} \end{cases}$
Changing from the overtaking lane to the driving lane	$y = \begin{cases} d + \frac{d}{2}\cos\left(\frac{\pi t}{T_{change}}\right) & T_{change} \geq t > 0 \\ \frac{d}{2} & t > T_{change} \end{cases}$

3.2 MDP modeling of the decision-making system

MDP can be expressed by a list $\{S, A, r, P, \eta\}$, which the state space is S, the behavior space is A, the reword function is $r: S \times A \mapsto R$, the state transformation matrix is P, and the decision-making objective function is η.

Vehicle's current state can be described by: the lane where the vehicle named l, the longitudinal speed of the vehicle named v_0, the distance and the speed of the vehicle which in front of this one on the driving lane named d_1 and v_1, the distance and the speed of the vehicle which behind this one on the driving lane named d_2 and v_2, the distance and the speed of the vehicle which in front of this one on the overtaking lane named d_3 and v_3, the distance and the speed of the vehicle which behind this one on the overtaking lane named d_4 and v_4.

Considering the process of driving by human being, we need to synthesize the relationship of the speed and the distance of the environmental vehicle. So we select the remnant reaction time of each lane as the integrated variable:

$$\begin{cases} t_1 = (d_1 + d_{s1} - d_{s0}) / v_0 \\ t_2 = (d_2 + d_{s0} - d_{s2}) / v_2 \\ t_3 = (d_3 + d_{s3} - d_{s0}) / v_0 \\ t_4 = (d_4 + d_{s0} - d_{s4}) / v_4 \end{cases} \tag{14}$$

Among them, d_{s0}、d_{s1}、d_{s2}、d_{s3} and d_{s4} are the minimal safety distance at the speed v_0、v_1、v_2、v_3 and v_4. So the state of MDP can be predigested as:

$$S = \{(l, t_1, t_2, t_3, t_4)\} \tag{15}$$

To avoid making different decisions too frequently in the process of changing lanes, the vehicle will keep the decisions made in previous time if $l = 1.5$ (the vehicle both in the driving lane and the overtaking lane); and if $l \neq 1.5$, the aggregate of the action of MDP is $A = \{a_1, a_2\}$, which is shown in Table 3

TABLE 3. THE ACTION OF MDP

Action	Name of the action	The description of the action
a_1	Driving on the driving lane	Following in limited speed v_{task} driving on the driving lane
a_2	Driving on the overtaking lane	Following in limited speed $v_{overtake}$ driving on the overtaking lane

Following in limited speed means that the vehicle take the smaller one of v_{plan} and v_{task} (or $v_{overtake}$) as traveling speed. In the process of change lane, the vehicle follows the one which has the least reaction time in front of it.

The performances of the driving process include safety, smoothness and quickness. The corresponding reward functions are defined as follow:

$$r_{safe} = \begin{cases} t_1 & l = 1, d_1 > 3, d_2 > 3 \\ t_3 & l = 2, d_3 > 3, d_4 > 3 \\ -100 & others \end{cases} \tag{16}$$

$$r_{smooth} = -\sum |\Delta v_0| \tag{17}$$

$$r_{quick} = \begin{cases} v_0 - v_{task} & v_0 < v_{task}, l = 1 \\ v_0 - v_{task} - 0.2 & v_0 < v_{task}, l = 2 \\ 0 & v_0 \geq v_{task}, l = 1 \\ -0.2 & v_0 \geq v_{task}. l = 2 \end{cases} \tag{18}$$

Among them, $l = 1, d_1 > 3, d_2 > 3$ means the vehicle is driving on the driving lane and the distance with others is more than 3m; $l = 2, d_3 > 3, d_4 > 3$ means the vehicle is driving on the overtaking lane and the distance with others is more than 3m, $\sum |\Delta v_0|$ denotes the accumulation of the longitudinal speed changing over time at the process of taking action.

The total reward function is defined as the summation of these three functions.

3.3 RL to solve the MDP problem

In this paper, we choose Least Square Policy Iteration

(LSPI) method to solve the Autonomous Vehicle problem. In order to avoid 'curse of dimensionality', linear approximation is used to approach the state-action function $\hat{Q}^{\pi[t]}(s,a)$. The state-action function $\hat{Q}^{\pi[t]}(s,a)$ can be approached by basic function with weight as follow:

$$\hat{Q}^{\pi[t]}(s,a,w) = \phi(s,a)^T \omega \tag{19}$$

Among them:

$$\phi(s,a) = \left(\underbrace{0,\cdots,0}_{M(l-1)}, \phi_1(s), \phi_2(s), \cdots, \phi_M(s), \underbrace{0,\cdots,0}_{M(N_a-1)} \right) \tag{20}$$

N_a is the number of actions, action a is defined as l, $\{\phi_i(s)\}$ is the basic function, $\omega = (\omega_1, \omega_2, \cdots, \omega_{M\times Na})^T$ is the weight vector.

For a sample set $D = \{(s_i, a_i, s_i', r_i) \mid i = 1,2,\cdots,L\}$ defined

$$\Phi = \begin{pmatrix} \phi(s_1,a_1)^T \\ \cdots \\ \phi(s_i,a_i)^T \\ \cdots \\ \phi(s_L,a_L)^T \end{pmatrix} \quad \Phi' = \begin{pmatrix} \phi(s_1',\pi[t](s_1'))^T \\ \cdots \\ \phi(s_i',\pi[t](s_i'))^T \\ \cdots \\ \phi(s_L',\pi[t](s_L'))^T \end{pmatrix} \quad R_e = \begin{pmatrix} r_1 \\ \cdots \\ r_i \\ \cdots \\ r_L \end{pmatrix}$$

The solution of Least Square Policy Iteration is :

$$\begin{cases} \omega^{\pi[t]} = \left(\Phi^T(\Phi - \gamma\Phi')\right)^{-1} \Phi^T R_e \\ \pi[t+1](s) = \arg\max \phi(s,a)^T \omega^{\pi[t]} \end{cases} \tag{21}$$

The LSPI Algorithm can be presented as follows.

LSPI ($\pi_0, \varepsilon, \omega_0$) Algorithm:

initialization: $\tilde{A} \leftarrow 0$, $\tilde{b} \leftarrow 0, \omega' \leftarrow \omega_0, \pi' \leftarrow \pi_0$
repeat
  $\omega \leftarrow \omega'; \pi \leftarrow \pi';$
  for each $(x_i, a_i, x_{i+1}, r_i) \in S$
    $\tilde{A} \leftarrow \tilde{A} + \phi(x_i,a_i)(\phi(x_i,a_i) - \gamma\phi(x_{i+1}, \pi(x_{i+1}))^T$
    $b \leftarrow b + \phi(x_i,a_i) r_i.$
  end
  $\omega' \leftarrow \tilde{A}^{-1}\tilde{b};$
  $\pi' = \arg\max \phi^T(x,a)\omega$
until ($\| \omega - \omega' \| < \varepsilon$ or $\pi \approx \pi'$)
return π.
export π

3.4 The simulation result of the automatic driving

In the simple condition which is shown in Figure 5, we test the strategy which is obtained by RL. In this condition, environmental vehicles keep traveling on the original lane at limited speed, and smart vehicle makes decision based on RL and current road conditions. In Figure 5, (a) is the state of the initial moment; (b) to (h) are the states of the driving process.

At the initial moment, the Smart vehicle ① is located on the driving lane and the longitudinal displacement is 0, environmental vehicles ② ~ ④ are located in front of the smart vehicle. The initial speed of the Smart vehicle and environmental vehicles ② and ③ are based on their tasks, the initial speed of the environment vehicle ④ is the overtaking speed. After 4, 8, 12, 16, 20, 24 and 28 minutes, the smart vehicle completed the overtaking action and the entering traffic action in the simulation environment, as shown in Figure 5. It can be seen, the strategies based on RL can safely complete the decision-making tasks.

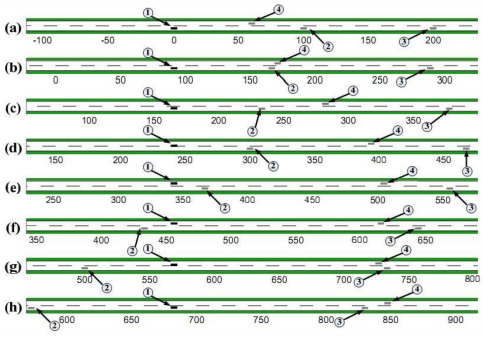

Figure 5. Intelligent vehicle driving process in simulated conditions

4. Conclusion

In this paper, we establish the simulation model with 14 degrees of freedom for the driving problem on highway and combine the RL with this model to build a decision-making system of the highway conditions. Firstly, based on the dynamic characteristic of the vehicle and the data-driven method, the dynamic model is established to provide a convenient and effective simulation tool for the study of automatic driving technology in highway environment. The testing comparison between the HQ430 and the simulation model shows that the simulation model can effectively reflect the dynamic characteristics of the real vehicle. We build up the decision-making system, combining the RL, under the simulation conditions by selecting appropriate state, action and reward. Simulation results show that the decision-making system is effective and provides an

important foundation to the real decision-making problem.

Acknowledgement

This paper is supported by National Natural Science Foundation of China under Grant 61075072, & 91220301, the Program for New Century Excellent Talents in University under Grant NCET-10-0901.

Reference

[1] Bertozzi M, Broggi A. Gold: "A Parallel Real-time Stereo Vision System for Generic Obstacle and Lane Detection" [J]. IEEE Transactions on Image Process, Vol 7, No 1:62-81, 1998.

[2] Kwon W, Lee S, "Performance Evaluation of Decision Making Strategies for An Embedded Lane Departure Warning System" [J]. Journal of Robotic System, Vol 19, No 10: 499-509, 2002.

[3] Amditis A, Bimpas M, Thomaidis G, Tsogas M, Netto M, Mammar S, Beutner A, Mohler N, Wirthgen T, Zipser S, Etemad A, Lio D M, Ciclloni R. "A Situation-adaptive Lane-keeping Support System:Overview of The Safelane Approach" [J]. IEEE Transactions on Intelligent Transportation Systems, Vol 11, No 3: 617-629, 2010.

[4] Lee J W, Kee C D, Yi U K, "A New Approach For Lane Departure Identification" [C]// Proceedings of IEEE Intelligent Vehicles Symposium. Columbus, OH, 100-105, 2003.

[5] Salvucci D D. "Inferring Driving Intent: A Case Study in Lane-change Detection" [C]// Proceedings of Human Factors Ergonomic Society 48th Annual Meeting. New Orleans, LA, Vol 48, No 19: 2228-2231, 2004

[6] Mccall J C, Trivedi M M. "Video-based Lane Estimation and Tracking for Driver Assistance: Survey, System, and Evaluation" [J]. IEEE Transactions on Intelligent Transportation Systems, Vol 7, No 1: 20-37, 2006.

[7] Zhang J M, Ren D B, Cui S M, Zhang J Y: "Direct Adaptive Control for Lane Keeping in Intelligent Vehicle Systems". Journal of Harbin Institute of Technology (New Series), Vol 16, No 6, 2009.

[8] Rajami R: "Vehicle Dynamics and Control". New York: Springer, 2006.

[9] Yang X H, Gao F, Liu G L, Wang G F, Xu G Y: "Development of An Intelligent Vehicle Experiment System." Chinese Journal of Mechanical Engineering. Vol 23, No 6, 2010

[10] Li Li, Jingyan Song, Fei-Yue Wang, Wolfgang Niehsen, Nan-Ning Zheng: "IVS 05: new developments and research trends for intelligent vehicles", IEEE Transacctions on Intelligent Systems, Vol 20, No 4:10-14, 2004.

[11] K. LEE;H. PENG: "Evaluation of automotive forward collision warning and collision avoidance algorithms" Vehicle System Dynamics, Vol 43, No 10:735-751, 2005.

[12] Stall D A, Bourne S: "The National Advanced Driving Simulator Potential Applications to ITS and AHS Research" [C]//1996 Annual Meeting of ITS America, Houston, Texas, America, April 15-18:1-15, 1996.

[13] Sutton R., Barto A.: "Reinforcement learning: an introduction." Neural Networks, Vol 13, No 1: 133-135, 2000

[14] Xin Xu, Dewen Hu, "Kernel-Based Least Squares Policy Iteration for Reinforcement Learning" IEEE Trans on Neural Networks, Vol 18, No 4:973–992, July 2007

[15] Xu X., et al.: "Efficient Reinforcement Learning Using Recursive Least-Squares Methods." Journal of Artificial Intelligence Research. Vol 16:259–292, 2002