2014 11th IEEE International
Conference on Control & Automation (ICCA)
June 18-20, 2014. Taichung, Taiwan

Freeway Ramp-Metering Control based on Reinforcement Learning

Ahmed Fares§ and Walid Gomaa†
Computer Science and Engineering
Egypt-Japan University for Science and Technology (E-JUST)
(†Currently on leave from the faculty of Engineering, Alexandria University)
Alexandria , Egypt
Email: {§ahmed.fares, †walid.gomaa}@ejust.edu.eg

***Abstract*—Random occurrences of traffic congestion on freeways lead to system degradation over time. If no smart control measures are applied, this degradation can lead to accumulated congestion which can severely affect other parts of the traffic network. Consequently, the need for an optimal and reliable traffic control has become more critical. The aim of this research is to control the amount of vehicles entering the mainstream freeway from the ramp merging area, i.e., balance the demand and the capacity of the freeway . This keeps the freeway density below the critical density. Consequently, this leads to maximum utilization of the freeway without entering in congestion while maintaining the optimal freeway operation. The Reinforcement learning based density control agent (RLCA) is designed based on Markovion modeling with an associated Q-learning algorithm in order to address the stochastic nature of the traffic situation. Extensive analysis is conducted in order to assess the proposed definition of the (state,action) pairs, as well as the reward function. We experiment with two case studies with two different network structures and demands. The first case study, which is the benchmark network used in literature, is the network with dense demand. Whereas the other one is the network with light demand. RLCA shows a superior response with respect to a predetermined reference points especially in terms of freeway density, flow rate, and total travel time.**

***Index Terms*—Ramp metering, Q-learning, traffic control, intelligent control, freeway, agent-based system, intelligent transportation system, sequential decision problem.**

I. Introduction

The huge increase in the number of vehicles all over the world leads to random occurrences of congestion. The sophistication of traffic network demands, as well as their severity, have also increased recently. Consequently the need for an optimal and reliable traffic control, both for urban and freeway networks, has become more and more critical.

Generally, there are four freeway control strategies: on-ramp control, mainline control, dynamic speed limits control , and corridor control. Controlling the number of vehicles entering the freeway from the ramp is called *the rate of ramp metering*, which could be measured by a ramp metering device. This is a device that generates only two signals: red and green (no yellow), controlled by a smart or basic controller that regulates the flow of traffic entering the freeway according to the current traffic volume. By using optimal ramp metering control, the freeway flow can be improved. As a result, the overall total travel time (TTT), fuel consumption, accidents, and air population are all improved in quality.

Challenges of the traffic control problem come from the inherent dynamic nature, nonlinearity, and uncertainty of the traffic network system. Hence, an optimal control sequence over all control decisions points is needed. Congestion on a particular section of the freeway is reached when the demand flow exceeds the capacity of this section. It has been generally agreed that this complex problem can not be represented exactly using a mathematical model. This is mainly due to the tremendous complexity of the problem and the huge amount of parameters that can contribute to the control actions. Several different techniques have been recently applied including: intelligent control theory, fuzzy control [1], [2], neural network control [3], [4], [5], [6], and even some hybrids of them [7]. Although there have been some progress with such line of research, these techniques need some expertise to extract the parameters and rules and huge amount of training data.

Ramp metering algorithms have developed considerably in recent years. They evolved from conventional approaches to intelligent ones based on intelligent control technologies. In [2] an adaptive fuzzy logic technique was proposed to regulate the dynamic speed limits (DSL) in the freeway. In this approach, a genetic-fuzzy logic on-ramp controller was designed based on traffic information such as density, speed, and ramp queue length. In [1] a genetic algorithm is used to calibrate and tune the fuzzy sets parameters in the fuzzy controller. Genetic fuzzy logic control (GFLC) was applied to the Amisun microscopic traffic simulator. As in [2] the genetic algorithm is applied to improve the performance of the fuzzy control system, by comparing the performance of GFLC and the fuzzy logic control (FLC) with respect to TTT. They showed that the GFLC provides an improvement over FLC. In [8] coordination of the ramp metering and the variable speed limits in a freeway traffic network was proposed. The objective of this kind of control is to minimize the TTT vehicles spent in the network. They used model predictive control (MPC) to solve the problem, where the macroscopic traffic simulation METANET is used. The controller receives input from the traffic system that consists of the flow, density, and speed; and its output to the traffic system consists of ramp metering and speed limits. They compare the performance

Authorized licensed use limited to: Universite Gustave Eiffel. Downloaded on November 12,2024 at 14:22:04 UTC from IEEE Xplore. Restrictions apply.

of the system before and after replacing the DSL (dynamic speed limits) by the main-stream metering and showed that the choice between the two approaches should be based on traffic demands. In [6] a nonlinear feedback method was proposed for on-ramp metering by using neuron adaptive control algorithm. They proposed two control approaches: neuron adaptive PID and neuron adaptive PSD, where the control parameters were impeded in the learning rate of the neuron. The parameters of the controller are adjusted by using the weight update rule of the neuron. They compare the performance with ALINE [9] which is a traditional method for feedback ramp metering. They showed that their algorithms have a better response in terms of traffic density and (robustness, instance response and control precision) of the ramp metering. In [3] a PID controller was developed whose parameters were adjusted by using a back-propagation (BP) neural network. The results are applied to freeway on-ramp metering. The controller inputs were the density and the desired density and the controller output is the ramp metering rate.

In this paper, we propose a new model of ramp metering based on modeling by *Markov Decision Process* and an associated Q-learning algorithm. In order to avoid: the computational complexity and the risk of being trap in local optimum of the MCP [10] and the solid knowledge of the system considered to extract the rules in *fuzzy control system*.

Extensive analysis is conducted in order to assess the proposed definition of the (state and action) pairs, as well as the reward function. Assuming dynamic traffic demand, with the aim of satisfying the maximum utilization of the freeway capacity, and with avoiding congestion our approach can satisfy the need of freeway optimal control.

The rest of the paper is organized as follows. Section I is an introduction. Section II describes the freeway traffic flow model and shows the relationship between the flow and the density of the freeway. In section III we describe the reinforcement learning (RL) approach, particularly the Q-learning algorithm. In section IV we model the on-ramp traffic control problem using Markov decision process on which the Q-learning algorithm is to be applied. Additionally, the reinforcement learning based density control agent for ramp metering is designed. Section V presents the results of applying reinforcement learning for ramp metering control using benchmark networks (dense and light) and compares its performance to the no-control case. Finally, conclusions and directions for future work are summarized in Section VI.

II. Freeway Traffic Flow Model

Let N be the total number of vehicles in the freeway area of interest during the interval Δt. Via the conservation principal, the total number of vehicles in the area of interest in the freeway section at time $t+\Delta t$ is $N(t+\Delta t)$ can be calculated using the following equation:

$$N(t+\Delta t) = N(t) + \Delta t \times [\lambda q_u(t) - \lambda q_d(t) + q_r(t)] \quad (1)$$

where (see Figure(1)) λ is the number of lanes in the freeway section. By assumption there are only one lane in the ramp and λ lanes in the freeway section. Assume $\lambda = 3$. $q_u(t)$ is the upstream traffic flow, which is defined as the number of vehicles passing the reference points X_1, X_2, X_3 per unit of time ($vehicles/hour$). $q_d(t)$ is the downstream traffic flow, which is defined as the number of vehicles leaving the reference points X_5, X_6, X_7 per unit of time ($vehicles/hour$). $q_r(t)$ is the ramp inflow which is defined as the number of vehicles passing the reference points X_4 per unit of time ($vehicles/hour$).

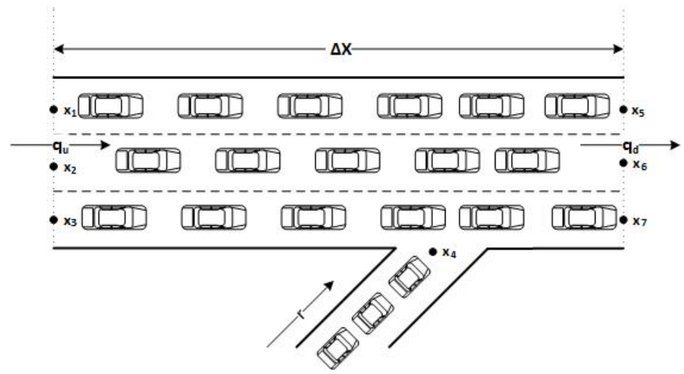

Fig. 1. Uncontrolled isolated on-ramp section

Dividing both sides of equation (1) by $\lambda\Delta x$, where Δx is the length of the area of interest in the freeway section; the result is the following equation which describes the density of the freeway at time $t + \Delta t$.

$$\rho(t+\Delta t) = \rho(t) + \frac{\Delta t}{\Delta x}[q_u(t) - q_d(t) + \frac{q_r(t)}{\lambda}] \quad (2)$$

where $\rho(t) = \frac{N(t)}{\lambda \Delta x}$ is the traffic density on the freeway (which is described as the number of vehicles (N) over the length of the freeway (ΔX) ($vehicles/KM$)) at time t.

The following equation shows the relationship between the outflow and density.

$$q_d(t) = v_{ff}(1 - \frac{\rho(t)}{\rho_{jam}})\rho(t) \quad (3)$$

where V_{ff} is the free flow velocity on the freeway and ρ_{jam} is the jam density, which is the the maximum density achieved under congestion. The above relation is called the *fundamental diagram of the traffic flow*. Figure (2) illustrates the graph of relationship which has the following propensities.

- At the jam density the flow is zero.
- At the critical density the flow is maximum.
- Below the critical density the flow increases with the density and above the critical density the flow decreases with the density.

Equations (1), (2) and (3) complete the description of the freeway model. Equation (3) states that $q_d = f(\rho)$ which can be substituted into (1) and (2). Therefore, the objective of the ramp metering control is to keep the freeway density around the critical density $\rho \in [\rho_c - \varepsilon, \rho_c + \varepsilon]$.

Authorized licensed use limited to: Universite Gustave Eiffel. Downloaded on November 12,2024 at 14:22:04 UTC from IEEE Xplore. Restrictions apply.

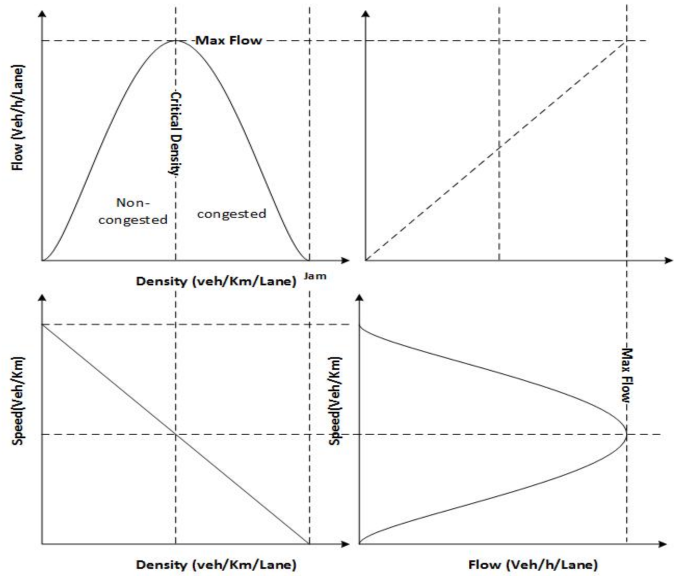

Fig. 2. Fundamental diagram of traffic flow

III. Reinforcement Learning

Reinforcement learning (RL) is a machine learning technique that addresses the question of how an autonomous agent, that perceives its environment through sensors and acts in it through actuators, can learn to choose optimal actions to achieve its goals. This is called *goal directed learning*. RL acts without supervision [11]. The control agent needs to optimize its behavior according to some objective function. This introduces the concept of *reward* received from the environment.

As shown in Figure (3), the agent interacts with its environment which is described by a set of states S.

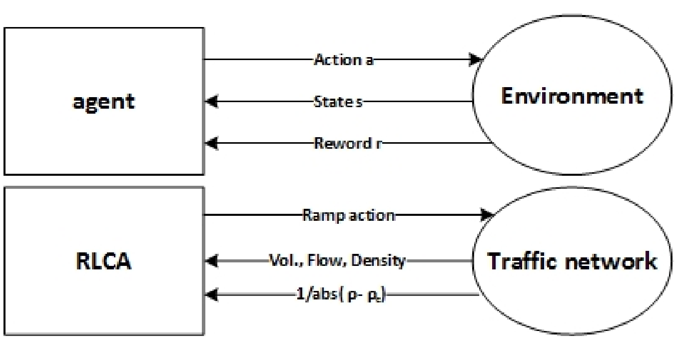

Fig. 3. RL Modeling of ramp metering

The agent can choose any action from a set of actions A to perform. An agent is learning by trial and error interactions with the surrounding environment. Every time the agent chooses an action a_t to be performed in some state s_t, the environment responds by a reward or penalty r_i as an evaluation of the quality of this immediate transition. These interactions lead to a sequence of state-action pairs (s_i, a_i) and immediate rewards r_i [12]. The reinforcement learning-based density control agent (RLCA) is shown in Figure (3) interacting with the traffic network. The agent receives the traffic state from the network detectors, such state consists of the number of vehicles N_t and the current density ρ_t. Based on these information the RLCA chooses an action a_t . The action space will be described in details in section IV-B. As a consequence of a_t the agent receives a reward r_t that keeps the density of the network around a small neighborhood of the critical density. The reward function will be described in details in section IV-C. The aim of the agent is to learn a policy which maps an arbitrary state into an optimal action $\pi : S \longrightarrow A$. An optimal action is an action that optimizes the long-term cumulative discounted reward, thus we optimize over an infinite horizon. Given a policy π, the value function corresponding to π is:

$$V^{\pi}(s_t) = r_t + \gamma r_{t+1} + \gamma^2 r_{t+2} + ... = \sum_{i=1}^{\infty} \gamma^i r_{t+i} \qquad (4)$$

where $\gamma \in (0, 1)$ is the *discount factor* and is necessary for the convergence of the previous formula. A policy π^* is optimal if its corresponding value function is optimal, that is, for any control policy π and any state s we have:

$$V^{\pi^*}(s) \geq V^{\pi}(s) \qquad (5)$$

(Here we assume the objective function is to be maximized.)

A. Q-learning

There are many types of reinforcement learning algorithms that use the same principle. The Q-learning algorithm of Watkins [13], that is assumed in current article, is similar to the dynamical programming paradigm. A Q-learning agent could learn based on experienced action sequence actuated in a Markovian environment. In fact, Q-learning (see Algorithm 1) is a kind of a Markov decision process (MDP). In his proof of convergence Watkins assumed lookup table to represent the value function [13].

Algorithm 1 Q-learning

Initialize the lookup table $\hat{Q}(s, a)$ (for example, by random small values);
for all s ∈ S , a ∈ A **do**
  repeat
    Start in an initial state s;
    repeat
      Choose action a and execute;
      Receive reward r and successor state s';
      $\hat{Q}(s, a) := r(s, a) + \gamma \, max_{a'} \hat{Q}(a', s')$;
      $s := s'$;
    until s is a terminal state or time limit reached
  until $\hat{Q}$ converges
end for

In Q-learning, the agent first observes the current state s_t, and then it chooses an action and executes it. It receives an immediate reward from the environment and observes the next

Authorized licensed use limited to: Universite Gustave Eiffel. Downloaded on November 12,2024 at 14:22:04 UTC from IEEE Xplore. Restrictions apply.

state s_{t+1}. Finally, it updates the Q values based on this last experienced interaction with the environment. This algorithm is guaranteed to converge to the optimal Q-values with a probability of one under some conditions: the Q-function is represented exactly using a lookup table and each state-action pair is visited infinitely often. After convergence to the optimal Q-function Q^*, the optimal control policy can be extracted as follows:

$$\pi^*(s) = \underset{a \in A}{argmax}\, Q^*(s,a) \tag{6}$$

And the optimal value function (the function that gives a valuation of the states; it can be viewed as a projection of the Q-function over the action space):

$$V^*(s) = V^{\pi^*}(s) = \max_a Q^*(s,a) \tag{7}$$

IV. Reinforcement Learning Density Control Agent

In the following subsections we describe all the components of the Markov decision process that corresponds the on-ramp metering control.

A. State space

The state of the freeway network mainly depends on traffic measurements and demands at a given time. The state space is three-dimensional where each state vector s_{t+1} at time $t+1$ consists of the following components:

1) The number of vehicles in the mainstream $N(t+1)$ (see Equation (1)).
2) The number of vehicles that entered the freeway from ramp $\Delta N(t+1)$ during the last time step.
3) The ramp traffic signal at the previous time step $Ts(t)$.

B. Action space

In order to *optimize the density of the freeway* the RLCA agent should select the optimal actions sequence based on the Q-learning algorithm. By placing the ramp metering on the on-ramp the red and green phases of the ramp metering change in order to control the flow entering the freeway from the merging point. So the action space is modeled as consisting only of two actions: red and green.

The optimal action is then chosen based on Equation (6).

C. Reward function

The reward function may either reward or punish the agent's last action. Since the RLCA agent's goal is to keep the freeway density ρ around a small neighborhood of the critical density ρ_{cr}, the reward function is designed so as to depend on the current freeway density ρ and how much it deviates from the critical density ρ_{cr}. Hence, the reward of taking an action a in state s is designed to be:

$$r(s,a) = \frac{1}{|\rho - \rho_{cr}|} \tag{8}$$

As long as the Q value (as a function of current state and action) is maximized, the difference between the freeway density ρ and the critical density ρ_{cr} is minimized. Therefore, the best utilization of the freeway is achieved without entering in congestion.

V. Experiments: Results and Analysis

We will demonstrate two case studies with two different network structures and demands. The first case is a dense network, which shows the need for smart control that regulates the flow from on-ramp in order to keep the density of the freeway close to the critical value. The second case study is a light network that needs no control to prove the concept that our RLCA always does the optimal action.

A. First case study

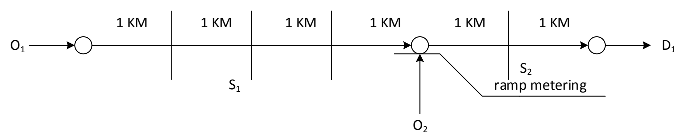

Fig. 4. The benchmark network

The benchmark network for the first case study is shown in Figure (4). This network consists of a mainstream freeway with two lanes and metered on-ramp with one lane. The network consists of two sources of inflow, the first one is O_1 for the mainstream flow, and the second source is O_2 for the on-ramp flow; it has only one discharge point D_1 with unrestricted outflow. The mainstream freeway consists of two sections S_1 and S_2. The first section S_1 lies before the ramp, it is 4 km long and consists of 4 segments each is 1 km long. The second section S_2 lies after the ramp, it is 2 km long and consists of 2 segments each is 1 km long. The mainstream freeway capacity is 4000 veh/h for both lanes, and the on-ramp capacity is 2000 veh/h. this The benchmark network structure can be found in [8], [2].

The network parameters are taken as follows: $V_{ff} = 60$ km^2/h, $\rho(0) = 40$ $veh/km/lane$, $\rho_{jam} = 180$ $veh/km/lane$, $\rho_{cr} = 37.5$ $veh/km/lane$, and $\Delta X = 6$ km. These network parameters can be found in [14].

As shown in Figure 5, the demand scenario follows the following distribution: the mainstream demand remains constant for 2 hours at high level (3500 veh/h) near the capacity and finally drops to low level (1000 veh/h) in the last 15 minuets. The on-ramp demand begins at low level (500 veh/h), in 7.5 minuets increases to high level (1500 veh/h) near the capacity, remains constant for 15 minuets and finally drops to low level (500 veh/h) in 7.5 minuets and remains constant for 2 hours. This demand scenario gives us the opportunity to study the effect of the on-ramp smart control. That is because in the first half hour the total demand exceeds the capacity, as shown in Figure 7, and we enter the congestion area in the fundamental diagram in the 'no control' case, therefor the TTT increases. As this demand scenario is dense, the RLCA can prevent a traffic breakdown by adjusting the metering rate such that the density remains near the critical value as shown in Figure 6.

Authorized licensed use limited to: Universite Gustave Eiffel. Downloaded on November 12,2024 at 14:22:04 UTC from IEEE Xplore. Restrictions apply.

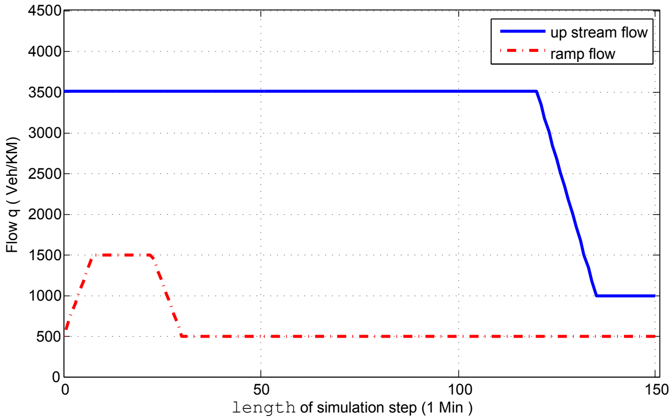

Fig. 5. The demand scenario used in the first simulation

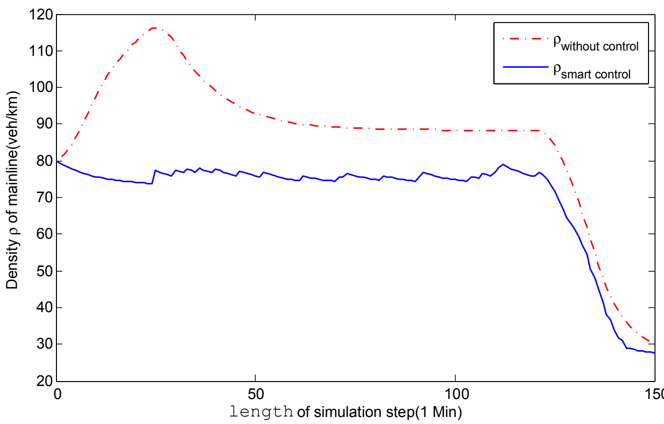

Fig. 6. Density of main-stream

The RLCA objective function is to optimize the density ρ of the mainstream int order to maintain the optimal freeway operation. As shown in Figure 6 in the 'no control' case the density ρ exceeds the critical value ρ_{cr}, but in the smart control case the densityρ is optimal over all simulated period. Figure 8 give us the optimal action sequence which keeps flow q closed to the capacity q_{max}. Thats because it keeps the density ρ of the mainstream relatively equal to the critical density ρ_{cr}.

B. Second case study

Figure 9 shows the second case study network structure. This network consists of mainstream freeway of 500 m long and one metered on-ramp. The network has two sources of inflow O_1 and O_2 and one source of outflow D_1. O_1 is for the mainstream flow, O_2 is for the on-ramp flow. The freeway has 3 lanes and the ramp has only one lane.

The network parameters are as follows: $V_{ff} = 80\ km^2/h$, $\rho(0) = 30\ veh/km/lane$, $\rho_{jam} = 110\ veh/km/lane$, $\rho_{cr} = 55\ veh/km/lane$, and $\Delta X = 500\ m$. These network parameters can be found in [6]

The demand scenario follows the following Gaussian function with center $c = 250$ and spread $\sigma = 275$:

$$f(t; \sigma, c) = e^{\frac{-(t-c)^2}{2\sigma^2}} \tag{9}$$

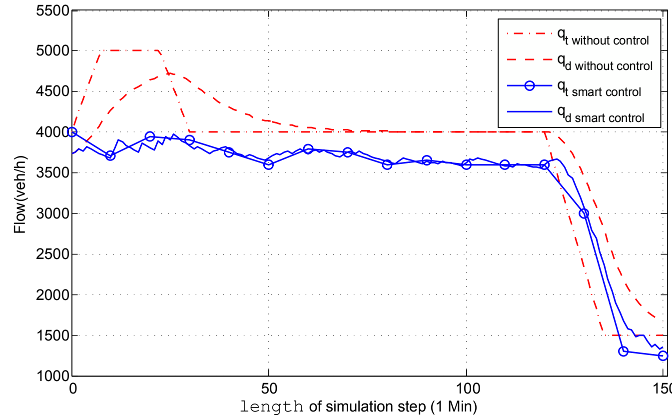

Fig. 7. Flow of mainstream

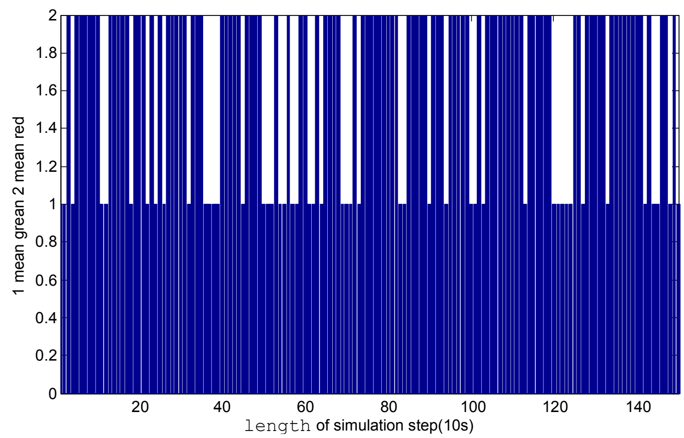

Fig. 8. Ramp metering Optimal action sequence

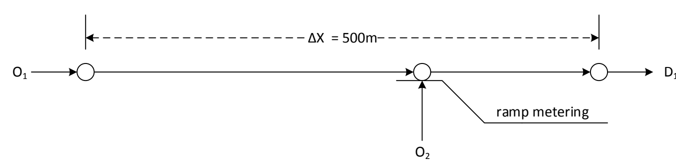

Fig. 9. The network of the second case study

For the mainstream demand we multiply each value by 1800, and for the on ramp demand we multiply each value by 1000, in order to reach the peek values of 1800 and 1000 for the mainstream and on-ramp demands respectively; see Figure 10.

This demand scenario gives us the opportunity to test our RLCA against a light network with density ρ always below the critical density ρ_{cr} as shown in Figure 11.

Since the demand flow always gives us a density ρ below the critical density ρ_{cr}, the optimal control sequence of actions of the RLCA are always green.

Figure 12 shows the optimal action sequence. In this case the demand flow does not exceed the capacity, which allows us to prove the concept that the RLCA always does the right action. That's is green over all decision time.

Authorized licensed use limited to: Universite Gustave Eiffel. Downloaded on November 12,2024 at 14:22:04 UTC from IEEE Xplore. Restrictions apply.

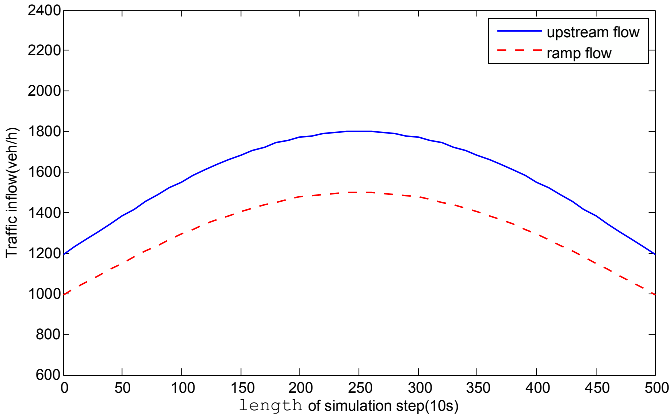

Fig. 10. The demand scenario used in the simulation

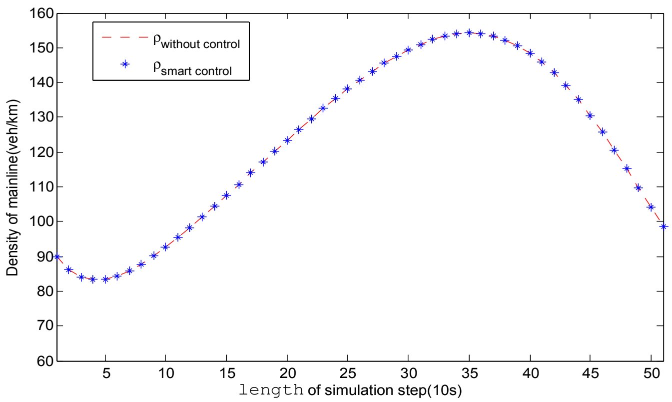

Fig. 11. Density of main-stream

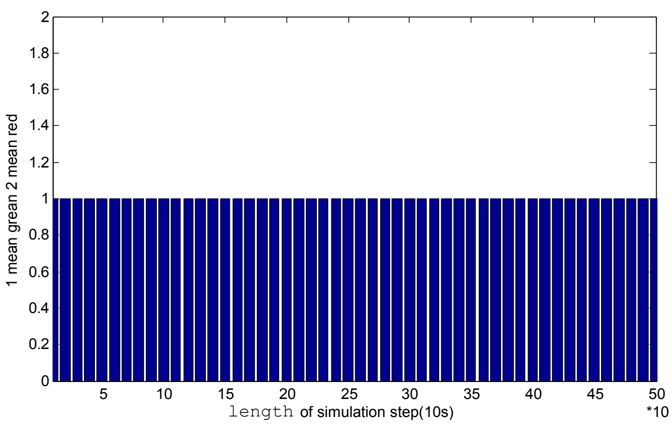

Fig. 12. Ramp metering Optimal action sequence

VI. Conclusion

In this paper, a reinforcement learning based density control agent (RLCA) is proposed. The RCLA objective function is to optimize the density of the freeway mainstream in order to maximize the flow and minimize the total travel time (TTT). RLCA is tested against two different case studies with two different network architectures and demands.

In the first case study - the dense network: the proposed RLCA guarantees a good performance in terms of keeping the flow close to the freeway capacity. That is because it keeps the density of the mainstream relatively equal to the critical density. The control technique we use is a simple version of Q-learning. The optimal action sequence is obtained and tested against two reference points: freeway capacity and critical density.

In the second case study - the light network: the optimal control sequence of actions of the RLCA are always green. This case study illustrates that our RLCA always does the optimal action.

In the future we plan to extend this work to more complicated networks. Such networks might contain several ramps, dynamic speed limits, adaption to changing road conditions such as sudden occurrence of accidents and weather hazards.

Acknowledgment

This research is supported by the Ministry of Higher Education (MoHE) of Egypt through PhD fellowships. Our sincere thanks to E-JUST University for guidance and support.

References

[1] X. Yu, W. Xu, F. Alam, J. Potgieter, and C. Fang, “Genetic fuzzy logic approach to local ramp metering control using microscopic traffic simulation,” in *Mechatronics and Machine Vision in Practice (M2VIP), 2012 19th International Conference*, 2012, pp. 290–297.

[2] A. Ghods, A. Kian, and M. Tabibi, “Adaptive freeway ramp metering and variable speed limit control: a genetic-fuzzy approach,” *Intelligent Transportation Systems Magazine, IEEE*, vol. 1, no. 1, pp. 27–36, 2009.

[3] X. Liang and Y. Fan, “Freeway ramp pid controller regulated by bp neural network,” in *Intelligent Vehicles Symposium, 2009 IEEE*, 2009, pp. 713–717.

[4] X. Liang, J. Li, and N. Luo, “Single neuron based freeway traffic density control via ramp metering,” in *Information Engineering and Computer Science (ICIECS), 2010 2nd International Conference on*, 2010, pp. 1–4.

[5] J. Li and X. Liang, “Freeway ramp control based on single neuron,” in *Intelligent Computing and Intelligent Systems, 2009. ICIS 2009. IEEE International Conference on*, vol. 2, 2009, pp. 122–125.

[6] C. Qi, Z. Hou, and X. Li, “Freeway feedback ramp metering based on neuron adaptive control algorithm,” in *Intelligent Computation Technology and Automation (ICICTA), 2008 International Conference on*, vol. 1, 2008, pp. 349–353.

[7] C. Feng, J. Yuanhua, L. Jian, Y. Huixin, and N. Zhonghai, “Design of fuzzy neural network control method for ramp metering,” in *Measuring Technology and Mechatronics Automation (ICMTMA), 2011 Third International Conference on*, vol. 1, 2011, pp. 966–969.

[8] A. Hegyi, B. D. Schutter, and H. Hellendoorn, “Model predictive control for optimal coordination of ramp metering and variable speed limits,” *Transportation Research Part C: Emerging Technologies*, vol. 13, no. 3, pp. 185 – 209, 2005. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0968090X05000264

[9] M. Papageorgiou, H. Hadj-Salem, and J.-M. Blosseville, “Alinea: A local feedback control law for on-ramp metering,” *Transportation Research Record*, no. 1320, 1991.

[10] D. Ernst, M. Glavic, F. Capitanescu, and L. Wehenkel, “Reinforcement learning versus model predictive control: a comparison on a power system problem,” *Systems, Man, and Cybernetics, Part B: Cybernetics, IEEE Transactions on*, vol. 39, no. 2, pp. 517–529, 2009.

[11] T. M. Mitchell, *Machine Learning*, 1st ed. New York, NY, USA: McGraw-Hill, Inc., 1997.

[12] R. S. Sutton and A. G. Barto, *Reinforcement learning: An introduction*. Cambridge Univ Press, 1998, vol. 1, no. 1.

[13] C. J. Watkins and P. Dayan, “Q-learning,” *Machine learning*, vol. 8, no. 3-4, pp. 279–292, 1992.

[14] A. Kotsialos, M. Papageorgiou, and A. Messmer, “Optimal coordinated and integrated motorway network traffic control,” in *14th International Symposium on Transportation and Traffic Theory*, 1999.

Authorized licensed use limited to: Universite Gustave Eiffel. Downloaded on November 12,2024 at 14:22:04 UTC from IEEE Xplore. Restrictions apply.