공학석사 학위논문

Developing Variable Speed Limit Control and Ramp Metering Strategy for Freeways Using Deep Reinforcement Learning

강화학습을 활용한 고속도로 가변제한속도 및 램프미터링 전략 개발

2022년 2월

서울대학교 대학원
공과대학 건설환경공학부

조 정 훈

Developing Variable Speed Limit Control and Ramp Metering Strategy for Freeways Using Deep Reinforcement Learning

지도 교수 김 동 규

이 논문을 공학석사 학위논문으로 제출함

2021년 12월

서울대학교 대학원
공과대학 건설환경공학부

조 정 훈

조정훈의 공학석사 학위논문을 인준함

2022년 2월

위 원 장 고 승 영 (인)

부위원장 이 청 원 (인)

위 원 김 동 규 (인)

Abstract

Recently, to resolve societal problems caused by traffic congestion, traffic control strategies have been developed to operate freeways efficiently. The representative strategies to effectively manage freeway flow are variable speed limit (VSL) control and the coordinated ramp metering (RM) strategy. This paper aims to develop a dynamic VSL and RM control algorithm to obtain efficient traffic flow on freeways using deep reinforcement learning (DRL). The traffic control strategies applying the deep deterministic policy gradient (DDPG) algorithm are tested through traffic simulation in the freeway section with multiple VSL and RM controls. The results show that implementing the strategy alleviates the congestion in the on-ramp section and shifts to the overall sections. For most cases, the VSL or RM strategy improves the overall flow rates by reducing the density and improving the average speed of the vehicles. However, VSL or RM control may not be appropriate, particularly at the high level of traffic flow. It is required to introduce the selective application of the integrated control strategies according to the level of traffic flow. It is found that the integrated strategy can be used when including the relationship between each state detector in multiple VSL sections and lanes by applying the adjacency matrix in the neural network layer. The result of this study implies the effectiveness of DRL-based VSL and the RM strategy and the importance of the spatial correlation between the state detectors.

Keyword : Traffic Control Strategy, Variable Speed Limit Control, Ramp Metering, Deep Reinforcement Learning

Student Number : 2020-23170

Table of Contents

List of Tables

List of Figures

Chapter 1. Introduction

The recent trends of the increasing density of the urban population and the increasing number of people living in metropolitan areas have led researchers to explore how to use existing roads more efficiently rather than constructing new roads since the available land is limited. In addition, the recent dramatic growth of autonomous vehicles and the advent of smart infrastructure diversify the types of vehicles on the road and they also change the road's infrastructure to a different form. In addition, autonomous driving keeps transforming the road infrastructure that we have maintained for such a long time by proposing an intelligent infrastructure.

The increase in the density of the population results in traffic congestion problems in urban areas, and the strategies that have been mainly developed to eliminate the shock waves on highways. Variable speed limit (VSL) control and ramp metering (RM) commonly have been applied to ensure efficient traffic flow on highways. However, with the recent introduction of autonomous driving, the control of individual vehicles has become possible, and related research is being actively conducted. VSL control is the method of controlling the flow of traffic that uses variable speed limits on each lane according to real-time traffic and weather conditions. RM refers to the traffic signals installed on the ramps that allow access to the freeway, and its purpose is to control the rate at which vehicles enter the congested traffics. RM is known to work well in breaking up the traffic flow that makes it difficult to merge onto the mainstream.

Machine learning techniques, including reinforcement learning

(RL), have been driving this development. Due to the stochastic and continuous nature of traffic flows, reinforcement learning can improve the traffic flow continually (Kušić et al., 2020). The RL-based VSL approach has been applied successfully, and it has been proven to be effective (Li et al., 2020; Schmidt-Dumont and Van Vuuren, 2015; Wang et al., 2019). In addition, it is more efficient than convolutional neural network or a fully connected network if we look at the effects of several ramps, detectors, and VSL sections on the freeway and reflect the spatial correlation between them (Lv et al., 2018; Wu et al., 2019). It is an advantage of the graphical structure of a graph neural network, and it is known that spatial correlation is well reflected in both traffic flow and in discrete demand prediction (Cho et al., 2021; Kim et al., 2019).

Also, the traffic simulation tools that imitate varying scenarios and cases on the road have enabled the iterative learning process for efficient control algorithms. SUMO, an open-source traffic simulator, is used extensively in proving the effectiveness of VSL or RM control (Wu et al., 2020, 2019) because it addresses the complex traffic dynamics, such as traffic jams and changing lanes (Kheterpal et al., 2018; Wu et al., 2017). It also is used to investigate the microscopic situations in signal traffic control (Akyol et al., 2019). Wu et al. (2020) proposed a novel actor-critic architecture to learn discrete differential variable speed limits on freeway bottlenecks for different reward signals (Wu et al., 2020). However, there is an increasing concern about the situation when there are multiple ramps and VSL controls in highways and the relationship between the several sections and VSL controls in each lane. It is also demanding to deal with how VSL controls react well in cooperative control with ramp metering.

The aim of this study was to develop a dynamic variable speed limit control and ramp metering control algorithm by applying reinforcement learning algorithms in order to eliminate shock waves on freeways with multiple VSL and ramps, thereby allowing efficient traffic flow and road safety, and to determine whether GCN layers with an adjacency matrix of geographical distance help detect a spatial correlation of multiple VSL detectors.

The remainder of this paper is organized as follows. A literature review on related works is provided in the following section. The next section discusses the methods that were used in this study, and the detailed results and discussion of these methods are presented. The paper concludes by discussing potential future work.

Chapter 2. Literature Review

Basically, there are two types of traffic jams, i.e., 1) traffic jams with the head fixed at the bottleneck and 2) traffic jams with the head and tail moving opposite to the direction of travel, which is easier to handle than the former (Hegyi et al., 2008). VSL can solve the traffic jam problems by using dynamic speed limits. VSL control mainly aims at two different objectives, i.e., 1) homogenizing traffic to reduce the differences in the speeds of the vehicles (Kuhne, 1991; Smulders, 1990) and 2) reducing the flow to prevent breakdown or jams (Hegyi et al., 2008, 2005; Popov et al., 2008; Zhang et al., 2005). VSL also has benefits in that it improves the safety of the road and has its environmental benefits derived from combustion (Abdel-Aty et al., 2006; Khondaker and Kattan, 2015; Wang et al., 2019; Zegeye et al., 2010). Furthermore, VSL helps solve the shockwave problem on the freeway (Hegyi et al., 2008), and it is known that it can reduce the total travel time by about 20 – 30% (Carlson et al., 2010; Hegyi et al., 2005; Wang et al., 2019).

Recently, the accelerated development of computing algorithms and resources has enabled the application of various machine learning techniques for controlling traffic. The continuous learning Q-learning-based VSL control outperformed the feedback-based VSL strategy (Li et al., 2017), and other advanced and efficient algorithms have been proposed with deep Q networks and actor-critic algorithms (Wu et al., 2020). Also, there have been various approaches to set the reward function, one of the most critical parts of the RL algorithm. Wang proposed cooperative VSL to maximize the

two different reward controls, i.e., freeway traffic mobility and safety benefits (Wang et al., 2019). Wu has set its objectives in total travel time, bottleneck speed, emergency braking, and vehicular emissions (Wu et al., 2020). However, VSL control is reported to be unsuccessful in homogenization and to reduce lane-changing activity on the low-speed segments of the highway (Soriguera et al., 2017).

Another method to control traffic flow effectively is the ramp-metering strategy. RM controls entering the mainstream freeway from the on-ramp. It is beneficial in balancing the demand and the capacity of the traffic flow. Despite the queues that are generated on the ramp, the greater throughput attained by avoiding congestion outweighs the delays on the ramp queue. RM can lead to a dramatic enhancement of freeway traffic by maintaining capacity flow on the mainstream and increasing the outflow, thus decreasing the total time spent (TTS) in the system (Papageorgiou and Kotsialos, 2002). A well-known feedback-based strategy, ALINEA, is used often in local-level traffic. HEuristic Ramp-metering coOrdination (HERO), developed by Papamichail and Papageorgiou, outperforms the more simple ALINEA model (Papamichail et al., 2008). RM also utilizes reinforcement learning algorithms to the freeway ramp metering to control the density below the critical density, thereby preventing the breakdown of the system (Fares and Gomaa, 2014). In this paper, we used Markovian modeling with an associated Q-learning. It also is important to consider the equity and balance of traffic efficiency between vehicles in establishing a coordinated RM strategy. RL-based coordinated ramp metering systems considering equity were applied, and they performed to reduce TTS and to maintain equally distributed total waiting time with a low standard deviation (Lu et al., 2017).

To integrate the advantages of both strategies, i.e., VSL and RM control, cooperative freeway traffic control is proposed. Second order model for freeway control optimization of integrated variable speed limit control and ramp metering model was developed with extended Kalman filter for state estimation (Alessandri et al., 1997). Hegyi reported that the choice between the speed limits and the mainstream metering should be made considering the mainstream demand and the on-ramp demand by applying predictive model control (Hegyi et al., 2005). This study reported that VSL can prevent traffic breakdown and maintain a higher outflow, which significantly reduce congestion and lower total time spent, while RM is unable to prevent congestion. Primary effect of VSL is only limited to mainstream flow, and the range of VSL and RM control is completely different. Traffic flow efficiency is reported to be improved substantially in the integration of VSL and coordinated RM (Papamichail et al., 2008). ALINEA and genetic-fuzzy ramp metering and variable speed limits showed superior performance in controlling the congestion on the freeway while keeping the computational simplicity (Ghods et al., 2009). The integrated control in the bottleneck or virtual lane reduction and determination of RM and VSL both contribute to improvements in the throughput (Lu et al., 2011). Lu also reported that RM only control could not significantly improve the bottleneck situation, while VSL and integrated VSL and RM control significantly improve the efficacy. When multiple strategies are applied simultaneously, coordinating between the strategies also is important. For example, the integrated control must apply knowledge-based learning and access the data provided by multiple VSL sections or state loop detectors. To accomplish this, a method is required that considers the spatial relationship of the detectors.

The basics of traditional RL frameworks are for the agent to interact with the environment to increase estimated values through actions. In addition, with the development of deep learning, deep reinforcement learning (DRL) has developed, and the deep neural network is embedded in the RL algorithm. For instance, deep Q networks (DQN) are based on Q learning (Mnih et al., 2015). DRL has been applied and advanced successfully in various fields, such as playing online games, robotics, and autonomous driving (Gu et al., 2017; Sallab et al., 2017; Silver et al., 2016). Policy gradient methods include A3C (Mnih et al., 2016) and DDPG (Lillicrap et al., 2015).

Kipf first proposed the graph convolution network (GCN) (Kipf and Welling, 2017), and several disciplines, including smart mobility systems, have applied GCN (Cho et al., 2021; Kim et al., 2019; Yu et al., 2020). Knowledge sharing graph convolutional networks (KS-GCN) use knowledge GCN to capture the spatial dependency between different sensors fully and learn the meaningful spatial pattern from traffic states (Wu et al., 2019). Hence, there is in need to develop the traffic control strategy for efficient traffic flow by applying the DDPG algorithm and a GCN that can reflect spatial correlation.

Chapter 3. Methods

3.1. Study Area and the Collection of Data

The scope of this study is a freeway in the United States, on ramps appear repeatedly, and traffic congestion occurs often in the ramp section. The selected interstate freeway, I-880, also called the Nimitz Freeway, is located in the San Francisco bay area of California in the United States, and it serves as a connection to the auxiliary interstate in the north to south direction. The selected region is the southbound section of interstate highway 880 in Oakland, California. It is a section that is 2,656.90 m downstream from where it meets the I-980 freeway in Old City, shown in **Figure 3.1**. There are three on-ramps in between, of which ramp metering control currently is in operation at the on-ramp near Oak Street.

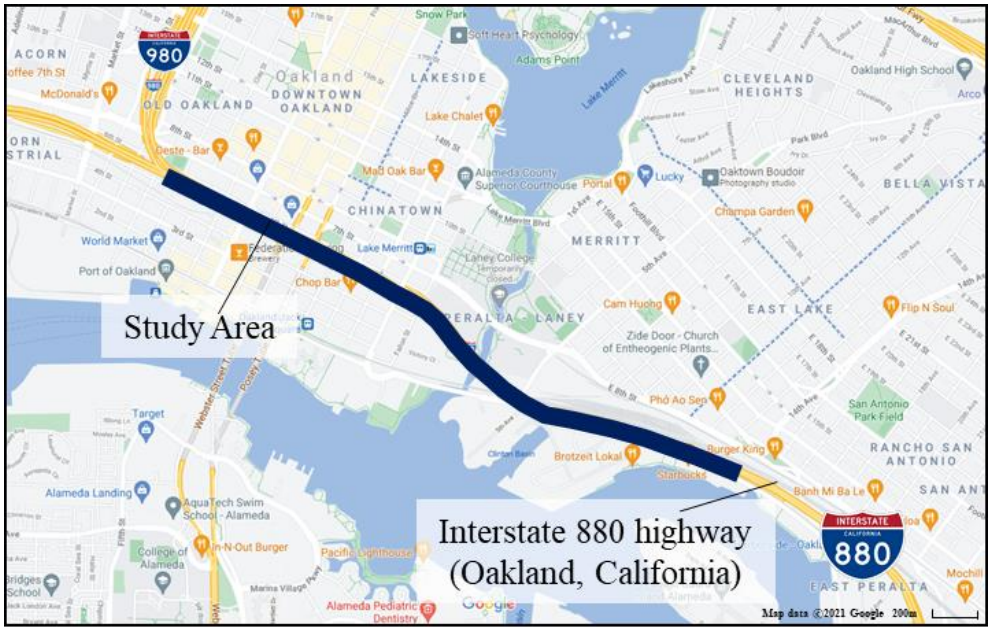


Figure 3.1 Study area of the I-880 S freeway

We utilized the loop detector data provided by PeMS in the California Department of Transportation (Caltrans, 2021). The data that we used were acquired from between 7:00 and 11:00 AM on Wednesday, September 18, 2019, before the COVID-19 pandemic. The selected time range includes the morning peak-hour commuting vehicles, and the section where RM is now operating was selected. In this study, VSL control was designed to occur in two areas in the corresponding section, and ramp metering control also was implemented (Figure 3.2). In addition, loop detectors that can detect the traffic state were installed before and after each control section.

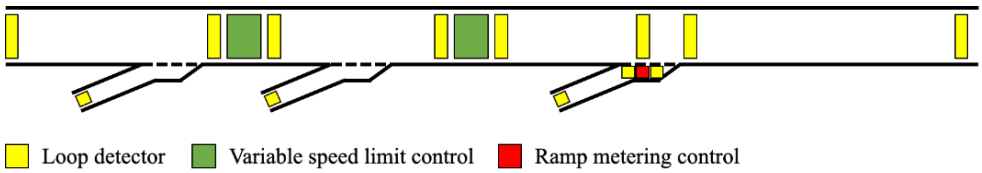
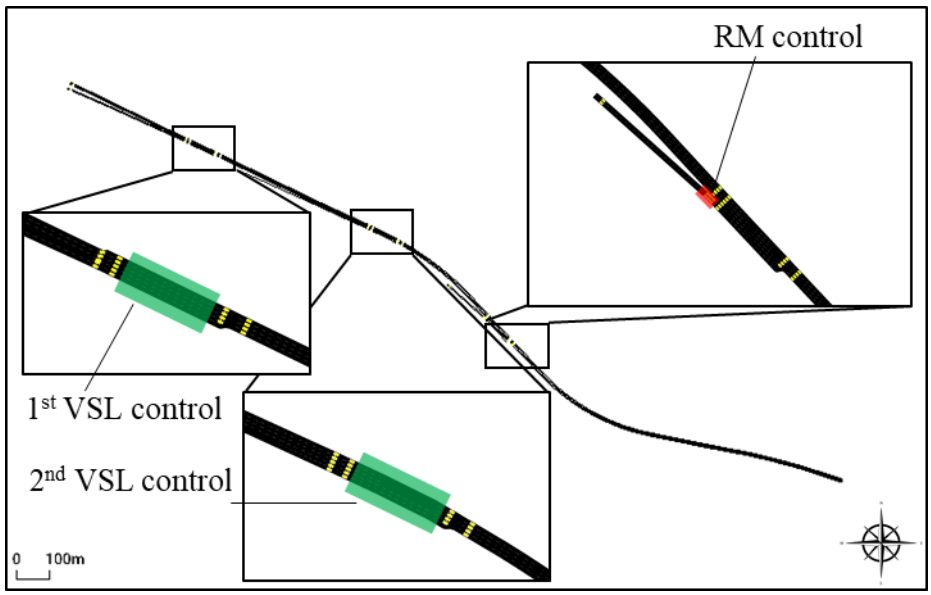


Figure 3.2 Location of the loop detectors, the variable speed limit control, and the ramp metering control

3.2. Simulation Framework

In order to analyze the effects of the strategies on traffic flow, it was very important to simulate the freeway situation properly. In this study, the freeway section was constructed as a simulation road network. The road network was imported into the SUMO network using the osmWebWizard function, which helps export the road network provided by OpenStreetMap, which is available at openstreetmap.org (OpenStreetMap, n.d.). As shown in Figure 3.3, we leave only necessary road information and exclude the roads with few relationships, i.e., the imported road network and loop detectors, to get information from the simulation. The information on the vehicle running on the road also must be determined. In this study, we selected the open-source software, SUMO, and the Traffic Control Interface (TraCI) packages to develop speed limits for each lane. The default "LC 2013" model in the simulation network was applied for the lane-changing model. The vehicles were expected to reflect the car-following model of IDM. The type of vehicle assumes that automobiles and trucks constituted 85% and 15%, respectively. Simulation runs were conducted for 4 hours, and the strategies of VSL were adjusted by the algorithm every minute, and those of RM were changed every second. If the simulation is set to make an accident, the moment and the length of the accident are generated randomly within the simulation horizon, i.e., in the range of at least 1 second to 10 minutes.



Notes. VSL = Variable speed limit control; RM = Ramp metering control

Figure 3.3 Simulated network of the study area of the I-880 S freeway

3.3. Trip Generation and Route Choice

The simulation environment utilizes real-world traffic data from the loop detector in the PeMS system (Caltrans, 2021). The trip distribution was performed on the assumption that the volume of traffic in the relevant section is distributed in proportion to the amount of incoming traffic, considering the volume of traffic of the inflow and outflow sections. After calculating the route of the vehicle in the section, the demand of the route was estimated based on this assumption. Since the PeMS data that were collected were aggregated in units of 5 minutes, it is assumed that the vehicle departs randomly at a Poisson rate at 5-minute intervals.

3.4. Deep Deterministic Policy Gradient (DDPG) Algorithm

The actor−critic framework consists of the actor that generates actions that can increase the value function and the critic network that evaluates the actor's policy. The DDPG algorithm was developed to find a parameterized policy, π_θ, that maximizes its expected reward return over the controlling period (Lillicrap et al., 2015). It utilizes the stochastic behavior policy for exploration, and its exploration is independent of the learning algorithm since it is an off−policy algorithm. The final estimated deterministic target policy, $\pi_\theta(s_t)$, for state s_t derives a deterministic action, a_t, for discrete time step t.

The critic function is used to evaluate the policy, where the estimated value function is parameterized by the parameter of the critic, θ^Q (Equation 3.1).

$$\pi_\theta(s_t) = \underset{a_t}{\operatorname{argmax}}\, Q(s_t, a_t|\theta^Q) \tag{3.1}$$

As stated earlier, the model is designed to maximize the actual value function Q, which is the function of reward selection (Equation 3.2).

$$Q(s_t, a_t) = E\left[\sum_{k=0}^{\infty} \gamma^k r_{t+k}(s_t, a_t)\right] = r_t + \gamma^k Q(s_{t+1}, a_t) \tag{3.2}$$

where r_t is the reward function, and γ is the discount factor.

The sum of the immediate reward and the outputs of the target actor and critic network, y_i, and the loss function for the critic model, $L(Q, \theta^Q)$, are stated as Equations 3.3 and 3.4, respectively.

$$y_i = r(s_i, a_i) + Q\left(s_{i+1}, \pi_{\hat{\theta}^\pi}(s_{i+1}) | \hat{\theta}^Q\right) \quad (3.3)$$

$$L(Q, \theta^Q) = \frac{1}{N}\sum_i \left(y_i - Q(s_i, a_i | \theta^Q)\right)^2 \quad (3.4)$$

where θ^π denotes the parameter of the actor, $\hat{\theta}^\pi$ and $\hat{\theta}^Q$ are the weights of the target actor and the critic networks, respectively, and N is the number of samples.

The loss function for the actor model, $L(\pi, \theta^\pi)$, and the gradient used to update the weight θ^π are stated as Equations 3.5 and 3.6, respectively.

$$L(\pi, \theta^\pi) = -\frac{1}{N}\sum_i Q\left(s_i, \pi_{\theta^\pi}(s_i)\right) \quad (3.5)$$

$$\nabla_{\theta^\pi} = \frac{1}{N}\sum_i \nabla_a Q(s_i, a_i | \theta^Q)|_{a=\pi_{\theta^\pi(s_i)}} \nabla_{\theta^\pi} \pi_{\theta^\pi(s_i)} \quad (3.6)$$

Replay buffer R, a finite-sized cache, is used to store transitions sampled from the environment in exploration policy and the tuple (s_i, a_i, r_i, s_{i+1}) (Lillicrap et al., 2015). The probability p_t of transition is $\frac{1}{\text{rank}(t)}$, where $\text{rank}(i)$ is the rank of transition i, sorted accordingly to the absolute value of the TD error signal.

The detailed process of the DDPG algorithm starts with randomly initializing the actor-network, Q_{θ^Q}, and the critic network, π_{θ^π}, with parameters θ^Q and θ^π, respectively. The weights of the target networks Q' and π' also are initialized with $\hat{\theta}^Q$ and $\hat{\theta}^\pi$, respectively. For each episode, the model initializes the random

process for action exploration and receives initial observation of the state s_1 from the simulation environment. The noise x for action exploration is modeled as the Laplacian process, i.e., $\mathrm{L}(x|b)\sim\frac{1}{2b_t}\exp\left(-\frac{x}{b_t}\right)$, where b_t decays with the learning rate in every time step. For the length of the time period, the algorithm selects action according to the current policy and the exploration ($a_t = int(f_{\hat{\theta}^\pi}(s_t)+x_t)$). The reward r_t and the next state s_{t+1} are observed from the environment, and the tuple (s_t, a_t, r_t, s_{t+1}) is stored in replay buffer R. Using the transition probability, p_t, the minibatch is sampled randomly. The model updates the critic to minimize the critic loss and also updates the actor by the sampled gradient. The TD error computed by the critic updates the transition probability, and the target network of the actor and critic is updated accordingly.

3.5. Graph Convolution Network (GCN) Layer

The related work applied the knowledge to the model because of the necessity to consider the spatial correlation of loop detectors and ramps (Wu et al., 2019). Furthermore, GCN considers relationships among nodes represented in a graphical structure. Based on this, a graph neural network layer was applied with an adjacency matrix of the geographical distance between state detectors of each lane in the control section (Figure 3.4). This allows the model to learn functions that map graph signals asynchronously to traffic control signals.

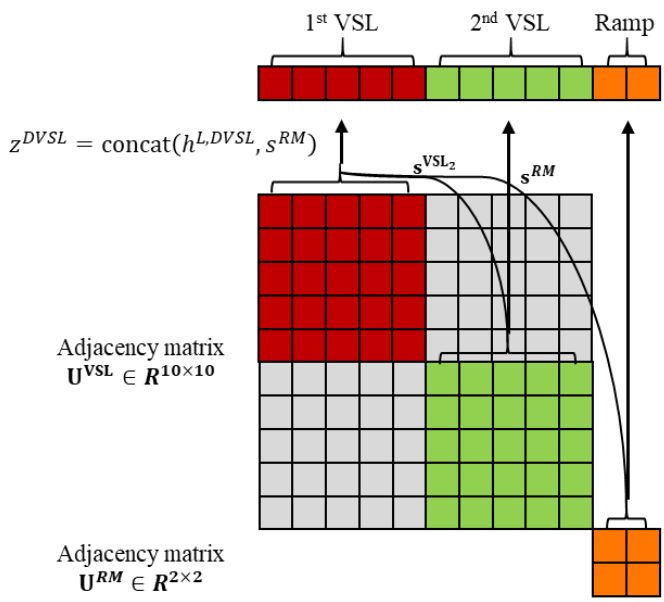


Figure 3.4 Description of the concept on the application of graph convolution network layer

3.6. RL Formulation

Loop detectors are installed in every section of the freeway, and they acquire various information, such as the number of vehicles that pass the section, their occupancy rates, and their average speed. Fifteen detectors were installed, and they can measure the state before and after the VSL and RM sections assumed in this study. There are three detectors upstream of the first VSL section, four between the two VSL sections, four downstream of the second VSL section, and four downstream of the RM sections. The model gets information every second from this state detector. Accordingly, the number of dimensions of the entire state is 15.

Actions that the model can take are divided into VSL and RM controls. As for the VSL, since the continuous speed limit is not that useful in actual operation, the results of the action must be converted to a deterministic speed limit. There are ten lanes in the VSL section, and each of the actions includes the determination of the speed limits every 10 mph in the range of 40 to 80 mph. Each lane is conferred the variable speed limit every minute. It was assumed that the most appropriate action was selected from the results of the model calculation, and the speed limit was changed every minute considering the driver's perception time and operation and the control situation. Ramp metering gave the green-light phase when the value between 0 and 1, calculated as a result of the model, was close to 1, and a red-light phase occurred when the value was close to 0.

The agent was considered as a single agent, i.e., a one road operator. However, now that the connected vehicle is not fully implemented, it is assumed that there is only one road operator who decides to use the VSL strategy and RM on the road.

Reinforcement learning defines a function called reward and uses a strategy to improve and maximize this reward. Therefore, it is very important to set the reward in the RL environment properly. In Wu 2020, several reward functions were reviewed, and each had different traffic engineering meanings, such as safety, efficiency, and eco-friendliness (Wu et al., 2020). In this paper, we set the total outflow and total average speed as reward functions and examined the effect between them.

Chapter 4. Results

4.1. VSL and RM

To compare the effectiveness of each strategy, we examined the uncontrolled situation with no action, the VSL only application, the RM application only, and the VSL and coordinated RM applications. Here, the DDPG algorithm was applied to all strategies, and a situation with two rewards was tested. The learning was made up of 200 episodes, and the learning rate of actor 0.0002, the learning rate of the critic 0.0005, gamma 0.999, and tau 0.005 were used as the hyperparameters for the learning process. During the learning process, it was found that there was a slight change in the fundamental of traffic flow through the episodes. Overall, the average speed showed a tendency to increase, and the density showed a tendency to decrease slightly. The fast learning rate may have led to the tendency to find solutions quickly at the beginning of the episode.

There are several performance indicators to compare the efficacy of each strategies: traffic flow rate at the state detector, average vehicle speed on the freeway section, throughput of the section, and the vehicle speed calculated on the bottleneck state detectors. First of all, traffic flow rate refers to the number of passing vehicles on the state detectors, and this study utilize this performance indicator as a reward function. In addition, average vehicle speed, also used as a reward function, represents the average speed of each vehicles on the whole freeway section. Thirdly, throughput is calculated by the number of vehicles present at the

start plus those attempting to enter and successfully entering the system during the analysis period. Lastly, the bottleneck speed stands for the average speed of each vehicles passing the bottleneck state detectors. Figure 4.1 presents the detailed location of the calculating performance indicators.

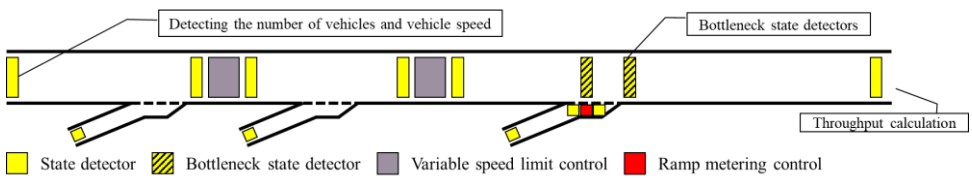


Figure 4.1 Detailed location for the calculation of the performance indicators

During the learning process, the result of the actions ordinarily tends to set high-level speed limits and open ramps. It can be seen that it is more effective not to reduce the traffic flow by not restricting the speed or ramp metering than the earnings obtained by lowering the speed of the main vessel by performing VSL or RM. In other words, it is more efficient to increase the overall speed of the vehicle rather than to have the vehicle delayed as the result of preemptively reducing the speed of the vehicle ahead of the congestion section and the improvement of the overall flow by joining the vehicle to the main lane is better than the vehicle delays in the mainline caused by the vehicle ramping on, which is seen for the traffic flow to be more efficient.

TABLE 4.1-2 summarize the results on the traffic flow for the uncontrolled situation, only VSL is used one, one that only RM is applied, and when VSL and RM are integrated. Overall, the results show that VSL, RM, and VSL and RM integrated control improve most

performance indicators than uncontrolled cases. For the case using flow rate as a reward, the VSL and RM integrated control result shows the best in all performance measures (TABLE 4.1). In addition, RM control shows better performance than VSL control in terms of reward. On the other hand, the average speed as a reward reveals that VSL control shows better performance than RM control in terms of reward (TABLE 4.2). It implies that different objectives could lead to different effects on the traffic flow from the control based on reinforcement learning. Therefore, it must be analyzed while comparing different rewards.

TABLE 4.1 Model comparison for uncontrolled, VSL, RM, and integrated control under the reward as the flow

Strategies	Reward		
	Reward	Throughput	Bottleneck speed
Uncontrolled	10,002.85	26,251.5	27.4838
VSL Control	10,083.21* (0.80%)	26,457.90* (0.79%)	27.9967* (1.87%)
RM Control	10,586.81* (5.84%)	27,782.38* (5.83%)	27.6226* (0.51%)
VSL and RM control	10,640.66* (6.38%)	27,928.57* (6.39%)	28.1723* (2.51%)

Notes. * = statistically significant difference in 95% confidence level; VSL = Variable speed limit control; RM = Ramp metering control.

TABLE 4.2 Model comparison for uncontrolled, VSL, RM, and integrated control under the reward as the average speed

Strategies	Reward		
	Reward	Throughput	Bottleneck speed
Uncontrolled	9,644.63	26,239.21	27.4781
VSL Control	9,976.64* (3.44%)	26,479.69* (0.92%)	28.0060* (1.92%)
RM Control	9,701.15* (0.59%)	27,689.20* (5.53%)	27.5922* (0.42%)
VSL and RM control	10,036.61* (4.06%)	27,880.73* (6.26%)	28.1564* (2.47%)

Notes. * = statistically significant difference in 95% confidence level; VSL = Variable speed limit control; RM = Ramp metering control.

Figure 4.2-9 compare the average speed and flow rate of vehicles in the freeway section under uncontrolled, VSL control, RM control, and the integrated control of VSL and RM situations. On the I-880 southbound freeway, VSL and RM control were selected for the section where persistent congestion occurred or shifted overall traffic flow.

The average speed of each link could reveal the effect of VSL (**Figure 4.2 and 4.3**). The average speed of the link before 1st VSL section increased by 0.37% and that of 1st VSL applied link by 0.96%. The average speed ascended by 0.29% before the second VSL link and 1.72% after the second VSL section. On the other hand, the effect of RM could be found in **Figure 4.2 and 4.4**. The average speed of on-ramp link increased by 385.58%, and that of post-RM section

decreased by 2.01%. **Figure 4.2 and 4.5** show that the application of integrated VSL and RM resulted in increase in speed around 1st and 2nd VSL sections and on-ramp link. Also, the congestion in the on-ramp section was alleviated partially, and the average speed increased in the overall section.

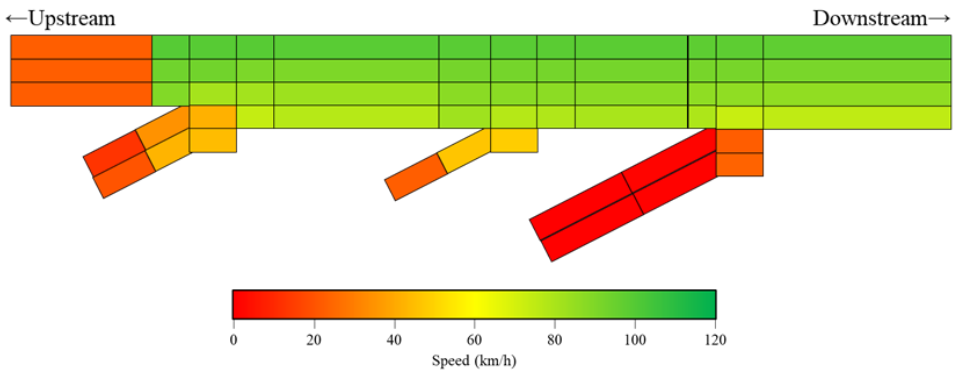
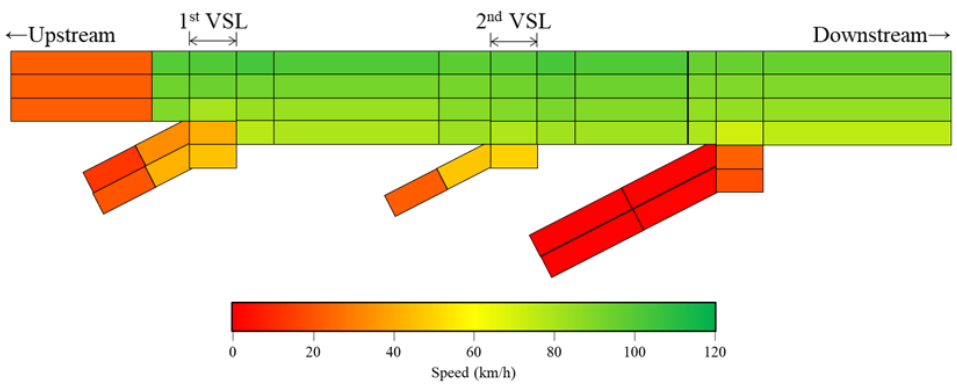
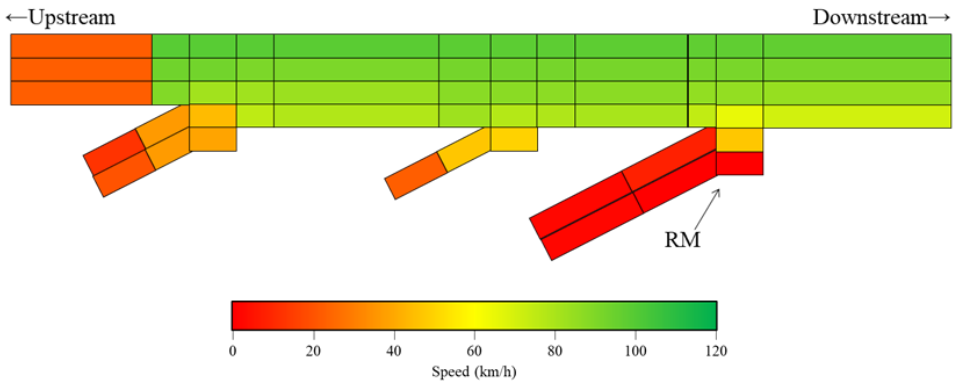


Figure 4.2 Heatmap of average vehicle speed by freeway section under uncontrolled situations



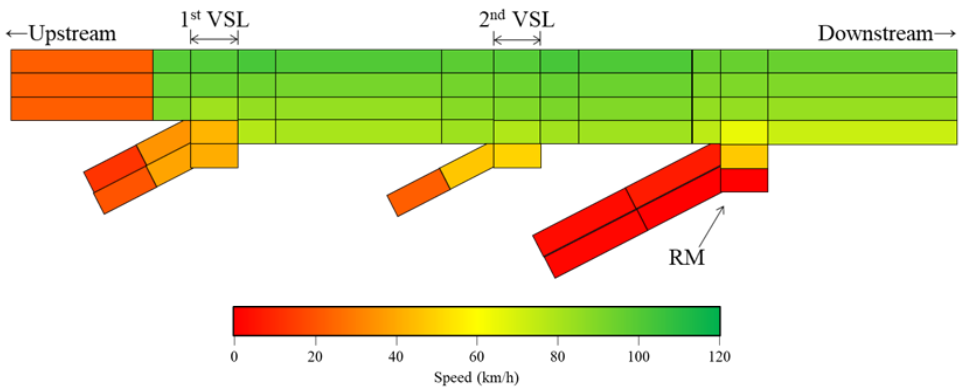
Notes. VSL = Variable speed limit control

Figure 4.3 Heatmap of average vehicle speed by freeway section under VSL controlled situations



Notes. RM= Ramp metering control

Figure 4.4 Heatmap of average vehicle speed by freeway section under RM controlled situations



Notes. VSL = Variable speed limit control; RM= Ramp metering control

Figure 4.5 Heatmap of average vehicle speed by freeway section under VSL and RM controlled situations

Furthermore, comparison of the flow rate at each link enable the estimating efficacy of control strategies (Figure 4.6–9). Figure 4.6 and 4.7 shows that flow rate at the 1st and 2nd VSL section increased by 2.09% and 1.66%, respectively. VSL control reduced the traffic flow rate on curbside of the main line. Flow rate at the link right after

the RM control, and at the on-ramp link increased by 5.55% and 88.00%, respectively (**Figure 4.6 and 4.8**). It seems that RM control could also alleviate congestion and contribute to the improvement of flow. The results of the integrated VSL and RM control indicate that flow rate increasement was higher than RM only control (**Figure 4.8 and 4.9**).

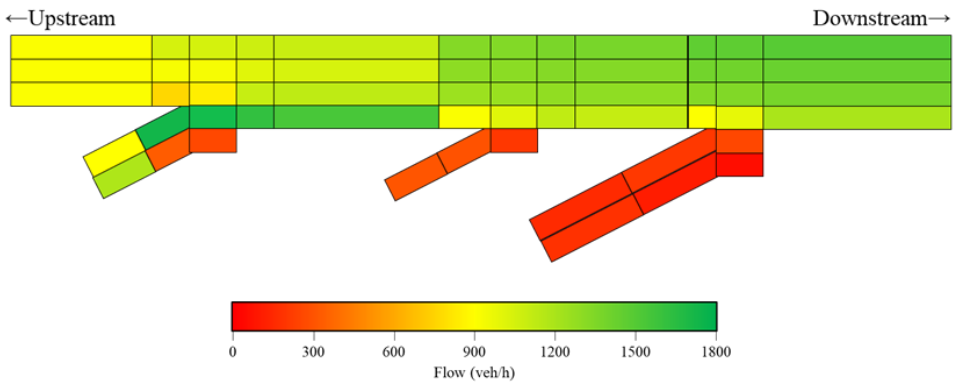
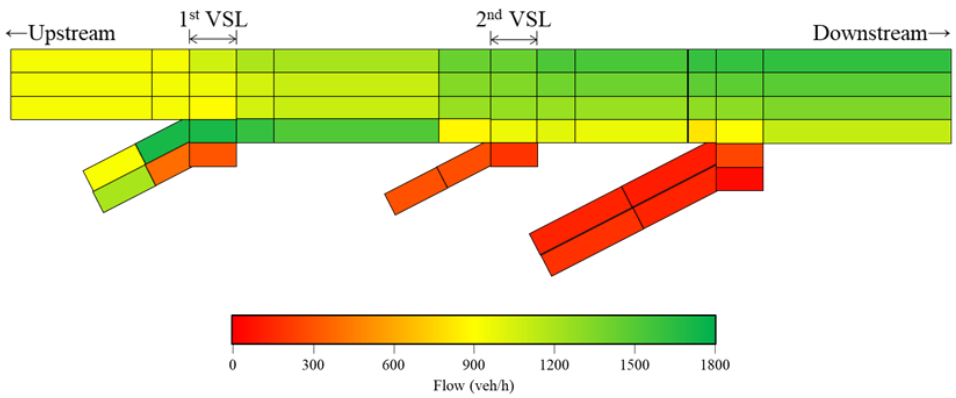
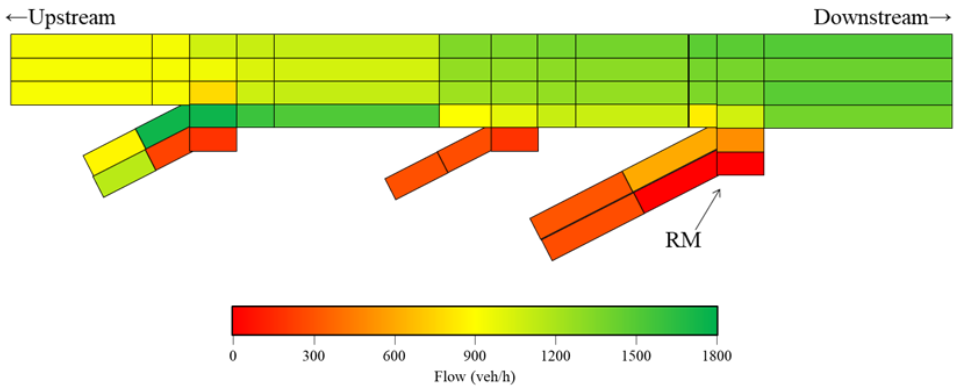


Figure 4.6 Heatmap of flow rate by freeway section under uncontrolled situations



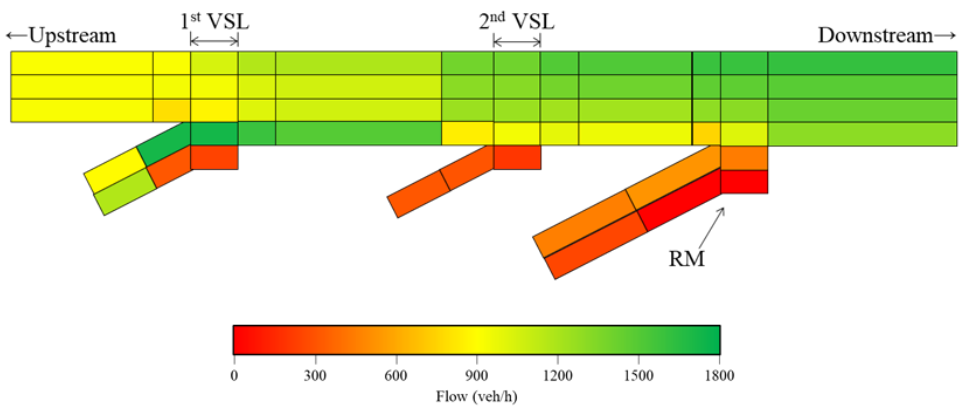
Notes. VSL = Variable speed limit control

Figure 4.7 Heatmap of flow rate by freeway section under VSL controlled situations



Notes. RM= Ramp metering control

Figure 4.8 Heatmap of flow rate by freeway section under RM controlled situations



Notes. VSL = Variable speed limit control; RM= Ramp metering control

Figure 4.9 Heatmap of flow rate by freeway section under VSL and RM controlled situations

4.2. Efficiency according to the flow rate

There are various traffic conditions on the highway throughout the day. There is a low level of traffic flow in the early morning and afternoon, and there is a high traffic flow rate during peak hours of commuting. In various traffic flow situations, the manager who is operating the freeway must apply VSL and RM properly. Therefore, it was confirmed whether the VSL and RM strategies were effective under traffic flow rate conditions. The concept of simulated traffic volume ratio used in this study is used to generate simulation data of various traffic flow conditions by multiplying the actual traffic volume data collected through VDS data by the ratio and then apply it to the experiment. It enables the simulation test on various road condition from uncongested to highly congested freeway conditions. **Figure 4.10-11** shows the traffic flow condition applied with the simulated traffic volume ratio of 0.2 to 2.5 times the usual morning rush hour, and the effect was investigated by applying the integrated model of VSL and RM using the DDPG algorithm, respectively.

As the simulated traffic volume ratio increases, the density does not always increase, and the flow does not continue to increase (**Figure 4.10**). In addition, the average vehicle speed increased under the low simulated traffic volume ratio due to increasing the speed limit after implementing the algorithm. **Figure 4.11** implies that the average density decreased around a certain flow rate level. Also, the higher the simulated traffic volume ratio, the higher the flow dispersion.

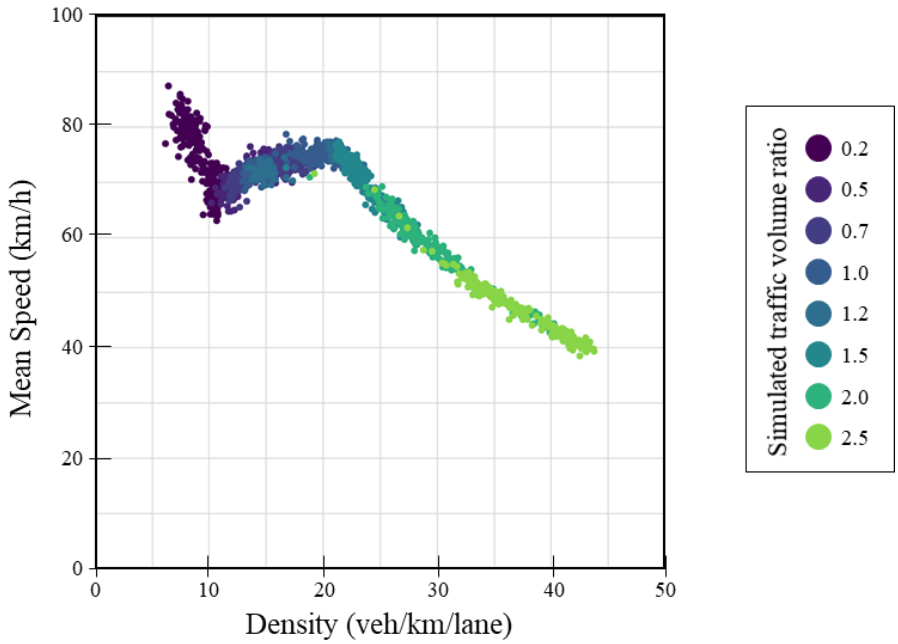


Figure 4.10 Fundamental diagram of traffic flow (mean speed-density) after implementation of VSL and RM control

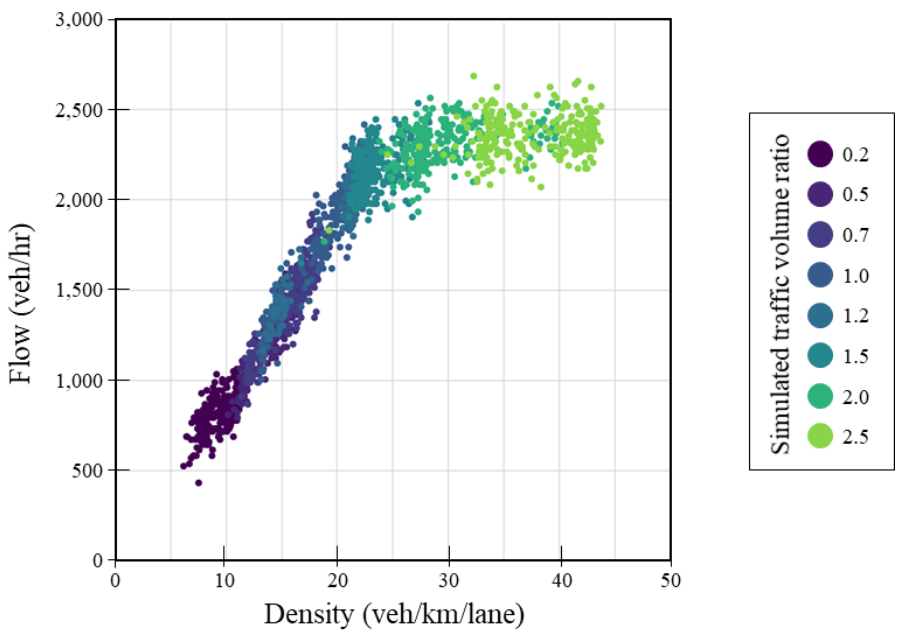


Figure 4.11 Fundamental diagram of traffic flow (flow-density) after implementation of VSL and RM control

TABLE 4.3 shows the change of the traffic flow fundamentals (i.e., flow, density, and speed) according to simulated traffic volume rates before and after applying the algorithm. The results indicate that the change of the overall flow rate is not statistically significant under the low volume ratio. The decrease in traffic flow became statistically significant when the simulated traffic volume ratio was 1.2 or higher. This means that when the traffic flow is low, the difference in overall traffic volume is not large by applying the algorithm. On the other hand, in terms of density and speed, most of the cases (except 1.0) show insignificance for density reduction. It could be seen that traffic control algorithm can cause higher density in the congested situations. The simulated results where simulated traffic volume ratio is 1.0 or higher show that vehicle speed increased significantly.

TABLE 4.3 Comparison before and after application of the strategy in terms of traffic flow, density, and speed

Simulated traffic volume ratio	Flow (veh/hr)		Density (veh/km/lane)		Speed (km/h)	
	Before learning	After learning	Before learning	After learning	Before learning	After learning
0.2	816.30 (6.76)	816.12 (6.76)	9.2338 (0.08865)	9.1987 (0.1212)	74.4545 (0.3593)	74.6997 (0.4776)
0.5	**1,243.07 (11.00)**	**1,249.41* (7.36)**	14.0150 (0.2164)	14.1246* (0.07844)	71.8978 (0.5639)	71.7007 (0.1856)
0.7	1,442.85 (8.80)	1,448.41 (11.04)	15.7210 (0.1320)	15.8778* (0.1026)	73.2295 (0.3172)	72.9123* (0.1326)
1.0	1,825.66 (9.07)	1,755.87* (3.23)	**24.6045 (0.2062)**	**21.5344* (0.1433)**	**62.6941 (0.5318)**	**66.8590* (0.2255)**
1.2	**1,940.14 (8.53)**	**1,953.98* (8.77)**	20.2854 (0.2661)	20.3652 (0.1035)	**73.6894 (0.7086)**	**74.0164* (0.1787)**
1.5	**2,162.61 (11.30)**	**2,169.39* (7.96)**	24.0321 (1.3585)	23.2786* (0.2558)	**68.5855 (2.2215)**	**70.9967* (0.6609)**
2.0	**2,296.86 (8.30)**	**2,299.57 (10.96)**	32.0159 (1.6543)	30.1979* (0.7932)	**54.0429 (2.2101)**	**57.2798* (1.3253)**
2.5	**2,348.39 (9.97)**	**2,354.21* (7.30)**	39.5380 (1.5829)	37.6978* (0.7894)	**44.1639 (1.9156)**	**46.4071* (1.0100)**

Notes. * = statistically significant difference in 95% confidence level; Bold for enhancement; Standard deviation in parentheses.

These observations imply that the VSL and RM control can significantly change flow fundamentals and be effective at the appropriate level of traffic flow. However, in a traffic flow situation that exceeds the capacity, the congestion level of the mainline was relatively severe. Therefore, it leads to performing the VSL or RM strategy to increase the density and lower overall speed. In this case, the flow also decreased.

In the learning process through episodes, there are the points at which the speed limit changes for each lane. Initially, the model tries random speed limits and RM strategies via exploration after which it changes the speed limit to 70 mph, followed by increasing the speed limit to 80 mph. The model ultimately raises the speed limits to ameliorate the overall flow, closer to the critical point, unless there has been an accident. However, the pronounced periods for increasing the speed differ from each lane and the flow rates.

4.3. Effectiveness of the GCN Layer

It is significant to include the spatial correlation between the state detectors that gather information continuously. In that end, the adjacency matrix, including the geographical distance data of each state detector, was put into the dense layer of the critic model and was multiplied by the existing matrix in the corresponding layer. As shown in TABLE 4.4, it was observed that the improvement of reward was statistically insignificant compared to the strategy of the existing model under the same traffic flow level and learning process with an average speed as a reward function. Although it is difficult to see a prominent difference, it is meaningful that the model could be improved just by embedding the geographic location of the state detector.

TABLE 4.4 Comparison of the models for the effectiveness of the graph convolution network layer

Strategies	Reward (Flow)		
	Reward	Throughput	Bottleneck speed
Baseline	10,640.66	27,928.57	28.1723
GCN layer	10,633.77	27,909.37	28.1756
Strategies	Reward (Average speed)		
	Reward	Throughput	Bottleneck speed
Baseline	10,036.61	27,880.73	28.1564
GCN layer	10,048.15	27,907.38	28.1701

Notes. GCN = Graph Convolution Network.

Chapter 5. Conclusion

Prior work has documented the enhanced performance of the differential variable speed limit control by the deep RL algorithm, and the effects of different rewards were verified (Wu et al., 2020). Previous studies were able to verify and confirm the effect of various approaches to VSL, but they did not confirm the effect of integrating several control strategies. To resolve this gap, the aim of this study was to confirm the effectiveness of this integrated strategy in various aspects. Overall, the VSL or RM strategy improves the overall flow rates by reducing the density of vehicles and improving the vehicles' average speed. However, particularly at a high level of traffic flow, VSL or RM control may not be appropriate. In addition, it was found that the integrated strategy can be used when including the relationship between each state detector in multiple VSL sections and lanes by applying the adjacency matrix in the neural network layer. The improvement of flow fundamentals was analyzed at the link level based on the simulation results. It is also remarkable that the action results obtained as a result of this strategy indicated that there is an association between the front and the back when there are several VSL sections and that a different speed limit for each lane is effective. It is meaningful that real-world data of one of the freeways used extensively in the United States and a simulation environment similar to reality were used to reflect the realistic environment. Therefore, this study proves the excellent performance of the strategies of VSL and RM in improving freeway traffic flow by reducing congestion and increasing the average speed of vehicles. Most notably, the results

of this study broaden the view to integrating multiple traffic control strategies. Furthermore, it was shown that problems that may arise in the integration process could be solved by including a model that reflects the spatial correlation.

Our results further provide useful information when learning-based models are included in strategies to control traffic. The traffic management systems vary in their form and efficacy based on the different flow rate. In the future, if autonomous vehicles are introduced to roads and connected vehicles are implemented so that vehicles and road infrastructure can communicate with each other, this study can make it possible to advance VSL and RM further. However, some limitations are worth noting, i.e., some values were missing in the data, and the data collected by VDS could not be considered to be entirely reliable. Also, there is a limitation in that the data were aggregated in 5-minute increments, which did not properly reflect the traffic flow. In addition, there is a part where it was assumed that the microscopic simulation and its link level flow fundamental analysis worked well and was similar to the actual traffic flow. Therefore, future work should include follow-up work for multi-agent reinforcement learning, considering each agent as an independent, but connected, agent. With the introduction of autonomous driving, control of individual agents has become somewhat easier, and many studies have been proposed recently. It is important to remember that human-driven vehicles and autonomous vehicles have different operating properties, and their complex influence and relationship should be identified in mixed flow. Also, proactive traffic control algorithm considering whole traffic network is needed. Useful traffic management system should include proactively managing traffic to prevent congestion by understanding

wide range of traffic network and its data. At this time, it is multi-agent reinforcement learning and graph based neural network that can best reveal this situation and find the optimal strategy, and follow-up studies on this should be conducted.

Bibliography

Abdel-Aty, M., Dilmore, J., Dhindsa, A., 2006. Evaluation of variable speed limits for real-time freeway safety improvement. Accid. Anal. Prev. 38, 335–345. https://doi.org/10.1016/j.aap.2005.10.010

Akyol, G., Silgu, M.A., Celikoglu, H.B., 2019. Pedestrian-friendly traffic signal control using eclipse sumo. Epic Ser. Comput. 62, 101–106. https://doi.org/10.29007/c6k6

Alessandri, A., Di Febbraro, A., Ferrara, A., Punta, E., 1997. Optimal control of freeways via speed signalling. ECC 1997 - Eur. Control Conf. 6, 1403–1408. https://doi.org/10.23919/ecc.1997.7082297

Caltrans, 2021. Performance Measurement System (PeMS) [WWW Document]. URL http://pems.dot.ca.gov (accessed 5.1.21).

Carlson, R.C., Papamichail, I., Papageorgiou, M., Messmer, A., 2010. Optimal mainstream traffic flow control of large-scale motorway networks. Transp. Res. Part C Emerg. Technol. 18, 193–212. https://doi.org/10.1016/j.trc.2009.05.014

Cho, J.-H., Ham, S.W., Kim, D.-K., 2021. Enhancing the Accuracy of Peak Hourly Demand in Bike-Sharing Systems using a Graph Convolutional Network with Public Transit Usage Data. Transp. Res. Rec. J. Transp. Res. Board 036119812110120. https://doi.org/10.1177/03611981211012003

Fares, A., Gomaa, W., 2014. Freeway ramp-metering control based on Reinforcement learning, in: 11th IEEE International Conference on Control & Automation (ICCA). IEEE, pp. 1226–1231. https://doi.org/10.1109/ICCA.2014.6871097

Ghods, A., Kian, A., Tabibi, M., 2009. Adaptive freeway ramp metering and variable speed limit control: a genetic–fuzzy approach. IEEE Intell. Transp. Syst. Mag. 1, 27–36. https://doi.org/10.1109/MITS.2009.932718

Gu, S., Holly, E., Lillicrap, T., Levine, S., 2017. Deep reinforcement learning for robotic manipulation with asynchronous off–policy updates, in: 2017 IEEE International Conference on Robotics and Automation (ICRA). IEEE, pp. 3389–3396. https://doi.org/10.1109/ICRA.2017.7989385

Hegyi, A., De Schutter, B., Hellendoorn, H., 2005. Model predictive control for optimal coordination of ramp metering and variable speed limits. Transp. Res. Part C Emerg. Technol. 13, 185–209. https://doi.org/10.1016/j.trc.2004.08.001

Hegyi, A., Hoogendoorn, S.P., Schreuder, M., Stoelhorst, H., Viti, F., 2008. Specialist: A dynamic speed limit control algorithm based on shock wave theory. IEEE Conf. Intell. Transp. Syst. Proceedings, ITSC 827–832. https://doi.org/10.1109/ITSC.2008.4732611

Kheterpal, N., Parvate, K., Wu, C., Kreidieh, A., Vinitsky, E., Bayen, A., 2018. Flow: Deep Reinforcement Learning for Control in SUMO, in: EPiC Series in Engineering. pp. 134–115. https://doi.org/10.29007/dkzb

Khondaker, B., Kattan, L., 2015. Variable speed limit: A microscopic analysis in a connected vehicle environment. Transp. Res. Part C Emerg. Technol. 58, 146–159. https://doi.org/10.1016/j.trc.2015.07.014

Kim, T.S., Lee, W.K., Sohn, S.Y., 2019. Graph convolutional network approach applied to predict hourly bike–sharing demands considering spatial, temporal, and global effects. PLoS One 14,

1–16. https://doi.org/10.1371/journal.pone.0220782

Kipf, T.N., Welling, M., 2017. Semi−supervised classification with graph convolutional networks. 5th Int. Conf. Learn. Represent. ICLR 2017 − Conf. Track Proc.

Kuhne, R.D., 1991. Freeway Control Using a Dynamic Traffic Flow Model and Vehicle Reidentification Techniques. Transp. Res. Rec. 251–259.

Kušić, K., Ivanjko, E., Gregurić, M., Miletić, M., 2020. An overview of reinforcement learning methods for variable speed limit control. Appl. Sci. 10. https://doi.org/10.3390/app10144917

Li, Z., Liu, P., Xu, C., Duan, H., Wang, W., 2017. Reinforcement Learning−Based Variable Speed Limit Control Strategy to Reduce Traffic Congestion at Freeway Recurrent Bottlenecks. IEEE Trans. Intell. Transp. Syst. 18, 3204–3217. https://doi.org/10.1109/TITS.2017.2687620

Li, Z., Xu, C., Pu, Z., Guo, Y., Liu, P., 2020. Reinforcement Learning−Based Variable Speed Limits Control to Reduce Crash Risks near Traffic Oscillations on Freeways. IEEE Intell. Transp. Syst. Mag. 1–1. https://doi.org/10.1109/MITS.2019.2907631

Lillicrap, T.P., Hunt, J.J., Pritzel, A., Heess, N., Erez, T., Tassa, Y., Silver, D., Wierstra, D., 2015. Continuous control with deep reinforcement learning. 4th Int. Conf. Learn. Represent. ICLR 2016 − Conf. Track Proc.

Lu, C., Huang, J., Deng, L., Gong, J., 2017. Coordinated ramp metering with equity consideration using reinforcement learning. J. Transp. Eng. 143, 1–11. https://doi.org/10.1061/JTEPBS.0000036

Lu, X.Y., Varaiya, P., Horowitz, R., Su, D., Shladover, S.E., 2011.

Novel freeway traffic control with variable speed limit and coordinated ramp metering. Transp. Res. Rec. 55–65. https://doi.org/10.3141/2229-07

Lv, Z., Xu, J., Zheng, K., Yin, H., Zhao, P., Zhou, X., 2018. LC-RNN: A Deep Learning Model for Traffic Speed Prediction, in: Proceedings of the Twenty-Seventh International Joint Conference on Artificial Intelligence. International Joint Conferences on Artificial Intelligence Organization, California, pp. 3470–3476. https://doi.org/10.24963/ijcai.2018/482

Mnih, V., Badia, A.P., Mirza, M., Graves, A., Lillicrap, T.P., Harley, T., Silver, D., Kavukcuoglu, K., 2016. Asynchronous Methods for Deep Reinforcement Learning. 33rd Int. Conf. Mach. Learn. ICML 2016 4, 2850–2869.

Mnih, V., Kavukcuoglu, K., Silver, D., Rusu, A.A., Veness, J., Bellemare, M.G., Graves, A., Riedmiller, M., Fidjeland, A.K., Ostrovski, G., Petersen, S., Beattie, C., Sadik, A., Antonoglou, I., King, H., Kumaran, D., Wierstra, D., Legg, S., Hassabis, D., 2015. Human-level control through deep reinforcement learning. Nature 518, 529–533. https://doi.org/10.1038/nature14236

OpenStreetMap, n.d. OpenStreetMap [WWW Document]. URL https://www.openstreetmap.org/ (accessed 5.1.21).

Papageorgiou, M., Kotsialos, A., 2002. Freeway Ramp Metering: An Overview. IEEE Trans. Intell. Transp. Syst. 3, 271–281. https://doi.org/10.1109/TITS.2002.806803

Papamichail, I., Kampitaki, K., Papageorgiou, M., Messmer, A., 2008. Integrated Ramp Metering and Variable Speed Limit Control of Motorway Traffic Flow, IFAC Proceedings Volumes. IFAC. https://doi.org/10.3182/20080706-5-kr-1001.02384

Popov, A., Hegyi, A., Babuška, R., Werner, H., 2008. Distributed controller design approach to dynamic speed limit control against shockwaves on freeways. Transp. Res. Rec. 93–99. https://doi.org/10.3141/2086-11

Sallab, A., Abdou, M., Perot, E., Yogamani, S., 2017. Deep Reinforcement Learning framework for Autonomous Driving. Electron. Imaging 2017, 70–76. https://doi.org/10.2352/ISSN.2470-1173.2017.19.AVM-023

Schmidt-Dumont, T., Van Vuuren, J.H., 2015. Decentralised reinforcement learning for ramp metering and variable speed limits on highways. IEEE Trans. Intell. Transp. Syst. 14, 1–10.

Silver, D., Huang, A., Maddison, C.J., Guez, A., Sifre, L., van den Driessche, G., Schrittwieser, J., Antonoglou, I., Panneershelvam, V., Lanctot, M., Dieleman, S., Grewe, D., Nham, J., Kalchbrenner, N., Sutskever, I., Lillicrap, T., Leach, M., Kavukcuoglu, K., Graepel, T., Hassabis, D., 2016. Mastering the game of Go with deep neural networks and tree search. Nature 529, 484–489. https://doi.org/10.1038/nature16961

Smulders, S., 1990. Control of freeway traffic flow by variable speed signs. Transp. Res. Part B 24, 111–132. https://doi.org/10.1016/0191-2615(90)90023-R

Soriguera, F., Martínez, I., Sala, M., Menéndez, M., 2017. Effects of low speed limits on freeway traffic flow. Transp. Res. Part C Emerg. Technol. 77, 257–274. https://doi.org/10.1016/j.trc.2017.01.024

Wang, C., Zhang, J., Xu, L., Li, L., Ran, B., 2019. A New Solution for Freeway Congestion: Cooperative Speed Limit Control Using Distributed Reinforcement Learning. IEEE Access 7, 41947–41957. https://doi.org/10.1109/ACCESS.2019.2904619

Wu, C., Kreidieh, A., Parvate, K., Vinitsky, E., Bayen, A.M., 2017. Flow: A Modular Learning Framework for Autonomy in Traffic. arXiv 1–18.

Wu, Y., Tan, H., Jiang, Z., Ran, B., 2019. ES-CTC: A deep neuroevolution model for cooperative intelligent freeway traffic control. arXiv.

Wu, Y., Tan, H., Qin, L., Ran, B., 2020. Differential variable speed limits control for freeway recurrent bottlenecks via deep actor-critic algorithm. Transp. Res. Part C Emerg. Technol. 117, 102649. https://doi.org/10.1016/j.trc.2020.102649

Yu, B., Lee, Y., Sohn, K., 2020. Forecasting road traffic speeds by considering area-wide spatio-temporal dependencies based on a graph convolutional neural network (GCN). Transp. Res. Part C Emerg. Technol. 114, 189–204. https://doi.org/10.1016/j.trc.2020.02.013

Zegeye, S.K., De Schutter, B., Hellendoorn, J., Breunesse, E.A., 2010. Variable speed limits for area-wide reduction of emissions. IEEE Conf. Intell. Transp. Syst. Proceedings, ITSC 507–512. https://doi.org/10.1109/ITSC.2010.5625032

Zhang, J., Boitor, A., Ioannou, P., 2005. Design and evaluation of a roadway controller for freeway traffic. IEEE Conf. Intell. Transp. Syst. Proceedings, ITSC 2005, 543–548. https://doi.org/10.1109/ITSC.2005.1520106

Abstract

최근에는 교통혼잡으로 인한 사회적 문제를 해결하기 위해 고속도로를 효율적으로 운영하기 위한 교통통제 전략이 다양하게 개발되고 있다. 고속도로 교통류를 효과적으로 관리하기 위한 대표적인 전략으로는 차로별 제한속도를 다르게 적용하는 가변 속도 제한(VSL) 제어와 진입 램프에서 신호를 통해 차량을 통제하는 램프 미터링(RM) 전략 등이 있다. 본 연구의 목표는 심층 강화 학습(deep reinforcement learning)을 활용하여 고속도로의 효율적인 교통 흐름을 얻기 위해 동적 VSL 및 RM 제어 알고리즘을 개발하는 것이다. 고속도로의 여러 VSL과 RM 구간에서 시뮬레이션을 통해 심층 강화학습 알고리즘 중 하나인 deep deterministic policy gradient (DDPG) 알고리즘을 적용한 교통류 제어 전략을 검증한다. 실험 결과, 강화학습 기반 VSL 또는 RM 전략을 적용하는 것이 램프 진입로 구간의 혼잡을 완화하고 나아가 전체 구간의 혼잡을 줄이는 것으로 나타났다. 대부분의 경우 VSL이나 RM 전략은 본선과 진입로 구간의 밀도를 줄이고 차량의 평균 통행 속도를 증가시켜 전체 교통 흐름을 향상시킨다. VSL 또는 RM 전략들은 높은 수준의 교통류에서 적절하지 않을 수 있어 교통류 수준에 따른 전략의 선택적 도입이 필요하다. 또한 검지기간 지리적 거리와 관련한 인접 행렬을 포함하는 graph neural network layer이 여러 지점 검지기의 공간적 상관 관계를 감지하는 데 이용될 수 있다. 본 연구의 결과는 강화학습 기반 VSL과 RM 전략 도입의 필요성과 지점 검지기 간의 공간적 상관관계의 중요성을 반영하는 전략 도입의 효과를 시사한다.

주요어 : 교통류제어 전략, 가변제한속도제어, 램프미터링, 강화학습
학 번 : 2020-23170