Final Project Report

Drivers' Attitudes Towards Rerouting: Impacts on Network Congestion

Prepared for **Teaching Old Models New Tricks (TOMNET) Transportation Center**

By

Ziming Liu
Email: ziming_liu@gatech.edu

Jorge A. Laval
Email: jorge.laval@ce.gatech.edu

Georgia Institute of Technology
School of Civil and Environmental Engineering
790 Atlantic Drive, Atlanta, GA 30332

August 2023

TECHNICAL REPORT DOCUMENTATION PAGE

1. Report No. N/A	**2. Government Accession No.** N/A	**3. Recipient's Catalog No.** N/A
4. Title and Subtitle Drivers' Attitudes Toward Rerouting: Impacts on Network Congestion		**5. Report Date** August 2023
		6. Performing Organization Code N/A
7. Author(s) Ziming Liu, Jorge A. Laval, https://orcid.org/0000-0002-0986-4046		**8. Performing Organization Report No.** N/A
9. Performing Organization Name and Address School of Civil and Environmental Engineering Georgia Institute of Technology 790 Atlantic Drive, Atlanta, GA 30332		**10. Work Unit No. (TRAIS)** N/A
		11. Contract or Grant No. 69A3551747116
12. Sponsoring Agency Name and Address U.S. Department of Transportation, University Transportation Centers Program, 1200 New Jersey Ave, SE, Washington, DC 20590		**13. Type of Report and Period Covered** Research Report (2021 – 2023)
		14. Sponsoring Agency Code USDOT OST-R
15. Supplementary Notes N/A		

16. Abstract

Bifurcation, a phenomenon that always occurs in the transition phase between free flow and congestion in the macroscopic fundamental diagram (MFD), have been suspected of decreasing network efficiency. Some studies have shown that rerouting behavior will effectively postpone the occurrence of bifurcation. On the other hand, deep neural networks within reinforcement learning algorithms have been used in traffic control in recent years and produced important breakthroughs. However, recent work has found that it is difficult to learn something in extremely congested networks due to the congested network property.

This report investigates how rerouting behavior affects the bifurcation phenomenon by using deep reinforcement learning (DRL) in drivers' behavior guidelines. We show that i) the learning efficiency of DRL is affected heavily by the congestion level, and ii) the result of the convergence process indicates that DRL didn't take good effect at both low- and high-density levels. These findings lead to a contradiction. We used two different models related to drivers' rerouting tendency to ascertain the key factors in the training process. The study demonstrated the weakness of DRL on drivers' behavior control, and we hope it will be a useful reference to better understand the effect of DRL on transportation and to advance research in this area.

17. Key Words Bifurcation; Rerouting; Deep reinforcement learning; High density		**18. Distribution Statement** No restrictions.	
19. Security Classif.(of this report) Unclassified	**20. Security Classif.(of this page)** Unclassified	**21. No. of Pages** 20	**22. Price** N/A

DISCLAIMER

The contents of this report reflect the views of the authors, who are responsible for the facts and the accuracy of the information presented herein. This document is disseminated in the interest of information exchange. The report is funded, partially or entirely, by a grant from the U.S. Department of Transportation's University Transportation Centers Program. However, the U.S. Government assumes no liability for the contents or use thereof.

ACKNOWLEDGMENTS

This study was funded by a grant from A USDOT Tier 1 University Transportation Center, supported by USDOT through the University Transportation Centers program. The authors would like to thank the TOMNET and USDOT for their support of university-based research in transportation, and especially for the funding provided in support of this project.

TABLE OF CONTENTS

LIST OF TABLES

No table of figures entries found.

LIST OF FIGURES

EXECUTIVE SUMMARY

The existing literature on deep reinforcement learning (DRL) has revealed numerous unresolved issues regarding its applicability in the context of urban network control. The relationship between network properties and learning performance has not been fully explained, and the specific effects of factors such as congestion levels, rerouting tendency, and flow loading rates on learning outcomes are still unknown. Additionally, there is no conclusive evidence in the current body of literature to demonstrate the effectiveness of DRL in highly congested network environments.

This project aimed to explore potential network properties by investigating the impact of rerouting behavior on network efficiency, comparing the performance of a reinforcement learning model and a deep reinforcement learning model. The project also aimed to determine how different variables, such as congestion level and rerouting tendency, affect rerouting decisions and whether machine learning can effectively address the bifurcation phenomenon and increase network efficiency.

In this report, we explore the impact of rerouting behavior and the limitations of the DRL policy in controlling it. Our investigation reveals that the DRL policy faces difficulties in delaying the onset of gridlock and providing satisfactory training results in high-density environments. We focus on the relationship between bifurcation and rerouting behavior and find that training the rerouting behavior with DRL policy can either postpone or eliminate the occurrence of bifurcations. We also find that the convergence results are slow and need to be further analyzed in light of the known weaknesses of DRL. In contrast to traffic signal control, machine learning applied to drivers' adaptive behavior has proved to be more challenging in both low-density and high-density traffic scenarios, but moderate density environments have shown more promising outcomes.

Based on the findings of this study, drivers' adaptive behavior remains a promising factor for traffic control and warrants further research. Behavioral research is needed to better understand the aversion to rerouting decisions from certain segments of the population. Although this project shows that learning these mechanisms from simulation data has proven challenging, further research is needed in this direction as it should provide a solid framework to tackle this important problem.

INTRODUCTION

In the realm of urban transportation networks, traffic congestion can emerge from a variety of unexpected events, defined as changes in traffic state or information unbeknownst to certain drivers, such as incidents, road work, or traffic restrictions. As a result, the localized congestion can propagate throughout the network, culminating in a veritable gridlock, a state of the system where all roads become entirely congested, rendering vehicle movement impossible, which was called "complete jam" or "collapse of the network" by Daganzo (1996) and Gayah and Daganzo (2011). This situation is infrequent in low-flow traffic, as drivers can perceive or be informed of significant changes in travel time and cost before they encounter congestion, providing alternative routes. Thus, the gridlock is unlikely to form in a brief period. In contrast, high-flow traffic and simulation work can quickly generate gridlock when drivers' rerouting behavior is inadequate, leading to deteriorated simulation outcomes.

The Macroscopic Fundamental Diagram (MFD) exposes a relationship between flow and density in a road network, where the bifurcation phenomena refer to the split of the MFD curve at a particular density or flow point. The bifurcation point creates two or more potential equilibrium states for a given density, which can prompt distinct traffic patterns and flow regimes. This can have noteworthy implications for traffic management and control, necessitating the shift of the bifurcation point to optimize flow and diminish congestion. Daganzo et al. (1996) discovered that increasing driver adaptation to real-time traffic conditions can postpone the bifurcation critical density. As such, discovering an effective way to define drivers' rerouting behavior has become a critical problem to reduce the effect of bifurcation, thus improving simulation outcomes.

The development of machine learning (ML) as a dominant paradigm in contemporary decision-making has led to the widespread deployment of various ML methodologies, including the popular technique of reinforcement learning (RL), across a wide range of domains (Mnih et al. 2013). Within this fertile research landscape, the application of RL to teach drivers how to effectively reroute under different congestion scenarios has emerged as a promising area of investigation. However, Laval and Zhou (2019) recent studies have revealed an unexpected and puzzling discovery - existing deep reinforcement learning (DRL) models demonstrate a notable inability to learn effectively in congested network environments due to the problem of vanishing gradients. This finding, while intriguing, must be considered within the context of the underlying assumption of complete driver adaptation, where all drivers have a strong ability to adapt their routing strategies. Otherwise, the validity of this conclusion may be subject to a potentially significant interpretive bias.

Upon thorough investigation, the existing literature on deep reinforcement learning (DRL) has revealed numerous unresolved queries regarding its applicability in congested networks. The relationship between network properties and learning performance has not been fully explained, and the specific effects of factors such as congestion levels, rerouting tendency, and flow loading rates on learning outcomes are still unknown. Additionally, there is no conclusive evidence in the current body of literature to demonstrate the effectiveness of DRL in highly congested network environments.

This project aims to explore the potential network properties by investigating the impact of rerouting behavior on network efficiency, comparing the performance of reinforcement learning model and deep reinforcement learning model. The project also aims to determine how different

variables, such as congestion level and rerouting tendency, affect rerouting decisions and whether machine learning can effectively address the bifurcation phenomenon and increase network efficiency.

To achieve these objectives, the report is structured as follows. First, the background knowledge of macroscopic fundamental diagram in urban networks, reinforcement learning, and the current application of RL in traffic control is presented. Next, the problem setup is defined, and deep q-learning experiments are conducted to identify the impact of density level and rerouting tendency on learning performance. Finally, the report concludes with a discussion and outlook section.

LITERATURE REVIEW

Bifurcation phenomena in MFD

After the Macroscopic Fundamental Diagram (MFD) in network has been verified in previous studies (Daganzo, 2007, Geroliminis and Daganzo, 2008)**,** the utility of the MFD has also been demonstrated with real-world data in a recent publication (Redmond and Mokhtarian, 2001, Geroliminis and Daganzo, 2008, Buisson and Ladier, 2009). As a result, macroscopic models that provide a robust description of the multifarious interrelationships between variables across a plethora of lanes in sprawling networks have gained significant popularity, surpassing their prior status as mere pedagogical curiosities.

According to the literature, the Macroscopic Fundamental Diagram (MFD) can be accurately defined when congestion is homogeneous, and trips remain time-invariant throughout the network. This assumption requires that no specific location within the network experiences extreme congestion while other areas remain free-flowing. For each lane in the network, it is assumed that the kinematic wave model is followed (Lighthill and Whitham, 1955, Richards, 1956), with a common fundamental diagram (Daganzo and Geroliminis, 2008, Laval and Castrillon, 2015). However, it is challenging to observe a well-defined MFD for a particular network in real-life scenarios, as it is rare for all links to exhibit similar congestion levels.

Previous research has observed a phenomenon where the Macroscopic Fundamental Diagram (MFD) curve bifurcates at the transition point from the free flow branch to the congested branch, regardless of the loading or unloading periods (Daganzo et al. 2011, Mahmassani et al. 2013b). Researchers have attributed this to the instability of equilibrium patterns in the congested regime and identified random turning by drivers as an influential factor. Furthermore, it has been confirmed that at the high-density level, there exist multiple flow values corresponding to a specific density (Jin et al. 2013, Gan et al. 2017).

Bifurcation phenomena have been identified as the primary cause of low flows at high densities and the tendency of networks to jam, with their negative effects diminishing as drivers become more adapted to real-time traffic conditions (Daganzo et al. 2011, Mahmassani et al. 2013b). Empirical NMFDs reported in the literature have also been found to exhibit bifurcation, and the uneven distribution of congestion can negatively impact network performance (Ambuhl et al. 2017, Shim et al. 2019). For example, a recent empirical study examined the influence of detouring patterns, heterogeneity, commute trips, and trip completion rates on bifurcations in a real-world dataset (Shim et al. 2019).

Deep Reinforcement Learning

(1) Reinforcement learning (RL)

Reinforcement learning (RL) is a type of learning that maps environmental states to actions, with the goal of maximizing the cumulative reward received during interactions with the environment (Sutton and Barto, 2018). Markov Decision Process (MDP) is a common framework used to model RL problems, and is typically defined as a quadruple (A, S, R, P):

1. **A:** A collection of all action that agent could execute. $a_t \in A$ means the action of *agent* at

time *t*.

2. **S:** A collection of all environmental states. $s_t \in S$ means the state of *agent* at time *t.*
3. **R**: $S \times A \to R$ is the reward function. $r_t \sim R(s_t, a_t)$ means the reward value when *agent* do action a_t in state s_t.
4. **P:** $S \times A \times S \to [0,1]$ is the State transfer probability distribution function. $s_t + 1 \sim P(s_t, a_t)$ means the probability of *agent* transfer to next state $s_t + 1$ when *agent* do action a_t in state s_t.

In RL, policy $\pi: S \to A$:is a mapping from state space to action space. It is expressed as the agent chooses an action a_t in state s_t , performs the action and moves to the next state $s_t + 1$ with probability $f(s_t, a_t)$, while receiving a reward r_t from environmental feedback. Assuming that the immediate reward at each time step in the future must be multiplied by a discount factor γ, the sum of rewards from the start of time t to the end of the episode at time T is defined as:

$$R_t = \sum_{t'=t}^{T} \gamma(t' - t) r_t' \tag{1}$$

The state-action value function $Q^{\pi}(s, a)$ refers to perform action a in the current state s and following the policy $\boldsymbol{\pi}$ until the end of the episode, and the cumulative reward obtained by the agent in this process is expressed as:

$$Q^{\pi}(s, a) = E[R_t | s_t = s, a_t = a] \tag{2}$$

For all state-action pairs, if the expected return of a strategy π^* is greater than or equal to the expected return of all other strategies, then the strategy π^* is called the optimal strategy. There may be more than one optimal policy, but they share a state-action value function:

$$Q^*(s, a) = max_{\pi} E[R_t | s_t = s, a_t = a] \tag{3}$$

Eq. (3) is called the optimal state-action value function, and the optimal state-action value function follows the Bellman optimally equation. which is

$$Q^*(s, a) = E(s' \sim S)[r + \gamma max_a' Q(s', a') | s, a] \tag{4}$$

However, the combination of RL and deep neural network may have problems such as algorithm instability (Tsitsiklis and Van Roy, 1996), which has always hindered the development and application of DRL.

(2) Deep Reinforcement learning (DRL)

Before the emergence of DRL, preliminary work had been conducted. However, due to the shortage of training data and computing power, these studies only utilized deep neural networks to decrease the dimension of high-dimensional input data so that traditional RL algorithms could process it. Riedmiller (2005) was the first to use a multilayer perceptron to estimate the Q-value function and proposed the Neural Fitted Q Iteration (NFQ) algorithm. Lange and Riedmiller (2010) combined DL model and RL method and introduced a deep auto-encoder (DAE) model. However,

DAE was only suitable for control problems with visual perception as the input signal and the state space dimension is small. Abtahi and Fasel (2011) applied deep belief network as a function approximator in traditional RL, which significantly enhanced the agent's learning efficiency and successfully applied it to the character segmentation task of license plate images. Lange and Riedmiller (2012) proposed the Deep Fitted Q-Learning (DFQ) algorithm and employed it in vehicle control. Koutnik et al. (2014) merged the Neural Evolution (NE) method with the RL algorithm and utilized it in a video racing game to achieve the automatic driving of the racing car.

The application of DRL in transportation

As per the methodological background of DRL, DRL is a combination of deep learning and RL that enables the agent to learn complex, high-dimensional input-output mappings without any prior knowledge of the system. In transportation, DRL has been applied to rerouting behavior in two application domains: Traffic signal control and Vehicle routing optimization.

According to the literature, DRL has been applied to various types of traffic control, including adaptive traffic signal control, speed limit control, lane pricing, and ramp metering. Among them, adaptive traffic signal control has been more extensively studied, and the use of DRL has shown promising results. By using DRL, the signal phase can be changed based on real-time traffic conditions in the lanes that the signal controls. This problem is modeled as a sequential decision-making problem. Various DRL methods, such as DQN, DDPG (Casas, 2017), A2C (Coskun et al. 2018, Chu et al. 2019) and PPO (Lin et al. 2018) have been used to control both single intersections and coordinated intersections in a network. Some literature has also applied LSTM with policy networks to solve partially observable environments (Shi and Chen 2018, Chu et al. 2019).

Classical route optimization problems like the travelling salesman problem and vehicle routing problem traditionally use static or dynamic traffic designs. However, with the increasing use of DRL in transportation, vehicle routing optimization has become another popular research area. c. Bello et al. (2016), Khalil et al. (2017) and Kool et al. (2018) utilized DRL to solve the travelling salesman problem, and other studies have applied it to vehicle routing problems (Nazari et al. 2018, Balaji et al. 2019, Kullman et al. 2019, Zhao et al. 2020, Zhang et al. 2020, Peng et al. 2019, James et al. 2019 and Chen et al. 2019). In practical applications, DRL has been used in three areas: urban freight delivery, on-demand ridesharing for passengers, and vehicle holding control.

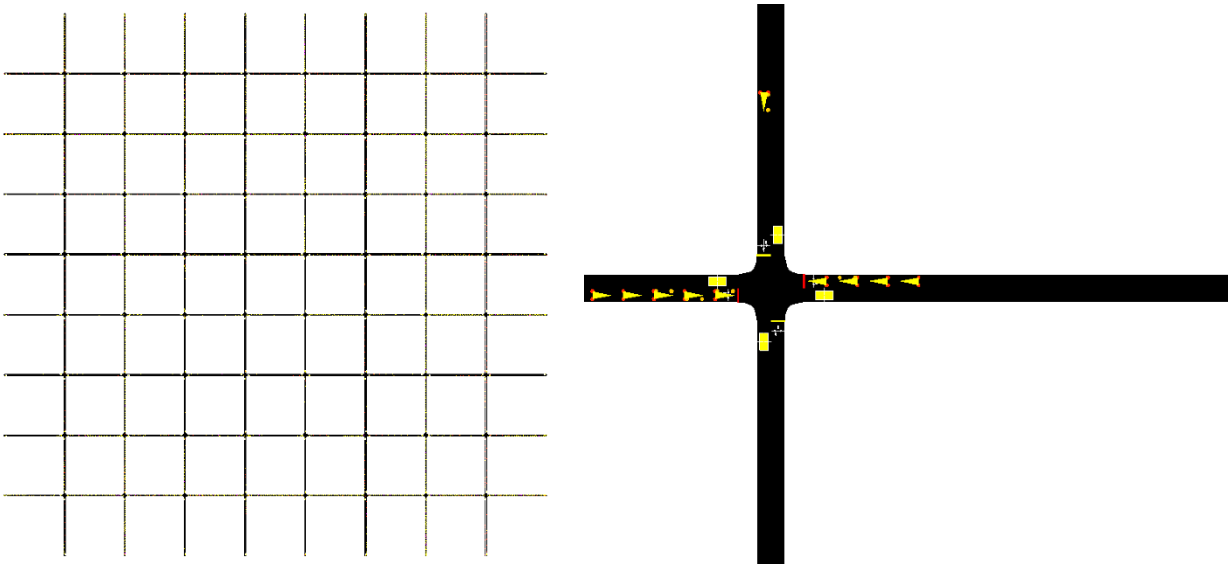
According to the aforementioned review, it can be observed that DRL has shown promising results in signal control, but its application in rerouting behavior has been limited. Previous studies on DRL-based routing problems have mainly focused on pre-assignment, with related parameters being the characteristics of vehicle parameters. In other words, DRL has only been used to determine a route before entering the network, without adjusting a route while the vehicle is running in the network. Recently, Mushtaq et al. (2021, 2022) proposed traffic management systems that combine signal control and rerouting behavior. However, DRL was only applied to signal control, and a predetermined policy was used to reroute vehicles in different traffic situations. In one system, an algorithm was used to estimate the waiting time before entering the intersection and then perform rerouting behavior. In another system, the focus shifted to lanes, where the policy detected the real-time density or number of vehicles on the lane and changed the route for the vehicles that would enter it. Therefore, further research is needed to investigate the application of DRL policies in training rerouting behavior.

DATA

Simulation Environment

In the simulation work, we aimed to building a realistic traffic environment, with the following setting:

1) Network topology: A homogeneous 9x9 grid network with open boundaries was created in SUMO to ensure all intersections had the same layout. This is illustrated in **Figure 1** (a).
2) Intersection configurations: The configurations across all intersections were kept the same, with each edge having one lane. The network was homogeneous with no arterial or minor roads present. This is illustrated in **Figure 1** (b).
3) O-D: In this project, all vehicles had a fixed origin and destination, with drivers initially adopting the shortest route. They would then either continue on the original route or select a new one when they arrived at each intersection.
4) Traffic flow: Traffic flow was evenly distributed throughout the network to avoid any 'hot spots.' Each lane produced the same traffic flow at the same time, with each vehicle randomly choosing a target lane as their destination before entering the network.
5) Traffic signals: Each traffic light had the same signal assignment, with a green light duration of 35 seconds for phase 1 and phase 3 in the N-S direction and E-W direction, respectively, and a red-light duration of 5 seconds for phase 2 and phase 4.



(a) Network Layout (b) Intersection

Figure 1 Network layout and intersection configurations

The DRL Framework

In the simulation work, we assign a specific vehicle as **an agent** that learns from the current environment. **The observed state** of each agent is an $5 \times n$ matrix of bits that includes the agent's direction, the destination's direction, and the three levels of density of the three lanes that the vehicle can enter. **The action** for each agent has three possible options: turn left, go straight, turn right, or take a U-turn. **The reward** at time t, R_t is defined as a combination of the score of decreased density and increased route length. For example, if a vehicle reroutes and enters a new

lane, the score will increase if the new environment satisfied the requirement in different situation. On the other hand, the score will be decreased.

The policy for the vehicle agent is similar to that of a deep neural network, comprising four layers with a Rectified Linear Unit (ReLU) non-linearity activation function. The input layer consists of five nodes that represent the observed state of the agent, while the output layer produces the probabilities of the three possible action options for the agent. Furthermore, the two hidden layers in this study both contain nine nodes to extract and process data before providing the output.

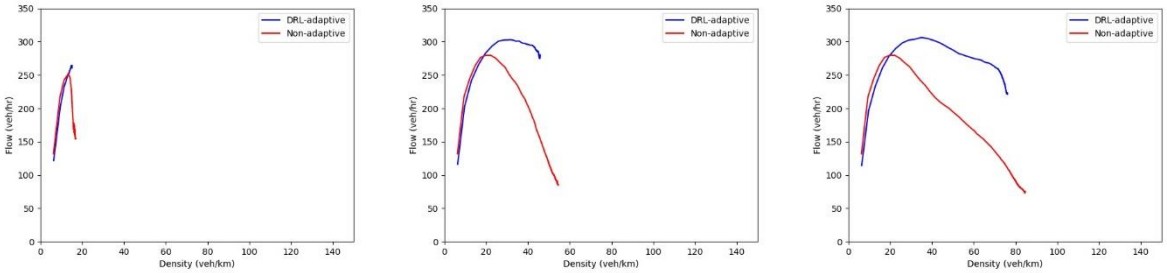
As the network is homogeneous and contains identical intersections and signal assignments, there is no need to use different policies. Therefore, we can use one policy to train a single agent and apply the same parameters and trained matrix to other agents. This training process not only avoids more explicit coordination but also ensures that the states of all agents are determined solely by the training policy.

ANALYSIS

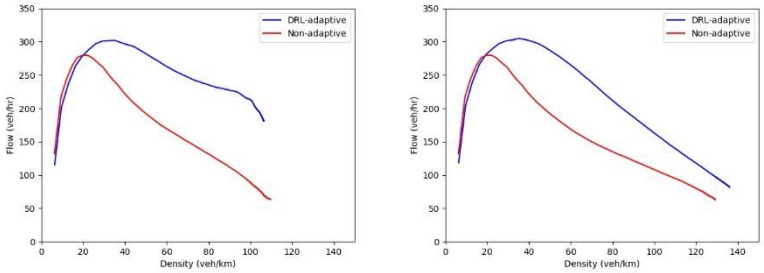
Impact of Driver's adoption on Network efficiency

We compared the MFD curve of non-adaptive driving behavior and random rerouting behavior in different density level.

The comparison between non-adaptive driving behavior and random rerouting behavior in different density levels was conducted by examining the relationship between average flow and average density, as depicted in **Figure 2**. The MFD curve without any rerouting behavior is represented by the red line, while the curve with rerouting behavior under DRL policy is represented by the blue line. Notably, a significant bifurcation phenomenon was observed in each non-adaptive MFD curve, even in different density levels. Surprisingly, the bifurcation occurred earlier than expected when the density level was only equal to 0.1, as shown in **Figure 2** (a). However, the bifurcation did not exhibit a substantial difference in high-density levels, as depicted in **Figure 2** (d) and **Figure 2** (e).



$(a)\ Density = 0.1$ $(b)\ Density = 0.3$ $(c)\ Density = 0.5$



$(d)\ Density = 0.7$ $(e)\ Density = 0.9$

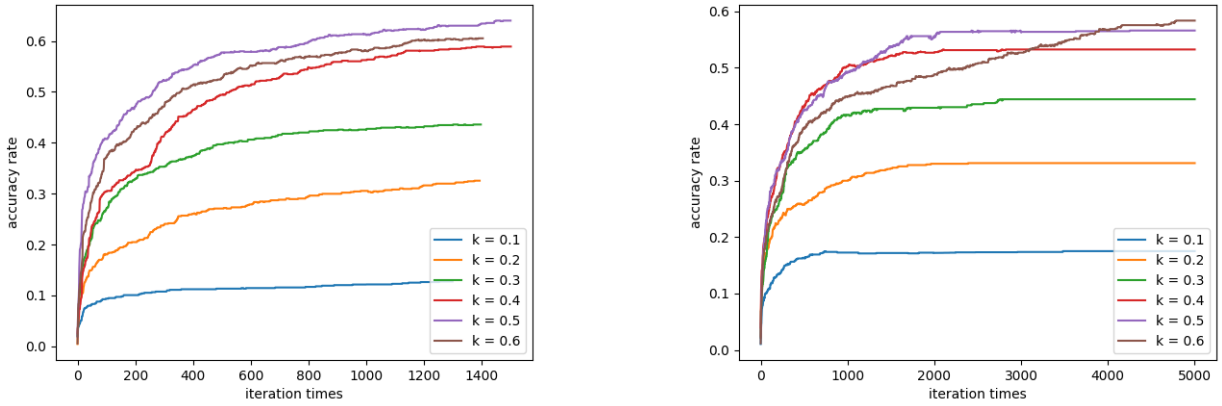
Figure 2 Impact of DRL of rerouting behaviour in different density level

Impact of Driver's rerouting tendency

We classified driver rerouting tendencies into two types: those who prioritize distance from their destination and those who prioritize the density of the next lane. These two types of driver attitudes may have different impacts on the learning behavior and resulting MFD curve. Additionally, we trained these two models in various density levels.

Both models demonstrate that the accuracy of learning generally increases as the network density

increases. Moreover, the convergence speed in each density level is rapid, but achieving a high accuracy requires numerous iterations. However, there are some differences between the two models. The model that prioritizes distance from the destination indicates that the learning results do not improve beyond a density level of 0.5. Conversely, the accuracy with $k = 0.5$ remained superior to $k = 0.6$, see **Figure 3** (a). Meanwhile, the accuracy with $k = 0.5$ and $k = 0.4$ in the model that prioritizes density is better than $k = 0.6$ initially, but after 3000 iterations, the accuracy with $k = 0.6$ surpasses the previous two results, see **Figure 3** (b).



(a) Tend to low density (b) Tend to destination

Figure 3 The Convergence speed of two models in different density levels

RESULTS

Impact of Driver's adoption on Network efficiency

Upon careful comparison of the bifurcation curves across disparate density levels, it becomes apparent that the judicious deployment of rerouting behavior may serve as a potent panacea for the prevention and/or postponement of bifurcation events. However, this salutary effect gradually attenuates with the concomitant increase in density level, such that the desirable impact of rerouting is diminished when operating in high-density conditions. A cursory perusal of **Figure 2** (e) and (a) reveal that the former exhibits a more modest offset as compared to the latter, thus underscoring the rather unsurprising conclusion that random rerouting behavior is unlikely to furnish any significant increase in the capacity flow.

An intriguing revelation emanates from the fact that even in the presence of rerouting behavior, bifurcation events can still manifest, as borne out by the bifurcation curves in **Figure 2** (b), (c) and (d). It is notable that the flow in these instances exhibits a stark reduction at the preeminent locale of target network density. Remarkably, however, bifurcation is entirely absent in **Figure 2** (a) and (e), implying that the density variable may not be the unequivocal causal factor in engendering these discernible outcomes.

Impact of Driver's rerouting tendency

To expound on the phenomenon of model accuracy increasing with density before k=0.5, we must consider the fundamental learning principle of models, which is to analyze all possible observations of an agent and calculate the probabilities of all available choices. Owing to the uneven distribution of vehicles in the network, agents experience fewer potential states at extremely low or high densities, as opposed to moderate densities. However, in a free-flowing network, it is irrelevant to consider traffic conditions. Nevertheless, the outcomes of the study demonstrate that the DRL algorithm cannot train a rerouting behavior model in a congested urban network of significant size. This suggests a need for further research to address the limitations of the DRL algorithm in training models for adaptive behavior in congested urban networks.

CONCLUSIONS AND POLICY IMPLICATIONS

In this report, we explore the impact of rerouting behavior and the limitations of the DRL policy in controlling it. Our investigation reveals that the DRL policy faces difficulties in delaying the onset of gridlock and providing satisfactory training results in high-density environments. We focus on the relationship between bifurcation and rerouting behavior and find that training the rerouting behavior with DRL policy can either postpone or eliminate the occurrence of bifurcation. However, we also find that the convergence results are not up to our expectations in both free-flow and high-density environments, which raises concerns about the effectiveness of the DRL policy in the training process.

We propose that one possible explanation for this phenomenon is the need to reconsider the reward and state design and modify the episode training experiment to improve the training outcome. Alternatively, it could be that the DRL policy cannot learn to control drivers' behavior in high-density environments, creating a self-contradictory situation where we need to avoid gridlock while promoting driver adaptation to delay or prevent congestion. To enhance the training efficiency of the DRL policy, we must address this issue and find ways to balance the two goals. Laval and Zhou (2019) have uncovered the inadequacy of current DRL signal control methods in high-level congestion. Their findings also suggest that traffic signal control has a negligible impact on traffic congestion. In contrast, machine learning applied to drivers' adaptive behavior has yielded unsatisfactory results in both low-density and high-density traffic scenarios, but moderate density environments have shown more promising outcomes, which is in stark contrast to the machine learning applied to traffic signal control. However, based on the current work, drivers' adaptive behavior is a promising factor for traffic control, which needs further research.

Regarding the first explanation for the unsatisfactory results, we need to find more evidence to prove that DRL for rerouting behavior control can be successful in free-flow situations. Then, we can investigate the effect of DRL policy in high-level density. On the other hand, the results have shown that the DRL policies lead to better results than random policy in experiments. Thus, we cannot consider using a random policy to train the drivers' behavior as it conflicts with real traffic situations. As mentioned earlier, we hope the simulation could provide similar results to real traffic work.

When drivers make rerouting decisions, besides considering the density of the original target lane, the estimated travel distance is also a crucial factor for drivers, which cannot be considered by a random policy. Therefore, we need to consider the travel distance factor while designing the DRL policy. Based on the discussion above, we hope this report will serve as a valuable source of inspiration for further investigations of DRL policy designing and the effectiveness of DRL in high-level density. Also, we aim to provide a better understanding of the potential of machine learning in addressing the challenges of urban transportation networks and expect it will have significant implications for the development of efficient and sustainable transportation systems in urban areas.

REFERENCES

[1] Daganzo, C. F. (1996). The nature of freeway gridlock and how to prevent it. Transportation and traffic theory, 629-646.

[2] Gayah, V. V., & Daganzo, C. F. (2011). Clockwise hysteresis loops in the macroscopic fundamental diagram: an effect of network instability. Transportation Research Part B: Methodological, 45(4), 643-655.

[3] Daganzo, C. F., Gayah, V. V., & Gonzales, E. J. (2011). Macroscopic relations of urban traffic variables: Bifurcations, multivaluedness and instability. Transportation Research Part B: Methodological, 45(1), 278-288.

[4] Mnih, V., Kavukcuoglu, K., Silver, D., Graves, A., Antonoglou, I., Wierstra, D., & Riedmiller, M. (2013). Playing atari with deep reinforcement learning. arXiv preprint arXiv:1312.5602.

[5] Laval, J. A., & Zhou, H. (2019). Large-scale traffic signal control using machine learning: some traffic flow considerations. arXiv preprint arXiv:1908.02673.

[6] Daganzo, C. F., Gayah, V. V., & Gonzales, E. J. (2011). Macroscopic relations of urban traffic variables: Bifurcations, multivaluedness and instability. Transportation Research Part B: Methodological, 45(1), 278-288.

[7] Geroliminis, N., & Daganzo, C. F. (2008). Existence of urban-scale macroscopic fundamental diagrams: Some experimental findings. Transportation Research Part B: Methodological, 42(9), 759-770.

[8] Redmond, L. S., & Mokhtarian, P. L. (2001). The positive utility of the commute: modeling ideal commute time and relative desired commute amount. Transportation, 28, 179-205.

[9] Buisson, C., & Ladier, C. (2009). Exploring the impact of homogeneity of traffic measurements on the existence of macroscopic fundamental diagrams. Transportation Research Record, 2124(1), 127-136.

[10] Lighthill, M. J., & Whitham, G. B. (1955). On kinematic waves II. A theory of traffic flow on long crowded roads. Proceedings of the Royal Society of London. Series A. Mathematical and Physical Sciences, 229(1178), 317-345.

[11] Richards, P. I. (1956). Shock waves on the highway. Operations research, 4(1), 42-51.

[12] Daganzo, C. F., & Geroliminis, N. (2008). An analytical approximation for the macroscopic fundamental diagram of urban traffic. Transportation Research Part B: Methodological, 42(9), 771-781.

[13] Laval, J. A., & Castrillón, F. (2015). Stochastic approximations for the macroscopic fundamental diagram of urban networks. Transportation Research Procedia, 7, 615-630.

[14] Mahmassani, H. S., Saberi, M., & Zockaie, A. (2013). Urban network gridlock: Theory, characteristics, and dynamics. Procedia-Social and Behavioral Sciences, 80, 79-98.

[15] Jin, W. L., Gan, Q. J., & Gayah, V. V. (2013). A kinematic wave approach to traffic statics and dynamics in a double-ring network. Transportation Research Part B: Methodological, 57, 114-131.

[16] Gan, Q. J., Jin, W. L., & Gayah, V. V. (2017). Analysis of traffic statics and dynamics in signalized networks: a poincaré map approach. Transportation science, 51(3), 1009-1029.

[17] Ambühl, L., Loder, A., Menendez, M., & Axhausen, K. W. (2017). Empirical macroscopic fundamental diagrams: New insights from loop detector and floating car data. In TRB 96th Annual Meeting Compendium of Reports (pp. 17-03331). Transportation Research Board.

[18] Shim, J., Yeo, J., Lee, S., Hamdar, S. H., & Jang, K. (2019). Empirical evaluation of influential factors on bifurcation in macroscopic fundamental diagrams. Transportation Research Part C: Emerging Technologies, 102, 509-520.
[19] Sutton, R. S., & Barto, A. G. (2018). Reinforcement learning: An introduction. MIT press.
[20] Tsitsiklis, J., & Van Roy, B. (1996). Analysis of temporal-diffference learning with function approximation. Advances in neural information processing systems, 9.
[21] Riedmiller, M. (2005). Neural fitted Q iteration–first experiences with a data efficient neural reinforcement learning method. In Machine Learning: ECML 2005: 16th European Conference on Machine Learning, Porto, Portugal, October 3-7, 2005. Proceedings 16 (pp. 317-328). Springer Berlin Heidelberg.
[22] Lange, S., & Riedmiller, M. (2010, July). Deep auto-encoder neural networks in reinforcement learning. In The 2010 international joint conference on neural networks (IJCNN) (pp. 1-8). IEEE.
[23] Abtahi, F., & Fasel, I. (2011). Deep belief nets as function approximators for reinforcement learning. RBM, 2, h3.
[24] Lange, S., Riedmiller, M., & Voigtländer, A. (2012, June). Autonomous reinforcement learning on raw visual input data in a real world application. In The 2012 international joint conference on neural networks (IJCNN) (pp. 1-8). IEEE.
[25] Koutnik, J., Schmidhuber, J., & Gomez, F. (2014). Online evolution of deep convolutional network for vision-based reinforcement learning. In From Animals to Animats 13: 13th International Conference on Simulation of Adaptive Behavior, SAB 2014, Castellón, Spain, July 22-25, 2014. Proceedings 13 (pp. 260-269). Springer International Publishing.
[26] Casas, N. (2017). Deep deterministic policy gradient for urban traffic light control. arXiv preprint arXiv:1703.09035.
[27] Coşkun, M., Baggag, A., & Chawla, S. (2018, November). Deep reinforcement learning for traffic light optimization. In 2018 IEEE International Conference on Data Mining Workshops (ICDMW) (pp. 564-571). IEEE.
[28] Chu, T., Wang, J., Codecà, L., & Li, Z. (2019). Multi-agent deep reinforcement learning for large-scale traffic signal control. IEEE Transactions on Intelligent Transportation Systems, 21(3), 1086-1095.
[29] Lin, Y., Dai, X., Li, L., & Wang, F. Y. (2018). An efficient deep reinforcement learning model for urban traffic control. arXiv preprint arXiv:1808.01876.
[30] Shi, S., & Chen, F. (2018). Deep recurrent Q-learning method for area traffic coordination control. Journal of Advances in Mathematics and Computer Science, 27(3), 1-11.
[31] Bello, I., Pham, H., Le, Q. V., Norouzi, M., & Bengio, S. (2016). Neural combinatorial optimization with reinforcement learning. arXiv preprint arXiv:1611.09940.
[32] Khalil, E., Dai, H., Zhang, Y., Dilkina, B., & Song, L. (2017). Learning combinatorial optimization algorithms over graphs. Advances in neural information processing systems, 30.
[33] Kool, W., Van Hoof, H., & Welling, M. (2018). Attention, learn to solve routing problems!. arXiv preprint arXiv:1803.08475.
[34] Nazari, M., Oroojlooy, A., Snyder, L., & Takác, M. (2018). Reinforcement learning for solving the vehicle routing problem. Advances in neural information processing systems, 31.

[35] Balaji, B., Bell-Masterson, J., Bilgin, E., Damianou, A., Garcia, P. M., Jain, A., ... & Ye, C. (2019). Orl: Reinforcement learning benchmarks for online stochastic optimization problems. arXiv preprint arXiv:1911.10641.

[36] Kullman, N. D., Mendoza, J. E., Cousineau, M., & Goodson, J. C. (2019). Atari-fying the Vehicle Routing Problem with Stochastic Service Requests. arXiv preprint arXiv:1911.05922.

[37] Zhao, J., Mao, M., Zhao, X., & Zou, J. (2020). A hybrid of deep reinforcement learning and local search for the vehicle routing problems. IEEE Transactions on Intelligent Transportation Systems, 22(11), 7208-7218.

[38] Zhang, K., He, F., Zhang, Z., Lin, X., & Li, M. (2020). Multi-vehicle routing problems with soft time windows: A multi-agent reinforcement learning approach. Transportation Research Part C: Emerging Technologies, 121, 102861.

[39] Peng, B., Wang, J., & Zhang, Z. (2020). A deep reinforcement learning algorithm using dynamic attention model for vehicle routing problems. In Artificial Intelligence Algorithms and Applications: 11th International Symposium, ISICA 2019, Guangzhou, China, November 16–17, 2019, Revised Selected Reports 11 (pp. 636-650). Springer Singapore.

[40] James, J. Q., Yu, W., & Gu, J. (2019). Online vehicle routing with neural combinatorial optimization and deep reinforcement learning. IEEE Transactions on Intelligent Transportation Systems, 20(10), 3806-3817.

[41] Chen, I. M., Zhao, C., & Chan, C. Y. (2019, October). A deep reinforcement learning-based approach to intelligent powertrain control for automated vehicles. In 2019 IEEE Intelligent Transportation Systems Conference (ITSC) (pp. 2620-2625). IEEE.

[42] Mushtaq, A., Haq, I. U., Imtiaz, M. U., Khan, A., & Shafiq, O. (2021). Traffic flow management of autonomous vehicles using deep reinforcement learning and smart rerouting. IEEE Access, 9, 51005-51019.

[43] Mushtaq, A., Sarwar, M. A., Khan, A., & Shafiq, O. (2022). Traffic Management of Autonomous Vehicles using Policy Based Deep Reinforcement Learning and Intelligent Routing. arXiv preprint arXiv:2206.14608.