2022 17th International Conference on
Control, Automation, Robotics and Vision (ICARCV)
December 11-13, 2022. Singapore

2022 17th International Conference on Control, Automation, Robotics and Vision (ICARCV) | 978-1-6654-7687-4/22/$31.00 ©2022 IEEE | DOI: 10.1109/ICARCV57592.2022.10004227

Reinforcement Learning method with Dynamic Learning Rate for Real-Time Route Guidance based on SUMO

Yuzhen Li[1], Jiawen Tang[2], Han Zhao[1] and Ruikang Luo[1]

***Abstract*— The increasing number of vehicles and dynamic changes in traffic situations make real-time route planning strongly necessary. The route-guiding method is supposed to cope with dynamic traffic situations. In addition, the ability to adapt to the second fastest route is very important when traffic congestion suddenly occurs on the fastest path. This paper proposes a method of using reinforcement learning to solve dynamic route planning problems, and the adaptation from a static learning rate to a dynamic learning rate enhances the capability to deal with emergent congestion. Meanwhile, the waiting time before each traffic light also is considered as a reward factor in the proposed algorithm. Contrast experiments have been conducted on the simulation network by SUMO, which has demonstrated well that our proposed method has better performance than other methods.**

I. INTRODUCTION

Considering economic, comfort and safety reasons, the autonomous car developed rapidly, which is able to sense their environment and move with litter or no human input [1] [2]. Ideally, fully autonomous vehicles are capable of moving from place A to place B without human interference, regardless of how complicated are the severe weather and traffic situations. A key challenge for the autonomous car is efficient and dynamic route planning, and many unpredictable factors cause traffic congestion, usually followed by a series of issues like traffic collision, environmental disruption, etc. As the report shows from the Texas Traffic Institute [3], 4.2 billion hours, 2.9 billion gallons of fuel, and a total cost of 78 billion dollars were wasted by US drivers in 2005, owing to traffic delays. Nowadays, an exponential increase of vehicles in use aggravates the waste. Apart from these problems, previous papers [4]–[6] discussed many other issues caused by congestion comprehensively. Many predictable and unpredictable factors will cause traffic congestion such as driver behaviors, weather conditions, machine breakdown, etc. Several researchers are focusing on the traffic prediction problem to predict the short-term traffic flow in the future [7]–[13]. However, traffic congestion is hard to be prevented and traffic prediction is not sufficient enough to alleviate traffic congestion. Besides traffic prediction, choosing an alternative route is an efficient solution when congestion occurs. The navigating method plays an important role in sudden congestion, but the optional route is hard to plan if vehicles cannot get access to real-time data. Nowadays, many countries support the development of intelligent transportation from economy to policy, which motivates route guidance technologies. With the advancements in 5th Generation Mobile Communication Technology (5G) and sensor technologies, the highly-frequent and real-time data exchange is achieved between vehicles, pedestrians, networks and infrastructure. Based on 5G and Vehicular-to-everything methods, real-time information, like mean speed and halting vehicles can be updated frequently. There are two main directions for automatic vehicle development. Firstly, Some companies recommend that every car can solely achieve autonomous driving. In this way, each car only relies on the internal sensors to perceive the information about the surrounding environment, which means that the real-time information about traffic hardly shares with each car and traffic center, thereby increasing the difficulty of scheduling the route quickly and accurately. Secondly, others advocate the intelligent transportation system (ITS), also named intelligent Road-Vehicle Collaboration. ITS analyze the information from not only internal sensors but also these data from external sensors or equipment, like those installed on the roadside or used by pedestrians. It is imperative and significant to construct the digital infrastructure, which is the key to increasing real-time data sharing. This new traffic model will rebuild our driving experience through the

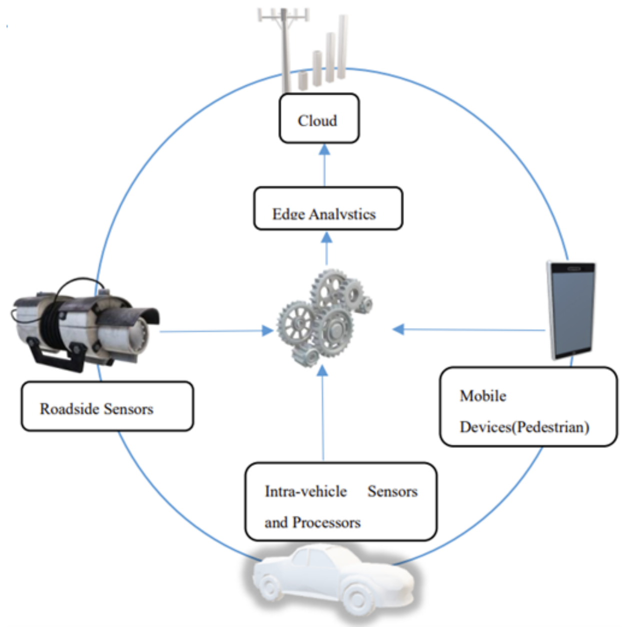

Fig. 1. The Architecture of Intelligent Transportation System

[1]School of Electrical and Electronic Engineering, Nanyang Technological University, Singapore. E-mail: yli083@e.ntu.edu.sg
[2]School of Physics Electronic Science, Changsha University of Science Technology E-mail: tang.lucy@foxmail.com

data communication between people, vehicles, roads, and environments [14]. Figure 1 depicts the overall architecture of the intelligent transportation system.

II. Related Work

Conventional graph-based methods, like Dijkstra and the A* algorithm, perform well in solving route-planning problems. Dijkstra algorithm is a basic greedy algorithm. It gradually decides the next action and also guarantees that each step is the shortest. Before going next node, this algorithm will check whether the distance between the arrived node and other remaining nodes is the shortest. The time complexity of the Dijkstra method will exponentially rise with the number of junctions and roads so its performance on large-scale networks is not acceptable. Compared with the Dijkstra algorithm, A* algorithm combines the idea of the greedy algorithm with some plans for future development, which means that A* not only considers the distance between the beginning node and the arrived node but also the distance between the destination and the arrived node. After analyzing these two algorithms, they are more suitable for the situation with static route-planning problems and small road networks. Although paper [6] adapts the A* algorithm for dynamic problems, the computation time increases in turn and this algorithm needs to satisfy FIFO (first-in-first-out) property, which is hard to achieve in real situations. In addition, many proposed papers improved the A* algorithm, like Block A* [15] and D* [16], which has been successfully applied to automatic vehicle guidance and achieved remarkable performance. However, drivers prefer to choose the route with less time, instead of a shorter path. Path planning according to travelling time is a kind of dynamic problem, and those traditional methods are only good at static issues, like distance-based problems. Therefore, we put attention to reinforcement learning for time-based route planning.

With the rapid development of reinforcement learning and associated infrastructure, papers [17] [18] proposed methods to plan the fastest route based on data provided by vehicles, detect devices, etc. Paper [19] applied the Q-learning algorithm to the shortest path planning based on Simulation of Urban Mobility (SUMO). Paper [20] proposed a method for multi-agent to find the shortest path. Paper [21] proposed a deep reinforcement learning method to solve EV dispatching problem. However, these proposed methods ignore the traffic light information and also are hard to offer an optional route when a traffic jam suddenly occurs in the fastest path. Both traffic lights and sudden congestion have significant impacts on deciding the fastest path. Although some method is suitable for the dynamic situation which combines improved A* and reinforcement learning to handle multi-agent route guidance, it still ignores the effect of traffic lights. Unlike the existing reinforcement learning method, the proposed method uses a dynamic learning rate which achieves a better performance on normal path planning and plans the second fastest path when sudden congestion occurs. For basic reinforcement learning methods (Q-learning and Sarsa), a Q-table is used to store the reward for doing different actions at different states. The dimension of the Q-table will increase with the size of networks, and the traditional method failed in large-scale problems. The deep Q network (DQN) method was proposed by Mnih for high dimensional problems, which combines deep neural network and reinforcement learning [22]. After the proposal of DQN, many improved methods were proposed [23] [24] [25], such as double DQN, Dueling DQN and distributional DQN, but these methods hardly make an optimal decision when they face a situation not included in training datasets. Compared with other route-planning methods, the proposed method has a better performance and the theorem is more reasonable. Firstly, it considers waiting time at traffic lights as a factor during planning routes. Besides the ability to plan a route under normal situations, it is also good at planning routes when traffic congestion suddenly appears.

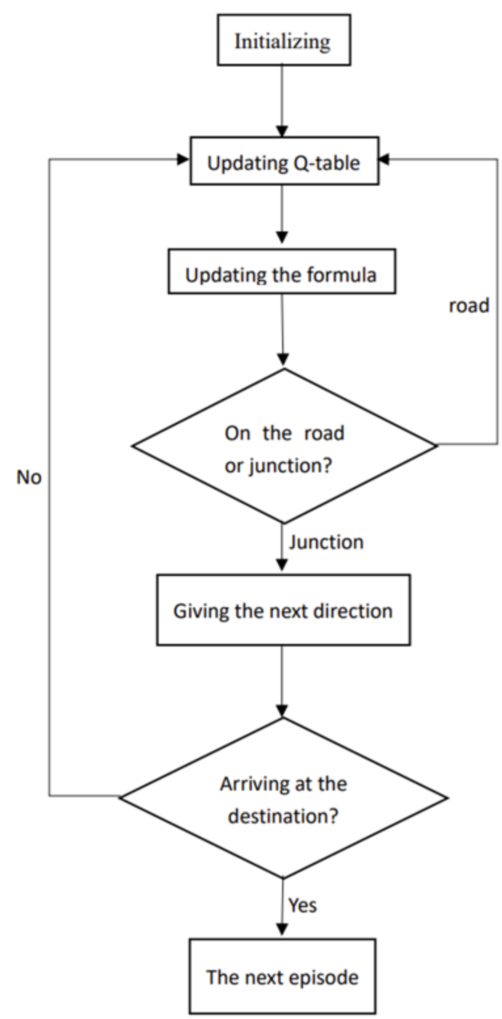

Fig. 2. The Flow Chart of RL updating process in real-time route guidance problem

III. Methodology

Paper [11] assumes that there is a traffic center on each junction responsible for measuring real-time information on

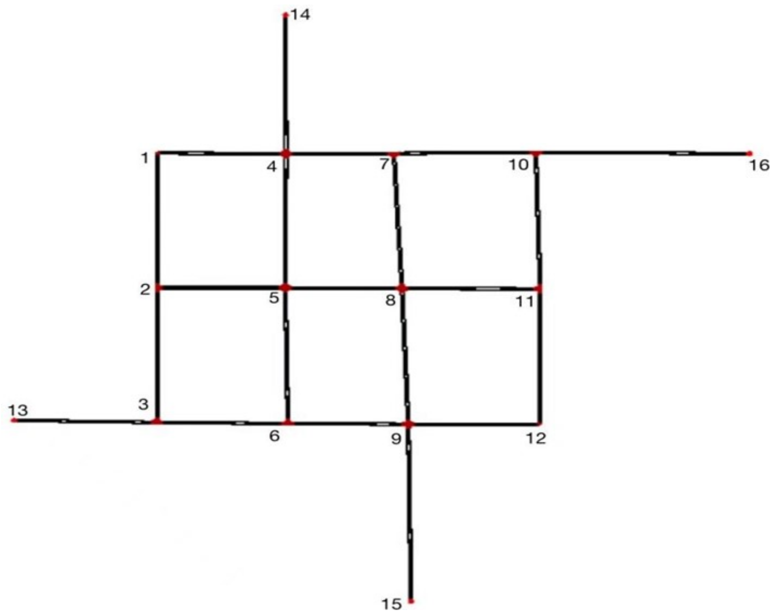

Fig. 3. The topological structure of the simulation network

neighboring roads and updating the measured data every second. In our problem, Figure 4 shows the modelling of the traffic center in each junction. Take traffic center i as an example, it will detect the mean travelling time of road from i to p and road from i to j. After measuring the mean traveling time, it will update the information every second and each car can use real-time information. The specific steps of the proposed reinforcement learning method are proposed below and the flow chart is shown in Figure 2:

We initialize all Q-values in the Q-table to 0, dynamic learning rate, discount factor ($\gamma = 1$) and greed policy rate ($\epsilon = 0.6$). The Conventional method set the learning rate to 0.6. The following steps are repeated until the destination episode: Firstly, we measure the mean travelling time of each neighboring road and update the real-time data. We use the following formula to update the Q-table every second based on the real-time data from the previous step.

$$Q(i,j) \leftarrow Q(i,j) + \alpha \left[t_{ij} + \gamma \min \left[Q(i,k) + t_{light}(j,k) \right] - Q(i,j) \right] \quad (1)$$

$$\alpha = \frac{max(abs(t_{last} - t_{travel})}{max(t_{last}, t_{travel}), 0.6)} \quad (2)$$

In Equation 1 and 2, t_{last} is the mean travelling time of the updated road section on the last time step. t_{travel} is the mean travelling time of the updated road section on the current time step and $t_{light}(j,k)$ represents the waiting time in junction j when the vehicle comes from junction i and assumes driving towards junction k. t_{ij} is the travel time of junction i to junction j. The greedy policy rate ϵ is minus by 0.02 for each step until it equals 0. If the vehicle closes to a junction, the corresponding traffic center gives the next direction according to the latest Q-table. In contrast, this step should pass when the vehicle still drives along the road. At last, if the vehicle arrives at the destination node, this episode is terminated. Different from the conventional reinforcement learning method, this proposed method uses a dynamic learning rate and also takes the waiting time before the signalized junction into account.

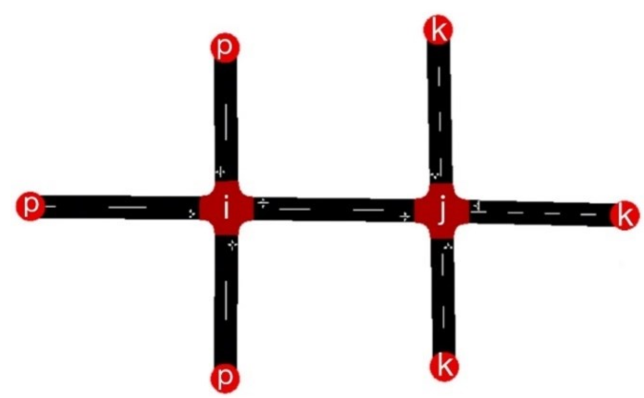

Fig. 4. A diagram of traffic center model, each junction is regarded as a traffic center

IV. EXPERIMENTS

The experiment simulates through the Simulation of Urban Mobility (SUMO) application. In this part, we will compare the total travelling time from the original point to the destination point of two methods (static and dynamic learning rate) in normal situations and traffic congestion situations. Experiments are carried out on the above network and vehicle flow is randomly produced. This network totally contains 16 junctions, and a vehicle tries to find the fastest path from original node 13 to destination node 16. In order to test these algorithms under emergent traffic congestion, setting the limit speed of one section on the fastest path to 0.5m/s imitates the traffic jam. In addition, the vehicle departure time from junction 13 sets to 100 seconds, which guarantees that the car flows in the simulation drive stably. The comparison explicitly shows the proposed method has better performance in normal and congested situations. Method 1 and method 2 are based on papers [14] [16].Table I shows the specific information about the simulation road network.

TABLE I
LINK NUMBER AND LINK LENGTH IN THE SIMULATION ROAD NETWORK

Link No.	Link length (m)	Link No.	Link length (m)
Link 1-2	200.00	Link 6-9	141.00
Link 1-4	150.87	Link 7-8	200.79
Link 2-3	150.87	Link 7-10	166.33
Link 2-5	150.87	Link 8-9	201.93
Link 3-6	150.87	Link 8-11	161.61
Link 3-13	167.43	Link 9-12	153.97
Link 4-5	199.11	Link 9-15	263.10
Link 4-7	126.23	Link 10-11	202.80
Link 4-14	206.22	Link 10-16	250.01
Link 5-6	201.79	Link 11-12	201.32
Link 5-8	136.60		

A. *Experiment result under normal traffic condition*

As shown in Figure 6, the episodes used to converge are 7 iterations and 1 iteration, respectively for method 2 and the proposed method. The fastest path and total travelling time calculated by these two methods are the same. The fastest path is junction 13-3-2-1-4-7-10-16, and the traveling time takes 53 seconds. In conclusion, the proposed method performs best in normal traffic situations. Figure 5 shows the result of method 1, and it needs more iterations to converge compared with the other two methods.

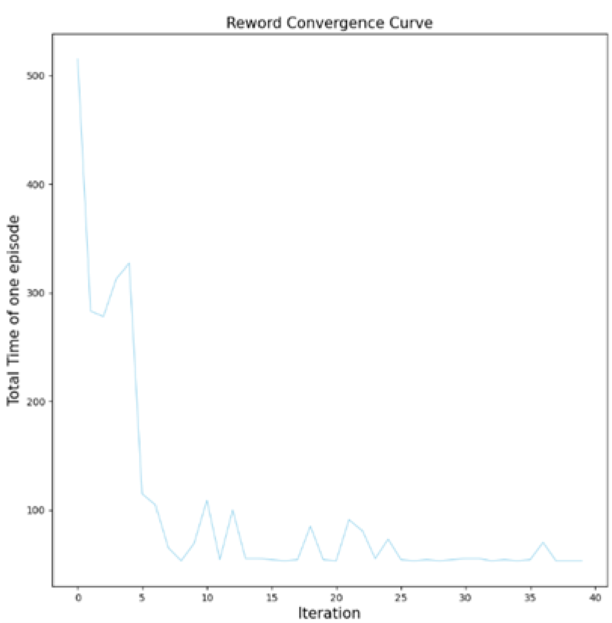

Fig. 5. Experimental Result of the traveling time with method 1

B. *Experiment result of the traffic congestion situation*

To produce sudden congestion conditions, both of these two methods will set the limit speed of 0.5 m/s on the

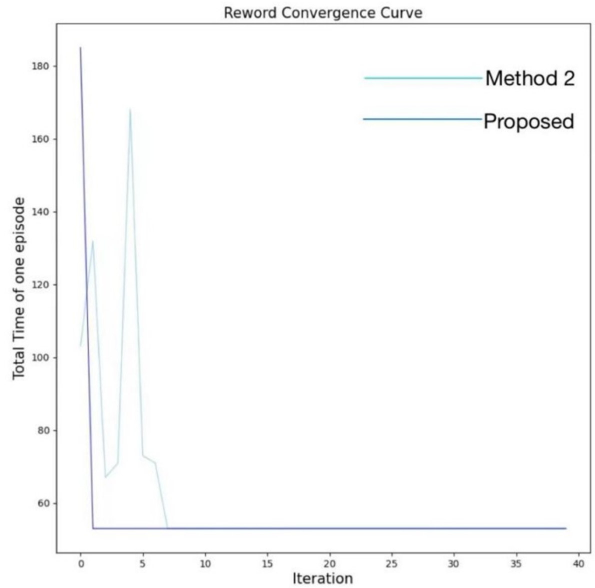

Fig. 6. Simulation Result of Normal Situation

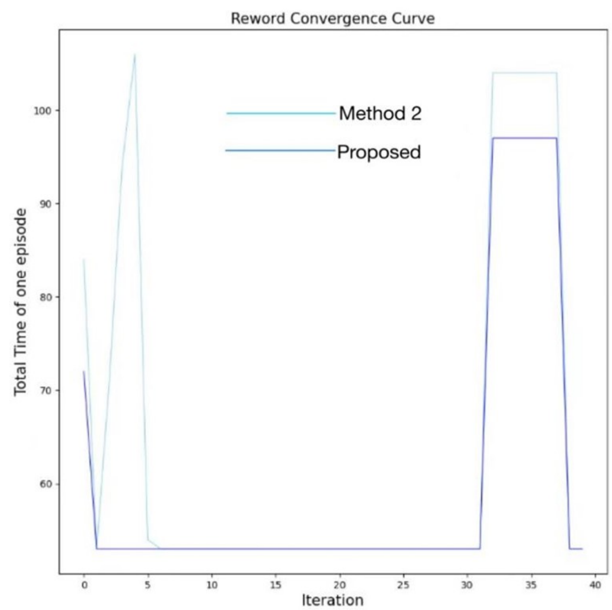

Fig. 7. Simulation Result of Sudden Traffic Congestion

section from junction 4 to junction 7. The congestion is set at episodes 32 to 37, and it occurs at time step 130 seconds which is the time when the vehicle arrives at junction 4. From Figure 7, both method 2 and the proposed method immediately choose another path as an option when traffic congestion occurs, but method 2 cannot find the optimal path that needs over 100 seconds. The proposed method not only converges into an optional path but also picks up the second-fastest route. Figure 8 shows the result of method 1 under sudden traffic congestion. From these results, method 1 costs more iterations to converge and it also cannot find out the second shortest path.

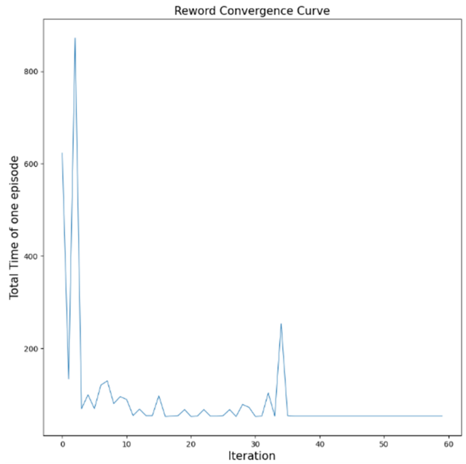

Fig. 8. Experimental Result of Method 1 in traffic congestion condition

V. CONCLUSIONS

In this paper, an adaptive reinforcement learning method is proposed to solve the route guidance problem. Although conventional reinforcement learning method is capable of dealing with route-planning under many situations, the dynamic learning rate performs better from converging time to handling sudden emergent traffic jam. Many papers also use deep reinforcement learning to plan route, but these methods cannot use the real-time data to choose a route, which hardly satisfy the actual requirements. Besides, a new parameter added in conventional equation to consider the waiting time caused by traffic light and this proposed method has a good performance based experimental results. In the future, this method should be tested under large scale network.

ACKNOWLEDGMENT

This research is supported by A*STAR under its RIE2020 Advanced Manufacturing and Engineering (AME) Industry Alignment Fund C Pre-Positioning (IAF-PP) (Award A19D6a0053).

REFERENCES

[1] R. Schmidt, H. Weisser, P. Schulenberg, and H. Goellinger, "Autonomous driving on vehicle test tracks: overview, implementation and results," in *Proceedings of the IEEE Intelligent Vehicles Symposium 2000 (Cat. No. 00TH8511)*. IEEE, 2000, pp. 152–155.

[2] J. Hu, P. Bhowmick, F. Arvin, A. Lanzon, and B. Lennox, "Cooperative control of heterogeneous connected vehicle platoons: An adaptive leader-following approach," *IEEE Robotics and Automation Letters*, vol. 5, no. 2, pp. 977–984, 2020.

[3] F. M. Padron, *Traffic congestion detection using VANET*. Florida Atlantic University, 2009.

[4] T. Taleb, A. Benslimane, and K. B. Letaief, "Toward an effective risk-conscious and collaborative vehicular collision avoidance system," *IEEE Transactions on Vehicular Technology*, vol. 59, no. 3, pp. 1474–1486, 2010.

[5] W. Wu, J. Zhang, A. Luo, and J. Cao, "Distributed mutual exclusion algorithms for intersection traffic control," *IEEE Transactions on Parallel and Distributed Systems*, vol. 26, no. 1, pp. 65–74, 2014.

[6] I. Chabini and S. Lan, "Adaptations of the a* algorithm for the computation of fastest paths in deterministic discrete-time dynamic networks," *IEEE Transactions on intelligent transportation systems*, vol. 3, no. 1, pp. 60–74, 2002.

[7] L. Huang, S. Zhao, R. Luo, R. Su, M. Sindhwani, S. K. Chan, and G. R. Dhinesh, "An incremental map matching approach with speed estimation constraints for high sampling rate vehicle trajectories," in *2022 IEEE 17th International Conference on Control & Automation (ICCA)*. IEEE, 2022, pp. 758–765.

[8] R. Luo and R. Su, "Traffic signal transition time prediction based on aerial captures during peak hours," in *2020 16th International Conference on Control, Automation, Robotics and Vision (ICARCV)*. IEEE, 2020, pp. 104–110.

[9] R. Luo, Y. Zhang, Y. Zhou, H. Chen, L. Yang, J. Yang, and R. Su, "Deep learning approach for long-term prediction of electric vehicle (ev) charging station availability," in *2021 IEEE International Intelligent Transportation Systems Conference (ITSC)*. IEEE, 2021, pp. 3334–3339.

[10] R. Luo, Y. Song, H. Zhao, Y. Zhang, Y. Zhang, N. Zhao, L. Huang, and R. Su, "Dense-tnt: Efficient vehicle type classification neural network using satellite imagery," *arXiv preprint arXiv:2209.13500*, 2022.

[11] R. Luo, Y. Song, L. Huang, Y. Zhang, and R. Su, "Stgin: A spatial temporal graph-informer network for long sequence traffic speed forecasting," *arXiv preprint arXiv:2210.01799*, 2022.

[12] H. Zhao, H. Yang, Y. Wang, D. Wang, and R. Su, "Attention based graph bi-lstm networks for traffic forecasting," in *2020 IEEE 23rd International Conference on Intelligent Transportation Systems (ITSC)*. IEEE, 2020, pp. 1–6.

[13] H. Zhao, H. Yang, Y. Wang, R. Su, and D. Wang, "Domain-adversarial-based temporal graph convolutional network for traffic flow prediction problem," in *2021 IEEE International Intelligent Transportation Systems Conference (ITSC)*. IEEE, 2021, pp. 1365–1370.

[14] A. Ferdowsi, U. Challita, and W. Saad, "Deep learning for reliable mobile edge analytics in intelligent transportation systems: An overview," *ieee vehicular technology magazine*, vol. 14, no. 1, pp. 62–70, 2019.

[15] P. Yap, N. Burch, R. C. Holte, and J. Schaeffer, "Block a*: Database-driven search with applications in any-angle path-planning," in *Twenty-Fifth AAAI Conference on Artificial Intelligence*, 2011.

[16] A. Stentz, "Optimal and efficient path planning for partially known environments," in *Intelligent unmanned ground vehicles*. Springer, 1997, pp. 203–220.

[17] S. Yu, J. Zhou, B. Li, S. Mabu, and K. Hirasawa, "Q value-based dynamic programming with sarsa learning for real time route guidance in large scale road networks," in *The 2012 International Joint Conference on Neural Networks (IJCNN)*. IEEE, 2012, pp. 1–7.

[18] Z. Zhang and J.-M. Xu, "A dynamic route guidance arithmetic based on reinforcement learning," in *2005 International Conference on Machine Learning and Cybernetics*, vol. 6. IEEE, 2005, pp. 3607–3611.

[19] S. S. Koh, B. Zhou, P. Yang, Z. Yang, H. Fang, and J. Feng, "Reinforcement learning for vehicle route optimization in sumo," in *2018 IEEE 20th International Conference on High Performance Computing and Communications; IEEE 16th International Conference on Smart City; IEEE 4th International Conference on Data Science and Systems (HPCC/SmartCity/DSS)*. IEEE, 2018, pp. 1468–1473.

[20] M. Z. Arokhlo, A. Selamat, S. Z. M. Hashim, and M. H. Selamat, "Route guidance system using multi-agent reinforcement learning," in *2011 7th International Conference on Information Technology in Asia*. IEEE, 2011, pp. 1–5.

[21] Y. Song, H. Zhao, R. Luo, L. Huang, Y. Zhang, and R. Su, "A sumo framework for deep reinforcement learning experiments solving electric vehicle charging dispatching problem," *arXiv preprint arXiv:2209.02921*, 2022.

[22] V. Mnih, K. Kavukcuoglu, D. Silver, A. Graves, I. Antonoglou, D. Wierstra, and M. Riedmiller, "Playing atari with deep reinforcement learning," *arXiv preprint arXiv:1312.5602*, 2013.

[23] H. Van Hasselt, A. Guez, and D. Silver, "Deep reinforcement learning with double q-learning," in *Proceedings of the AAAI conference on artificial intelligence*, vol. 30, no. 1, 2016.

[24] Z. Wang, T. Schaul, M. Hessel, H. Hasselt, M. Lanctot, and N. Freitas, "Dueling network architectures for deep reinforcement learning," in *International conference on machine learning*. PMLR, 2016, pp. 1995–2003.

[25] W. Dabney, M. Rowland, M. Bellemare, and R. Munos, "Distributional reinforcement learning with quantile regression," in *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 32, no. 1, 2018.