J Inf Process Syst, Vol.21, No.1, pp.43~51, February 2025
https://doi.org/10.3745/JIPS.04.0334
ISSN 2092-805X (Electronic)

JOURNAL OF INFORMATION PROCESSING SYSTEMS JIPS

Optimizing Traffic Signal Control Using LLM-Driven Reward Weight Adjustment in Reinforcement Learning

Sujeong Choi[1] and Yujin Lim[2,*]

Abstract

With advancements in information technology, traffic signal control has become a crucial component of smart transportation systems, and research based on reinforcement learning (RL) for this purpose is being actively conducted. However, tuning the weights of a multi-objective reward function remains a challenging task. This paper proposes an algorithm that leverages a large language model (LLM) to dynamically adjust the weights of the RL reward function in real time, enabling efficient traffic signal control at intersections. We compare the performance of dynamic weight adjustment via LLM and evaluate the signal control efficiency of the proposed model under various weather conditions.

Keywords

Large Language Model, Reinforcement Learning, Traffic Signal Control

1. Introduction

Advancements in information technology are driving innovation in modern transportation systems. Smart transportation systems recognize IT as a core enabler for optimizing urban traffic flow, enabling safer and more efficient mobility. In particular, traffic signal control is critical to reducing traffic congestion and travel time, with a significant impact on traffic flow in complex urban environments.

As intelligent traffic signal control systems for intersections advance, various studies have applied fuzzy logic [1] and genetic algorithms [2]. However, these approaches often struggle to adapt to sudden traffic changes, such as unexpected volume increases or incidents, which may result in congestion and delays. This highlights a growing demand for signal control strategies capable of dynamically adapting to real-time traffic conditions. To address these challenges, this paper proposes a traffic signal control method based on reinforcement learning (RL). Additionally, we aim to improve the performance of signal control by leveraging a large language model (LLM) to analyze the intersection environment in real time and adjust the weights of the multi-objective reward function as needed. Building upon our previous work [3], this study expands the approach by implementing LLM-driven weight adjustment decisions, enabling real-time adjustments in response to dynamic traffic conditions.

This paper is organized as follows. Section 2 reviews related work on traffic signal control using RL and LLMs. Section 3 describes the dueling double deep q-network (D3QN) architecture, the Markov

Manuscript received November 13, 2024; first revision December 9, 2024; accepted December 11, 2024.
* **Corresponding Author:** Yujin Lim (yujin91@sookmyung.ac.kr)
[1] Dept. of IT Engineering, Sookmyung Women's University, Seoul, Korea (suzzang77@sookmyung.ac.kr)
[2] Division of Artificial Intelligence Engineering, Sookmyung Women's University, Seoul, Korea (yujin91@sookmyung.ac.kr)

decision process (MDP) design of the proposed model, and the LLM and prompt configurations applied. In Section 4, we compare the performance of traffic signal control under LLM-based weight adjustments and evaluate the efficiency of the proposed model under various weather conditions. Finally, Section 5 presents the conclusion and discusses directions for future research.

2. Related Work

Table 1 summarizes related studies and their limitations. First, [4] formulated the traffic signal control problem as an MDP and solved it using deep Q-network (DQN), emphasizing the impact of reward function weights and introducing a single-objective reward function. In [5], a traffic signal control model using deep deterministic policy gradient (DDPG) was developed for real-world intersections, evaluating performance metrics such as waiting time, queue length, and passing vehicles to determine the optimal reward function. These studies used single-objective reward functions, which limited their ability to account for multiple factors simultaneously, resulting in challenges in balancing performance. Notably, [4] highlighted the difficulty of setting appropriate weights for the reward components. To address these challenges, LLMs have been increasingly applied in traffic signal control to support refined decision-making under real-time conditions. In [6], light GPT was used to analyze intersection traffic flow and adjust signal timings. However, due to the limitations of LLMs in learning behavior-reward feedback, it is difficult for LLMs alone to learn an optimal policy. [7] integrated LLM feedback with proximal policy optimization (PPO)-based RL for traffic signal control. Nevertheless, the feedback generated by the LLM was not directly incorporated into policy updates, limiting the model's ability to optimize the policy effectively.

This study addresses these limitations by applying an LLM to D3QN-based traffic signal control, proposing a novel algorithm that dynamically adjusts the weights of a multi-objective reward function in real time. This approach aims to reflect the complex interactions within intersections and achieve efficient traffic signal control.

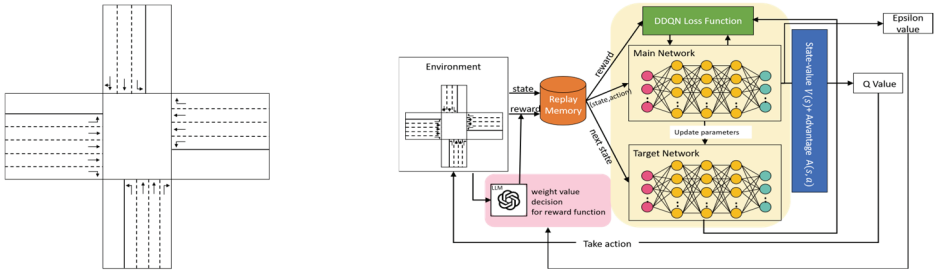
Table 1. Related works

Study	Algorithm	Action space	Explanations	Limits
Zheng et al. [4]	DQN	Discrete	Proposed MDP design for effective traffic signal control	Challenges in setting appropriate weights for reward function components Single-objective reward function
Lee et al. [5]	DDPG	Continuous	Definitions of three reward functions and their comparative analysis regarding waiting time, queue length, and throughput	Single-objective reward function
Lai et al. [6]	LLM	Adaptive	Study of LLM-based traffic signal control	Challenges in LLM's independent policy learning
Pang et al. [7]	PPO + LLM	Discrete + Adaptive	LLM feedback on RL agent actions and signal adjustments	Limitations of feedback on actions in improving a policy

3. System Model

3.1 D3QN-LLM Framework

In this study, a D3QN-based RL agent is deployed at a single intersection to observe real-time traffic conditions and control signals accordingly. Fig. 1(a) illustrates the single-intersection environment used in this study, where each lane supports straight, left-turn, and right-turn movements. The lane configuration follows that proposed in [7] and is designed to enhance the realism of the traffic signal control experiment by considering directional differences in traffic volume. The proposed model, D3QN-LLM, integrates D3QN with an LLM, and its framework is shown in Fig. 1(b). While previous studies have primarily used DQN or PPO algorithms, PPO often incurs higher overhead in discrete and simple action spaces. Consequently, this study employs D3QN, which mitigates the overestimation issues associated with DQN. Fig. 2 illustrates the process within an intersection environment.



(a) (b)

Fig. 1. (a) Single intersection and (b) D3QN-LLM framework.

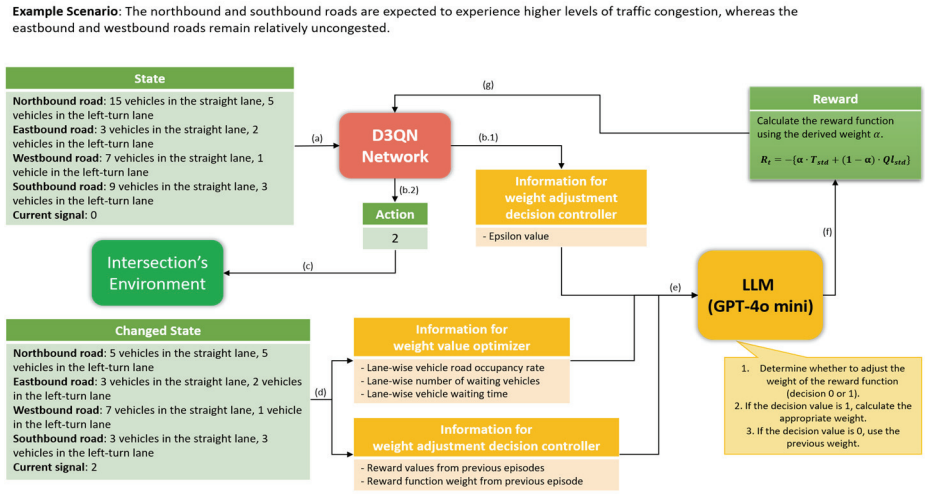


Fig. 2. Flow of the proposed methodology: (a) state information is used as input for (b.2) the D3QN network, which selects an action. This action is then reflected in (c) the intersection state, which is subsequently updated. Based on the updated intersection state, (d) information for weight adjustment is generated. The data derived during (b.1) the learning process of the D3QN network is passed to (e) the LLM to determine the appropriate weight for the reward function. Subsequently, (f) the reward function is calculated using the determined weight, and (g) it is used to update the Q-values. These Q-values indirectly optimize the policy responsible for selecting the optimal action. This process is repeated continuously.

3.2 MDP

This study proposes a D3QN-LLM algorithm to optimize traffic signal control at a single intersection, aiming to minimize average travel time. The definitions of state (S_t), action (A_t), and reward (R_t) used in this study are as follows.

The state S_t at time t represents the number of vehicles in each lane (straight and left-turn) for all directions (east, west, south, north) at the intersection. It also includes the currently assigned signal P_t for the intersection. This study focuses on controlling straight and left-turn lanes while excluding right-turn lanes, which typically have minimal impact on traffic flow. For the east direction, the number of vehicles in the straight lane is denoted by $N_{east\_s,t}$, and the number in the left-turn lane is denoted by $N_{east\_l,t}$. For the west direction, the number of vehicles in the straight lane is denoted by $N_{west\_s,t}$, and the number in the left-turn lane by $N_{west\_l,t}$. For the south direction, the number of vehicles in the straight lane is denoted by $N_{south\_s,t}$, and the number in the left-turn lane by $N_{south\_l,t}$. For the north direction, the number of vehicles in the straight lane is denoted by $N_{north\_s,t}$, and the number in the left-turn lane by $N_{north\_l,t}$. Therefore, the state S_t is defined as shown in Eq. (1):

$$S_t = \{N_{east\_s,t}, N_{east\_l,t}, N_{west\_s,t}, N_{west\_l,t}, N_{south\_s,t}, N_{south\_l,t}, N_{north\_s,t}, N_{north\_l,t}, P_t\}. \quad (1)$$

The agent considers the current state S_t to select and perform the appropriate action A_t. The action A_t represents the green signal assigned to the intersection at time t. Therefore, the action A_t is defined as shown in Eq. (2), and Fig. 3 illustrates the directions to which the green signal is assigned.

$$A_t = \{0,1,2,3\}. \quad (2)$$

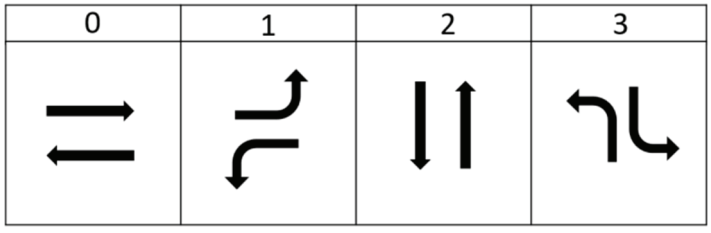


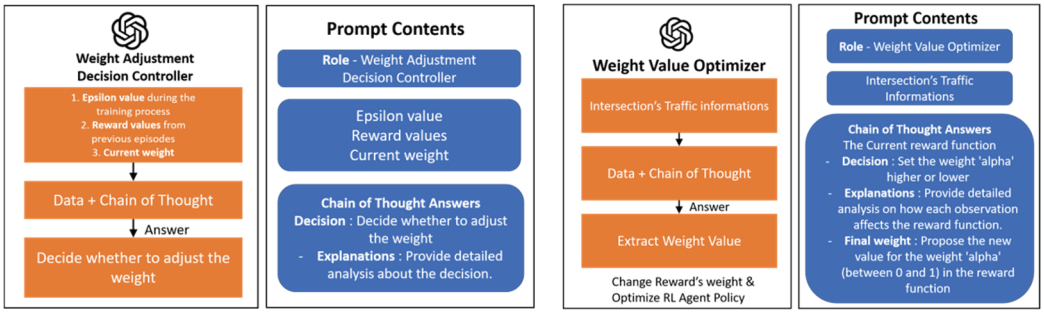
Fig. 3. Action space.

The agent receives a reward R_t after performing action A_t. This study aims to minimize average travel time at intersections. The reward function, defined in Eq. (3), is based on waiting time and queue length, which are key factors affecting travel time, as referenced in [4]. Waiting time refers to the duration a vehicle remains stopped at a red signal, and queue length indicates the number of stopped vehicles. These metrics are measured within a 3 km radius of the intersection to calculate travel time. To ensure fair signal distribution, standard deviations T_{std} and Ql_{std} are used, where T_{std} measures variations in waiting time and Ql_{std} measures variations in queue length by direction. These standard deviations help prevent signal bias in favor of specific directions. Additionally, α, a weight parameter ranging from 0 to 1, is dynamically adjusted according to traffic conditions.

$$R_t = -\{\alpha \cdot T_{\text{std}} + (1-\alpha) \cdot Ql_{\text{std}}\}. \quad (3)$$

3.3 LLM and Prompt Design

Effective learning requires an appropriate weight α, but manual tuning is complex and hinders the learning of an optimal policy. This study employs an LLM to assess the need for weight adjustment and derive the optimal weight based on post-action traffic conditions. Open AI's GPT-4o mini, a lightweight version of GPT-4o, was selected as the LLM for its balance between performance and efficiency. Fig. 4 presents the LLM prompt for weight adjustment, designed with reference to [7]. The weight adjustment decision controller determines the need for weight adjustment by analyzing the epsilon value and reward values from previous episodes. The epsilon value reflects the exploration-exploitation balance in D3QN, while past rewards evaluate the impact of weight adjustments. Based on this decision, the weight value optimizer analyzes traffic conditions to derive the optimal weight α.



(a) (b)

Fig. 4. (a) Weight adjustment decision controller and (b) weight value optimizer.

4. Performance Analysis

4.1 Experimental Setting

The intersection traffic dataset for the experiments was generated using Simulation of Urban Mobility (SUMO) [8], with performance metrics such as:

- Average travel time (ATT): The average time a vehicle takes to pass through the intersection, measuring overall efficiency;
- Average queue length (AQL): The average number of vehicles waiting at the intersection, indicating traffic congestion;
- Average queue length (AWT): The average time vehicles wait at traffic signals, emphasizing efforts to reduce driver delays.

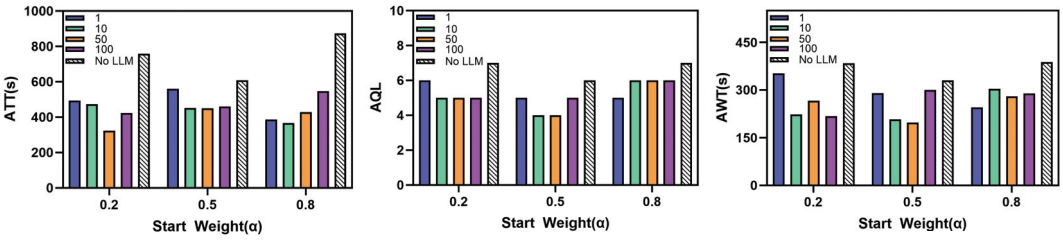
In the first experiment, we compared fixed-weight cases with those where the weights were dynamically adjusted by the LLM, under the assumption of periodic adjustment. In the second experiment, the LLM selectively adjusted the weights based on necessity, evaluating the effectiveness of weight adjustment decisions at specific episodes. In the third experiment, the experimental environment was divided into three weather conditions, as shown in Table 2, and SUMO parameters were configured for each condition [9] to compare the performance of the proposed model with existing methods.

Table 2. SUMO parameter settings based on weather

Weather of environment	Accel	Decel	eDecel	sDelay
Sunny	2.60	4.50	9.00	0.00
Rainy	0.75	3.50	6.00	0.25
Snowy	0.50	1.50	2.00	0.50

4.2 Experimental Result

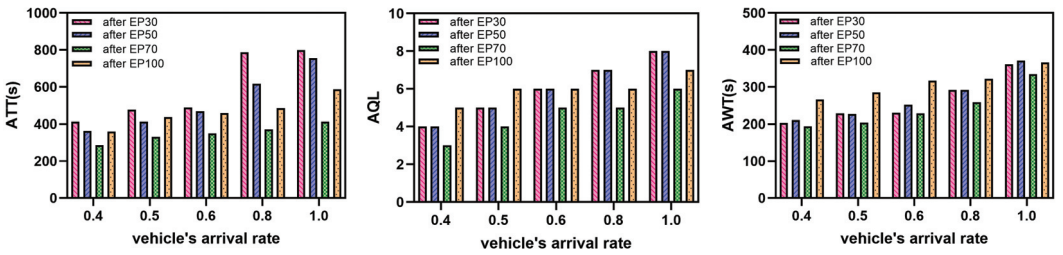
Fig. 5 illustrates the results of the first experiment. When the weight α was set to 0.5 and the adjustment interval was 10 or 50, stable performance was observed for ATT, AQL, and AWT. Specifically, when the initial weight was set to 0.5 and dynamically adjusted using the LLM, ATT improved by approximately 25%, AQL by 33%, and AWT by 40% compared to fixed weights. Overall, dynamic adjustment using the LLM significantly enhanced performance compared to fixed weights.



(a) (b) (c)

Fig. 5. Comparison of (a) ATT, (b) AQL, and (c) AWT performance with initial weight settings (0.2, 0.5, and 0.8): fixed weight (No LLM) vs. dynamic weight adjustment every 1, 10, 50, and 100 episodes.

Fig. 6 illustrates the results of the second experiment. Unlike the first experiment with fixed weight adjustment intervals, the LLM assessed the need for weight adjustment at each episode and adjusted the weights accordingly. Adjusting the weights using the LLM was found to be effective during episodes 30 to 100, where the exploration rate decreased and the exploitation rate increased. Overall, as vehicle arrival rates increased, ATT, AQL, and AWT rose. However, from episode 70 onward, these metrics remained relatively low due to dynamic weight adjustment. Based on these findings, the proposed model was developed.



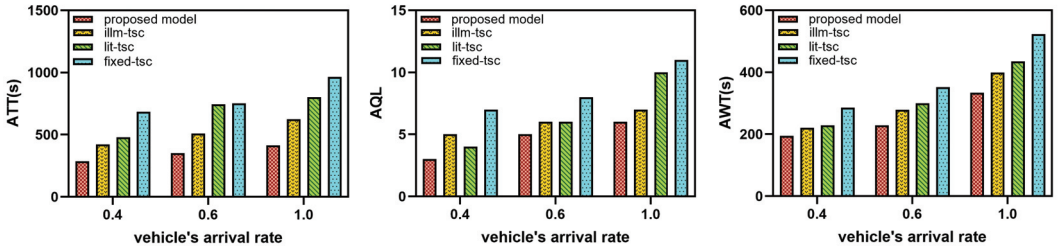
(a) (b) (c)

Fig. 6. Comparison of (a) ATT, (b) AQL, and (c) AWT performance based on vehicle arrival rate and weight adjustment initiation episode (after episode 30, 50, 70, and 100).

Figs. 7–9 illustrate the performance comparison between the proposed model and other methods under different vehicle arrival rates in sunny, rainy, and snowy weather conditions, respectively.

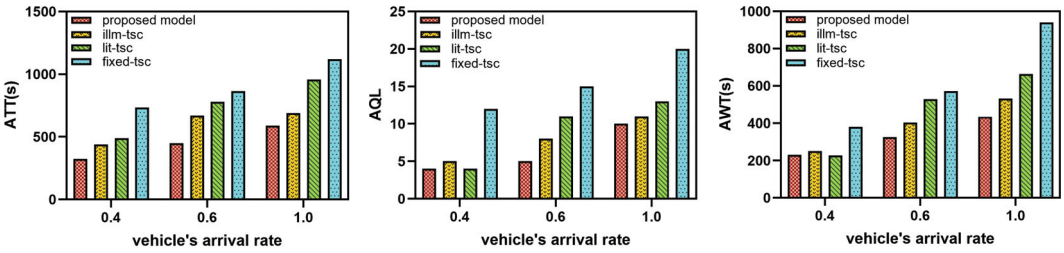
The other methods used for comparison are as follows.

- ILLM-TSC [7]: ILLM-TSC, a PPO-based traffic signal control model, minimizes vehicle waiting time with a single-objective reward function and guides the RL agent's actions through LLM feedback.
- LIT-TSC [4]: LIT-TSC is a DQN-based traffic signal control model that uses queue length in the reward function to minimize travel time.
- FIXED-TSC: FIXED-TSC is a traffic signal control model that uses a fixed sequence of green signals.



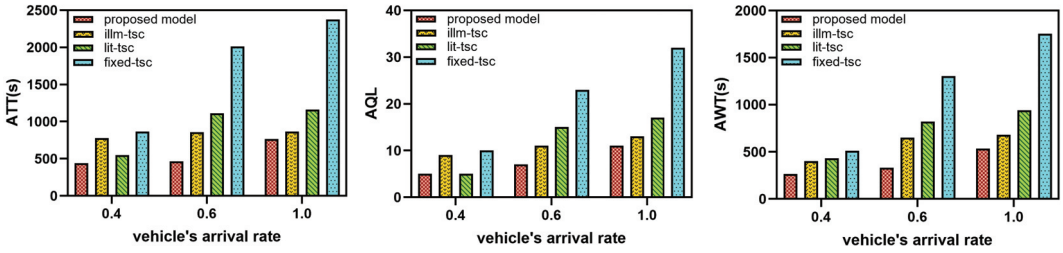
(a) (b) (c)

Fig. 7. Performance comparison of proposed model, ILLM-TSC, LIT-TSC, and FIXED-TSC in sunny weather: (a) ATT, (b) AQL, and (c) AWT by vehicle arrival rate.



(a) (b) (c)

Fig. 8. Performance comparison of proposed model, ILLM-TSC, LIT-TSC, and FIXED-TSC in rainy weather: (a) ATT, (b) AQL, and (c) AWT by vehicle arrival rate.



(a) (b) (c)

Fig. 9. Performance comparison of proposed model, ILLM-TSC, LIT-TSC, and FIXED-TSC in snowy weather: (a) ATT, (b) AQL, and (c) AWT by vehicle arrival rate.

Experimental results show that the proposed model outperformed comparative models, with improvements of 21%–29% over ILLM-TSC, 27%–40% over LIT-TSC, and 46%–56% over FIXED-TSC. As weather transitioned from sunny to snowy conditions, traffic control became more challenging. LIT-TSC demonstrated sensitivity to weather changes, while ILLM-TSC exhibited limitations under low vehicle arrival rates in snowy conditions. These findings emphasize the limitations of single-objective reward functions. In contrast, the proposed model maintained stable performance across varying conditions and demonstrated adaptability. Additionally, using an LLM to adjust reward weights was more effective than feedback-based action adjustments in fundamentally improving the RL policy.

5. Conclusion

This study proposes a D3QN-based traffic signal control algorithm that uses an LLM to dynamically adjust multi-objective reward weights based on traffic conditions in real time. Experimental results showed that LLM-driven weight adjustment enabled balanced optimization of queue length and waiting time or allowed a focus on specific objectives to ensure stable performance, demonstrating high adaptability. Future research will expand this approach to multi-intersection environments, where the LLM will analyze neighboring traffic conditions to prioritize RL-based signal control.

Conflict of Interest

The authors declare that they have no competing interests.

Funding

This work was supported by the Institute of Information & Communications Technology Planning & Evaluation–ICT Challenge and Advanced Network of HRD grant funded by the Korea government (Ministry of Science and ICT) (No. IITP-2025-RS-2022-00156299).

References

[1] B. P. Gokulan and D. Srinivasan, “Distributed geometric fuzzy multiagent urban traffic signal control,” *IEEE Transactions on Intelligent Transportation Systems*, vol. 11, no. 3, pp. 714-727, 2010. https://doi.org/10.1109/TITS.2010.2050688

[2] H. Ceylan and M. G. H. Bell, “Traffic signal timing optimisation based on genetic algorithm approach including drivers’ routing,” *Transportation Research Part B: Methodological*, vol. 38, no. 4, pp. 329-342, 2004. https://doi.org/10.1016/S0191-2615(03)00015-8

[3] S. Choi and Y. Lim, “Reinforcement learning-based traffic signal control using large language models,” *Annual Conference of KIPS,* vol. 31, no. 2, pp. 672-675, 2024. https://doi.org/10.3745/PKIPS.y2024m10a.672

[4] G. Zheng, X. Zang, N. Xu, H. Wei, Z. Yu, V. Gayah, K. Xu, and Z. Li, “Diagnosing reinforcement learning for traffic signal control,” 2019 [Online]. Available: https://arxiv.org/abs/1905.04716.

[5] H. Lee, Y. Han, Y. Kim, and Y. H. Kim, “Effects analysis of reward functions on reinforcement learning for traffic signal control,” *PLoS ONE*, vol. 17, no. 11, article no. e0277813, 2022. https://doi.org/10.1371/journal.pone.0277813

[6] S. Lai, Z. Xu, W. Zhang, H. Liu, and H. Xiong, “Large language models as traffic signal control agents: Capacity and opportunity,” 2023 [Online]. Available: https://arxiv.org/abs/2312.16044.

[7] A. Pang, M. Wang, M. O. Pun, C. S. Chen, and X. Xiong, “iLLM-TSC: Integration reinforcement learning and large language model for traffic signal control policy improvement,” 2024 [Online]. Available: https://arxiv.org/abs/2407.06025.

[8] P. A. Lopez, M. Behrisch, L. B. Walz, J. Erdmann, Y. P. Flotterod, R. Hilbrich, L. Lucken, J. Rummel, P. Wagner, and E. Wiessner, “Microscopic traffic simulation using SUMO,” in *Proceedings of the 2018 21st International Conference on Intelligent Transportation Systems (ITSC)*, Maui, HI, USA, 2018, pp. 2575-2582. https://doi.org/10.1109/ITSC.2018.8569938

[9] L. Da, M. Gao, H. Mei, and H. Wei, “Prompt to transfer: sim-to-real transfer for traffic signal control with prompt learning,” *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 38, pp. 82-90, 2024. https://doi.org/10.1609/aaai.v38i1.27758

Sujeong Choi **https://orcid.org/0009-0000-3559-5436**

She received her B.S. degree in IT Engineering from Sookmyung Women's University in 2020. Since September 2022, she has been a Master's student. Her current research interests include reinforcement learning and traffic signal control.

Yujin Lim **https://orcid.org/0000-0002-3076-8040**

She received B.S., M.S., and Ph.D. degrees in Computer Science from Sookmyung Women's University, Korea, in 1995, 1997 and 2000, respectively, and Ph.D. degree in Information Sciences from Tohoku University, Japan, in 2013. From 2004 to 2015, she was an associate professor in Department of Information Media, Suwon University, Korea. She joined the faculty of IT Engineering at Sookmyung Women's University, Seoul, in 2016, where currently she is a professor. Her research interests include edge computing, intelligent agent system, and artificial intelligence.