ENB-RL: Multi-Agent Reinforcement Learning with Explicit Neighbourhood Backtracking for Cooperative Traffic Signal Control

Yilong Ren[a,b,c], Yizhuo Chang[a,b], Zhiyong Cui[a,*], Xiao Chang[a], Han Jiang[a], Haiyang Yu[a,b,c], Yinhai Wang[d]

[a] <State Key Laboratory of Intelligent Transportation System, Beihang University>, <Beijing>, <China>
a.yilongren@buaa.edu.cn, b.chyzh1121@buaa.edu.cn, c.zhiyongc@buaa.edu.cn, f.hyyu@buaa.edu.cn
[b] <School of Transportation Science and Engineering, Beihang University>, <Beijing>, <China>
d.changxiao@buaa.edu.cn, e.buaajh@buaa.edu.cn
[c] <Zhongguancun Laboratory>, <Beijing>, <China>
[d] <Civil and Environmental Engineering, University of Washington>, <Seattle>, <USA>
g.yinhai@uw.edu
[*] Corresponding author

Extended abstract submitted for presentation at the Conference in Emerging Technologies in Transportation Systems (TRC-30)
September 02-03, 2024, Crete, Greece

July 12, 2024

Keywords: Cooperative traffic signal control, multi-agent reinforcement learning

1 INTRODUCTION

In recent years, urbanization and the rise in car ownership have significantly amplified traffic congestion, turning it into a pressing societal issue. The implementation of advanced Cooperative Traffic Signal Control (CTSC) strategies is generally recognized as an effective way to improve traffic efficiency and alleviate urban traffic congestion. Multi-agent reinforcement Learning (MARL) has been empirically demonstrated as a highly promising paradigm for the CTSC of urban road networks. However, some recent literature has found a counter-intuitive phenomenon that well-designed RL-based CTSC models are rather less effective than independent control of each intersection in the multi-intersection scenario (Mei *et al.*, 2023).

To explain this phenomenon, we analyze the source code of dozens of open-source RL-based ATSC methods. We note that mainstream traffic simulators often provide an API to easily obtain the traffic status of the entire entry lane. In such an ideal setting, the observability of the intersection agents has not changed and the commonly used POMDP still can be applied for modeling of the CTSC, the interaction between agents is substantially weakened. On this basis, we propose a two-stage hypothesis: firstly, the setup of surveillance zone length, an easily overlooked factor may fundamentally determine whether a MARL-based CTSC algorithm is effective or not; further, in multi-intersection scenarios, the interactions between intersections are time-lagged.

2 METHODOLOGY

Considering the incompleteness of the surveillance area in real scenarios, we propose **ENB-RL**, a MARL model containing an Explicit Neighborhood Backtracking (ENB) module, a Double Deep Q-Network (DDQN), and a noisy net. The entire architecture of our proposed ENB-RL is shown in Fig 1. The ENB module is the core of our paper , which consists of a neighborhood backtracking stack to store and update neighborhood intersections' historical throughput in a segmented weighted way and a multi-head attention model for spatio-temporal differentiated input. Such explicit and precise inputs can improve the agent's observations in incomplete perceptual environments. Meanwhile, the multi-head attention model can enhance the ability of the agent to mine the internal correlation of its input states, so that the agent can better refine the key input state information which is most helpful for decision-making, and improve the ability of the algorithm to adapt to different environments. The throughput released by the lane at the neighborhood intersection is defined as:

$$tp^{j}_{t,t-1} = \sum_{lane \in lane^{j}_{i,conn}} throughput_{lane,t-1} \tag{1}$$

The agent stack the neighbor throughput information within the backtracking range to obtain the final neighborhood backtracking stack.

$$backtrack^{i,j}_{t} = \left(tp^{j}_{t,t-1}, ..., tp^{j}_{t,t-k}, ..., tp^{j}_{t,t-K}\right), i \in neighbor_j, k = 1, 2, ..., K \tag{2}$$

Since the backtracking information in the neighborhood stack has different impacts on the current intelligence under different environments, we introduce the multi-attention mechanism to enhance the ability of the agent to improve the ability of the algorithm to adapt to different environments.

$$o^{back}_{t} = MultiAttention\left(o_t, backtrack^{i,j}_{t}\right) \tag{3}$$

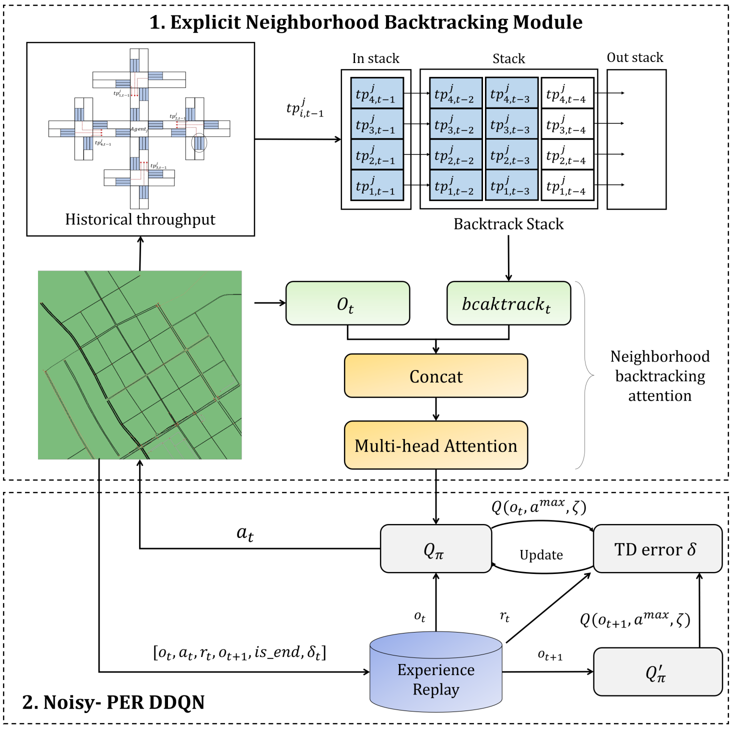
1. Explicit Neighborhood Backtracking Module
Historical throughput
$tp^{j}_{i,t-1}$
In stack
Stack
Out stack
$tp^{j}_{4,t-1}$ $tp^{j}_{4,t-2}$ $tp^{j}_{4,t-3}$ $tp^{j}_{4,t-4}$
$tp^{j}_{3,t-1}$ $tp^{j}_{3,t-2}$ $tp^{j}_{3,t-3}$ $tp^{j}_{3,t-4}$
$tp^{j}_{2,t-1}$ $tp^{j}_{2,t-2}$ $tp^{j}_{2,t-3}$ $tp^{j}_{2,t-4}$
$tp^{j}_{1,t-1}$ $tp^{j}_{1,t-2}$ $tp^{j}_{1,t-3}$ $tp^{j}_{1,t-4}$
Backtrack Stack
O_t
$bcaktrack_t$
Concat
Multi-head Attention
Neighborhood backtracking attention
a_t
$Q(o_t, a^{max}, \zeta)$
Q_π
Update
TD error δ
o_t
r_t
$Q(o_{t+1}, a^{max}, \zeta)$
$[o_t, a_t, r_t, o_{t+1}, is\_end, \delta_t]$
Experience Replay
o_{t+1}
Q'_π
2. Noisy- PER DDQN

Figure 1 – *Architecture of the proposed ENB-RL, which consists of an explicit neighborhood backtracking module, the PER DDQN and the Noisy-Net.*

Table 1 – *Performance on synthetic road network under different surveillance zone.*

Surveillance zone	50m		150m		300m	
Metrics	ATT	AWT	ATT	AWT	ATT	AWT
iDDQN	592.55	297.81	571.73	286.13	**540.52**	**271.64**
DDQN-Glo	691.48	387.21	727.11	435.93	780.56	523.59
DDQN-Nei	588.73	296.55	590.25	302.66	592.55	314.39
MDDQN	629.17	336.83	651.99	358.18	686.24	397.56
ENB-RL	**536.76**	**254.69**	**570.12**	**285.75**	593.22	314.45

Further, considering that historical backtracking information may lead to convergence instability, we introduce the Noisy-Net, a stochastic Gaussian noise-based exploration method to create uncertainties that allow the agent to choose different actions with probability, which can improve the efficiency and stability of exploration.

3 RESULTS

We perform comprehensive experiments on a synthetic grid 5x5 and a Jinan real-world road network with the Simulation of Urban Mobility (SUMO) platform to verify the authenticity, effectiveness, benefits, and generalizability of our method. We introduce three most effective and representative evaluation metrics to evaluate different methods. **Average travel time (ATT):** The average travel time for all vehicles from the time they enter the road network to the time they leave. **Average waiting time (AWT):** The average time that vehicles stop for traffic congestion and signal control, where the waiting is defined as the vehicle speed below $0.1 m/s$. **Best convergence episode (BCE):** It indicates the sampling efficiency of a DRL algorithm. For comparison purposes, we use base 5 integers.

Firstly, we examine the performance of various DDQN-based methods over different surveillance zone lengths to verify our proposed hypothesis. In both synthetic and real-world road networks, we will consider different radii of the surveillance zone (50 m, 150 m, 300 m). Experimental results are shown in Tab. 1, the length of surveillance zones does affect the performance of cooperative control models. Additionally, our method is superior to other cooperative control methods and independent control methods when the surveillance zone is relatively short.

To demonstrate the performance of the proposed ENB-RL, three conventional TSC methods and three state-of-the-art (SOTA) RL models: HD-RMPC (Ren *et al.*, 2022a), IntelliLight (Wei *et al.*, 2018), MAPPO (Luo *et al.*, 2024), MASAC (Mao *et al.*, 2022) are employed for comparison. Experimental results are shown in Tab. 2, ENB-RL has the best convergence performance on both synthetic and real-world datasets, and outperforms other SOTA models. As far as we know, there is no CTSC strategy has achieved better performance in Jinan road network than our own, including the latest RL-based methods (Liu *et al.*, 2023) and large language model-based methods (Lai *et al.*, 2023).

Considering that ENB-RL mainly consists of the ENB module, DDQN, and Noisy-Net, we construct ablation experiments to verify the effectiveness of each module and find that each module of our method makes sense in terms of performance or convergence speed. The results are shown in Tab. 3. Moreover, we individually embed the ENB module into three commonly used RL models for plug-and-play testing and find that it can work in other models.

Table 2 – *Performance of advanced ATSC algorithms on synthetic and real-world road network.*

Dataset	Grid 5 × 5			Jinan		
Metrics	ATT	AWT	BCE	ATT	AWT	BCE
Fixed-time	679.72	361.18	-	295.75	111.78	-
Max-pressure	662.23	355.29	-	280.42	111.11	-
HD-RMPC	613.84	303.48	-	235.87	68.84	-
IntelliLight	663.12	382.98	90	244.89	75.08	60
MAPPO	681.93	371.79	90	241.95	43.44	75
MASAC	597.25	275.20	80	247.90	77.58	120
ENB-RL	**536.76**	**254.69**	**75**	**201.52**	**35.44**	**55**

Table 3 – *Results of the ablation comparison.*

Dataset	Grid 5 × 5			Jinan		
Metrics	ATT	AWT	BCE	ATT	AWT	BCE
PER-DDQN	592.55	297.81	115	207.24	39.56	100
Noisy-PER-DDQN	592.35	298.86	80	207.35	40.21	70
ENB(MLP)+Noisy-PER-DDQN	569.82	283.94	80	205.61	39.33	70
ENB-RL	**536.76**	**254.69**	**75**	**201.52**	**35.44**	**60**

References

Lai, Siqi, Xu, Zhao, Zhang, Weijia, Liu, Hao, & Xiong, Hui. 2023. Large language models as traffic signal control agents: Capacity and opportunity. *arXiv preprint arXiv:2312.16044*.

Liu, Yiling, Luo, Guiyang, Yuan, Quan, Li, Jinglin, Jin, Lei, Chen, Bo, & Pan, Rui. 2023. GPLight: grouped multi-agent reinforcement learning for large-scale traffic signal control. *Pages 199–207 of: Proceedings of the Thirty-Second International Joint Conference on Artificial Intelligence.*

Luo, Haoqing, Bie, Yiming, & Jin, Sheng. 2024. Reinforcement Learning for Traffic Signal Control in Hybrid Action Space. *IEEE Transactions on Intelligent Transportation Systems*, 1–17.

Mao, Feng, Li, Zhiheng, Lin, Yilun, & Li, Li. 2022. Mastering arterial traffic signal control with multi-agent attention-based soft actor-critic model. *IEEE Transactions on Intelligent Transportation Systems*, **24**(3), 3129–3144.

Mei, Hao, Lei, Xiaoliang, Da, Longchao, Shi, Bin, & Wei, Hua. 2023. Libsignal: An open library for traffic signal control. *Machine Learning*, 1–37.

Ren, Yilong, Jiang, Han, Zhang, Le, Liu, Runkun, & Yu, Haiyang. 2022a. HD-RMPC: A Hierarchical Distributed and Robust Model Predictive Control Framework for Urban Traffic Signal Timing. *Journal of Advanced Transportation*, **2022**.

Ren, Yilong, Jiang, Han, Ji, Nan, & Yu, Haiyang. 2022b. TBSM: A traffic burst-sensitive model for short-term prediction under special events. *Knowledge-Based Systems*, **240**, 108120.

Song, Xiang Ben, Zhou, Bin, & Ma, Dongfang. 2024. Cooperative traffic signal control through a counterfactual multi-agent deep actor critic approach. *Transportation Research Part C: Emerging Technologies*, **160**, 104528.

Wei, Hua, Zheng, Guanjie, Yao, Huaxiu, & Li, Zhenhui. 2018. Intellilight: A reinforcement learning approach for intelligent traffic light control. *Pages 2496–2505 of: Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining.*