GESA: A GEneral Scenario-Agnostic Reinforcement Learning for Traffic Signal Control*

Haoyuan Jiang[1], Ziyue Li[2,*], Zhishuai Li[1], Lei Bai[3], Hangyu Mao[1], Wolfgang Ketter[2] and Rui Zhao[1]

[1]Sensetime Research, China
[2]University of Cologne, Germany
[3]Shanghai AI Lab, China

Abstract

Reinforcement learning (RL) can automatically learn a better policy through a trial-and-error paradigm and has been adopted to revolutionize and optimize traditional traffic signal control systems that are usually based on handcrafted methods. However, most existing RL-based models are either based on a single scenario or multiple independent scenarios, where each scenario has a separate simulation environment with predefined road network topology and traffic signal settings. These models implement training and testing in the same scenario, thus being strictly tied up with the specific setting and sacrificing model generalization heavily. While a few recent models could be trained by multiple scenarios, they require a huge amount of manual labor to label the intersection structure, hindering the model's generalization. In this work, we aim at a *general* framework that could eliminate heavy labeling and model a variety of scenarios *simultaneously*. To this end, we propose a general Scenario-Agnostic (GESA) reinforcement learning framework for traffic signal control with: (1) A general plug-in module to map all different intersections into a unified structure, freeing us from the heavy manual labor to specify the structure of intersections; (2) A unified state and action space design to keep the model input and output consistently structured; (3) A large-scale co-training with multiple scenarios, leading to a generic traffic signal control algorithm. GESA can automatically handle various structured intersections from various cities without human labeling, and it co-trains a generalist agent to control traffic signals for multiple cities together, which also demonstrates superior transferability in zero-shot settings. In experiments, we demonstrate our algorithm as the first one that can be co-trained with seven different scenarios without manual annotation and gets **13.27%** higher rewards than baselines. When dealing with a new scenario, our model can still achieve **9.39%** higher rewards. The code, scenarios, and demos are available here. The full paper is available at [1].

Keywords

Traffic signal control, Reinforcement learning, A generalist agent, Zero-shot transfer

Reinforcement learning (RL) [2, 3, 4] has been preferably adopted into the TSC domain since it is a learning-based method with higher automation. Such a trial-and-error paradigm based on the traffic simulator has demonstrated better performance than transport engineering-based methods [5]. The recent RL-based TSC models can be roughly divided into two categories based on the scenarios where the training and testing are conducted. A scenario is usually a simulation environment that contains a set of intersections: (1) **Single-scenario RL**, as the majority, its training and testing need to be on the same scenario [2, 6, 7]. However, the model will be unusable or perform badly in a new scenario. For example, in Fig. 1(b) top, these methods might be trained and tested in the same scenario with 5×5 four-approach intersections of Fig. 1(a2), but these methods will ill-perform in another new scenario with mixed intersections of Fig. 1(a1) and 1(a2). (2) **Multi-scenario RL**, as shown in Fig. 1(b) bottom, where training is conducted in multiple scenarios, and testing could be in different scenarios. For example, [8, 9, 10] are proposed to train a TSC system with multi-scenarios. However, in the training stage, the existing multi-scenario RL models need heavy manual labor to annotate the structure of intersections, such as the direction of each entering approach, the number of entering lanes of each entering approach, the traffic movement of each entering lane, etc. Moreover, they either achieve multi-scenario co-training in a sequential manner, one scenario after another, leading to a rather unstable learning curve and slower convergence [8, 10] or only narrow the scale of a scenario to only one intersection in one scenario, which heavily limits the model's generalization [9].

Moreover, current RL-based TSC methods are trained with several pre-defined and fixed scenarios, whereas they cannot gain generalization capability without labeling, which limits the application of RL-based methods in the real world. These methods can exploit the various traffic flows generated by the simulator to make the model effective in training scenarios, but finding a low-cost universal method with promising transferability meanwhile is still a research gap. As a result, the existing methods still face tremendous challenges in jumping out from the simulation and implementing them in real cities. This is known as *sim2real* challenge. The challenges mainly come from the wide gap between the real complex cities and the simplified simulation systems. In the real world, the intersection structure could be rather versatile in terms of different settings of approaches (*i.e.*, north, south, east, west), movements (*i.e.*, left, right, through), and lanes (*e.g.*, two through lanes, one right-through lane). As shown in Fig. 1(a), an intersection could have a different amount of approaches. Within an approach, there can be different combinations of movements; A lane could also combine different movements. However, most of the existing methods only consider a standard simulation intersection with four approaches and three lanes (right, through, and left) within each approach. This largely limits the model generalization.

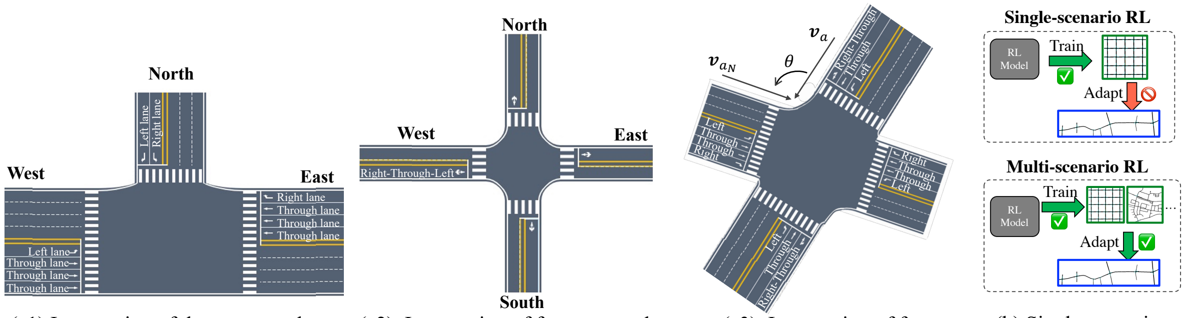
To conclude, a qualified TSC approach needs high generalization and effectiveness: it should handle various intersections and be able to transfer to other unseen targets easily and with low cost. In this paper, we aim to answer three questions: (1) How do we co-train an RL with multi-scenarios without labeling, given the diverse intersection structures? (2) Will multi-scenario co-training improve the TSC? If yes, why? And the more scenarios, the better? (3)

STRL'24: Third International Workshop on Spatio-Temporal Reasoning and Learning, 5 August 2024, Jeju, South Korea

* You can use this document as the template for preparing your publication. We recommend using the latest version of the ceurart style.

*Corresponding author.
jianghaoyuan@zju.edu.cn (H. Jiang); zlibn@wiso.uni-koeln.de (Z. Li)

© 2024 Copyright for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).



(a1) Intersection of three approaches, with 2-4 entering lanes, and 2 movements on each approach

(a2) Intersection of four approaches, with 1 entering lane of right-through-left movement on each approach

(a3) Intersection of four approaches, with 4 entering lanes, and 3 movements, yet with an arbitrary angel

(b) Single-scenario Training *v.s.* Multi-scenario Training
Training scenario
Adapting scenario

Figure 1: (a) Three intersections with different structures in terms of approach, movement, and lane; (b) Single-scenario RL *v.s.* Multi-scenario RL.

Does the co-trained RL model still perform well in the new scenario?

To narrow the *sim2real* gap significantly and get more ready to be deployed in real cities, in this paper, we provide a **GE**neral **S**cenario-**A**gnostic (GESA) reinforcement learning framework for the TSC task. To our best knowledge, GESA is the first work that pursues high generability and co-trains multiple scenarios without labels: it automatically handles various scenarios; the reinforcement learning is designed accordingly to achieve generalization; it is co-trained with multiple scenarios simultaneously and demonstrates high transferability. Specifically, to co-train in multiple scenarios with various intersections, the vectors with approach spatial information are employed to map shape-odd and complex intersections into the standard intersection. Then, the mapped intersections are used to generate the characteristic information of each traffic movement and the phase of the traffic lights in a specific order. Finally, we extend the original FRAP [6] to a policy gradient-based framework, which can facilitate the model coverage and is compatible with different intersections.

The contributions are summarized in three-fold: (1) We present a general plug-in module to map the intersections into a unified structure, freeing us from the heavy manual labeling work to specify the intersection structure and enabling large-scale co-training under multiple different scenarios. (2) Accordingly, we design a unified state and action space to keep the model input and output structure consistent for more general capabilities. Moreover, the GESA can adapt to various unseen scenarios and achieve promising performance without re-training. (3) We build two real-world scenarios using the real city road map and the real traffic dynamics, together with five public scenarios, where we co-train and validate the GESA with prudent experiments. All these lead us closer to the ultimate goal: to implement RL-based TSC in real cities.

References

[1] H. Jiang, Z. Li, Z. Li, L. Bai, H. Mao, W. Ketter, R. Zhao, A general scenario-agnostic reinforcement learning for traffic signal control, IEEE Transactions on Intelligent Transportation Systems (2024) 1–15.

[2] E. Van der Pol, F. A. Oliehoek, Coordinated deep reinforcement learners for traffic light control, Proceedings of Learning, Inference and Control of Multi-agent Systems (at NIPS 2016) 1 (2016).

[3] H. Wei, G. Zheng, H. Yao, Z. Li, Intellilight: A reinforcement learning approach for intelligent traffic light control, in: Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining, 2018, pp. 2496–2505.

[4] H. Wei, C. Chen, G. Zheng, K. Wu, V. Gayah, K. Xu, Z. Li, Presslight: Learning max pressure control to coordinate traffic signals in arterial network, in: Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining, 2019, pp. 1290–1298.

[5] K.-L. A. Yau, J. Qadir, H. L. Khoo, M. H. Ling, P. Komisarczuk, A survey on reinforcement learning models and algorithms for traffic signal control, ACM Computing Surveys (CSUR) 50 (2017) 1–38.

[6] G. Zheng, Y. Xiong, X. Zang, J. Feng, H. Wei, H. Zhang, Y. Li, K. Xu, Z. Li, Learning phase competition for traffic signal control, in: Proceedings of the 28th ACM International Conference on Information and Knowledge Management, 2019, pp. 1963–1972.

[7] L. Zhu, P. Peng, Z. Lu, Y. Tian, MTLight: Efficient multi-task reinforcement learning for traffic signal control, in: ICLR 2022 Workshop on Gamification and Multiagent Solutions, 2022.

[8] X. Zang, H. Yao, G. Zheng, N. Xu, K. Xu, Z. Li, Metalight: Value-based meta-reinforcement learning for traffic signal control, in: Proceedings of the AAAI Conference on Artificial Intelligence, volume 34, 2020, pp. 1153–1160.

[9] A. Oroojlooy, M. Nazari, D. Hajinezhad, J. Silva, Attendlight: Universal attention-based reinforcement learning model for traffic signal control, Advances in Neural Information Processing Systems 33 (2020) 4079–4090.

[10] M. Wang, Y. Xu, X. Xiong, Y. Kan, C. Xu, M.-O. Pun, Adlight: A universal approach of traffic signal control with augmented data using reinforcement learning, arXiv preprint arXiv:2210.13378 (2022).