Real World Traffic Optimization by Reinforcement Learning: A Concept [*]

Henri Meess[1], Jeremias Gerner[2], Daniel Hein[3], Stefanie Schmidtner[2], and Gordon Elger[1]

[1] Fraunhofer IVI, Ingolstadt, Germany
{henri.meess,gordon.elger}@ivi.fraunhofer.de
[2] Technische Hochschule Ingolstadt, Ingolstadt, Germany
{jeremias.gerner,stefanie.schmidtner}@thi.de
[3] GEVAS software GmbH, Munich, Germany daniel.hein@gevas.de

1 Issues in Real Traffic Light Systems

Due to the growing urban population [22], the existing infrastructure and traffic control are successively reaching their limits, making an optimization of the traffic flow by intelligent control of Traffic Lights (TL) increasingly important. Previous research has already shown the basic suitability of Deep Reinforcement Learning (DRL) methods for TL control, for both, the optimization of single intersections [13, 14] and the optimization of traffic networks using Multi Agent Reinforcement Learning (MARL) [1, 19, 10, 15, 2, 11, 7]. A major gap in research concerning this area is the training and usage in real-life systems due to several challenges [18, 20, 23]: (1) Training in real systems is difficult since agents cannot perform unrestricted arbitrary actions. (2) It cannot always be guaranteed that the learned policies are sufficiently robust. (3) DRL controllers must ensure that existing safety and operational constraints are enforced at all times. Thus, DRL-based TL controllers have been implemented mostly simulation-based [23]. However, these simulation-based approaches can only be transferred to reality to a limited extent since [6]:

- *Multimodality:* In most simulations only car traffic was simulated, neglecting other traffic participants.
- *Baselines:* DRL methods have rarely been compared with state-of-the-art traffic-actuated controls that are already used in the real-world.
- *State and actions:* In reality, data collection is considerably more difficult. Furthermore, the action space is not correctly aligned with current TL control units.
- *Simulation environments:* Most implementations were only done in symmetric networks with distribution-based traffic demand, oversimplifying real traffic situations. Traffic state information was directly taken from simulation, neglecting the lack of this data in real systems. Therefore, training on simulations and inference in reality can lead to inconsistencies [24, 8].

[*] This work was funded by the German Federal Ministry of Transport and Digital Infrastructure (BMVI) as part of the project "KIVI - Artificial Intelligence in the traffic system of Ingolstadt [Germany]".

This work picks up the concept, presented in [16], to combine state-of-the-art DRL methods with existing traffic engineering methods to overcome the issues mentioned above and extends it by a description how to specifically close the gap between simulation and reality.

2 Closing the Simulation to Reality Gap

To enable a system that can be used effectively in real-world traffic systems, the RL framework to optimize the traffic lights is intended to build upon and extend established systems and procedures of traditional traffic engineering. In particular, this concerns the extraction of state and reward data and the actions definition. For the extraction of the state and reward data it is planed to use a traffic estimation model named DRIVERS [9]. DRIVERS creates a microscopic state representations based on raw detector data from real world traffic systems. Therefore it generates Origin-Destination (OD) matrices from the detector data which highly correspond to the real traffic on a macroscopic level. Based on the OD matrices traffic is simulated by a microscopic simulation model. This makes it possible to obtain microscopic state information which is coherent to the real-life traffic. Figure 1 shows the planned structure of the system for training and operation. In the core RL-System DRIVERS serves as the main data source for the state and reward information. The RL components and DRIVERS are closely linked, because the state and reward definitions are based on the DRIVERS output and the DRIVERS model is therefore an inherent part of the learned policy. To achieve a save operation of the system it will be first trained and evaluated in a simulation-based environment. In this environment a simulation serves as a surrogate for the real traffic network. This allows to train and optimize the core RL-System without affecting the real traffic network. To enable a realistic traffic representation in the simulation, a second DRIVERS instance is used. This instance is used to generate OD matrices from the sensor data of the real network, which serve as the demand definition for the simulation.

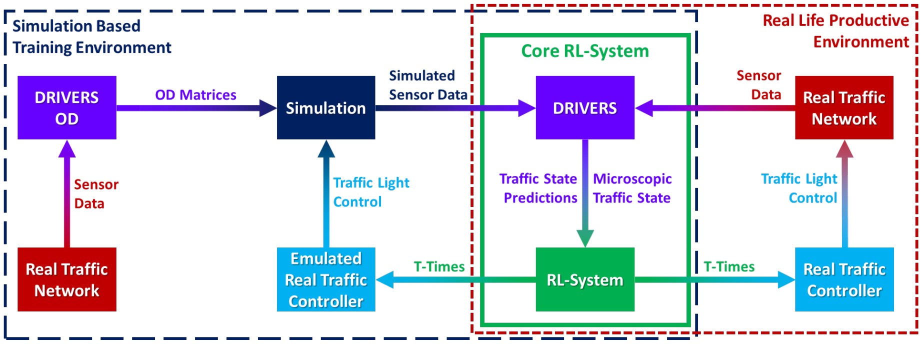

Fig. 1. Architecture of the proposed framework

After the configuration of the core RL-System has been optimized and tested in the simulation framework, it can seamlessly be integrated in the real traffic environment by switching the sensor data source and the control interface for the core RL-System from the simulation to the real system.

For both, the simulation based training and real life productive system, microscopic traffic generated by DRIVERS is transferred to the RL-System and refined through state representation methods known from the literature. Specifically Discrete Traffic State Encoding (DTSE) [5, 19, 4] and feature-based representation [12, 6] methods will be critically evaluated, especially for their theoretical justification and applicability in a real-world setting. For a multimodal optimization, all traffic participants are to be represented for this purpose, enabling a fair distribution of green times for all road users. Another important point for the applicability of RL systems in reality is the action definition. To achieve a compatibility with real control systems the action definition will be based on the widely spread time gap control [17]. This control method is based on frame signal plans, witch consist of T-times. T-times are lower and upper time limits, in which the local controller can independently switch the phases. For each phase i from crossing k there is a minimum ${T_{min}}_i^k$ and maximum ${T_{max}}_i^k$ admissible T-time. We define the set of actions $A^k = \{a_{i,min}^k, a_{i,max}^k\}$ for a single agent at crossing k with the following condition: ${T_{min}}_i^k \leq a_{i,min}^k \leq a_{i,max}^k \leq {T_{max}}_i^k$. After $a_{i,min}^k$ has been exceeded and no further vehicles are registered for a defined time, or if $a_{i,max}^k$ is reached, the traffic controller switches to the next phase $i+1$. As the actions can only vary in an interval that takes the minimum and maximum admissible T-times of the phases into account, the safe operation and a minimal performance of the traffic lights is guaranteed with all possible combinations. The RL agent's goal is to find the optimal T-times for the given traffic situation. Thus, we obtain a continuous action space with two values per phase where all actions follow the reasonable phases and transitions of the existing systems.

Such an action space comes with several challenges: (1) a (dis-) continuous action space; the majority of papers deal with discrete action spaces [3] (2) a constrained action space (3) actions depend on other actions that are defined at the same time (4) a high number of actions at the same time; compared to different approaches.

To ensure that the constrains hold and generic Actor Critic methods can be used we define the actions as:

$$a_{i,min}^k = rnd(sig_{out i,min}^k \times ({T_{max}}_i^k - {T_{min}}_i^k)) + {T_{min}}_i^k \tag{1}$$

$$a_{i,max}^k = rnd(sig_{out i,max}^k \times ({T_{max}}_i^k - a_{i,min}^k)) + a_{i,min}^k \tag{2}$$

Where sig_{out} is the respective outcome of the last actors NN-Layer with a sigmoid activation function for the distinctive element of the agent's action set. Based on domain knowledge and theoretical considerations derived from traffic engineering, this approach aims to achieve the following:

1. The constrained action space can ensure that agents achieve a minimum level of performance in all situations. Even completely unknown traffic scenarios

do not lead to a full failure of the system while safety is secured by the T-times concept.
2. The occurrence of a green wave is simplified by the specification of allowed T-times.
3. This approach can be ported directly into the real application. Following the concept stated in Figure 1.

3 Cooperative Optimization

To ensure goal-directed cooperative optimization, an incentive for cooperation must be created. Usually, common rewards or reward sharing between the neighbors [21] are used. Furthermore, the state space can get enriched with relevant information of the neighbors [2]. To extend this basic setup, new approaches for the cooperation of multiple agents will be explored. These apply different concepts of shared critics, e.g.: (1) The actors and critics outputs are fed into a shared critic. The actors updates are based on a weighted gradient of own and shared critic. (2) The actors output are fed into their respective and a shared critic. Additionally, the shared critic gets superordinate state representations. The actors updates are based on a weighted gradient of own and shared critic. Additionally, we will investigate to what extent a benefit is created by providing information about outflowing edges to overcome deadlocks caused by not sufficient informed policies[4]. By this, streets or regions shall be jointly optimized as clusters or common routes. We thereby encourage direct cooperation as a shared critic directs the gradients for optimization.

4 Outlook

In this paper, we outlined a concept to bring RL from simulative applications to real use in the field. To solve the stated problems we propose a detailed consideration of individual intersections, multimodality, and specific configurations of MARL for practical implementation. Through the consideration and combination with current techniques for traffic control we increase the applicability of our concept for real-world traffic networks. To ensure compatibility we train in simulations on real data derived by online traffic estimations as well as random generated traffic. We use DRIVERS in the simulation to estimate the traffic behavior even though the actual traffic is available in the simulation and further add a simulation of the actual traffic controller. By this, we strongly adapt to the later in-field implementation even while training and try to overcome the simulation to reality gap in this field. Finally, the real deployment in Ingolstadt's road network is planned, where we after all want to prove the applicability of RL for real-world traffic optimization.

[4] A more detailed overview and presentation of the concepts will be provided in the full version of this work.

References

1. Abdoos, M., Mozayani, N., Bazzan, A.L.C.: Traffic light control in non-stationary environments based on multi agent q-learning. In: 2011 14th International IEEE Conference on Intelligent Transportation Systems (ITSC). pp. 1580–1585 (2011). https://doi.org/10.1109/ITSC.2011.6083114
2. Chu, T., Wang, J., Codecà, L., Li, Z.: Multi-agent deep reinforcement learning for large-scale traffic signal control (2019)
3. El-Tantawy, S., Abdulhai, B., Abdelgawad, H.: Design of reinforcement learning parameters for seamless application of adaptive traffic signal control. Journal of Intelligent Transportation Systems **18** (06 2014). https://doi.org/10.1080/15472450.2013.810991
4. Gao, J., Shen, Y., Liu, J., Ito, M., Shiratori, N.: Adaptive traffic signal control: Deep reinforcement learning algorithm with experience replay and target network. arXiv preprint arXiv:1705.02755 (2017)
5. Genders, W., Razavi, S.: Using a deep reinforcement learning agent for traffic signal control. arXiv preprint arXiv:1611.01142 (2016)
6. Haydari, A., Yilmaz, Y.: Deep reinforcement learning for intelligent transportation systems: A survey. IEEE Transactions on Intelligent Transportation Systems (2020)
7. Hussain, A., Wang, T., Jiahua, C.: Optimizing traffic lights with multi-agent deep reinforcement learning and v2x communication. arXiv preprint arXiv:2002.09853
8. Kadian, A., Truong, J., Gokaslan, A., Clegg, A., Wijmans, E., Lee, S., Savva, M., Chernova, S., Batra, D.: Sim2real predictivity: Does evaluation in simulation predict real-world performance? IEEE Robotics and Automation Letters **5**(4), 6670–6677 (2020)
9. Kemper, C.: Dynamische Simulation des Verkehrsablaufs unter Verwendung statischer Verflechtungsmatrizen. Ph.D. thesis, Hannover: Gottfried Wilhelm Leibniz Universität Hannover (2006)
10. K.J., P., A.N, H.K., Bhatnagar, S.: Multi-agent reinforcement learning for traffic signal control. In: 17th International IEEE Conference on Intelligent Transportation Systems (ITSC). pp. 2529–2534 (2014). https://doi.org/10.1109/ITSC.2014.6958095
11. Kuyer, L., Whiteson, S., Bakker, B., Vlassis, N.: Multiagent reinforcement learning for urban traffic control using coordination graphs. In: Joint European Conference on Machine Learning and Knowledge Discovery in Databases. pp. 656–671. Springer (2008)
12. Li, L., Lv, Y., Wang, F.Y.: Traffic signal timing via deep reinforcement learning. IEEE/CAA Journal of Automatica Sinica **3**(3), 247–254 (2016)
13. Liang, X., Du, X., Wang, G., Han, Z.: A deep reinforcement learning network for traffic light cycle control. IEEE Transactions on Vehicular Technology **68**(2), 1243–1253 (2019)
14. Lin, Y., Dai, X., Li, L., Wang, F.Y.: An efficient deep reinforcement learning model for urban traffic control (2018)
15. Medina, J.C., Benekohal, R.F.: Traffic signal control using reinforcement learning and the max-plus algorithm as a coordinating strategy. In: 2012 15th International IEEE Conference on Intelligent Transportation Systems. pp. 596–601 (2012). https://doi.org/10.1109/ITSC.2012.6338911
16. Meess, H., Gerner, J., Hein, D., Schmidtner, S., Elger, G.: Reinforcement learning for traffic signal control optimization: A concept for real-world implementation - extended abstract, accepted at AAMAS 2022

17. Oertel, D.I.R., Krimmling, I.J., Körner, D.I.M., et al.: Verlustzeitenbasierte lsasteuerung eines einzelknotens (2011)
18. Parvez Farazi, N., Zou, B., Ahamed, T., Barua, L.: Deep reinforcement learning in transportation research: A review. Transportation Research Interdisciplinary Perspectives **11**, 100425 (2021). https://doi.org/https://doi.org/10.1016/j.trip.2021.100425, https://www.sciencedirect.com/science/article/pii/S2590198221001317
19. Van der Pol, E., Oliehoek, F.A.: Coordinated deep reinforcement learners for traffic light control. Proceedings of Learning, Inference and Control of Multi-Agent Systems (at NIPS 2016) (2016)
20. Rasheed, F., Yau, K.L.A., Noor, R.M., Wu, C., Low, Y.C.: Deep reinforcement learning for traffic signal control: A review. IEEE Access **8**, 208016–208044 (2020). https://doi.org/10.1109/ACCESS.2020.3034141
21. Salkham, A., Cunningham, R., Garg, A., Cahill, V.: A collaborative reinforcement learning approach to urban traffic control optimization. In: 2008 IEEE/WIC/ACM International Conference on Web Intelligence and Intelligent Agent Technology. vol. 2, pp. 560–566 (2008). https://doi.org/10.1109/WIIAT.2008.88
22. United Nations, D.o.E., Social Affairs, P.D..: World urbanization prospects: The 2018 revision. custom data acquired via website. (2018), https://population.un.org/wup/DataQuery/
23. Wei, H., Zheng, G., Gayah, V., Li, Z.: Recent advances in reinforcement learning for traffic signal control: A survey of models and evaluation. ACM SIGKDD Explorations Newsletter **22**, 12–18 (01 2021). https://doi.org/10.1145/3447556.3447565
24. Zhang, F., Leitner, J., Milford, M., Upcroft, B., Corke, P.: Towards vision-based deep reinforcement learning for robotic motion control. arXiv preprint arXiv:1511.03791 (2015)