Reinforcement Learning for Traffic Signal Control Optimization: A Concept for Real-World Implementation

Extended Abstract

Henri Meess
Fraunhofer IVI
Ingolstadt, Germany
henri.meess@ivi.fraunhofer.de

Jeremias Gerner
Technische Hochschule Ingolstadt
Ingolstadt, Germany
jeremias.gerner@thi.de

Daniel Hein
GEVAS software GmbH
Munich, Germany
daniel.hein@gevas.de

Stefanie Schmidtner
Technische Hochschule Ingolstadt
Ingolstadt, Germany
stefanie.schmidtner@thi.de

Gordon Elger
Fraunhofer IVI
Ingolstadt, Germany
gordon.elger@ivi.fraunhofer.de

ABSTRACT

The improvement of traffic control is one of the most important but also most ambitious goals in the field of urban traffic today. Due to the domain's long history, new methods must first establish themselves in practice and, above all, demonstrate reasonable and robust optimization results. Potential methods which came up in recent years are Reinforcement Learning (RL) and Multi-Agent Reinforcement Learning (MARL). In research, the use of RL/MARL for traffic optimization is widely spread. However, it has not yet managed to make it from simulations into practical implementation due to: (1) The lack of real data for the online estimation of the state space. (2) The compatibility to real controllers. (3) The necessity of guarantees to ensure the resilience of the controls. Enabled by a project developing testbeds and practical approaches to optimize traffic through AI, we present a concept to close this gap to the online control of real networks and to overcome the stated issues.

KEYWORDS

Multi-Agent Reinforcement Learning in real-world; MARL; traffic optimization; multimodal traffic; DRL

ACM Reference Format:

Henri Meess, Jeremias Gerner, Daniel Hein, Stefanie Schmidtner, and Gordon Elger. 2022. Reinforcement Learning for Traffic Signal Control Optimization: A Concept for Real-World Implementation: Extended Abstract. In *Proc. of the 21st International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2022), Online, May 9–13, 2022*, IFAAMAS, 3 pages.

1 THE ISSUE IN REAL SYSTEMS

Due to predicted and observed increase of the urban population [26], the existing infrastructure and traffic control are increasingly reaching their limits. To avoid a cost-intensive expansion of the infrastructure in urban regions, an optimization of the traffic flow by intelligent control of Traffic Lights (TL) is necessary. Previous research has already shown the basic suitability of Deep Reinforcement Learning (DRL) methods for TL control, for both, the optimization of single intersections [14, 16] and the optimization of traffic networks using MARL [1, 2, 8, 11, 12, 18, 27]. A major gap in research concerning this area is the training and usage in real-life systems due to several challenges [21, 22, 28]: (1) Training in real systems is difficult because 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 [28]. However, these simulation-based approaches can only be transferred to reality to a limited extent since [7]: *Simulation environments* - Most implementations were only done in symmetric networks with distribution-based traffic demand, oversimplifying real traffic situations. *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.

The goal of this work is to combine state-of-the-art DRL methods with existing traffic engineering methods to overcome the issues mentioned above and enable the transfer of DRL to real traffic systems. The real-world setup is provided by the KIVI project (Ingolstadt, Germany), which supplies a High Definition Testbed (HDT) containing high-quality sensor data from intersections. These intersections serve as a starting point for the real-world implementation of DRL algorithms for TL control.

2 REAL-WORLD TRAFFIC OPTIMIZATION

Aim of KIVI is to develop a RL setting that can be used in real-world traffic networks and adapts to current traffic engineering solutions. To enable this, the KIVI-RL framework is intended to build upon and extend established systems and procedures of traditional traffic engineering. In the first step, a traffic estimation model named DRIVERS [10] is used to generate microscopic state representations based on real-life data. DRIVERS generates Origin-Destination (OD) matrices from real-world detector data which highly correspond to the real traffic on a macroscopic level. Based on the (OD) matrices traffic can be simulated. This enables to obtain microscopic state information which is coherent to the real-life traffic. The microscopic traffic generated by DRIVERS is transferred to state representation

Proc. of the 21st International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2022), P. Faliszewski, V. Mascardi, C. Pelachaud, M.E. Taylor (eds.), May 9–13, 2022, Online.

methods known from the literature. Specifically Discrete Traffic State Encoding (DTSE) [4, 5, 27] and feature-based representation [7, 13] methods will be critically evaluated, especially for their theoretical justification and applicability in a real-world setting. Past research missed to include traffic participants outside of individual motorized traffic [7], which requires further investigations for a real-world implementation. Challenges here include the extension of the state representation to the other traffic participants and setting up the reward structure for fair optimization. Although the goal is a network-wide optimization of the over 100 actuated intersections in Ingolstadt by cooperative MARL, research regarding state and reward will first be conducted on a single intersection. This can generate deep knowledge for a small setting and DRIVERS can be critically compared with the real traffic captured by the HDT. Since the multi-agent setting is an aggregation of individual agents, the gained knowledge can be transferred effectively.

In order to create a RL system that is transferable to real-world applications, it is necessary to select the action space analogous to the specifications of actual traffic controllers. Therefore our action definition is based on the widely spread time gap control [20]. Here, frame signal plans based on T-times are defined, containing lower and upper limits, in which the local controller can switch phases. Basically, for each phase i from crossing k there is a minimum ${T_{min}}_i^k$ and maximum ${T_{max}}_i^k$ admissible T-time. After ${T_{min}}_i^k$ has been exceeded and no further vehicles are registered for a defined time, or if ${T_{max}}_i^k$ is reached, the traffic controller switches to the next phase. The agent's goal is to find the optimal T-times for the given traffic situation. We define the set of actions $A^k = \{a_{i,min}^k, a_{i,max}^k\}$ for a single agent k with the following condition: ${T_{min}}_i^k \leq a_{i,min}^k \leq a_{i,max}^k \leq {T_{max}}_i^k$. Thus, we obtain a continuous action space with two values per phase. All actions follow the reasonable phases and transitions of the existing systems. Thereby we force that the predefined safeties hold. 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.

Within the project three of the most common algorithms for continuous action spaces will be investigated as baseline [15] [6] [25]. To ensure that such generic continuous actors 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. Furthermore, various extensions like Recurrent Neuronal Networks [23] and Working Memory Graphs [17] will be introduced to simplify the solving of this complex task. 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.

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 [24] 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 work with different concepts of shared critics: (1) The actors and critics are fed into a shared critic. The actors updates are based on a weighted gradient of own and shared critic. (2) The actors are fed into their respective and a shared critic. Additionally, the shared critic gets superordinate state representations (SSR). The actors updates are based on a weighted gradient of own and shared critic. (3) Each actor is fed into a shared critic who gets a SSR based on which it evaluates the chosen actions respective to the ordinate states. (4) A system of multiple critics or multiple critic outputs will be investigated to motivate Pareto-optimal solutions by using different reward goals. 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. 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. Compared to the usage of one agent for the whole traffic network, we reduce the complexity especially in the process of acting while keeping a tendency for direct cooperation encouraged by the shared critic. A more detailed explanation and overview of the whole system can be found in [19]. Training on simulation and inference in reality can lead to certain issues [9, 29]. 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. We further add a simulation of the actual traffic controller. By this, we strongly adapt to the later in-field implementation even while training.

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. Within the project, we will successively transfer findings from the individual intersection consideration into the MARL system. 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.

ACKNOWLEDGMENTS

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]".

REFERENCES

[1] Monireh Abdoos, Nasser Mozayani, and Ana L. C. Bazzan. 2011. Traffic light control in non-stationary environments based on multi agent Q-learning. In *2011 14th International IEEE Conference on Intelligent Transportation Systems (ITSC)*. 1580–1585. https://doi.org/10.1109/ITSC.2011.6083114

[2] Tianshu Chu, Jie Wang, Lara Codecà, and Zhaojian Li. 2019. Multi-Agent Deep Reinforcement Learning for Large-scale Traffic Signal Control. arXiv:1903.04527 [cs.LG]

[3] Samah El-Tantawy, Baher Abdulhai, and Hossam Abdelgawad. 2014. 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] Juntao Gao, Yulong Shen, Jia Liu, Minoru Ito, and Norio Shiratori. 2017. Adaptive traffic signal control: Deep reinforcement learning algorithm with experience replay and target network. *arXiv preprint arXiv:1705.02755* (2017).

[5] Wade Genders and Saiedeh Razavi. 2016. Using a deep reinforcement learning agent for traffic signal control. *arXiv preprint arXiv:1611.01142* (2016).

[6] Tuomas Haarnoja, Aurick Zhou, Pieter Abbeel, and Sergey Levine. 2018. Soft Actor-Critic: Off-Policy Maximum Entropy Deep Reinforcement Learning with a Stochastic Actor. arXiv:1801.01290 [cs.LG]

[7] Ammar Haydari and Yasin Yilmaz. 2020. Deep reinforcement learning for intelligent transportation systems: A survey. *IEEE Transactions on Intelligent Transportation Systems* (2020).

[8] Azhar Hussain, Tong Wang, and Cao Jiahua. [n.d.]. Optimizing Traffic Lights with Multi-agent Deep Reinforcement Learning and V2X communication. *arXiv preprint arXiv:2002.09853* ([n. d.]).

[9] Abhishek Kadian, Joanne Truong, Aaron Gokaslan, Alexander Clegg, Erik Wijmans, Stefan Lee, Manolis Savva, Sonia Chernova, and Dhruv Batra. 2020. Sim2Real predictivity: Does evaluation in simulation predict real-world performance? *IEEE Robotics and Automation Letters* 5, 4 (2020), 6670–6677.

[10] Carsten Kemper. 2006. *Dynamische Simulation des Verkehrsablaufs unter Verwendung statischer Verflechtungsmatrizen.* Ph.D. Dissertation. Hannover: Gottfried Wilhelm Leibniz Universität Hannover.

[11] Prabuchandran K.J., Hemanth Kumar A.N, and Shalabh Bhatnagar. 2014. Multi-agent reinforcement learning for traffic signal control. In *17th International IEEE Conference on Intelligent Transportation Systems (ITSC)*. 2529–2534. https://doi.org/10.1109/ITSC.2014.6958095

[12] Lior Kuyer, Shimon Whiteson, Bram Bakker, and Nikos Vlassis. 2008. Multiagent reinforcement learning for urban traffic control using coordination graphs. In *Joint European Conference on Machine Learning and Knowledge Discovery in Databases*. Springer, 656–671.

[13] Li Li, Yisheng Lv, and Fei-Yue Wang. 2016. Traffic signal timing via deep reinforcement learning. *IEEE/CAA Journal of Automatica Sinica* 3, 3 (2016), 247–254.

[14] Xiaoyuan Liang, Xunsheng Du, Guiling Wang, and Zhu Han. 2019. A deep reinforcement learning network for traffic light cycle control. *IEEE Transactions on Vehicular Technology* 68, 2 (2019), 1243–1253.

[15] Timothy P. Lillicrap, Jonathan J. Hunt, Alexander Pritzel, Nicolas Heess, Tom Erez, Yuval Tassa, David Silver, and Daan Wierstra. 2019. Continuous control with deep reinforcement learning. arXiv:1509.02971 [cs.LG]

[16] Yilun Lin, Xingyuan Dai, Li Li, and Fei-Yue Wang. 2018. An Efficient Deep Reinforcement Learning Model for Urban Traffic Control. arXiv:1808.01876 [cs.AI]

[17] Ricky Loynd, Roland Fernandez, Asli Celikyilmaz, Adith Swaminathan, and Matthew Hausknecht. 2020. Working Memory Graphs. arXiv:1911.07141 [cs.LG]

[18] Juan C. Medina and Rahim F. Benekohal. 2012. 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*. 596–601. https://doi.org/10.1109/ITSC.2012.6338911

[19] Henri Meess, Jeremias Gerner, Daniel Hein, Stefanie Schmidtner, and Gordon Elger. 2022. Reinforcement Learning for Traffic Signal Control Optimization: A Concept for Real-World Implementation. *to be submitted at ABMUS2022* (2022).

[20] Dipl-Ing Robert Oertel, Ing Jürgen Krimmling, Dipl-Ing Matthias Körner, et al. 2011. Verlustzeitenbasierte LSA-Steuerung eines Einzelknotens. (2011).

[21] Nahid Parvez Farazi, Bo Zou, Tanvir Ahamed, and Limon Barua. 2021. Deep reinforcement learning in transportation research: A review. *Transportation Research Interdisciplinary Perspectives* 11 (2021), 100425. https://doi.org/10.1016/j.trip.2021.100425

[22] Faizan Rasheed, Kok-Lim Alvin Yau, Rafidah Md. Noor, Celimuge Wu, and Yeh-Ching Low. 2020. Deep Reinforcement Learning for Traffic Signal Control: A Review. *IEEE Access* 8 (2020), 208016–208044. https://doi.org/10.1109/ACCESS.2020.3034141

[23] David E Rumelhart, Geoffrey E Hinton, and Ronald J Williams. 1985. *Learning internal representations by error propagation.* Technical Report. California Univ San Diego La Jolla Inst for Cognitive Science.

[24] As'ad Salkham, Raymond Cunningham, Anurag Garg, and Vinny Cahill. 2008. 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. 560–566. https://doi.org/10.1109/WIIAT.2008.88

[25] John Schulman, Filip Wolski, Prafulla Dhariwal, Alec Radford, and Oleg Klimov. 2017. Proximal Policy Optimization Algorithms. arXiv:1707.06347 [cs.LG]

[26] Department of Economic United Nations and Population Division (2018). Social Affairs. 2018. World Urbanization Prospects: The 2018 Revision. *custom data acquired via website.* (2018). Retrieved October 5, 2021 from https://population.un.org/wup/DataQuery/

[27] Elise Van der Pol and Frans A Oliehoek. 2016. Coordinated deep reinforcement learners for traffic light control. *Proceedings of Learning, Inference and Control of Multi-Agent Systems (at NIPS 2016)* (2016).

[28] Hua Wei, Guanjie Zheng, Vikash Gayah, and Zhenhui Li. 2021. Recent Advances in Reinforcement Learning for Traffic Signal Control: A Survey of Models and Evaluation. *ACM SIGKDD Explorations Newsletter* 22 (01 2021), 12–18. https://doi.org/10.1145/3447556.3447565

[29] Fangyi Zhang, Jürgen Leitner, Michael Milford, Ben Upcroft, and Peter Corke. 2015. Towards vision-based deep reinforcement learning for robotic motion control. *arXiv preprint arXiv:1511.03791* (2015).