Deep Reinforcement Learning for Coordination in Traffic Light Control

Elise van der Pol [a]

[a] *University of Amsterdam*

Abstract

This thesis combines the Deep Q-learning algorithm with max-plus coordination in transfer planning in order to find optimal coordinated policies in traffic light control with multiple agents, an approach that outperforms earlier work.

1 Introduction

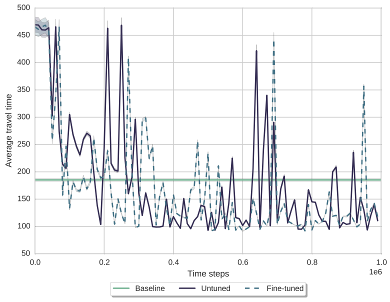
The cost of traffic congestion in the EU is estimated to be 1% of the EU's GDP [2], and good solutions for traffic light control may reduce traffic congestion, saving time and money and reducing pollution. To find optimal traffic light policies, reinforcement learning (RL) uses reward signals from the environment to learn to make optimal decisions. The Deep Q-learning algorithm [4] (DQN) has shown good results on Atari games, and we research whether it can also be succesfully applied to traffic light control. In particular, this thesis [6] applies DQN to the problem of traffic light control and extends earlier work [7] to the multi-agent case.

2 Deep Reinforcement Learning for Traffic Light Control

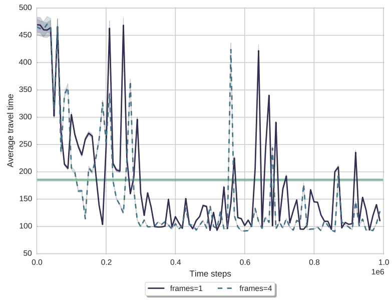
A traffic light agent is trained using the DQN algorithm. The state representation is a binary matrix of vehicle positions, which is fed into a convolutional network. Different settings are compared, by varying amongst others replay memory size, state representation and the use of prioritized experience replay [8]. While the DQN algorithm may suffer from stability issues, caused in part by its convolutional network struggling with non-i.i.d. datasets and moving targets, it can learn directly from image data and find very good policies. However, in traffic light control, taking suboptimal actions may lead to traffic jams, making exploration more costly than in Atari games. To deal with these issues, the best found settings are combined into a fine-tuned DQN agent, which is compared with the untuned agent and a baseline agent that uses a linear approximator in Figure 1a, on the average travel time found during policy evaluations at different points in the training process. While both DQN agents outperform the linear baseline, there is clear oscillation in the graph - the training process is not stable. Fine-tuning, especially adding more information to the state, helps - see for example Figure 1b, where adding past position matrices to the state improves stability.

3 Coordination of Deep Reinforcement Learners

First, the DQN algorithm is used to learn local joint Q-value functions of two neighbouring agents. Then, using *transfer planning* [5], the learned functions are used in the max-plus coordination algorithm [1] to make decisions in a three-agent chain and four-agent cycle. In Figure 2, the fine-tuned agent is compared to the untuned agent and an earlier RL approach [3], evaluated at the best policy found. In both the three-agent and four-agent case, the fine-tuned agent is most stable, while the best policies found by both agents are close in performance. However, in the four-agent scenario, the untuned agent has some huge travel time spikes, suggesting that the error in estimating the Q-value function stacks with the lack of guarantees for max-plus in cyclical graphs, resulting in suboptimal coordinated policies.

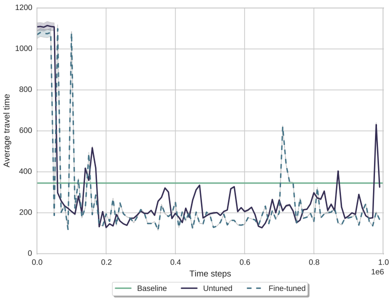


(a) Average travel time. Fine-tuned versus untuned DQN agent, linear baseline.

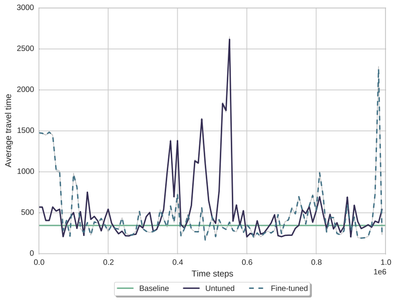


(b) Average travel time. One versus four state frames.

Figure 1



(a) Average travel time, three-agent scenario



(b) Average travel time, four-agent scenario

Figure 2

4 Conclusion

Traffic light control presents unique challenges not necessarily present in the benchmarks used in earlier work. Moreover, while it can outperform earlier work, DQN for traffic light control suffers from stability issues. However, in principle the approach of using deep reinforcement learning for learning source problems in transfer planning is promising and there are many directions for future research that can make this approach more reliable, both in general and for traffic light control.

Acknowledgements

The Tesla K40 used for this research was donated by the NVIDIA Corporation. The thesis was supervised by Frans A. Oliehoek.

References

[1] C. M. Bishop. *Pattern Recognition and Machine Learning*. Springer-Verlag New York, Inc., 2006. ISBN: 0387310738.

[2] Commission of the European communities. *White paper European transport policy for 2010: time to decide*. 2001.

[3] L. Kuyer et al. “Multiagent reinforcement learning for urban traffic control using coordination graphs”. In: *ECML-PKDD*. 2008.

[4] V. Mnih et al. “Human-level control through deep reinforcement learning”. In: *Nature* 518.7540 (2015), pp. 529–533.

[5] F. A. Oliehoek, S. Whiteson, and M. T. Spaan. “Approximate solutions for factored Dec-POMDPs with many agents”. In: *AAMAS*. 2013.

[6] E. van der Pol. “Deep reinforcement learning for coordination in traffic light control”. Master’s Thesis. University of Amsterdam, 2016.

[7] T. Rijken. “DeepLight: Deep reinforcement learning for signalised traffic control”. Master’s Thesis. University College London, 2015.

[8] T. Schaul et al. “Prioritized experience replay”. In: *ICLR*. 2016.