Heterogeneous Multi-Agent Deep Reinforcement Learning for Traffic Lights Control

Jeancarlo Josué Argüello Calvo
Supervisor: Ivana Dusparic
Master of Science in Computer Science Future Networked Systems
University of Dublin, Trinity College Dublin
2018

People are wasting time in traffic jams that often are caused by an inefficient control of traffic lights in intersections. In recent years several approaches have been taken in order to optimize traffic flow in junctions. The most promising technique has been Reinforcement Learning (RL) due to its capacity to learn the dynamics of complex problems without any human intervention. Different RL implementations have been applied in urban traffic control (UTC) to optimize towards single and multiple agents achieving collaboration. The main problem of these approaches is the curse of dimensionality that arises from the exponential growth of the state and action spaces because of the number of intersections.

RL had a breakthrough when it was combined with Neural Networks to implement a method so called Deep Reinforcement Learning (DRL) which enhances hugely the performance of RL for large scale problems. Cutting edge Deep Learning (DL) techniques have demonstrated to work very well for traffic lights in single agent environments. Nonetheless, when the problem scales up to multiple intersections the need for coordination becomes more complex, and as a result, the latest studies take advantage of the similarity of agents in order to train several agents at the same time. However, the usage of homogeneous junctions is not a real world scenario where a city has different layouts of intersections.

This thesis proposes to use Independent Deep Q-Network (IDQN) to train heterogeneous multi-agents to deal with both the curse of dimensionality and the need for collaboration. The curse of dimensionality is handled by using the Deep Q-Network technique, whose performance and stability is enhanced by using aggregated methods such as Dueling Networks for faster training by computing separately the value and the advantage functions, Double Q-Learning for selecting better Q-values by preventing overoptimistic value estimations and Prioritized Experience Replay for learning more efficient from the experience replay memory by sampling more frequently transitions from which there is a high expected learning progress. IDQN trains simultaneously and separately each agent which allow us to support heterogeneous agents. Unfortunately, this technique can lead to convergence problems because one agent's learning makes the environment appear non-stationary to other agents, and this problem conflicts with experience replay memory on which DQN relies. We address this issue by conditioning each agent's value function on a fingerprint that disambiguates the age of the data sampled from the replay memory.

The proposed solution is evaluated in the widely used open source SUMO simulator. We demonstrate that the proposed IDQN technique is suitable for optimization of traffic light control in a heterogeneous multi-agent setting with the usage of the proposed fingerprint technique that stabilizes the experience replay memory in order to deal with non-stationary environments. We show that it outperforms normal fixed-time and DQN without experience replay.