Traffic Management using Reinforcement Learning Algorithms on Reinflow Platform

Don Kushan Nilanga Athalage

Abstract

This proposal describes using multi-agent reinforcement learning (RL) to learn from dynamic traffic data to control traffic light to lower the wait time on traffic light(Wiering, 2000). RL systems are agents who interact with the environment to learn an optimal value function and behave in optimal policies to control the traffic flow(Wiering, 2000). In this proposal we will be investigating the viable reinforcement learning models that can be used to optimise the traffic flow and how to go about implementing said models inside Reinflow platform. Platform will first be trained on simulators with range of traffic scenarios to learn optimal value functions and implemented parallel to the existing systems(Ghods et al., 2010). With the functionality which have been implemented in the past we can set up and train a RL strategy easily.

Introduction

Traffic control has been on of the most challenging optimising problems. Developing an optimal traffic control policy takes a long time and effort of many engineers are needed. We can automate this process by using RL models and dynamic programming on simulations to build a good enough model to be deployed in real world and adapt and learn from real world scenarios. Since optimising traffic light behaviour manually is a tedious task, use of machine learning approach is the most suitable for problems like this(Wiering, 2000). Reinforcement Learning is a sub domain of machine learning where an agent make actions in an environment and learn from the given environment. There is a range of reinforcement learning methods such as Value iteration, Policy iteration and Policy optimisation using value functions. In the given proposal, a temporal difference (TD) method extension with asynchronous agents with state of the art optimisation algorithm Proximal Policy Optimisation will be used with multiple instances of Renflow Platform.

Reinforcement Learning

Reinforcement learning is a adaptive learning strategy which is emulates the learning strategy used by intelligent beings. Animals interact with the environment and get reward according to how they behave. This simple intuition can be further extended to create intelligent learning agents(Mnih et al., 2015). Reinforcement learning can be simply explained by learning to react to environment in order to maximise a reward.

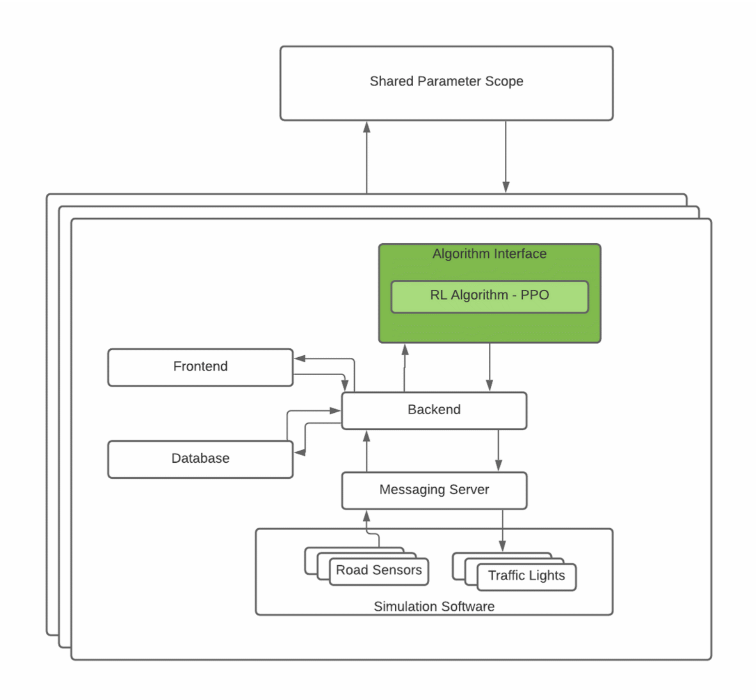
Policy Iteration

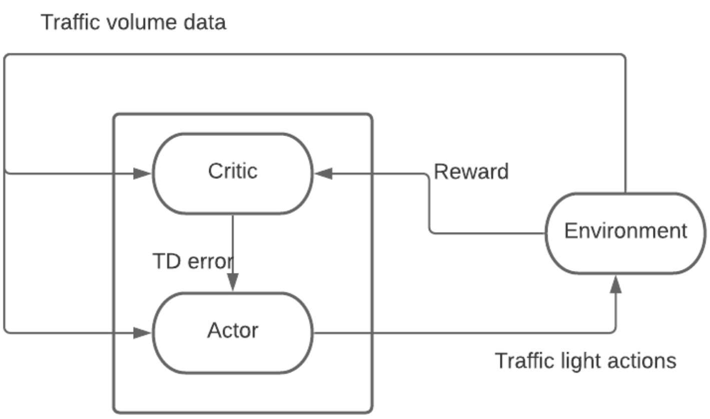
Policy iteration is a sub field of reinforcement learning. The proposed algorithms uses a state of the art policy iteration method. Policy iterations take place the state values(Traffic Sensor data) and provides action values(Traffic Light controls)(Sutton et al., 2000). The policy approximation method we are using is Proximal Policy Optimisation which is introduced by research team at OpenAI which is a breakthrough research founding(Schulman et al., 2017). This policy iteration uses small policy updates over a large training iterations to maintain a stable policy. By using this approach in traffic control our systems will learn to improve reward without making unstable changes.

Asynchronous Agents

Reinforcement learning models require a enormous amount of data. If we use single agent to learn everything the time it would take to learn a good policy will be long. To avoid that we use asynchronous agents. There will be multiple instances of the reinflow platform in different areas with many reinforcement learning agents so that they will share and update a shared parameter set. This will improve the training time and the stability of the model by a huge amount.

Stack Diagram





1.1 RL algorithm PPO

Bibliography

Ghods, A. H., Fu, L., & Rahimi-Kian, A. (2010). An efficient optimization approach to real-time coordinated and integrated freeway traffic control. *IEEE Transactions on Intelligent Transportation Systems*, *11*(4), 873–884.

Mnih, V., Kavukcuoglu, K., Silver, D., Rusu, A. A., Veness, J., Bellemare, M. G., Graves, A., Riedmiller, M., Fidjeland, A. K., & Ostrovski, G. (2015). Human-level control through deep reinforcement learning. *Nature*, *518*(7540), 529–533.

Schulman, J., Wolski, F., Dhariwal, P., Radford, A., & Klimov, O. (2017). Proximal policy optimization algorithms. *ArXiv Preprint ArXiv:1707.06347*.

Sutton, R. S., McAllester, D. A., Singh, S. P., & Mansour, Y. (2000). Policy gradient methods for reinforcement learning with function approximation. *Advances in Neural Information Processing Systems*, 1057–1063.

Wiering, M. A. (2000). Multi-agent reinforcement learning for traffic light control. *Machine Learning: Proceedings of the Seventeenth International Conference (ICML'2000)*, 1151–1158.