

Unconventional reinforcement learning on traffic lights with SUMO

Master Degree in Computer Science

Francesco Refolli

Supervisor: Prof. Giuseppe Vizzari



Outline

- 1 Introduction
- 2 Multi Agent Learning
- 3 Curriculum Learning
- 4 Experiments and findings

Problem statement

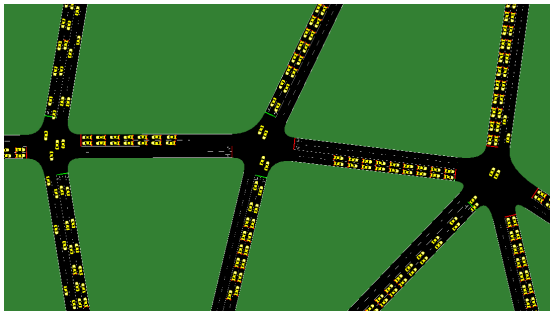
This thesis dealt with the Traffic Light Control problem (TCL) applying uncommon reinforcement learning techniques. In particular, the following research objective were pursuit in the performed experiments:

- Evaluating the effectiveness of Curriculum Learning
- Evaluating the effectiveness of Multi Agent Learning
- Evaluating the effectiveness of Self-Adaptive agents
- Comparing observation/reward functions
- Comparing tabular and deep learning models
- Comparing Reinforcement Learning with existing solutions
- Evaluating the role of hyperparameters



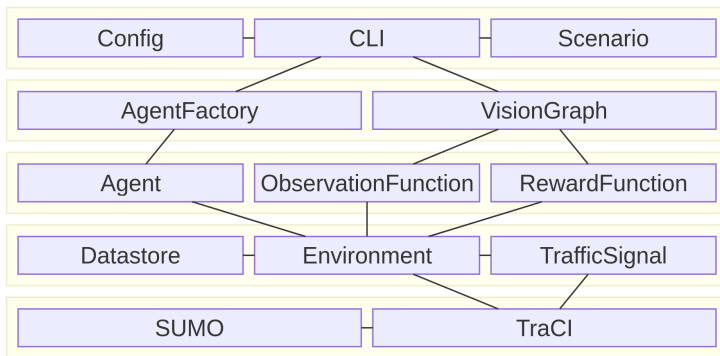
SUMO
SIMULATION OF URBAN MOBILITY

- Free and Open Source microscopic traffic simulator
- Developed at German Aerospace Center (DLR)
- Multimodal: cars, trams, bikes, pedestrians ...
- Highly customizable



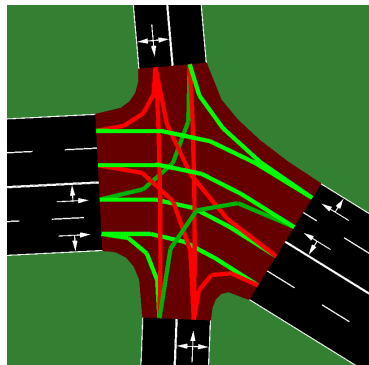
SUMO-RF: SUMO + Reinforcement Learning

A FOSS framework for Reinforcement Learning with SUMO developed as fork of *LucasAlegre/sumo-rl* with a focus on modularity, flexibility and Multi Agent Learning. It also contains several utilities for format conversions, metrics analysis and plot, schematic-based demand generation and more.

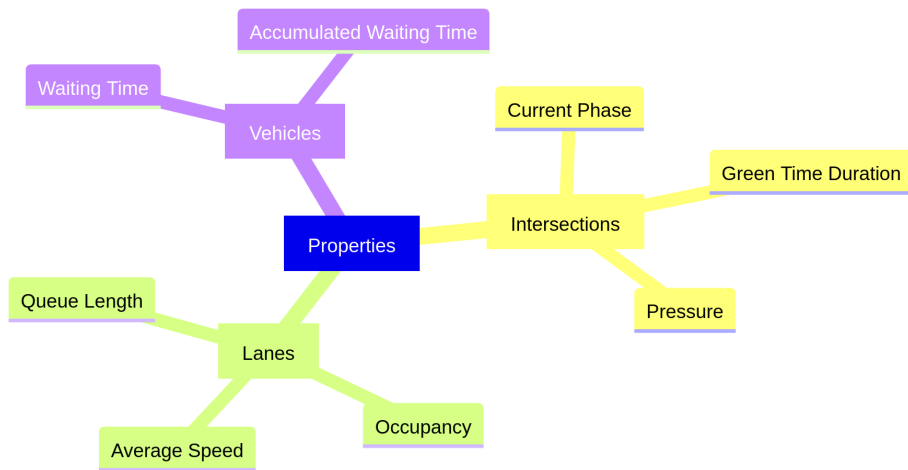


The Agent Model

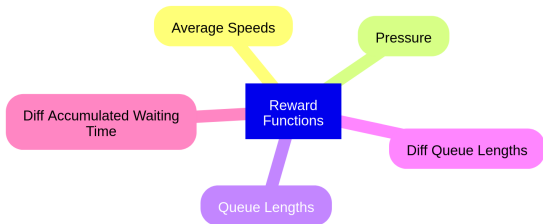
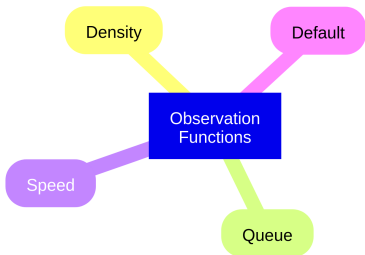
- Each agent can control one intersection and at each step (every 5 seconds) it can choose the next phase of the intersection.
- Every action is automatically enforced by TrafficSignal with also an intermediate yellow phase.
- It receives an observation of the current condition and a reward proportional to the goodness of its behaviour.
- If the agent is "smart", it uses the collected data to improve itself!



The Global State



Observing and Rewarding

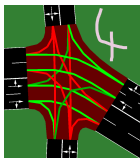
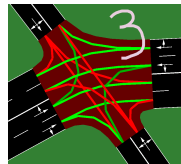
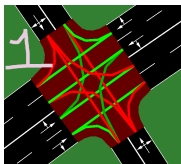


Multi Agent Learning

No agent is isolated, they are all part of a whole and they influence each other with their own behaviour.

What if an agent can *sense* its surrounding area by sharing observations with neighbours?

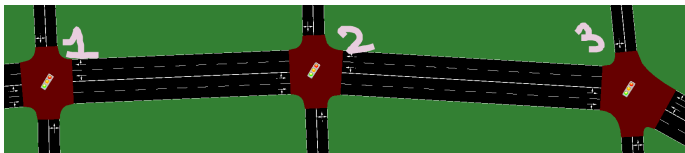
What if an agent is *rewarded* for its influence on surrounding area by sharing rewards with neighbours?



Observation sharing

- The state S of an agent is computed through two distinct observation functions f_{int} , f_{ext} meaning respectively the internal state and external state.
- The state of an agent x is a concatenation of $f_{int}(x)$ and $f_{ext}(y)$ for all $y \in N(x)$, where N is a function mapping an agent with its neighbours.
- By default, f_{int} is the Density function which is the one that without sharing gives best results.

Example:

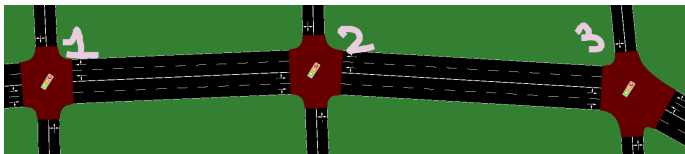


$$S(TLS_1) = f_{int}(TLS_1) \bullet f_{ext}(TLS_2) \quad S(TLS_2) = f_{int}(TLS_2) \bullet f_{ext}(TLS_1) \bullet f_{ext}(TLS_3)$$

Reward sharing

- The reward R of an agent is computed through one single reward function f representing the condition of an intersection.
- The state of an agent x is a sum of $f(x)$ and $f(y)$ for all $y \in N(x)$, where N is a function mapping an agent with its neighbours.

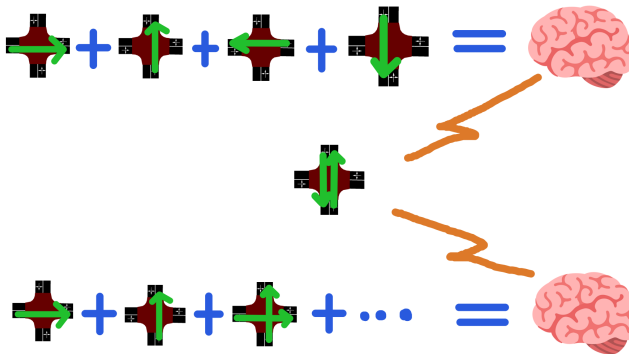
Example:



$$R(TLS_1) = f(TLS_1) + f(TLS_2)$$

$$R(TLS_2) = f(TLS_2) + f(TLS_1) + f(TLS_3)$$

Curriculum Learning



Monolithic Approach

Thank You