Self-Driving Agent

Summary

The intention of this exercise is to teach a self-driving car (Agent) to automatically drive to reach the final destination maximizing rewards. Q learning algorithm is implemented.

Key Factors:

Rewards:

Any time the agent acts, a reward is reached, the possible rewards are: - 10: Agent reaches destination - 2: Agent action is valid and equals the next*waypoint() policy - 1: Agent action is None - 0.5: Agent action is valid but is not the next*waypoint() policy - -1: Agent action is not valid

Planner:

Agent uses a planner to reach destination, this planner identifies the next waypoints agent should follow to reach destination in a grid environment. However planner is limited since the process to find the optimun waypoints firstly checks for optimal waypoints in the North-South direction and, in case None is found, then in the East-West direction. As a consequence, in cases where there are two optimal waypoints, only the waypoint in the North-South direction is reported.

Q-Learning:

In order to provide the ability to the agent of learning from previous states, a Q-matrix is generated. Each element of this Q matrix is an agent state&action which includes the following information: - Street light color - Oncoming traffic - traffic on the left - traffic on the right - Action taken

Each state&action in the the Q-matrix has a particular value that is updated accordingly to the regard obtained. At the very begining of the game, all possible pairs of states and actions had been assigned to a value of 0 but as the agent walks through the different states and takes different actions received rewards are used to update values. Values are updated with the received reward of the current state plus the action that maximizes Q-value for the next state.

Results:

In the following graph it is represented the total number of different rewards for a simulation of 100 trials for an "intelligent" and "non intelligent" agent. The intelligent agent uses the Q equation to identify the optimum actions while non-intelligent agent simply follows those actions towards the next waypoint indicated by the Planner.

In the way rewards are defined, since an action=None has a rewards of 1, while an action that is valid but not to the next_waypoint is rewarded with 0.5, it is not expected the intelligent agent to be faster than the non-trained agent reaching destination, but an overall increase of the rewards obtained, as the following graph shows:

Rewards in 100 trials for a trained and non trained Agent