Risk Averse Path Planning Using Lipschitz Approximated Wasserstein Distributionally Robust Deep Q-Learning

Cem Alptürk
Supervisor: Venkatraman Renganathan
Examiner: Anders Rantzer

June 2022

Reinforcement learning is a powerful tool to find solutions to problems where a clear answer may not be available. An agent can interact with its environment to learn how to solve a given problem over time by interpreting its past experiences. A common usage for reinforcement learning is achieving human-like performance in games.

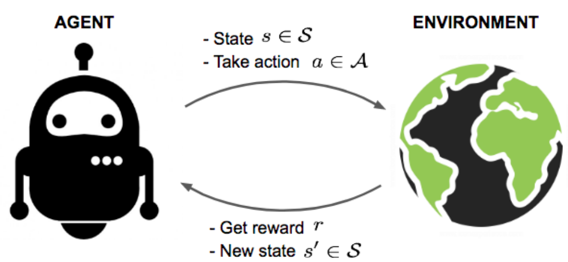

Figure 1: Illustration of reinforcement learning.

Another application is for path planning for an autonomous robot in an environment with obstacles. With the increase in autonomous driving and robotics, the need for safe path planning becomes more trivial. Reinforcement learning is becoming a popular tool in finding a path from point A to point B. However if the robot dynamics have some uncertainties involved such as process disturbances, it becomes even more important to find a safe path rather than the shortest path.

The agent is able to learn a path from any point in the environment to any goal position by learning the relationship between the actions it takes and the rewards it receives for these actions, where reaching the goal results in a positive reward, where colliding with an obstacle results in a negative reward. This can be compared to a child learning something new where good behavior is rewarded and undesired behavior is penalized. By learning what actions to take in order to get the best rewards, the agent is able solve a desired problem. From the robots perspective, the goal is to maximize the rewards it gets where this corresponds to finding a path to the goal position. Common reinforcement learning methods learn the average rewards, due to the randomness in the environment. However, the average reward is not enough to predict black swan events that occur rarely Since we desire a safe path, the robot tries to estimate the worst case rewards for taking an action at a certain position in the environment. By looking at the worst case rewards, it can become risk-averse, where getting too close to an obstacle may result in a collision and thus the robot decides to stay away from it as much as possible. Becoming risk-averse also results in taking sub-optimal paths to the goal, thus there is a trade off between safety and optimality.

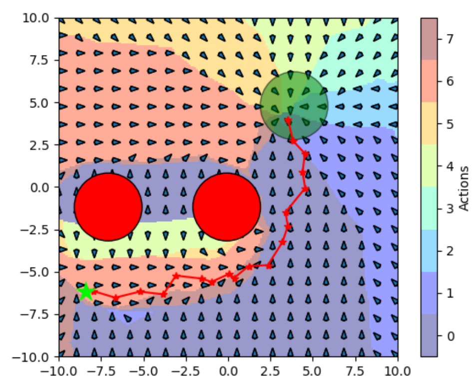

Figure 2: For any position in the environment the agent can find a safe path the goal (green) by following the arrows without colliding with the obstacles (red). The arrows around the obstacles and borders point away since the risk of collision becomes too high and causes the robot to have a longer path without getting too close to the obstacles.