Aman Singh

Project 2 Wrapper

Reinforcement learning is a powerful technique for problem-solving in environments with stochastic actions. As with any Markov Decision Process, the reward function dictates what is considered optimal behavior by an agent. Since a reinforcement learning agent is trying to find a policy that maximizes expected future reward, changing when and how much reward the agent gets changes its policy.

However, if the reward function is not specified correctly—rewards are not given for the appropriate actions in the appropriate states—the agent's behavior can differ from what is intended by an AI designer. Consider the boat racing game pictured above, where the goal, as understood by people, is to quickly finish the race. Humans have no difficulty playing the game and driving the boat to the end of the course. However, when a reinforcement learning agent learns how to play the game, it never completes the course. In fact, it finds a spot and goes in circles until time runs out. You can see the RL agent in action in this video: https://youtu.be/tlOIHko8ySg . The agent's reward function is the score the player receives while playing the game. Score is given for collecting power-ups and doing tricks, but no points are given to players for completing the course.

Question 1: Watch the video and explain why the agent's policy has learned this circling behavior instead of progressing to the end of the course. Explain the behavior in terms of utility and reward.

- Since points are achieved independently of whether the race is completed or not, the utility function simply tries to achieve as much point as fast as possible by circling doing tricks and picking up power-ups

Aman Singh

Question 2: When humans play, the rules for scoring are the same. Score is a way for games to give feedback about how well the player is doing. Why do humans play differently, always completing the course? That is, why don't humans circle in the same spot in the course if they are receiving the same score feedback as the agent?

- Because humans know that while having a higher score is nice, the race must be completed in order to end the game. Also, to win the race, the player must be first. Putting these two in mind, humans will generally try to finish the race first and put points as a nice bonus.

Question 3: The agent's original reward function is $R(s_t,a) = game\_score(s_t) - game\_score(s_{t-1})$. Describe—in terms of utility, reward, and score—**two (2)** ways one could modify the reward function to get the agent to behave more like a human player. That is, what do we need to change to make the agent complete the course every single time? Assume the agent has access to state information such as the position and speed of the boat and rival racers, but we cannot change how the game itself provides scores.

- One way we can modify the reward function is to have the relative position of the boat to rival racers be paramount to the game score. The reward function always prioritizes being in first place and tries to get more points after it has successfully reached first place.
- Another way we can modify the reward function is to reward it when it stays in the track and moves fast along the track. This rewards the agent to move along fast in the track, hopefully winning the race along the way, like a human would.

Question 4: Self-driving cars do not use reinforcement learning for a variety of reasons, including the difficulty of teaching RL agents in the real world (instead of a simulation or computer game). Suppose however, that you tried to make a reinforcement learning agent that drove a taxi. The agent is given reward based on how much fare is paid, including tips. Describe a scenario in which, after the taxi agent has learned a policy, the autonomous car might choose to do an action that could put either the rider, pedestrians, or other drivers in danger. If you think there is **not** such a scenario, explain how the reward function might be altered to cause the autonomous car to learn a policy that endangers the rider, pedestrians, or other drivers.

- A taxi might will most likely rewarded by its passengers for getting to their destinations faster. This can be detrimental in the long run, as the taxi might be driving dangerously fast to get to its destination, or maybe even ignore traffic rules that may slow the taxi down.