CS 3600 Project 2 Wrapper

CS 3600 - Fall 2023

Due October 13th 2023 at 11:59pm EST via Gradescope

Introduction

This Project Wrapper is composed of 4 questions, each worth 1 point. Please limit your responses to a maximum of 200 words. The focus of this assignment is to train your ability to reason through the consequences and ethical implications of computational intelligence, therefore do not focus on getting "the right answer", but rather on demonstrating that you are able to consider the impacts of your designs.

Context

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 (meaning rewards are not given for the appropriate actions in the appropriate states) the agent's behavior can differ from what is intended by the AI designer. Consider the boat racing game pictured above. 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 like we expect from a human player. Explain the behavior in terms of utility and reward.

Answer: The reinforcement learning agent's circling behavior in the boat racing game can be linked to a mismatch between the reward function and the intended goal of the game. Unlike humans, who can immediately grasp the objective of reaching the finish line, the agent's policy is shaped by the reward that it will receive. In this game, the only rewards provided are for collecting those turbo points, with no reward given for completing the course and reaching the finish line. As a result, the agent's policy prioritizes actions that maximize its reward within the constraints of the reward function. Circling in one spot and performing tricks, rather than completing the course, becomes the optimal strategy in the eyes of the agent because it continually accumulates rewards through these actions.

Question 2

When humans play, the rules for scoring are the same. Why do humans play differently then, always completing the course? Why don't humans circle in the same spot in the course endlessly if they are receiving the same score feedback as the agent?

Answer: Humans play differently from the reinforcement learning agent in the boat racing game because of the inherent understanding of the broader context. Humans' ability to infer the intended goal beyond the explicit rules of scoring eventually lead humans to finish the course regardless of the same score feedback. We bring common-sense knowledge and intuition to the game, realizing that the ultimate aim is to cross the finish line. Unlike the AI agent, we don't rely solely on immediate rewards and rather the overarching objective, which is to complete the race successfully. In essence, our cognitive abilities allows us to understand the game's true objective and lead us to act in a way that aligns with the intended experience.

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** 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 all rival racers, but we cannot change how the game itself provides scores through the call $game\_score(s_t)$.

Answer: To encourage the reinforcement learning agent to behave more like a human player and consistently complete the course, we can modify the reward function in two distinct ways while keeping the game's scoring mechanism, $game\_score(s_t)$, unchanged.

Course Completion: One approach is to introduce a reward that explicitly rewards the agent for reaching the finish line. We can add a reward component that only provides a significant positive reward when the agent crosses the course's endpoint. By doing this, the agent's policy will start associating course completion with higher rewards, aligning its behavior with human tendencies.

Time-Based: Another effective modification is to introduce a time-based reward. The agent should receive progressively diminishing rewards as time passes. This time penalty could even be added as a negative reward for each time step or as a decreasing reward. With this change, the agent would realize that taking longer to finish the course results in a lower overall score, encouraging it to complete the race as quickly as possible, much like a human player who inherently values speed in achieving the goal.

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, and the dangers of a taxi accidentally learning undesired policies as we saw with the boat game example. 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 for the ride, including tips given by the passenger. Describe a scenario in which, after the taxi agent has learned a policy, the autonomous car might choose to do an action that puts either the rider, pedestrians, or other drivers in danger.

Answer: If the reward for the taxi agent is solely based on the amount of fare collected without considering safety as a crucial factor, the agent may prioritize rapid and risky behavior to maximize profits. It might choose to speed through traffic, disregard traffic rules, or make sudden lane changes to get passengers to their destinations quicker. While these actions may result in higher fares and potentially larger tips, they come at the expense of safety. Such behavior could put passengers at risk by increasing the likelihood of accidents or injuries during the ride. Additionally, it poses a significant danger to pedestrians and other drivers on the road, as the agent's focus on revenue generation could lead to reckless actions that compromise overall road safety.