Yujin Hwang

Mark Riedl

CS 3600

10/10/2023

Wrapper 2

1) The agent's behavior of circling in the boat racing game instead of progressing to the end of the course can be attributed to the reward function. Because the agent's policy, which determines its actions, is shaped by the expected future reward, the reward function incentivizes actions that maximize the player's score, particularly by collecting power-ups and performing tricks. However, it fails to provide any reward for completing the course, so the agent simply follows the most reliable path to achieve its objective of maximizing reward: circling in one spot to repeatedly collect power-ups and perform tricks. This behavior allows the agent to accumulate rewards without the uncertainties associated with completing the course which does not promise any rewards which is why the agent does not act to progress the game like a human.

2) Human players are driven by an understanding of the game's objective of completing the race quickly. Our reward system is not solely based on the immediate rewards of power-ups and tricks but inherently values the satisfaction of finishing the race successfully. Our complex decision-making skills give humans the ability to weigh the short-term rewards such as collecting power-ups and doing tricks against the long-term goal: course completion. This decision-making is beyond the capability of most reinforcement learning agents, which often focus on maximizing immediate rewards. This fundamental difference in reward structures between humans and the reinforcement learning agent leads to agents circling in the same spot in the course endlessly while humans don’t even though the rules stay the same.

3) To modify the reward function to make the agent behave more like a human player and complete the course every single time, we need to align the rewards with the desired behavior of course completion or penalize the behavior of circling. First, we can modify the reward function to provide a positive reward when the agent completes the course. This change will motivate the agent to prioritize finishing the race. Additionally, the utility for other actions, such as collecting power-ups or performing tricks, will be lower, reflecting their importance compared to completing the course. Another approach is to penalize the agent for engaging in repetitive circling behavior. This can be achieved by introducing a negative reward for a lack of progress or excessive looping. In this modified reward function, a penalty term can be added in proportion to the increase in distance from the starting point. This encourages the agent to keep making forward progress and discourages looping behavior.

4) To earn more money quickly, the taxi agent can decide to aggressively overtake a slower-moving vehicle dangerously. This action could involve speeding, changing lanes without blinkers, or even running a red light and other illegal activities. The agent does this to get the passenger to their

destination more quickly and increase the fare, potentially earning a larger tip. Aggressive driving could cause discomfort or even injury to the passenger as well as pedestrians and other drivers. Sudden acceleration, sharp turns, and other dangerous driving actions can put the rider in danger. It can lead to accidents, collisions, and harm to individuals directly and not directly involved in the taxi ride.