Matthew Muller

Dr. Balogun

CS-370: Current/Emerging Trends in CS

16 April 2023

**Design Defense**

**Analyze the differences between human and machine approaches to solving problems**

If a human being were to attempt to solve the maze of this game, they would likely begin by following the longest visible pathway until they hit a wall. From here they would likely repeat this strategy on the longest pathways that they do not think they have explored yet, backtracking when no new continuation along the given path is available.

The intelligent agent that controls the pirate character of this game solves the maze by learning about its layout through a system of rewards and penalties. The agent is able to move in one of four directions: left, right, up, or down. For every action, the agent receives a reward of either positive or negative points depending on the result of the action. When the agent reaches the target, the reward is 1 point. Moving to an occupied cell results in a penalty of -0.75 points. Attempting to move outside the matrix boundary results in a penalty of -0.8 points and moving from a cell to an adjacent cell results in a penalty of -0.04 points, primarily to avoid the agent wandering within the maze.

One of the main differences between the way that a human would solve this maze compared to the intelligent agent is that the human would not be concerned with a penalty for exploring the maze. They would also not have to be concerned with penalties for moving outside the boundaries or into an occupied cell, as they would be able to see these obstacles and plan their course accordingly. While the two have the same pursuit of reaching the treasure, the intelligent agent would end up finding a much more efficient path because it is penalized as it makes moves that do not reach this goal.

**Assess the purpose of the intelligent agent in pathfinding**

In a pathfinding problem such as this, it is important to find the right balance between exploration and exploitation. Exploration is when the agent makes moves that it does not have prior knowledge of with the goal of learning something new about its environment. Exploitation is when the agent makes a move that it already knows the result of with the goal of maximizing the reward that it receives. For the intelligent agent in this problem, an exploration factor of 0.1 serves well to train the network towards a 100% win rate. This factor means that for each action, the agent will have a 90% chance to make a choice that it knows will result in the highest reward while 10% of the time it will attempt to learn a new path. With this exploration factor, my testing of the system found that the agent was able to reach a 100% win rate within a range of 288 to 1032 epochs.

Reinforcement learning helps the agent determine the path to the goal by providing a method for the agent to map the maze through hundreds/thousands of games based on the rewards system. Over many playthroughs, the agent gradually determines which paths are the most successful in reaching the treasure fastest because it is penalized for excessive or unsuccessful moves and rewarded for reaching the goal. With each new game, the agent uses the information that it gathered in previous games to inform its actions towards the goal.

**Evaluate the use of algorithms to solve complex problems**

The goal of the deep Q-learning algorithm in this game is to find a path through a maze, presented as an 8x8 matrix, and reach the treasure cell as fast as possible. To begin, the agent is placed in a random free cell. From here, an epsilon-greedy exploration strategy determines whether the agent selects the random action or draws from its existing knowledge to seek the highest reward. When the agent takes an action, the algorithm stores the state of the environment (envstate), the reward, and the game status. If the pirate has won the game (reached the treasure cell), a 1 is appended to the win history and a 0 is appended if the pirate loses. These numbers are used to determine the pirate’s win rate and the agent has completed its training when it is able to reach a 100% win rate.