**Project Two**

Tessa D. Parker

Department of Computer Science, Southern New Hampshire University

CS 370: Current and Emerging Trends in Computer Science

Prof. Timothy Alexander

December 10, 2023

In this project, we have been asked to design an intelligent agent for a gaming company’s treasure hunt maze. In this maze, the intelligent agent is a pirate that will be competing against a human player to find the treasure at the end of the maze. The scenario that we are presented with is to solve a pathfinding problem using a Deep Q Network (DQN) implementing reinforcement learning (RL) to train the agent to become a competitive player against its human counterparts.

In order to address the problem of artificial intelligence (AI) being competitive against a human player, we should consider how it may behave in comparison to a human player in its decision-making process. A human player when faced with a maze would certainly most strongly rely on their eyesight, intuition, and memory to make decisions. These factors would vary from individual to individual based on previous knowledge and experience. It would introduce a lot of inconsistency and uncertainty into the environment based on the player. In contrast, AI uses learning algorithms to mathematically deduce outcomes and results. This often happens very rapidly. While the number of options in many circumstances may be vast, a well-trained agent will find the correct outcome seemingly almost instantaneously in comparison to its human counterparts.

The maze is set up on a grid where the agent can move left, right, up, or down. The pirate agent is penalized for being blocked, revisiting squares, or making invalid moves. On the other hand, the agent is rewarded for reaching the treasure and does not receive a penalty for making valid moves. This positive feedback system informs the agent to make intelligent moves moving forward.

Human cognition is prone to inconsistency and loss. We tend to misremember details and our thinking is disorganized with many influences and outside factors. Intelligent agents function in a strictly rational, mathematical manner. This means that they will strategically work through a problem with little to no data loss if the neural network is correctly configured. Where humans have an advantage, however, is in intuition. A human can see ahead in a problem and may not walk blindly into an obstacle. Whereas the agent in this scenario will only have awareness of the current square on which it rests and will methodically explore all available options, even if it means running into a wall or a dead end.

Exploration and exploitation serve different purposes in the realm of reinforcement learning. Exploration allows the agent to travel paths not yet tried in the hopes of gaining new and useful information. This added information is then used to improve the long-term performance of the agent and aid in better decision making in the future. Exploitation is often referred to as the ‘greedy’ approach. It uses current knowledge to choose actions that will provide the highest reward and maximize the short-term performance of the agent.

Reinforcement learning informs the agent on how to optimally achieve a goal by offering rewards and penalties based on behaviors and outcomes achieved. This system trains the agent to complete the maze in the most efficient way possible. As was mentioned above, penalties are given for undesirable behaviors that do not move the pirate agent towards the goal. Rewards are given for completing the goal and no penalty is given when no undesired behavior is encountered or detected.

The Q-learning algorithm is a popular reinforcement learning technique designed to find the outcome with the maximum reward value. It operates by balancing exploration and exploitation to learn through trial and error. I implemented Q-learning in this project by developing a neural network that allows the agent to iteratively observe the state of the environment, choose the best course of action, perform the action and receive/observe rewards and state changes. This process is done repeatedly in each episode. Each epoch runs as many episodes as necessary until either a win or loss state is reached. Through these repeated iterations, the Q-values are gradually updated and converge to their optimal values. Eventually the best action to take in each state will become clear.

References:

Yang, A. (2022, July 25). *What is exploration vs. exploitation in reinforcement* *learning?*. Medium. <https://angelina-yang.medium.com/what-is-exploration-vs-exploitation-in-reinforcement-learning-a3b96dcc9503>

Hashemi-Pour, C., & Carew, J. M. (2023, August 16). *What is reinforcement* *learning?: Definition from TechTarget*. Enterprise AI. <https://www.techtarget.com/searchenterpriseai/definition/reinforcement-learning>

Singh, N. (2023, July 11). *Understanding Q-learning: A powerful reinforcement learning technique*. Medium. <https://medium.com/@navneetsingh_95791/understanding-q-learning-a-powerful-reinforcement-learning-technique-29a3da36f611>