**Matt Knutson**

**CS-370: Current & Emerging Trends In Computer Science**

**SNHU - Instructor Richard**

**October 20th, 2024**

**Project 2 - Design Defense**

Over the last couple of months, this class has covered the many similarities and differences that machines and humans have. Although they both have a lot in common, machines are really only barely scratching the surface of organic intelligence. But machines do have the advantage of computing a process, or scenario, at an exponentially quicker rate than humans. This allows them the benefit of analyzing a much larger set of targeted information, giving them insight that would take humans much longer to acquire.

In the Treasure Maze situation, a human would begin by moving around in the environment and keeping track of where they had already traveled. The person would navigate through the maze using their senses. If they visually came to a dead end, they would become flushed with panic realizing they were stuck. This would make them turn around and continue exploring the maze. In the case of coming in contact with an area associated with danger, the person would feel physical pain, or mental anxiety. The person uses their memory to navigate all of these obstacles, and their reward is eventually getting to the gold and freeing themselves from the maze.

Likewise, the A.I. agent begins by traveling around the maze. The map is broken into a grid for the agent, and different squares in the grid are associated with reward values based on actions taken in that state. The agent retains a memory of these action/state pairs and their rewards, and then uses them to determine its next move. Eventually, after a period of many runs, the agent can begin to reference entire attempts at completing the task, instead of individual steps, and use these to better understand how it will approach executing its next attempt at reaching the treasure.

Both the human and agent are using memory to complete the task, but the agent uses only memory and accumulated reward values to make its way through the maze. However, this memory is much different than the human memory, as there is a perfectly preserved copy of every game played that is completely accessible and understandable to the agent at all times. The human memory is foggy, highly suggestable, and its reliability can fall heavily on the human’s state of being at any given time. In a maze situation, there could be many downfalls to relying on one’s memory in a particularly hectic situation. But humans can still fall back on their physical senses, intuitions, emotions, and their overall awareness of existence to assist them in finding the treasure and escaping the maze.

In pathfinding, the intelligent agent is the entity responsible for exploring the environment. It is rewarded and penalized for its actions while in different states and keeps track of the outcomes in order to optimize its behavior. Exploitation and exploration are added into the agent's design to give it access to different approaches when exploring the maze. Exploitation points the agent in the direction of taking actions that have already been proven to work. While

exploration pushes the agent to try new actions and take chances, hopefully leading to the discovery of larger rewards. Both of these characteristics are implemented in one variable called epsilon, which is kept on a setting between 0 and 1. The recommended setting for epsilon is 0.1, where the agent is using proven tactics nine out of ten times, and exploring uncharted territory only one time out of those ten. My personal preference for epsilon would be 0.3, allowing the agent to explore a couple more times and giving them the chance to discover new possibilities. This intuitively makes sense to me, but I have not worked in Keras enough to provide evidence for it being a more efficient setting.

The game in this project uses a Q-value chart that is updated by each iteration the agent completes going through the maze. The values for each state and reward combination are documented in the chart and used by the agent as a reference. A designated number of iterations takes place to populate the chart and then that accumulated data becomes the Q-value for the agent's next attempt at the maze. Eventually, the pirate agent will have an optimized routine that will allow it to successfully complete the maze and get the treasure every time. This will also result in the agent being quick enough to beat its human counterpart on most occasions.

**Citations**

Dugmeci, D. (2021, November 10). Finding The Shortest Path Using Q-Learning Algorithm. Medium.

https://towardsdatascience.com/finding-shortest-path-using-q-learning-algorithm-1c1f39e89505

GeeksforGeeks. (2023, January 23). Deep Q-Learning. GeeksforGeeks.

https://www.geeksforgeeks.org/deep-q-learning

GeeksforGeeks. (2024, June 16). What Is Markov Decision Process (MDP) And Its Relevance To Reinforcement Learning. GeeksforGeeks. https://www.geeksforgeeks.org/what-is-markov-decision-process-mdp-and-its-relevance-to-reinforcement-learning/?ref=oin\_asr1

Introduction To Advantage Actor-Critic Method (A2C). PyLessons. (2020, March 20).

https://pylessons.com/A2C-reinforcement-learning

Juliani, A. (2017, June 30). Simple Reinforcement Learning With Tensorflow Part 8: Asynchronous Actor-Critic Agents (A3C). Medium.

https://awjuliani.medium.com/simple-reinforcement-learning-with-tensorflow-part-8-asynchronous-actor-critic-agents-a3c-c88f72a5e9f2

Kapoor, S. (2018, June 2). Policy Gradients In A Nutshell. Medium.

https://towardsdatascience.com/policy-gradients-in-a-nutshell-8b72f9743c5d

Karagiannakos, S. (2018, November 17). The Idea Behind Actor-Critics & How A2c & A3C Improve Them. AI Summer.

https://theaisummer.com/Actor\_critics

Mandy. (2018). DeepLizard Reinforcement Learning. YouTube.

https://www.youtube.com/channel/UC4UJ26WkceqONNF5S26OiVw

Reinforcement Learning: Exploring Policy vs. Value-Based Methods. Data Headhunters. (2024, January 7). https://dataheadhunters.com/academy/reinforcement-learning-exploring-policy-vs-value-based-methods

Tkhan. (2022, January 2). Reinforcement Learning – Exploration vs Exploitation Tradeoff. AI ML Analytics.

https://ai-ml-analytics.com/reinforcement-learning-exploration-vs-exploitation-tradeoff