CS - 370

Professor Wilson

Nicholas Altemoos

04/19/2025

**Project Two**

The way that humans and AI solve problems are inherently different. Humans will solve the maze with different stimuli than the AI would and utilize a combination of logic, memory, and instincts. Humans can perceive stuff visually and see walls, and paths. The initial choice is based on visible paths instead of trying each direction till it works and then moves through the maze. As they move through the maze humans keep track of what moves they have made and try to identify patterns to avoid the same mistakes and backtrack. Humans will repeat this until they solve the maze. AI on the other hand solves problems using Deep Q-Learning which is a repetitive and data driven process. The AI starts by taking the current state and then choosing an action either based off the best option available or random choice. From that action it receives a reward and stores the experience in the replay buffer and repeats over and over as it updates its q-network so it can predict better moves. As it learns it will reduce the amount of exploration it does as it becomes more confident in its policy to solve the problem. They are both similar in the sense that both parties refine their strategies with past actions and their results over time. The key difference is that humans have a higher level of reasoning and visualization, and machines have learning through trial and error and feedback.

The purpose of the intelligent agent in this problem is to function as a “pirate” and navigate the grid to find the “treasure.” With path-finding its purpose is to find an efficient strategy for finding the most rewarding path to the goal. While doing this the agent utilizes both exploration and exploitation. Exploration is when the agent tries new actions even when the action is uncertain. Exploitation is when the agent chooses an action it already knows the best rewards for based on past experiences. The main difference is that exploration is about the agent learning while exploiting is about getting the most rewards. There needs to be a balance between the two. In this problem we start with a high exploration rate in the beginning so the agent can gather information about the environment and then start to have less exploration and exploitation as the policy is being developed. Here we used a decaying e-greedy policy to implement this concept. Overall reinforcement learning helps the agent to find the treasure with the agent learning and reinforcing its policy. It updates its policy as it receives reward values from their actions and estimates future values and iterates over and over.

Algorithms are used to solve complex problems by giving step by step instructions that can be followed. In this scenario we used a Deep Q-learning algorithm, we use this algorithm because it lets the agent learn from past experiences and create a policy to estimate the best way forward instead of exhausting all the options. We start with a neural network that takes the current state and gives q-values for all actions. Next, we make a replay buffer for experiences that the agent can sample from past experiences. We also have a target network that is periodically updated to offset from the main network to add stability. A e-greedy policy is also implemented to help the rate of exploration decay as the AI learns and reinforces its policy. The AI then goes through iterations till it reaches the goal.

**References**

Liu, J. (2021). *Comparing Human and AI Behavior in 3D Navigation Environments*. https://www2.eecs.berkeley.edu/Pubs/TechRpts/2021/EECS-2021-111.pdf

Schoberg, S. (2020, April 14). *Introduction to Reinforcement Learning (Q-Learning) by Maze Solving Example*. Analytics Vidhya. <https://medium.com/analytics-vidhya/introduction-to-reinforcement-learning-q-learning-by-maze-solving-example-c34039019317>

Edwards, S., Dorn, B., & Sanders, D. (2012). *4. Problem Solving and Algorithms*. Vt.edu. http://sofia.cs.vt.edu/cs1114-ebooklet/chapter4.html