Afshin E. Ahvazi

Professor Dewin Andujar

CS-370-X6163

August 18, 2023

Project 2: Q-Learning in Path-Finding Game

The differences between a human’s and an AI’s ability to solve a problem becomes apparent when analyzing what approaches each agent takes to solve the problem. As humans, we would start by studying the environment. At this stage, a few logical questions should arise that need to be answered: what is the starting point and the goal of the game? What and where are the obstacles? And, most importantly, what are the possible solutions to the problem? In this case, the last question translates to, how could the pirate reach his treasure? Often, an obvious solution may not be immediately available. This is when a human considers all these questions and the available information to logically derive at a solution. We would start moving the pirate character step by step, navigating around obstacles, and making decisions based on the surroundings. If the path is blocked, we would backtrack a bit and explore alternative routes. Eventually, through trial and error, a human learns which paths lead to dead ends and which paths are more likely to lead to the treasure.

To solve this problem, an AI system will need data regarding the system. In this example, it uses an 8 by 8 grid of cells to simulate the environment, where each cell will have values that signify if they are occupied or free, consequently rewarding or punishing the system. The AI system also needs to define the state and action system by representing the pirate’s movement (left, up, right, and down) using pre-defined values (like 0, 1, 2, and 3). Initially, the AI’s algorithm has to choose between the exploration and exploitation methods of gathering the necessary feedback information. Exploration, as the name suggests, involves taking actions that might provide new information to improve the agent's understanding of the environment. Exploitation, on the other hand, means taking actions the AI believes would maximize the rewards based on existing information (Khan, 2022). The ideal proportion of exploitation and exploration depends on the stage of learning: generally, exploration is more beneficial early on because it allows the AI to discover the possible paths and rewards. However, as the AI gains more knowledge and is better capable of predicting rewarding paths, exploitation becomes more beneficial.

The AI will iteratively continue to train. It would receive a reward based on the new state that would then be used to hone the Coefficient values (Q-values) in the neural network (Violante, 2019). Gradually and through many iterations, this leads to a model that can move the pirate more accurately through the maze. Thus, AI approaches the problem systematically and learns through repeated interactions and reward feedback. Humans, on the other hand, rely on intuition, reasoning, and memory, and can adapt their strategy through trial and error.

My approach to solving this problem was using the Keras framework to train a neural network that approximates the optimal Q-values through many cycles. The game environment is represented by the “TreasureMaze” class, AI agent actions and outcomes are stored in the “GameExperience” class for experience replay. The “qtrain” function iterates through epochs, randomly selecting starting positions, executing actions based on exploration-exploitation strategy, and collecting experiences.

References

Khan, T. (2022, January 2). *Reinforcement learning – exploration vs exploitation tradeoff*. AI ML Analytics. <https://ai-ml-analytics.com/reinforcement-learning-exploration-vs-exploitation-tradeoff/>

Violante, A. (2019, July 1). *Simple reinforcement learning: Q-learning*. Medium. <https://towardsdatascience.com/simple-reinforcement-learning-q-learning-fcddc4b6fe56>