Nicholas Wyrwas

CS-370-16865-M01

Mr. Mike Prasad

04/27/2025

7-3 Project Two Design Defence

Although humans and intelligent agents share a problem-solving design objective they handle information processing through fundamentally distinct mechanisms. Humans solve maze puzzles through memory recall and reasoning alongside trial-and-error techniques while intelligent agents rely on structured learning processes that utilize algorithms and stored data. Intelligent agents maintain perfect information consistency during recall unlike human memory which might produce errors or generate false memories (Reuell, 2015). Intelligent agents demonstrate greater reliability because they utilize past experiences to solve problems without any deviation.

During the treasure hunt game human participants will probably figure out the maze by looking for paths visually while remembering dead ends and relying on intuitive adjustments to their movements. The pirate agent employs a deep Q-learning algorithm to systematically assess each potential action based on accumulated rewards and penalties. Learning from experience defines both methods although the agent depends solely on numerical feedback and optimization instead of emotion or instinct. The agent achieves high efficiency yet shows reduced adaptability when confronted with unforeseen changes unlike a human player.

This intelligent agent's goal during pathfinding tasks is to discover the treasure using the quickest and most resource-efficient route. Exploration and exploitation are key concepts here. Through exploration the agent tests new actions for improved pathways while exploitation enables it to follow previously successful routes based on existing knowledge. A successful training strategy would begin with exploration and gradually move to exploitation as the agent develops an accurate map of the maze. Maintaining this balance ensures that the agent does not become trapped following ineffective routes.

Reinforcement learning drives the agent by using four main components: policy, reward, value function, and environment model. The policy guides agent actions, rewards signal treasure discovery success; the value function estimates upcoming rewards while the environment model forecasts outcomes from specific actions (Simplilearn, 2022). These components combine to guide the pirate agent toward the most beneficial course while steering clear of penalties.

This project involved implementing deep Q-learning with a neural network designed to predict action Q-values for each state. Experience replay helped train the model to achieve stable and efficient learning processes. Through its repeated training on previous experiences the network refined its reward estimates and chose optimal actions. The pirate mastered an optimal path to find treasure through this method which proved deep reinforcement learning's effectiveness in solving complex problems.

**References**

Reuell, P. (2015, October 5). *How the brain builds new thoughts*. Harvard Gazette. <https://news.harvard.edu/gazette/story/2015/10/how-the-brain-builds-new-thoughts>

Simplilearn. (2022, August 29). *What is Q-learning: Everything you need to know*. Simplilearn. <https://www.simplilearn.com/tutorials/machine-learning-tutorial/what-is-q-learning>