Design Defense

CS 370 Current/Emerging Trends in CS

Nicholle Caudy

8/11/2025

**Differences Between Human and Machine Approaches to Solving Problems**

The steps a human being would take to solve the treasure hunt maze would be to analyze their surroundings from the starting point. Typically, we rely on trial and error to find a solution to a problem or in this case a maze, starting with which direction to move first. Certain individuals may choose to begin at the conclusion and proceed systematically towards the starting point as a means of streamlining the process. People may record notes or make a map to keep track of obstacle locations, dead ends, the number of steps taken between obstacles, and directions chosen at decision points.

The intelligent agent starts at the start point and decides which direction to move first like humans. This intelligent agent uses a reward approach that will gain or lose points based on the experiences throughout the maze. Storing the points earned or lost throughout the maze the intelligent agent learns the optimal path for completing the maze. This is also like a human’s approach of taking notes or making a map to remember different obstacles, out of bound areas, or dead ends within the maze. Humans do not have the mathematical ability to calculate each other their moves taken within the maze, unlike the agent player. Humans also have the advantage of identifying areas that would cause a loss and choose to move around it, though they may not make the right choice to win. (Ryther, 2021)

**Purpose of the Intelligent Agent in Pathfinding**

The difference between exploitation and exploration is the intelligent agent taking the risk of exploring when choosing its next step instead of playing it safe or exploitation. Exploration can come with greater rewards but also comes with higher risk and exploitation comes with easier rewards because it has lower risks but can lead to lower average gains. Finding the right ratio of exploration and exploitation can be tricky as you want the agent to complete the task quickly but also with high rewards. This problem is addressed by employing the Epsilon Greedy Algorithm, which utilizes a constant epsilon value between 0 and 1 as a threshold for determining when the agent will explore rather than exploit. (Ryther, 2021) I think the ideal proportion of exploration and exploitation for this path-finding problem is 20% exploration and 80% exploitation, this allows the agent to explore new paths and potentially learn a better solution but also keeps the agent progressing through the maze and potentially gaining greater rewards along the way.

Reinforcement learning can help to determine the path to the goal (treasure) by the agent (pirate) by first setting up the environment with rules and rewards set in place ensuring higher rewards to encourage exploration. The agent will explore the environment to find new paths and evaluate the path taken by adjusting its actions based on the rewards received. The agent will repeatedly run through environment using exploration and exploitation updating and improving with each iteration until it finds the optimal path to the treasure (win!).

**Evaluate the Use of Algorithms to Solve Complex Problems**

Deep Q-learning was implemented for this game by importing relevant libraries and constructing the environment with a matrix representation. Rewards and penalties were set. The agent then explores the environment transitioning from exploration to exploitation using these rewards as a guide to help the agent make the next best move. The Epsilon Greedy Algorithm was utilized to determine appropriate intervals for exploration versus continued exploitation.

while qmaze.game\_status() == 'not\_over':

previous\_envstate = envstate

#valid action

valid\_action = qmaze.valid\_actions()

#get action

if np.random.rand() < epsilon:

action = random.choice(valid\_action)

else:

action = np.argmax(experience.predict(envstate)

#take action

envstate, reward, game\_status = qmaze.act(action)

The agent was tested to ensure adherence to established rules and restrictions. After numerous iterations, the agent achieved a 100%-win rate at epoch 457 using this model.

A black and white crossword puzzle

AI-generated content may be incorrect.

References:

Ryther, J. (2021, December 4). *Treasure maze reinforced learning*. Medium. <https://medium.com/@jryther91/treasure-maze-reinforced-learning-744a9aa80a71>

Understanding exploration vs. exploitation in reinforcement learning (RL) - NOMIDL. (n.d.-b). https://www.nomidl.com/generative-ai/exploration-vs-exploitation-rl/