Project Two Design Defense

Mason K Utt

SNHU

Abstract

This paper presents a thorough analysis and defense of the approaches employed to solve the Treasure Hunt Game, a pathfinding problem in which an intelligent agent, represented as a pirate, seeks the most efficient route to locate a hidden treasure. The paper does this by analyzing the differences between human and machine approaches to solving problems, assessing the purpose of the intelligent agent in pathfinding, and evaluating the use of algorithms to solve complex problems.

# Analyze the differences between human and machine approaches to solving problems.

## Describe the steps a human being would take to solve this maze.

One way a human being would solve the maze is to first identify the location of the pirate, then the location of the treasure, then it would identify the location of all the obstacles. The human would then identify the areas in which the pirate would get trapped by obstacles and avoid going that direction. The human would then try to make the most optimal line from the start to the finish and move the least distance away from the treasure. The human would only go away from the treasure when blocked by obstacles. The human would do this by looking at different sections of the maze and tracking where the pirate would go with their eyes. This is in accordance with Zhao’s (2014) study on human strategies for maze solving (Zhao, 2014). Humans do this by having a great spatial memory and can almost instantly solve the maze from anywhere if they have done it once and it is a small enough maze. Figure 1 demonstrates how a human would solve a maze.

## Describe the steps your intelligent agent is taking to solve this pathfinding problem.

The machine learns randomly by making any of its possible actions and learning from that state. For example, if the agent starts at 0,0 it can move down or right. If the agent has an obstacle to its right, it will lose the game and will get a negative reward. If the agent moves down and has no obstacle it continues playing and will get a positive reward if it moves closer to the target and does not die. The agent learns from exploitation and exploration. Exploitation is making moves that the policy recommends based on previous runs. Exploration is randomly taking and different turn to see what else exists. In the agent I wrote I implemented a decaying epsilon which meant at first the agent will do a lot of exploring, 50% of the time, but each epoch it will explore 1/100th less until it only explores 1% of the time.

## What are the similarities and differences between these two approaches?

There are several similarities and differences between these two approaches.

|  |  |  |  |
| --- | --- | --- | --- |
| Aspect | Human | Intelligent Agent | Both |
| Pathfinding | Use spatial memory and visual perception | Use state-action-reward matrix | Attempt to find the optimal path to the treasure |
| Identifying Obstacles | Visually identifies the obstacle | Learns the obstacle’s location from negative rewards | Avoid obstacles in the maze |
| Speed | Quick at solving small mazes | Minutes for small mazes up to hours for larger mazes. Once trained extremely fast. | N/A |
| Consistency | Varies among each trial | Once trained the machine will take the same path according to its policy from any location every time | N/A |

# Assess the purpose of the intelligent agent in pathfinding.

## What is the difference between exploitation and exploration? What is the ideal proportion of exploitation and exploration for this pathfinding problem? Explain your reasoning.

Exploitation is when the agent refers to its memory and its policy. Exploration is when the agent disregards the approved policy action and chooses a random move. The ideal proportion is hard to pin down, but I chose to go with a decaying epsilon. This means that at the beginning of training the agent will be 50/50 with exploration and exploitation. Each epoch the agent will explore 1% less and will exploit 1% more until it is 99% exploit, 1% explore. I chose this method because it allows for the agent to learn what happens in the beginning and once it has started to learn it will be less likely to randomly change later on.

## How can reinforcement learning help to determine the path to the goal (the treasure) by the agent (the pirate)?

Reinforcement learning helps determine the path by creating the policy for the agent to exploit. In this instance, the RL used is a Markov Decision Process. This works by using a transition function to show what the next state will be after an action and a reward function that rewards the agent off of the state it is in and the action it took to get there. The model learns based off of its past memory if an action is good or bad and uses that to make an informed decision.

# Evaluate the use of algorithms to solve complex problems.

## How did you implement deep Q-learning using neural networks for this game?

I implemented deep Q-learning by training a model with an MDP and then running that episode in the neural network. The Q portion of the Q-Learning is the MDP is the quality function. That is the maximum total you can get with each action from a state. The score is gathered recursively using Bellman’s Equation. The image figure 2 below shows what the score would be for each action if the pirate started in the bottom left and wanted to get the treasure in the top right. Q(s,a)=R(s,a)+maxi=0,1,…,n−1Q(s′,ai),(where s′=T(s,a))

References

*Deep reinforcement learning for maze solving*. qmaze. (n.d.). Retrieved April 16, 2023, from https://www.samyzaf.com/ML/rl/qmaze.html

*Reinforcement Learning*. Reinforcement learning. (n.d.). Retrieved April 16, 2023, from https://www.cs.cmu.edu/~15281-s23/assignments/programming/reinforcement/index.html

Zhao, Min (2014). **Understanding humans’ strategies in maze solving from eye-hand coordination.**Retrieved from <https://doi.org/doi:10.7282/T38W3FX5>

Figures

A picture containing crossword puzzle, text, black, indoor

Description automatically generatedFigure 1. This image depicts how a human would look forward to see where they would go. In section 1, the human looks down and right towards where the treasure is. The human sees it is blocked off and knows not to look in the area anymore. In section 2, the human will try and go to the right where the treasure is because down did not work. In section 3, the human moves even closer to the goal moving along the same path they originally traveled. In section 4, the human can finally reach the Treasure.

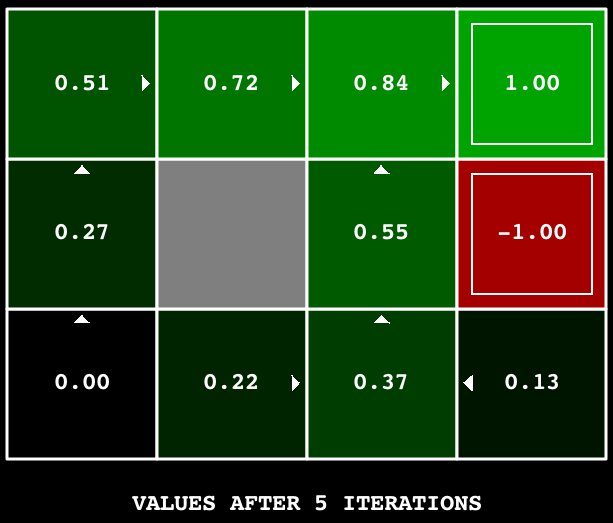


Figure 2. This image depicts the Bellman equation. It shows how the value of the reward if it moves in that direction. Q(s,a)=R(s,a)+maxi=0,1,…,n−1Q(s′,ai),(where s′=T(s,a))