CS 370 Module 7 Design Defense

James Nikolaou

CS 370

06/22/2024

* **Analyze the differences between human and machine approaches to solving problems.**
  + **Describe the steps a human being would take to solve this maze.**

The human will start at the top left corner of the maze. They will then identify what directions are available to them, and which of these directions which space is free or occupied. If they are occupied, they cannot move in that direction. If they are free, they are free to move to the next spot. This will be repeated until they reach their end goal or the end of the maze.

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

The intelligent agent is taking a different approach to solving the problem. They can’t see or rationally think like a person, so they will explore the maze by trial-and-error. The algorithm is rewarded and punished with a points system to help direct the path it should take. Reaching the target gives 1 point as a reward. If it navigates to a occupied space it loses .75 points, if it tries to move outside the maze it loses .8 points, and each step will lose .04 points to ensure it doesn’t wander forever.

* + **What are the similarities and differences between these two approaches?**

The obvious similarity is that they are trying to achieve the same end goal, finding the treasure (or end of the maze). Both are trying to move forward one step at a time, identifying the most efficient path to take. Their methodology of how they approach this task is different. The human will use visual queues to identify which way to move, or more importantly, which way not to move. The human knows the constraints of not moving off the maze and that blocks in the maze are not to be passed. They also know that moving randomly is not optimal, so they will try to find a linear path.

The algorithm is quite a clunky mess to start as it has no real knowledge of the rules and learns through trial and error. However, it does learn overtime what is good and bad based on the reward/penalty system incorporated. Eventually it will learn the path and become very efficient in navigating the maze.

* **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 defined as a greedy approach in which agents try to get more rewards by using estimated value but not the actual value. So, in this technique, **agents make the best decision based on current information. “** (Javatpoint, 2021). The rewards within this problem were as follows: Reaching the target gives 1 point as a reward. If it navigates to a occupied space it loses .75 points, if it tries to move outside the maze it loses .8 points, and each step will lose .04 points to ensure it doesn’t wander forever. These point values not only provide context to what is a good/bad move but rates the severity of the moves. Moving idly does provide a penalty to discourage movement, but moving out of bounds or an occupied space being much higher allows for a focused approach with rewards.

Exploration is defined as where the “agents primarily focus on improving their knowledge about each action instead of getting more rewards so that they can get long-term benefits. So, in this technique, **agents work on gathering more information to make the best overall decision.”** (Javatpoint, 2021). Exploration within this problem is to move between squares and identify what square is what, and eventually where the end goal is. It does help by even taking penalties with rewards, it helps provide a contextual layout to the reward system which then can exploit the best “rewarded” path.

Each of these have their respective goals, exploration for providing a layout and exploitation for utilizing this information for the highest pathing score possible. Considering the algorithm is moving through a maze, exploration should be weighted heavier than exploitation in this case. It’s better for the algorithm to identify what is a good spot and bad instead of randomly moving around due to it increasing the penalty with idle movement. Creating the environment for a clear and concise layout of rewards within the maze will help the algorithm identify the correct path with all the correct information more efficiently than blundering around until an optimal path is found.

* + **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 to the goal by the agent by providing context to its learning. A “bad” move is when the agent receives a penalty based on an action, and a “good” move is when it finds the end goal, the treasure. There is a penalty for all movements of .04, but in context this will really only help ensure no random movements, and by the end of the exploration phase, help the algorithm identify an optimal path with the least number of spaces possible.

* **Evaluate the use of algorithms to solve complex problems.**
  + **How did you implement deep Q-learning using neural networks for this game?**

Deep Q-learning is a combination of the usage of a neural network and a q-learning ideology. The q-learning ideology provides a concept of proving a “value” to the action/s from a state within the learning process which provides a matrix of values the agent can utilize to complete the objective. The neural network is utilized with the idea of where neural networks provide relevance to values within the q-ideology by providing a platform instead of a matrix to approximate the q-values for each action at any given state. Please reference the image below for an example of this:

A diagram of a computer system

Description automatically generated with medium confidence (Loeber, 2022)

This is far superior as it is far more scalable than a simple q-matrix. As long as the approximation value is accurate, it allows for much faster learning of the maze with the values associated with each position. The exploration phase provides reinforcement of the various q values assigned to each position within the maze. The exploitation portion of the reinforcement model takes advantage of these various q values within the neural network and finds the most optimal choice of actions to ensure the highest point score possible.

**Sources:**

*Exploitation and exploration in machine learning - javatpoint*. www.javatpoint.com. (2021). https://www.javatpoint.com/exploitation-and-exploration-in-machine-learning

Loeber, P. (2022, February 10). *Reinforcement learning with (deep) Q-learning explained*. News, Tutorials, AI Research. https://www.assemblyai.com/blog/reinforcement-learning-with-deep-q-learning-explained/#:~:text=Deep%20Q%20Learning%20uses%20the,action%20based%20on%20that%20state.