**Project Two: Design Defence**

Mitchell Lynds

Southern New Hampshire University

CS 370: Current and Emerging Trends in CS

Matthew McHann

February 20, 2023

In this maze navigation problem I believe a human player could look at this maze and determine the shortest route by simply counting out tiles to the goal repeatedly until the shortest route is found. A human should be able to reliably do this on this maze in just a few repetitions. There are two major problems with this strategy though.

In this problem I believe the only way that the agent can “see” the maze is by knowing whether a move in any given direction is valid from the current location. This means a human approaching this problem would need to memorize the layout of the maze. In that case the human would be essentially starting “in the dark” and would need to initially simply explore the maze to learn the shape and where the goal is. After a few games a human player would have the layout of our fairly simple test maze memorized and would be able to navigate to the goal along the shortest route reliably.

Another problem that a human trying to solve the problem would run into would be if the scale of the maze were to increase significantly. Starting each game at a random location within a very large and complex maze could make attempts to solve the maze by memorization impossible for a human. Given A pencil and sheet of grid paper the human may be able to slowly map the maze through exploration and then calculate the fastest routes.

The intelligent agent is solving this problem in much the same way that the human being would. It begins by exploring the maze at random until it learns which moves from which spaces result in completing the maze more quickly. After each iteration of the game the agent is trained on the results of its actions at each space. As the intelligent agent plays more repetitions of the maze game it makes better and better predictions when selecting the best move from that space.

The approaches used by a human agent and the artificial intelligent agent are quite similar but do have some key differences. Both would start by making nearly random moves until more information is learned about the maze.(Dugmeci, 2021) While the human agent can reliably calculate the shortest route as soon as the maze is sufficiently memorized the AI has to refine its route over many iterations even after it has visited every valid space in the maze several times. Other than requiring more iterations to create an effective model, the AI should have no trouble solving the maze regardless of scale, while a human player is limited somewhat by their memory of the maze.

Exploration refers to making random actions in order to discover new information while exploitation refers to making the best current action based on prior experience.(Lamba, 2018) For this problem we used a consistent 1:9 exploration to exploitation ratio and the model was able to be trained effectively fairly quickly. Ideally the agent would start by exploring almost exclusively and as it becomes better trained reduce the exploration factor eventually to zero.(Lamba, 2018) When exploration factor is too low early on in training it results in the actor repeating the same unproductive actions over and over again. When the exploration factor is too high late in training it can prevent the agent from successfully completing the maze despite having enough training data to do so.

The way reinforcement learning is used to determine the path to the goal is through Q-learning. This is a strategy where each action available at each state is assigned an expected reward value.(Lamba, 2018) The model then takes the action with the highest predicted reward at each state.(Dugmeci, 2021) At the end of the iteration the estimated reward values are updated to reflect the experience gained.(Lamba, 2018) In the specific case of the maze this means that if the agent is located at space (3,4) and the action for space (3,4) with the highest assigned reward value is to move to the right, then the agent will move to space (4,4). This is repeated until the reward estimates for each action at each space are sufficiently accurate.

Deep Q-learning is a variation of the expected reward system used in regular Q-learning. In deep Q-learning we replaced the table of expected reward values with a neural network.(Surma, 2019) The neural network is trained on random batches of experience data from previous games.(Surma, 2019) Rather than checking expected future reward values in a table, the neural network provides a prediction of which action has the highest value.(Surma, 2019) In this way the neural network is able to identify relationships within the data and learn the maze more quickly than plain Q-learning.

References:

Dugmeci, D. (2021, November 10). *Finding shortest path using Q-learning algorithm*. Medium. Retrieved February 26, 2023, from https://towardsdatascience.com/finding-shortest-path-using-q-learning-algorithm-1c1f39e89505

Lamba, A. (2018, September 3). *An introduction to Q-learning: Reinforcement learning*. Medium. Retrieved February 26, 2023, from https://medium.com/free-code-camp/an-introduction-to-q-learning-reinforcement-learning-14ac0b4493cc

Surma, G. (2019, November 10). *Cartpole - introduction to reinforcement learning (DQN - deep Q-learning)*. Medium. Retrieved February 26, 2023, from https://gsurma.medium.com/cartpole-introduction-to-reinforcement-learning-ed0eb5b58288