Kyle Lund

CS-370

# Design Defense

## Human and Machine Approaches

The way humans and AI approach solving a maze are similar but also different. Humans have logic, reasoning, and emotions that play part in exploring a maze. When looking at a maze a human will look at it and can eventually find the path based on visuals. An Ai first attempt through the maze will be slow as they do not rely on visual looks and tend to explore are possible paths until they reach the exit. As the Ai goes through more mazes it begins to learn and understand maze structure and if it reaches a dead end, it is faster at back tracking through the maze when compared to a human. The main similarity to the approaches is the trial and error it takes to complete a maze. Both parties are not perfect and know to back track when necessary and to choose a new unexplored path

## Pathfinding

Exploration is the act of trying new possibilities to find better rewards. Exploitation is performing the same action that has previously given significant rewards in the past. Exploration is the AI looking at new paths and making new decisions that it has never made before while looking at all the possible rewards it has collected. Exploitation takes the information the AI has learned from the maze and takes the best route to collect an almost guaranteed number of rewards that it has received before. The best balance between the two is the Epsilon strategy, this strategy has the Ai starting with an epsilon value of 1 where it is in full exploration mode. As the Ai moves through the maze that value decreases little by little as the Ai gets rewards it already knows of while still exploring until this value reaches 0. The exploitation as the epsilon value changes also changes as new rewards and more optimal routes are found. Reinforcement learning can assist the Ai in finding the treasure by rewarding the AI based on the states of the board. Reinforcement learning is based off “reward hypothesis”, where all goals are described by the maximization of all rewards. This can be simplified by rewarding the AI for taking good path decisions and punishing it with making bad decisions. This will make the AI find a path to maximize their reward.

## Algorithms

A process of the deep q learning creates a matrix for the agent to use to maximize its rewards in the long run. The AI agent will take advantage of the experience it has gained to find the best rewards. This algorithm will take a batch of its experience and learn from it and create different action-based policies to follow. These policies take on the dilemma of the exploration and exploitation in an efficient manner until it finds the best path forward.

Reference

*qmaze*. (n.d.). Www.samyzaf.com. https://www.samyzaf.com/ML/rl/qmaze.html

‌

Singhal, R. (2018, December 16). *Reinforcement Learning: A Beginner’s Guide*. Medium. https://medium.com/@rsinghal757/reinforcement-learning-a-beginners-guide-56de9c7fe1d9

‌

Makone, A. (2021, August 22). *Reinforcement Learning 6: Exploration vs Exploitation*. Medium. https://ashutoshmakone.medium.com/reinforcement-learning-5-exploration-vs-exploitation-c1bae5a2ea42

‌