CS-370

Zeb Berry

Design Defense

When a human solves a maze, the process is usually pretty intuitive. We look at the layout, plan a path in our head, and then make adjustments if we hit a dead end. Most people try to remember which turns worked and which did not, so they do not repeat the same mistakes. A lot of it is based on trial and error combined with memory and reasoning. If you were dropped into this pirate maze, you would probably start at the beginning, test a route, and backtrack when you get stuck until you eventually find the treasure.

My intelligent agent, on the other hand, does not see the maze in the same way. It uses reinforcement learning to break the task into states and actions. Each state is basically a snapshot of where the agent is, and the possible actions are which directions it can move. Instead of reasoning like a human, it learns from repeated episodes of trying things out. Every time it makes a move, it gets feedback in the form of a reward. Positive rewards come from reaching the treasure and negative rewards come from losing. Over time, it uses those rewards to build up a sense of which actions are best in which situations.

The similarity is that both a human and the agent rely on trial and error and gradually build knowledge of the maze. The difference is that a human can apply intuition and spatial reasoning, while the agent is purely mathematical. It is crunching numbers through neural networks and Q-values rather than thinking about the maze in a conscious way.

The purpose of this intelligent agent is to reliably solve the pathfinding problem. For it to work, it needs to balance exploration and exploitation. Exploration means trying out new actions to gather more information, while exploitation means using what it already knows to make the best choice. In the beginning, the agent needs a lot of exploration, so it does not miss potential solutions. Later on, it should lean more toward exploitation so it can consistently follow the best path it has learned. A good balance is to start with high exploration and gradually reduce it over time. For this maze problem, something like ninety percent exploration early on shrinking to around five to ten percent exploration later is ideal. This prevents the agent from getting stuck in a bad habit too early but still encourages it to settle on a winning strategy.

Reinforcement learning is what actually guides the pirate to the treasure. Each episode in the training process helps the agent connect the dots between states, actions, and outcomes. Over time, the Q-learning algorithm builds a value function that tells the agent the expected reward for each action in each state. That way, instead of guessing, the pirate can look up the best move and head straight for the treasure with high accuracy (Sutton & Barto, 2018).

To handle this, I implemented deep Q-learning using neural networks. The neural network serves as a function approximator that takes in the current maze state and predicts the Q-values for each possible action. Instead of storing a giant table of every state-action pair, which would get too large, the network generalizes and estimates the Q-values based on what it has learned. I used an experience replay buffer to store past moves and train the model in batches, which helps stabilize learning (Mnih et al., 2015). By running through thousands of episodes, adjusting weights based on the difference between predicted and actual rewards, the model gradually converged toward an effective strategy.

In short, the human way of solving the maze is intuitive and based on reasoning, while the agent’s way is mathematical and based on reinforcement. Exploration and exploitation help it learn efficiently, and deep Q-learning with neural networks gives it the power to solve the problem even when the maze is complex.

### References

Mnih, V., Kavukcuoglu, K., Silver, D., Rusu, A. A., Veness, J., Bellemare, M. G., … Hassabis, D. (2015). Human-level control through deep reinforcement learning. *Nature, 518*(7540), 529–533. <https://doi.org/10.1038/nature14236>