Project Two

Jordan Whitney

CS-370 Current/Emerging Trends in CS

Professor: Binyam Heyi

2/23/2025

The steps a human would take to solve this maze could differ depending on the person, how cautious the person is, how fast the person wants to complete the maze, and possibly want mental/emotional state the person is in. The first step a human would want to take is find the goal of the maze, then choose an initial direction to take, if they’ve run into an obstacle then there is a possibility to reverse directions, allowing them to reach the end goal. The steps my agent takes is observing the environment, then the agent picks a random action and moves in the chosen direction, it then receives a reward based on how they moved. That data is stored for training the neural network of past experiences, allowing the agent to update the Q-values. Then the agent leans by training choosing the best-known action relying on exploitation. Similarities between a human and the agent are they both aim to achieve the end goal whether that is finding the shortest or fastest path, both try different paths, a human will remember from past experiences (knowing where the dead ends are) and the agent also learns from its past training with reinforcement learning. Differences between the two would be the memory, depending on how good the human’s memory is, are they able to remember what paths they have already chosen, the agent stores the experiences in its memory for training. For a human the learning is more intuitive and immediate having to decide on which path to take, and for the agent it’s able to improve over reinforcement learning.

The difference between exploration and exploitation is that exploration is exploring new approaches or solutions to the problem, and exploitation is choosing the best actions from previous trails. Beysolow says,

“This is what at large is described as the exploration-exploitation trade-off, and it is controlled by the epsilon parameter. The key here is that the first possible path that might be utilized to reach a solution is not guaranteed to be the best path. With this being stated, it is unlikely that it will always be the case that if we keep searching, we will find a better solution than the current one, and therefore we abstain from solving the problem.”

I am not sure of the ideal proportion of exploitation and exploration, but in my efforts to train my agent I found that increasing the epsilon to 0.9 helped increase my agent reach the goal in 40 epochs. Reinforcement learning helps the agent reach the end goal by avoiding obstacles, minimizing the number of moves it takes to reach the end, making decisions based on the past trails it has completed.

I was able to build a neural network, store past experiences, randomly sample from past experiences, to help prevent bias. Training the agent that interacts with the maze collecting experience, updating the Q-values using Bellman’s equation. Using an epsilon-greedy strategy to help balance the learning, with completion allowing training to stop when it reaches the goal with a 90%+ win rate consistently.

References

Beysolow, I. T. (2019). *Applied reinforcement learning with python : With openai gym, tensorflow, and keras*. Apress L. P..

Lamba, A. (2018, September 3) An introduction to Q-Learning: reinforcement learning : <https://medium.com/free-code-camp/an-introduction-to-q-learning-reinforcement-learning-14ac0b4493cc>

Gulli, A., & Pal, S. (2017). *Deep learning with keras : Implementing deep learning models and neural networks with the power of python*. Packt Publishing, Limited.