Annamarie Cortes

Module 7

October 20, 2024

Project Two

Design Defense: Pirate Agent Using Deep Q-Learning

This treasure hunt game incorporates an intelligent agent, the pirate, as the competitor of the player. The player would need to compete and find the treasure first before the pirate does in order to win the game. Deep Q-learning is used to develop this intelligence agent. The design of the ‘pirate’ aims to navigate an 8x8 maze to find the hidden treasure. This design defense outlines the approach taken to solve the pathfinding problem, detailing how the intelligence agent operates, evaluating the chosen algorithm, and comparing machine and human approaches to problem solving.

Humans and machines differ in their approach to solving problems. A human would first observe the maze layout to identify potential obstacles and the treasure’s location. Then based on that understanding, humans would plan a route with the shortest distance possible. There is always trial and error in human approaches, if there are any obstacles and the first route fails, we adjust our strategy based on the feedback from our previous actions. Also, depending on the situation, humans can change the strategy in real time based on unexpected challenges faced. A machine or the intelligent agent, on the other hand, will perceive the maze as a grid where each cell is represented as a state. The machine will use reinforcement learning and updates its strategy based on received awards or penalties from its actions. It relies on algorithmic decision-making to calculate the expected utility actions through Q-values and aims to maximize the long-term rewards. It will store previous experiences in memory and relies on this to systematically explore the environment.

A human approach steps would be to observe the maze and identify obstacles as well as locate the hidden treasures. Then, plan out a route with the shortest distance possible. If an initial observation where the whole maze is seen, then human approach will be trial and error. They would first choose an initial direction and move through the maze. The human will formulate a strategy and adjust in real-time based on feedback from the obstacles faced.

A machine approach steps would be to initialize the Q-values for each state-action pair, then repeatedly select actions based on the policy. Then it will execute the action, observe the new state, and receive a reward or penalty. Then, update the Q-value for the taken action and store those experience in memory for future learning. It will gradually refine the policy through multiple episodes to maximize the treasure revival rate.

Both human and machine approaches involve a form of trial and error. They both learn from past experiences to refine their strategies. However, some key differences are that humans uses intuition and cognitive flexibility change strategies in real time as they see fit while machines rely on statistical models and data driven decisions. Machine operates through predefined algorithms and it lacks the adaptability and emotional intelligence humans posses.

The primary purpose of the intelligent agent in this game is to autonomously navigate the maze with the goal of finding the hidden treasure in the most efficient way possible. With the use reinforcement learning, the pirate learns optimal strategies. This is done by maximizing rewards that is associated with reaching the treasure while minimizing the penalties received from obstacles and inefficient moves. The pirate continuously improves its performance through self-learning and ultimately outperforms static or rule-based systems in dynamic environments (Sutton & Barto, 2018).

The difference between exploitation and exploration is that explorations refers to the agent trying new actions to better understand the environment, while exploitation involves using known information to maximize rewards. The ideal scenario would be to start with a higher exploration rate around 90% and then gradually transition to exploitation around 10% as the agent learns (Mnih et al., 2015). This balance allows the agent to gather sufficient data to make informed decisions while still discovering new strategies. A proper balance between exploration and exploitation is important in reinforcement learning because exploitation can lead to suboptimal policies while excessive exploration can prevent the agent from making effective decisions (Martin, 2020). Therefore, maintaining is essential for the pirate to effectively learn and develop strategies to find the treasure before the player does.

Reifnorcement learning is crucial for the pirate to find an optimized path to the treasure. With the use of rewards for successful actions and penalizing obstacles, the pirate develops a policy that maximizes expected rewards over time. The Q-learning algorithm updates the action-value estimates based on the feedback from the environment and allows the pirate to adapt its strategy (Watkins & Dayan, 1992). The pirate/agent will continually update its policy based on the actions it has taken, and the rewards and penalties it has received to make a robust framework for dynamic environments.

Algorithms provide systematic methods to solve complex problems such as path finding. Q-learning provides a structured way to learn optimal strategies in uncertain situations. By employing deep Q-learning with neural networks, the agent can handle more complex state spaces and allows it to generalize learning across different scenarios. “Deep reinforcement learning combines the best of both worlds: the power to deep learning and the explorations if reinforcement learning” (Kahn et., al, 2020). This indicates that with both deep learning and reinforcement learning, the agent has the ability to self-learn and form its own strategies and therefore find the most optimized path possible probably beating the human player over time.

Deep Q-learning was implemented using neural networks to predict the best actions for the pirate. The neural network consists of input layers representing maze states, hidden layers to process the information, and output layers provides Q-values for possible actions. To enhance learning, experience replay is included which allows the pirate to “learn from past experiences” and “improve its decision making” (Mnih et al., 2015). By doing so, the pirate is able to mimic the trial and error that human approach to problem solving have and is able to effectively make a strategy in finding the hidden treasure.

References

Khan, M., Sharma, A., & Sharma, S. (2020). Finding shortest path using Q-learning algorithm. Towards Data Science. https://towardsdatascience.com/finding-shortest-path-using-q-learning-algorithm-1c1f39e89505

Martin, R. (2020). Balancing exploration and exploitation. Medium. https://rogermartin.medium.com/balancing-exploration-and-exploitation-adb82146837

Mnih, V., Kavukcuoglu, K., Silver, D., Rusu, 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

Sutton, R. S., & Barto, A. G. (2018). Reinforcement learning: An introduction (2nd ed.). MIT Press.

Watkins, C. J. C. H., & Dayan, P. (1992). Q-learning. Machine Learning, 8(3-4), 279-292. https://doi.org/10.1007/BF00992698