Jamie Javis

CS - 370

06/25/2024

Design Defense

Humans and Machines think differently. In the case of the maze, humans would look at the path and then make moves based on real-time input and past experiences. Machines use probability and values to move with less of the ability to move would comparing values.

A human would look at the maze. A human would then use past experiences and intuition to calculate the path needed as well as make decisions based on the changes that could happen.

The agent utilizes the starting position and the maze map. When started, the agent would explore and then execute based on the test of the exploration. Based on the chosen action, the agent would update its position and receive feedback through the reward system. The agent then updates the value to reflect the observations on what worked and what didn't. The agent will continue until it gets to the treasure. Much like human intuition, this model will simulate that in machines to the best of its ability. It will test the moves much like how humans can remember what worked and what didn't. Unlike humans, the computer has to calculate moves to test if they would work when humans already have an idea based on past experiences and real-time feedback for tasks like the maze.

The agent's job is to be able to navigate a maze on its own. The agent's job is to also find the treasure as fast as possible. The purpose is to show the machine-learning techniques and see how fast it can happen. Exploitation is the agent choosing an action based on the values known to receive the reward and do it efficiently. Exploration is the agent's ability to discover different strategies while also working towards the goal of finding a potentially more efficient path. The balance between the two depends on the complexity of the problem. Higher epsilon would correlate to a higher exploration rate which is great for discovering paths. A lower epsilon would be more useful for exploitation while moving around the board. Reinforcement learning can help determine the path to the treasure because it will help the agent become quicker and make fewer errors over time. In this case, the error is less reward based on the value of Q. The Q-value is the estimation of the reward based on each given action. The agent is then able to update the policy based on the discoveries from the actions made and the rewards awaiting. The policy will be updated to reflect the improvements the agent has made based on decisions or actions over time.

The deep Q-learning has a lot of benefits for the AI. The agent uses the neural networks to implement new paths based on the observations from exploring. The agent will become faster and better at handling complex actions depending on the environment and how the agent has been able to move. One downside tends to be writing the right parameters and being careful about the values needed. If the agent has too much experience or too little it can create a bias based on the train it received. Deep Q-learning does provide a strong base for training agents, in this case, to navigate mazes. They can learn effective strategies and learn how to execute their moves faster and with more “intuitive” decisions.

References

*Actor-Critic Algorithm in Reinforcement Learning*. (2023, December 13). GeeksforGeeks.<https://www.geeksforgeeks.org/actor-critic-algorithm-in-reinforcement-learning/>

Deep Q-Learning. (2019, June 13). GeeksforGeeks.

<https://www.geeksforgeeks.org/deep-q-learning/>

Nuer, J. (2022, May 20). A Comprehensive Guide to Deep Q-Learning. Medium. <https://medium.com/@jereminuerofficial/a-comprehensive-guide-to-deep-q-learning-8aeed632f52f>

*PyLessons*. (n.d.). Pylessons.com. https://pylessons.com/CartPole-reinforcement-learning

*Reinforcement Learning: Exploring Policy vs. Value-Based Methods*. (n.d.). Dataheadhunters.com.<https://dataheadhunters.com/academy/reinforcement-learning-exploring-policy-vs-value-based-methods/>

‌Yoon, C. (2019, July 17). *Understanding Actor Critic Methods*. Medium.<https://towardsdatascience.com/understanding-actor-critic-methods-931b97b6df3f>

*What is the relation between Q-learning and policy gradients methods?* (n.d.). Artificial Intelligence Stack Exchange. https://ai.stackexchange.com/questions/6196/what-is-the-relation-between-q-learning-and-policy-gradients-methods