Breunna Bingham

Southern New Hampshire University

June 13, 2023

CS-370 Current/Emerging Trends in CS

Module Seven: Project Two

Humans and machines are very different when solving problems. Humans use their senses and neural pathways composed of nerves, synapses, dendrites, neurotransmitters and more to send a message from the sensory input to the brain. The brain reads that input and sends a signal back to the motor neurons telling it what to do from the decision it made. In this case, it is to solve the maze. Humans will need several tried to learn the maze from trial and error and the results vary from one human to the next, as far as how much time it would take to solve the maze. Past experiences would be remembered in the short-term memory and accessed every time the human needed to decide which route to take within the maze. The intelligent agent uses a lot more mathematical algorithms to form a decision based upon past experiences that have been stored to the memory. Depending on the type of neural networks that is being used, really depends on how fast or slow the intelligent agent learns the maze and can complete it. While both the human and the intelligent agent access information that was stored in the brain or the memory to compose their decision-making process, the intelligent agent relies heavily on mathematical computations to make a decision, and humans use more of a trial-and-error decision making.

Exploration and exploitation work hand-in-hand as some would say, but they serve very different purposes. “The exploration-exploitation trade-off is a fundamental dilemma whenever you learn about the world by trying things out. The dilemma is between choosing what you know and getting something close to what you expect (‘exploitation’) and choosing something you aren’t sure about and possibly learning more (‘exploration’) (Yang, 2023).” The ideal proportion of exploration vs exploitation for this problem is to allow the intelligent agent sufficient exploration time but increase the reward so that the exploitation is also beneficial. Exploration allows the intelligent agent a moment to occasionally make a choice in the maze, while exploitation is used to weigh a reward for the intelligent agent to consider and really make an informed decision when making that choice in the maze. The higher the reward the more weighted the decision becomes for the intelligent agent and vice versa. Reinforcement learning allows the intelligent agent, the pirate, to use the goal, the treasure, as the reward for a determination on how the pirate will make choices within the maze and maximize its chances of getting that reward faster.

Q-learning is a model free reinforcement learning algorithm. It takes the Bellman equation and updates values based on that. A Q-table is utilized in which all steps on the maze are marked as values and this allows the pirate to weigh out different options from that table, and ultimately make a decision. The pirate will reevaluate the reward from the decision it made and make changes to the Q-table as necessary. This allows the pirate to maximize the rewards for each state it is in and use that when deciding. Ultimately a pirate using the Q-learning algorithm will allow them to reach the reward in the shortest path.

References

Shyalika, C. (2021, December 12). A Beginners Guide to Q-Learning - Towards Data Science. *Medium*. https://towardsdatascience.com/a-beginners-guide-to-q-learning-c3e2a30a653c

Yang, A. (2023, March 17). What is Exploration vs. Exploitation in Reinforcement Learning? *Medium*. https://angelina-yang.medium.com/what-is-exploration-vs-exploitation-in-reinforcement-learning-a3b96dcc9503