Jarrale Butts

Southern New Hampshire University (CS-370)

Design Defense

**Human vs Machine Approach**

When attempting to solve a maze, a human would learn from life experiences and or use their “intuition” to solve the puzzle. If you run into a dead end, you do your best not to make that same mistake twice. Maybe you take random turns (left, right, up, or down) and at times go along with your intuition when choosing which direction to turn, if and until you feel as if your intuition may not be the best at making these choices. Rinse and repeat until you find the treasure.

My intelligent agent starts with a predetermined maximum number of chances to complete the maze which is 15000. During each game, the agent starts in a random empty cell location. It must take an action (move left, right, up, or down) based on a random number generated between 0 and 1. With the epsilon being initially set to 0.1, the agent will be more aggressive at choosing an action based on experiences which will change later when the epsilon is set to 0.5 allowing the agent to explore new paths more frequently. The agent is awarded points based on its actions, where it favors actions that award it more points. Rinse and repeat until it finds the treasure.

Similarities:

* Choose an action.
* Learn from experiences.
* Choose more actions based on exploration or exploitation.

Differences:

* Humans can choose actions based on intuition, not just exploiting experiences.
* The agent won’t forget its past experiences (less likely to repeat mistakes).
* There is no predetermined number to allow more exploration for humans (epsilon).

**Intelligent Agent Purpose**

Exploration vs exploitation refers to taking actions based on complete randomness or actions that will guarantee a positive reward for doing so. “Exploitation gives the agent a guarantee of reward, but it may not be the best reward possible in the environment.” (Makone, 2022, para. 2). “While in Exploration, agent tries to opt for unknown actions, which he hasn’t tried before which takes him to unknown states, this may lead to either disappointment or better reward.” (Makone, 2022, para. 2). The Epsilon greedy strategy is a great way to keep the balance between the two. The agent starts off making actions based on its past experiences, but once the agent has won a predetermined number of games (without losing) it can start to explore more, potentially finding a more optimal route than previously discovered.

Reinforcement learning refers to how an intelligent agent would make decisions to maximize rewards, ultimately achieving a goal. Reinforcement learning allows the agent to use the Markov Decision Process (MDP) which is “a set of states, actions, and rewards with rules for transitioning from one state to another” (Gulli & Pal, n.d., sec. Reinforcement Learning). The agent will use that set of states, actions, and rewards to determine the shortest route with the highest possible reward output toward it’s goal (the treasure).

**Algorithms Solving Complex Problems**

According to Gulli & Pal (n.d., sec. Q-Learning), “Q-learning can be used to find an optimal action for any given state in a finite MDP by maximizing the value of the Q-function which represents the maximum discounted future reward when an action is performed in a state”. The TreasureHuntGame utilized a function allowing the agent to store its past actions and their rewards (weighed against their discounted future reward), building on top of and learning from each memory creating a neural network. Learning from each memory during the MDP to find the most optimal route to the treasure is how the deep Q-learning algorithm was implemented using a neural work for this game.

References

Gulli, A., & Pal, S. (n.d.). *Deep Learning with Keras*. O’Reilly Online Learning. https://learning.oreilly.com/library/view/deep-learning-with/9781787128422/5bfa3e2c-5e85-4de0-b07c-96d1b8821094.xhtml

Makone, A. (2022, January 5). Reinforcement Learning 6: Exploration vs Exploitation. *Medium*. https://ashutoshmakone.medium.com/reinforcement-learning-5-exploration-vs-exploitation-c1bae5a2ea42