Human vs Machine in Learning

The treasure hunting game, from the perspective of the pirate, involves navigating the game world, overcoming obstacles, and finding the treasure before the human player does. When training an AI model to assume the role of the pirate or to teach a player how to navigate the game world in order to reach the treasure, both can be taught in a similar manner conceptually, using a type of reinforced learning.

A human would likely begin by making random movements to explore the environment around them to get a grasp on how the game works or feels. Their goal is to poke and prod the structure of the maze and where the hazards are contained in order to ascertain the best approach when navigating that will bring them the highest chance of finding the treasure. Using their memory of what paths may cause them to run into obstacles or encounter hazards, they may opt to go for a safer route or gamble on paths that may provide a greater reward and enhanced knowledge of the game, but perhaps a higher chance of failure. A human may also pull knowledge from other maze-like games they have played in order to apply that knowledge to the challenges they are encountering in the treasure hunt. Once a player has achieved a comfortable knowledge of the game and its mechanics, they can work off of this learned knowledge to apply their skills and win the game.

Machines are not totally dissimilar from humans in the way they learn how to approach a particular challenge. Using a deep q-learning technique, which is a type of reinforcement learning, an AI model can emulate some of the ways a human might approach the treasure hunt problem, but in a more mathematical and pointed way. While they must explicitly be trained on a particular subject and cannot pull from “life experiences” of previous games the way a human can, the same way a human might try random locations to explore how the game works or how the maze is structured, a model can do much the same way. As these actions are taken and data is gathered about the results over a long sequence of iterations, the model can eventually learn the most appropriate strategy involved in finding the treasure and reliably play the game successfully based on this trained understanding.

Purpose of the Intelligent Agent

When discussing exploitation and exploration, a simple way to discern the two concepts is that exploitation is taking advantage of known values to get the most gain, while exploration is improving knowledge about the environment in order to potentially net a more significant gain. In the example of our pirate searching for treasure, it may explore until it finds a strategy that can provide a reasonable chance of success. It will then exploit this known strategy in order to find the treasure and win the game reliably.

A common and logical method used in finding a balance between exploration and exploitation is to start with a high exploration rate at the start in order to rapidly discover viable techniques and then decay to a lower exploration rate as the model is trained on the actions that are more ideal for success in favor of exploiting these learned values. The particular model within this assignment's code had an exploration factor of 0.1, which means that it is mostly geared towards exploiting the current knowledge it has, while exploring new options every 9 or so movements.

The process of how reinforcement learning and the exploration/exploitation principle can be used to guide the pirate to the treasure has a number of components. These involve the state of the agent and the environment at various times during the game, the actions the agent has available to take, the rewards an agent might receive based on what actions it takes, and the policy that is developed based on all of these factors. The pirate, being the agent, explores various areas of the maze based on the exploration factor set, gets rewarded or penalized for the actions it takes, updates its knowledge of the environment or success of its actions, and continues through multiple iterations of this process until it has refined a method that brings the most reliable chances of success of finding the treasure and avoiding hazards.

The Use of Algorithms

The implementation of the Q-learning algorithm used to solve the treasure hunt problem was quite complicated, but overall followed many of the principles outlined in the preceding sections of this document. The algorithm establishes a few parameters, such as the exploration factor, the numbers of iterations, the structure of the maze, and more. Each iteration, or epoch, of the algorithm randomly places the agent in a free cell in the maze. As long as the agent has not entered a location that would cause the game to end for them, a movement is chosen based on two factors. If a random number that is generated is below the epsilon value of 0.1, the agent will explore randomly. This means that 1 in 10 movement will be a random, explorative movement. In most cases, this random number will be above the epsilon, so the agent will exploit instead, which means it will predict an appropriate move based on the previous experiences it has stored throughout the runs. The success of the agent is tallied based on a win rate and these runs are accumulated, stored, and modeled onto the screen.

Ultimately, the idea is to allow the pirate to explore occasionally to rapidly gain new information, exploit its current understanding of the environment, analyze its performance based on the rewards it is receiving, adjust its strategy, and ultimately accumulate a significant number of iterations of this data until a consistent, winning method is found.

References

IBM. (n.d.). *AI vs. Machine Learning vs. Deep Learning vs. neural networks: What's the difference?* IBM. Retrieved from https://www.ibm.com/cloud/blog/ai-vs-machine-learning-vs-deep-learning-vs-neural-networks

Lindwurm, E. (2019, November 2). *Intuition: Exploration vs exploitation*. Medium. https://towardsdatascience.com/intuition-exploration-vs-exploitation-c645a1d37c7a

Newman, T. (2017, December 7). *Neurons: What are they and how do they work?*. Medical News Today. https://www.medicalnewstoday.com/articles/320289

TensorFlow. (2019, May 09). *Machine Learning Zero to Hero (Google I/O'19) [Video]*. YouTube. https://www.youtube.com/watch?v=VwVg9jCtqaU

Thomas. (2022, July 5). *How similar are Neural Networks to our brains?*. Fast Data Science. https://fastdatascience.com/how-similar-are-neural-networks-to-our-brains/

Wang, M. (2020, November 17). *Deep Q-learning tutorial: Mindqn*. Medium. https://towardsdatascience.com/deep-q-learning-tutorial-mindqn-2a4c855abffc