Al Castle

Southern New Hampshire University

CS 370

Project two

This treasure hunt game is a pathfinding AI project where we have a player that needs to find the treasure before the AI agent does. Using deep Q-learning we created a program where the pirate learns the best way to find the treasure from any position on the map that we were given. If a human were to solve this problem, they would have to start by randomly picking a spot on the map, such as our agent does when starting the game. From here we would have to use a trail and error method to find which spot we can advance to. Picking a spot would either result in a successful or unsuccessful move, once the human has this knowledge, they will be able to make a new move, and process the possible output and make another decision.

The biggest difference between a human solving this problem and an intelligent agent solving this problem is that the intelligent agent does not know the correct path they need to take, so it chooses a spot at random. From here the intelligent agent will use multiple trials to see what works and what doesn’t. The agent will get the input data which will include the starting point in the maze (which will be random. Then when the algorithm is implemented, it will try and find the best path to take, the agent will accomplish this by doing multiple runs. Then the agent will print out the output which should be the shortest, most efficient pathway to the treasure.

One of the differences that were already mentioned is in the way the agent and the human will begin the treasure hunt. As stated above the human will be able to look at the map and pick a spot to start at based on the knowledge of the map they have, and the agents stated in our directions should start on the top left of the 8x8 matrix. One of the similarities between these two is that the both the human and the intelligent agent will know what the problem is by using whatever the output of their decision is. One of the major differences is going to be the time that it takes for agent and the human to finish the problem. The human for example, could make the same mistakes twice by natural human error, there could be other distraction in the environment that could lead to a human making a mistake. The intelligent agent on the other hand is only focusing on the problem at hand and it will be able to run through multiple games in a short amount of time. The agent will learn more from the mistakes that they make, in a quicker time, than compared to a human.

“In exploitation, an automation greedily peruses a path of learning that provides immediate rewards. In exploration however, an automation must decide to forego an immediate reward and selects instead a directionless exploration with the intent of discovering a greater reward elsewhere.” (Perez, 2017, para. 2) This could also be defined as exploitation will search all the sample inputs it is given and will test all these outputs for all possible outcomes of the solution. Exploration wont necessarily look for an immediate reward but will search test and improve all the possible solutions. If we look at our pathfinding problem, we should use some sort of combination of exploitation and exploration. Our agent will learn by going through the different pathways and will use exploitation to find an immediate reward which will be an advancement on the board, but the agent would also need some exploration because we need the agent to reach the end goal, and by the exploration definition that was stated above, we would also need the agent to look for better reward then the one it already has by advance one spot for example.

Reinforcement learning is something that we have used throughout this course and its were we have an agent that learns the environment that it is in, in this game the treasure hunt board is the environment, and our agent will be the one that makes the actions and will observe the reward it gets from these actions. Our RL agent receives a state in the environment and takes an action, as noted previously this will start out as a random process until there is more data that the agent collects, then the agent will use RL to maximize its reward and find the shortest path.

Q-learning from our textbook is “characterized by the fact that there are some police, which informs an agent of the actions to take in different scenarios.” (Beysolow, 2019, pg. 10) The goal of the deep Q learning implementation is to “find the best possible navigation sequence that results in reaching the treasure cell wile maximizing the reward.” First, we imported all the necessary libraries then create dour training environment for the agent. Next, we had to create our reward systems and create the learning agent. Next comes where we needed to implement the Q algorithm to learn the pathways this is done by having the agent choose and action and the preform it. Then it will measure the reward based on the move and then update our Q table and this process is done over and over until we have reached out win rate of 1.000

Citations

Lamba, A. (2018, August 27). *A brief introduction to reinforcement learning*. We’ve Moved to FreeCodeCamp.org/News. <https://medium.com/free-code-camp/a-brief-introduction-to-reinforcement-learning-7799af5840db>

Beysolow, II, Taweh. *Applied Reinforcement Learning with Python : With OpenAI Gym, Tensorflow, and Keras*, Apress L. P., 2019.*ProQuest Ebook Central*, <https://ebookcentral-proquest-com.ezproxy.snhu.edu/lib/snhu-ebooks/detail.action?docID=5880718>.

Perez, C. E. (2017, November 6). *Exploration, Exploitation and Imperfect Representation in Deep Learning*. Intuition Machine. <https://medium.com/intuitionmachine/exploration-exploitation-and-imperfect-representation-in-deep-learning-9472b67fdecd#:~:text=In%20exploitation%2C%20an%20automation%20greedily%20pursues%20a%20path>