Sammabine Admettre

Module 7

Project Two

CS-370

June 23, 2024

Someone requested me to come up with a strategy for removing a pirate labyrinth from the surroundings. For this assignment, I had to create a Deep Q Network (DQN) using Python. This DQN is responsible for the GameExperience and TreasureMaze classes (Deep Q Learning Algorithm for Simple Python Game Makes Player Stuck, n.d.). The pirate may move left, right, forwards, and backwards. The pirate agent was rewarded for successfully obtaining the booty, but they were also penalized for being stopped, entering the same location, and engaging in prohibited aggressive behavior (Deep Q Learning Algorithm for Simple Python Game Makes Player Stuck, n.d.). Using reinforcement learning, the pirate learnt the optimum way to behave in each state. Learning approaches are classified into two types: discovery and application.

Exploration is a reinforcement learning technique in which the agent choose activities that it has never done before or that it believes will help it learn something new and perform better in the long term (Weng, 2020). People utilize this method to decide what chores to do. The primary purpose of the research process is to learn new things and get a better understanding of them, which may lead to improved long-term decisions and performance (Weng, 2020).

Exploitation occurs when an agent behaves in a manner that it believes will result in the highest immediate payout based on what it knows so far (Weng, 2020). Exploitation attempts to induce the agent to perform its best in the short term by allowing it to make judgments based on its existing information.

Using speed to help you learn When the AI first tried to look into something, deterioration let it do a lot of research (Laskowski & Tucci, 2022). It's possible that it will use what it has learned in future projects. To find the alpha value, all you have to do is increase the time number ratio by the drop rate. If alpha is less than the cutoff epsilon (µ), the robot will stop wandering and use what it has learned (Laskowski & Tucci, 2022).

Since hyper parameter sensitization wasn't used, epsilon's value stayed at 0.1 throughout the trial. The only parameter that was changed by hand was the decay rate (Laskowski & Tucci, 2022).

At this stage, the haphazard journey and eventual reliance on information resemble the process of figuring out how to traverse a maze. Their first action may be to search for answers. Subsequently, they might use their acquired knowledge to develop exceptional concepts (2022). Humans and criminals possess contrasting perspectives and predictions. This hinders the pirate spy's ability to see activities occurring outside the designated region (2022). It may unintentionally encounter something, which people want to avoid. The use of the reinforcement learning (RL) technique, namely with the Deep Q-Network (DQN), does not need the use of a model (2022).

At this point, the route that is hard to predict and the final need for knowledge are like learning how to get through a maze. They might start by looking for answers, which would be their first move (Why Do You Not See Dropout Layers on Reinforcement Learning Examples?, n.d.). That being said, once it's done, they will be able to use what they have learned to create new ideas. People, including thieves, see and think about things in very different ways from other people. Therefore, it will be harder for the pirate spy to see things that are happening outside of the designated area (Why Do You Not See Dropout Layers on Reinforcement Learning Examples?, n.d.). By accident, it might come into contact with something that people are trying to stay away from. Reinforcement learning (RL) and deep learning (DQN) can work well together without using a model (Why Do You Not See Dropout Layers on Reinforcement Learning Examples?, n.d.).

Usually, exploration should take up the first phase of the learning process. This is done so the agent may investigate other routes and get knowledge about the surroundings. The agent may be able to apply their knowledge and perform better because of the shift in the balance toward exploitation brought about by their high degree of ability and understanding (Deep Reinforcement Learning Doesn’t Work Yet, n.d.).

Adjusting the decay rate allowed the different strategies used to be balanced. Eventually, the epsilon value of 0.1 was chosen since it was shown to have a 0.1 decay rate. This would enable up to ninety epochs of random learning also referred to as trial and error before implementing experience-based predictions. A simple labyrinth may not need ninety trials for someone who can forecast future events to successfully navigate it. It is hard to prove, nevertheless, whether such perspective is restricted to a single deed (Deep Reinforcement Learning Doesn’t Work Yet, n.d.).

***How can reinforcement learning help to determine the path to the goal (the treasure) by the agent (the pirate)?***

It has previously been shown that the agent can learn with experience. For each state, the user may get rules that relate it to a certain action (Ryther, 2021). When it comes to pathfinding, the states are the distinct locations on the map, and the actions are the many things the agent may perform to go between them (Ryther, 2021).

To achieve the long-term aim of improving the total expected return, the individual is either compensated or penalized for each successful activity (Ryther, 2021). According to what has been revealed so far, research and mining restricted the kind of activities that might be conducted inside the states (Ryther, 2021).

As the agent learns, it explores its environment and adjusts its action-value function depending on what it discovers (Ryther, 2021). As the agent gains experience, its action-value function will ultimately converge on the optimal values, and the policy it learns will become more accurate and effective in determining the best path to the objective (Ryther, 2021).

***How did you implement deep Q-learning using neural networks for this game?***

The agent is guided by these rewards as it moves from the exploration phase to the attack phase. In this way, the person can take the next best move that is possible. This result was based on the idea that learning functions are getting worse, which was talked about in the previous paragraph.

**Reference**

(2022). Bfskinner.org. <https://www.bfskinner.org/newtestsite/wp-content/uploads/2014/02/ScienceHumanBehavior.pdf>

*Deep Reinforcement Learning Doesn’t Work Yet*. (n.d.). Www.alexirpan.com. https://www.alexirpan.com/2018/02/14/rl-hard.html

*Deep Q Learning Algorithm for Simple Python Game makes player stuck*. (n.d.). Artificial Intelligence Stack Exchange. Retrieved July 1, 2024, from <https://ai.stackexchange.com/questions/13793/deep-q-learning-algorithm-for-simple-python-game-makes-player-stuck>

Laskowski, N., & Tucci, L. (2022, July 1). *What Is Artificial Intelligence (AI)?* TechTarget; TechTarget. https://www.techtarget.com/searchenterpriseai/definition/AI-Artificial-Intelligence

‌

Ryther, J. (2021, December 4). *Treasure Maze Reinforced Learning*. Medium. https://medium.com/@jryther91/treasure-maze-reinforced-learning-744a9aa80a71

‌

Weng, L. (2020, June 7). *Exploration Strategies in Deep Reinforcement Learning*. Lilianweng.github.io. https://lilianweng.github.io/posts/2020-06-07-exploration-drl/

‌

*Why do you not see dropout layers on reinforcement learning examples?* (n.d.). Artificial Intelligence Stack Exchange. Retrieved July 1, 2024, from https://ai.stackexchange.com/questions/8293/why-do-you-not-see-dropout-layers-on-reinforcement-learning-examples

‌

‌