Current/Emerging Trends in Computer Science

2/15/2023

Jad Alrehaoui

SNHU

Project 2

* **Describe the steps a human being would take to solve this maze.**

A human being will be trying to navigate in any direction, for them to find that a direction is effective will be decided by their brain and memory. When reaching a dead end, a human being will try to go back and memorize the way things looked or the direction they took, in order not to end up in the same dead end. Guessing is the hero of this game from a human being’s perspective.

* **Describe the steps your intelligent agent is taking to solve this pathfinding problem.**

The intelligent agent will process the entire environment at first and will keep it in memory. Then it will use an algorithm to find a solution by getting a reward when reaching the end and getting penalized when reaching a dead end.

* **What are the similarities and differences between those two approaches?**

The intelligent agent is replicating human behavior in a more efficient and faster way. Because a human brain functions as an analog device, the AI will surely be faster and more precise tackling the problem and finding the solution.

* **What is the difference between exploitation and exploration? What is the ideal proportion of exploitation and exploration for this pathfinding problem ? Explain your reasoning.**

Exploration: is more of a long-term benefit concept where it allows the agent to improve its knowledge about each action which could lead to long term benefits.

Exploitation: basically exploits the agent’s current estimated value and chooses the greedy approach to get the most reward.

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

The agent is the one navigating the environment, it is the one deciding what movement should it do. However the agent is being rewarded and penalized for the movements it makes. When the agent finds the treasure (the goal), it is being rewarded big. As mentioned earlier, having a greedy agent, will keep it running to get the biggest reward every time.

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

In deep Q-learning, we utilize a neural network to approximate the Q value function. The network receives the state as an input and output the Q values for all possible actions. The biggest output is our next action.

The agent holds a memory buffer with all past experiences. His next action will be decided by the maximum output of the network, the loss function is the mean squared error of the predicted Q-value and the target Q-value. From the Bellman equation we have that the target is R + g x max(Q) . The difference between the target and the predicted values is called temporal difference error (TDE).

References:

Karagiannakos, S. (2019, May 3). *Q Learning and Deep Q Networks - Towards Data Science*.

Medium. <https://towardsdatascience.com/q-learning-and-deep-q-networks-436380e8396a>

NA. (2022, January 2). *Reinforcement Learning – Exploration vs Exploitation Tradeoff*. AI ML Analytics. <https://ai-ml-analytics.com/reinforcement-learning-exploration-vs-exploitation-tradeoff/>