Janai Cano

Domonique.cano@snhu.edu

Module Seven

CS 370: Emerging Trends in Computer Science

Project Two

Design Defense for Deep Q-Learning Network

In this project, an intelligent “pirate” agent was developed and trained to find treasure at the far end of a maze environment. The maze environment is represented by the following 8x8 matrix, where 1 represents a free cell and 0 represents an occupied cell. The pirate starts in the top left space, while the treasure is in the bottom right cell, both represented by 1’s:

maze = np.array([  
 [1., 0., 1., 1., 1., 1., 1., 1.],  
 [1., 0., 1., 1., 1., 0., 1., 1.],  
 [1., 1., 1., 1., 0., 1., 0., 1.],  
 [1., 1., 1., 0., 1., 1., 1., 1.],  
 [1., 1., 0., 1., 1., 1., 1., 1.],  
 [1., 1., 1., 0., 1., 0., 0., 0.],  
 [1., 1., 1., 0., 1., 1., 1., 1.],  
 [1., 1., 1., 1., 0., 1., 1., 1.]  
])

The primary purpose of the intelligent agent here is to efficiently navigate the maze and find the treasure before the human player. The agent must gather information about the maze layout, including obstacles and potential paths. Then, the agent must decide the best action to take at each step. After deciding on the best action, the agent must then execute said action, assessing and updating its knowledge base as it goes.

If a human agent were dropped into a maze and told to run and find the treasure, the first move would probably be obvious (since DOWN is the only valid choice). As humans, we can rely on our five senses to intuitively gather information about our environment. This is a major difference in how humans and machines problem solve: humans rely heavily on intuition, experience, and reasoning to make decisions (Eluyode & Akomolafe, 2013). Machines, however, excel in processing vast amounts of data and performing repetitive tasks efficiently (Kao & Venkatachalam, 2018). They can learn from experiences as well, and optimize their actions based on predefined algorithms and training data (Kao & Venkatachalam, 2018).

Aware of these differences in learning, a deep Q-learning neural network that implemented the epsilon-greedy algorithm was developed to train the intelligent agent. The DQN takes advantage of exploration vs. exploitation, which is a technique that humans and animals use when learning unknown environments (Mnih et al., 2015). Exploration of an environment involves taking actions without knowing their outcomes. In this way, the agent can then learn the outcome of a certain action, record the outcome in memory, and therefore “learn” (Mnih et al., 2015). Exploitation makes use of these learned outcomes, evaluating the most rewarding outcomes in order to come up with the most optimal actions the intelligent agent should take.

For the maze problem, the implementation of deep Q-learning using neural networks allowed the agent to efficiently learn a policy used for decision-making. The epsilon value was continuously updated, allowing it to change based on positive or negative action outcomes. This provides a way for the agent to continuously learn and adapt to a changing environment after each action. After 43 epochs, the agent was able to find the treasure with 100% accuracy.

A screenshot of a computer screen

Description automatically generated

A screenshot of a computer program

Description automatically generated

References

Eluyode, O. S., & Akomolafe, Dr. D. T. (2013). *Comparative study of biological and artificial neural networks*. Scholars Research Library; Ondo State University of Science and Technology. <https://citeseerx.ist.psu.edu/document?repid=rep1&type=pdf&doi=0f545038de9dd4d0d23b99e2c1707210d288b321>

Kao, Y.-F., & Venkatachalam, R. (2018). Human and Machine Learning. *Computational Economics*, *57*(3), 889–909. <https://doi.org/10.1007/s10614-018-9803-z>

Mnih, V., Kavukcuoglu, K., Silver, D., Rusu, A. A., Veness, J., Bellemare, M. G., Graves, A., Riedmiller, M., Fidjeland, A. K., Ostrovski, G., Petersen, S., Beattie, C., Sadik, A., Antonoglou, I., King, H., Kumaran, D., Wierstra, D., Legg, S., & Hassabis, D. (2015). Human-level Control through Deep Reinforcement Learning. *Nature*, *518*(7540), 529–533. https://doi.org/10.1038/nature14236

‌