Michael Clerico

CS370-Emerging Trends

Professor O’Rourke

When we look at Human vs AI, we are looking at two different learning styles. Humans solve problems by using prior experiences and intuition. They also utilize heuristic strategies like always moving forward or following a specific wall (Right or Left). They will also try to avoid previously bad areas. Then lastly humans will spatially scan the grid and recognize patterns of a relationship that will dictate movement. Machines on the other hand are entirely driven by data. They utilize a computational method of reinforcement learning. Within this learning experience the program will be awarded with positive outcomes. The software may also be valued, positive and negative.

To solve this maze a human would look at the entire layout from start to treasure. They would then look at all the obstacles and plan a route around them that would get them closer to the treasure. At every decision-making fork, they would justify direction as getting closer to the treasure. If they ended on a dead end, they would then backtrack to ensure they did not duplicate the failure. When new information is collected, a human will adjust accordingly. The AI system would do things a little differently. The agent would select any free cell randomly. It would then observe the current state as an input to the network. The action it would take would be based on the probability of exploration vs exploitation. After it does said selection, it will receive a reward or value. The AI will save this information in its system so it can replay the experience. The batches of samples are then updated into the system as Q-values. It repeats this process over and over until it improves performance and policy. Both computers and humans utilize past experiences to dictate future decisions. Humans tend to rapidly adapt vs the computers gradual approach through trial and error. Humans are generally guided by curiosity and knowledge verses a computer which is guided by policy.

The purpose of the agent in this exercise was to navigate and find the treasure efficiently. The AI utilized deep Q learning and adapted its policies through trial and error within the environment. The more it interacted, the better it could make decisions. The biggest attribute this demonstrates is the agent’s ability to learn autonomously. It was able to adjust policy based off experience on its own. In the end we wanted a 100%-win rate and the most optimal path to take. The optimal path is found through rewards. The computer wants to achieve the biggest reward, so it must minimize penalties that will occur when it wanders or hits obstacles.

Deep Q-Learning (DQN) was the choice for this program because it handles complex problems like games and navigation very well. DQN is great at handling state space. DQN does this by learning a function approximator for the q-values, so it doesn’t have to get representation for every state. This way through approximation it can generalize when scaling. DQN then continues to learn from experience and does not require explicit requirements of the environment. This allows the agent to explore the unknown very well. Some areas that have to be considered are the qtrain loops that can require long training times. This is because the algorithm requires large amounts of data from interaction. You must find an optimal setting to minimize hypersensitivity. The program was rewarded with the reward structure() positively and rewarded negatively with movement(). These rewards then shape the way the agents move to encourage efficiency. The GameExperience class helps create stability my managing the sampling experience through replay and storage.

In the end DQN is the optimal framework to effectively learn navigating the maze because of its ability to handle state space and learn optimal movements through rewards. With the hyperparameter tuning and sample time, it still requires important fine tuning to optimize efficiency.

References

DeepMind. (n.d.). *Research*. <https://deepmind.com/research>

Silver, D., Hubert, T., Schrittwieser, J., Antonoglou, I., Lai, M., Guez, A., ... & Hassabis, D. (2018). *A general reinforcement learning algorithm that masters chess, shogi, and Go through self-play*. DeepMind. <https://deepmind.com/research/publications/AlphaZero-2018>

Sutton, R. S., & Barto, A. G. (2018). *Reinforcement Learning: An Introduction* (2nd ed.). MIT Press.