Adam Resemius

CS-370

12-09-2023

A Defensive Approach to Intelligent Agents through Deep Q-Learning in the Treasure Maze

In project two, we have learned enough information to create our very own artificial intelligence. This AI was trained using the deep Q-learning algorithm. This algorithm is a form of reinforcement learning algorithm that employs a deep neural network to estimate the Q-function, that assists with identifying the most advantageous action to take within a state. The goal of this AI is to create an intelligent agent (pirate) that must navigate a maze environment to find the treasure before the human player. This design defense will explain key concepts, techniques, and decisions made in the development of the intelligent agent.

Humans vs. Intelligent Agent Approach

To solve a maze, humans rely on sensory perception, special reasoning, and memory recall to make important decisions. Furthermore, while emotions are not traditionally considered as tools for solving mazes, certain emotional states and cues can influence decision-making and problem-solving strategies. These cues can be motivation, or being goal oriented, or tap into other emotional states to solve the puzzle.

An intelligent agent can create a representation of the maze where it finds patterns that are like solving the maze. It then uses exploration and self-play to randomly select paths, evaluate outcomes, and update the neural network based on results of the simulation. The intelligent agent then scores itself and creates a goal to adjust to ensure that it optimizes better decision making in next iterations.

AlphaGo Zero, an artificial intelligence program developed by DeepMind, made history by defeating the world champion Go player, Ke Jie, in a three-game match in May 2017. This victory marked a significant achievement as AlphaGo Zero demonstrated superior gameplay and strategic understanding compared to human champions. AlphaGo Zero was able to do this because they decided to train the data to rely on human patterns and intuition from datasets. Instead, it used reconstructive learning to learn the game by itself and the machine learned to create its own patterns.

The smart agent that we have developed does the same course of action, where it learns solely by movements that is provided. This ensures that it relies solely on patterns of the game, that her than patterns of human emotion and intuition. This allows for the machine to have an advantage us humans may not be able to see.

Understanding Exploitation and Exploration

In the Pirate Puzzle it is important to find a medium ground between Exploitation and Exploration. To understand how the Pirate Intelligent Agent takes advantage of both Exploitation and Exploration, it is important to understand how each method plays a role in machine learning.

Exploitation is a strategy that involves choosing actions that are currently believed to be the best path based on existing knowledge that an AI creates. In the context of the Pirate Puzzle, exploitation would be like the agent consistently choosing actions it deems most likely to succeed in finding the treasure based on its learned patterns. While this seems to be ideal, it is not motivating to look for a better path for a more optimal solution.

Exploration plays a role in this, as this strategy steers the model off course to potentially discover a better strategy. In the Pirate Puzzle, exploration would entail the agent occasionally deviating from its known paths to discover better routes to the treasure.

It is important to find an ideal proportion of both exploration and exploitation because heavy reliance on exploitation may lead to a lack of adaptability to unforeseen circumstances or changes in the maze. Excessive exploration may result in inefficiency and delayed progress. A common strategy is to incorporate exploration with a decreasing probability over time, allowing the agent to prioritize exploitation as it gains confidence in its learned strategies.

Deep Q-Learning in Maze Navigation

Algorithms play a great role in machine learning as they can be used to fine-tune machine learning and provide systematic and efficient methodologies to lead to solutions. Deep-Q learning specifically addresses complex problems that are dynamic and have uncertain environments by guiding decision-making through learned experiences. It is not valuable in scenarios where the optimal solution is not readily apparent.

With the implementation of Deep Q-Learning to train our AI mode, we were able to fine-tune various hyperparameters to enhance learning efficiency. Challenges included optimizing exploration rates to improve efficiency within the agent without sacrificing adaptability. In the program we found that an exploration rate of 33% to begin with was a sweet spot that allows for the AI to explore, but not completely find new paths at the beginning. This in turn, allowed our Agent to approach a 100%-win rate after 155 epochs.

Conclusion

This defense document highlights the importance of balancing exploration and exploitation in the development of an intelligent agent for maze navigation, specifically in the context of the Pirate Puzzle. By leveraging the deep Q-learning algorithm, our AI model demonstrated the ability to learn and adapt without relying on human patterns or intuition. Drawing inspiration from AlphaGo Zero's groundbreaking approach to self-learning, our intelligent agent focused on creating its own patterns through reconstructive learning. This strategic decision allowed the agent to surpass human gameplay and achieve a high level of efficiency in maze navigation. The exploration-exploitation dilemma was addressed through a carefully calibrated exploration rate, striking a balance between discovering new paths and exploiting learned strategies. The success of our AI model, reaching a 100%-win rate after 155 epochs, underscores the effectiveness of deep Q-learning in addressing complex, dynamic problems like maze navigation. This project contributes valuable insights into the development of intelligent agents and reinforces the significance of reinforcement learning to advance machine learning capabilities.

**References:**

Choudhary, A. (2023, August 21). *A hands-on introduction to deep Q-learning using openai gym in python*. Analytics Vidhya. https://www.analyticsvidhya.com/blog/2019/04/introduction-deep-q-learning-python/

Dong, X., Wu, J., & Zhou, L. (n.d.). Index terms—deep learning, geometry, AlphaGo Zero, Gan arXiv:1711 ... https://arxiv.org/pdf/1711.09091.pdf

GeeksforGeeks. (2023, January 23). *Deep Q-learning*. GeeksforGeeks. https://www.geeksforgeeks.org/deep-q-learning/

Silver, D. (2017, October 18). AlphaGo Zero: Starting from scratch. Google DeepMind. https://deepmind.google/discover/blog/alphago-zero-starting-from-scratch/