Pravishna Nand

CS 370

6/23/24

7-3 Project Two Submission

**Design Defense: Analyze Human vs. Machine Intelligence**

Humans and machines employ distinct approaches to solving problems like maze pathfinding, influenced by their respective cognitive and computational abilities. A human solving a maze typically starts with an overview of the maze layout, using visual perception and spatial reasoning to identify potential paths and obstacles. They might employ heuristic strategies, such as the right-hand or left-hand rule, to systematically explore the maze while keeping track of visited paths through memory and recognizing landmarks. This process involves trial and error, where humans dynamically adjust their path based on feedback and intuition, allowing for flexibility and quick adaptation to new information. In contrast, an intelligent agent powered by deep Q-learning begins with no prior knowledge of the maze and learns through systematic exploration. The agent collects data by randomly exploring paths, then uses this information to update its Q-table or neural network, which predicts the value of actions in different states. The agent gradually shifts from exploration to exploitation, optimizing its path based on learned policies to maximize rewards. Unlike humans, the agent relies on algorithmic processes and data-driven optimization, following a predefined strategy without intuition but achieving consistent efficiency over time. Both humans and machines seek to navigate the maze effectively, but while humans rely on cognitive shortcuts and adaptability, machines depend on rigorous data analysis and iterative learning to find the optimal solution. This results in humans being better at handling ambiguous or changing environments, whereas machines excel in systematic problem-solving within well-defined parameters.

**Design Defense: Purpose of Intelligent Agent**

The purpose of the intelligent agent in pathfinding is to autonomously navigate a maze to find the most efficient route to a goal, such as treasure, while adapting to the complexity of the environment. The agent leverages reinforcement learning, particularly deep Q-learning, to balance exploitation and exploration effectively. Exploitation involves using the agent’s current knowledge to follow known paths that have previously yielded high rewards, while exploration involves trying new paths to discover potentially better routes. An ideal proportion between exploitation and exploration is critical: in the early stages of learning, a higher rate of exploration (e.g., 70-80%) is beneficial to gather comprehensive information about the maze's layout and potential paths. As the agent’s understanding improves, the focus should gradually shift towards exploitation (e.g., reducing exploration to 20-30%), allowing it to capitalize on the knowledge acquired to consistently choose optimal paths. Reinforcement learning facilitates this process by enabling the agent to learn from the rewards and penalties received during exploration. The agent updates its Q-values or neural network based on these interactions, progressively refining its policy to maximize cumulative rewards. This iterative learning process helps the agent to effectively determine the optimal path to the goal by continuously adapting its strategy based on feedback from the environment, leading to improved navigation efficiency and decision-making over time.

**Design Defense: Evaluating Algorithms**

The use of algorithms, particularly deep Q-learning with neural networks, provides a powerful framework for solving complex problems like pathfinding in games by enabling agents to learn optimal strategies through interaction and feedback. To implement deep Q-learning for this game, we designed a neural network to approximate the Q-value function, which estimates the expected rewards for taking specific actions in given states. The process began by initializing a neural network with input nodes corresponding to the state representation of the maze and output nodes representing possible actions (e.g., move up, down, left, right). During training, the agent explored the maze and recorded transitions, including the current state, chosen action, received reward, and the subsequent state. This data was used to compute the Q-value updates, minimizing the difference between predicted and actual rewards (known as the Bellman equation) through a loss function. The neural network was trained using backpropagation and gradient descent to adjust its weights based on these updates. To improve stability and learning efficiency, techniques such as experience replay were employed, where the agent stored past experiences in a memory buffer and sampled them randomly for training, preventing correlations in the training data and smoothing out updates. Additionally, a target network was used to provide stable Q-value targets, updated periodically to match the primary network, further enhancing learning stability. This iterative process allowed the agent to refine its policy, balancing exploration of new paths and exploitation of known high-reward paths, ultimately enabling it to navigate the maze effectively by consistently selecting actions that maximize cumulative rewards. This approach illustrates the robustness of combining neural networks with reinforcement learning to tackle complex, dynamic environments where explicit programming of optimal strategies is impractical.

**References:**Lillicrap, T.P., Hunt, J.J., Pritzel, A., Heess, N.M., Erez, T., Tassa, Y., Silver, D., & Wierstra, D. (2015). Continuous control with deep reinforcement learning. *CoRR, abs/1509.02971*.

Yoon, C. (2019, July 17). *Double deep Q networks: Tackling maximization bias in deep Q-learning*. Towards Data Science. https://towardsdatascience.com/double-deep-q-networks-tackling-maximization-bias-in-deep-q-learning-32801bd7d459

Mnih, V., Puigdomenech Badia, A., Mirza, M., Graves, A., Lillicrap, T. P., Harley, T., ... & Kavukcuoglu, K. (2016). Asynchronous methods for deep reinforcement learning. In *International Conference on Machine Learning* (pp. 1928-1937). https://proceedings.mlr.press/v48/mniha16.pdf