Cody VanGosen

CS-370 Current/Emerging Trends in CS

Professor Morales

December 9, 2024

**Design Defense**

Humans approach solving a maze by leveraging heuristic reasoning, intuition, and cognitive flexibility. A person typically begins by visually scanning the environment to identify a starting point, the destination, and potential pathways. They might then attempt an initial route, making decisions at each junction based on perceived proximity to the goal. When a chosen path leads to a dead end, humans backtrack and explore alternative routes. This trial-and-error process is guided by memory, enabling them to avoid revisiting previously failed paths. Humans also use spatial reasoning and visual cues to prioritize pathways that seem more direct or likely to succeed (Jones, 2021). However, while human problem-solving is adaptable, it is not inherently systematic, which can result in inconsistent performance, particularly in complex environments with numerous branching options.

The intelligent agent developed for this project approaches the maze-solving problem using a fundamentally different methodology. It employs a deep Q-learning algorithm, a reinforcement learning technique that systematically evaluates the environment and iteratively refines its strategy. The agent begins by interacting with the maze in a random manner, collecting feedback in the form of rewards and penalties based on its actions. Over time, it learns to predict the optimal action for any given state by updating its Q-value estimates using the Bellman equation. Unlike humans, who rely on intuition and memory, the agent uses a mathematical framework to quantify the desirability of each action, ensuring consistent optimization (Sutton & Barto, 2018).

Despite these methodological differences, there are notable similarities between human and machine approaches. Both rely on feedback to adjust their strategies and improve performance. Just as humans learn from their mistakes, the agent uses penalties for invalid moves or suboptimal choices to refine its decision-making process. However, the machine’s reliance on systematic exploration and exploitation ensures it evaluates all possible options, often leading to more efficient and optimal solutions compared to a human’s heuristic approach.

The purpose of the intelligent agent in this project is to autonomously navigate the maze and find the treasure while adhering to predefined constraints. To achieve this, the agent must balance two competing objectives: exploration and exploitation. Exploration involves testing unknown or less familiar paths to discover potentially better strategies, while exploitation focuses on leveraging known pathways to maximize immediate rewards. Striking the right balance between these objectives is critical. If the agent over explores, it wastes resources by unnecessarily revisiting suboptimal paths. Conversely, excessive exploitation can prevent the agent from discovering more efficient routes. The epsilon-greedy strategy was employed to address this challenge, where the agent begins with a high exploration rate that gradually decreases as it learns more about the environment. This approach ensures that the agent collects sufficient information during early stages of training while converging toward optimal behavior in later stages (Tokic, 2010).

Reinforcement learning plays a central role in the agent’s ability to determine the optimal path to the treasure. By interacting with the environment, the agent learns to associate specific actions with rewards, gradually constructing a policy that maximizes its cumulative reward over time. This trial-and-error learning process mimics aspects of behavioral conditioning, enabling the agent to autonomously adapt its strategy based on its experiences. For example, the agent penalizes itself for invalid moves, such as attempting to pass through a wall or moving out of bounds, while rewarding itself for reaching the treasure. These rewards and penalties guide the agent toward more efficient navigation strategies, demonstrating the effectiveness of reinforcement learning in solving pathfinding problems (Sutton & Barto, 2018).

The implementation of deep Q-learning in this project involves several key components. The neural network model is structured with an input layer corresponding to the maze states, multiple hidden layers with ReLU activation functions, and an output layer that predicts Q-values for all possible actions. Experience replay is utilized to enhance the agent’s learning process by storing past episodes and reusing them during training. This technique prevents overfitting and ensures the agent generalizes well to new scenarios within the maze. The training process also incorporates an Adam optimizer to minimize the loss function and a discount factor to appropriately weigh future rewards. The result is a robust and efficient model capable of solving the maze consistently and autonomously (Mnih et al., 2015).

The agent’s performance was evaluated through systematic testing, achieving a 100%-win rate by epoch 260. This indicates that the agent successfully learned the optimal policy for navigating the maze, demonstrating the power of deep reinforcement learning in addressing complex decision-making tasks. The use of a neural network model allowed the agent to generalize its knowledge across different states, while the exploration-exploitation trade-off ensured it efficiently balanced learning and performance optimization.

Overall, this project highlights the potential of artificial intelligence to solve intricate problems that require systematic exploration and decision-making. By leveraging deep Q-learning and reinforcement learning principles, the intelligent agent demonstrates an ability to outperform human problem-solving in certain contexts, offering a scalable and reliable solution to the maze navigation challenge.

**References**

Jones, P. (2021). *Human problem-solving and heuristics in spatial navigation*. *Journal of Cognitive Research, 15*(4), 245–258.

Mnih, V., Kavukcuoglu, K., Silver, D., Rusu, A. A., Veness, J., Bellemare, M. G., ... & Hassabis, D. (2015). Human-level control through deep reinforcement learning. *Nature, 518*(7540), 529–533.

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

Tokic, M. (2010). Adaptive epsilon-greedy exploration in reinforcement learning based on value differences. In *KI 2010: Advances in Artificial Intelligence* (pp. 47–58). Springer.