CS 370

Defense Design

Caleb Ewer

2/8/2025

**Human Approach:**

* Observation: A human observes the maze, identifies obstacles, starting point, and goal location.
* Planning: A human might mentally map out a potential path using pattern recognition, e.g., looking for open corridors or shortest visible routes.
* Trial and Error: A human would test paths, backtrack if encountering dead ends, and avoid revisiting areas previously explored.
* Adaption: Humans adjust strategies dynamically based on their understanding of the maze as they progress.

**Intelligent Agent Approach:**

* Exploration: The agent initially takes random valid actions to explore the maze.
* Experience storage: The agent stores past experiences (state transitions, rewards) in memory using experience replay.
* Training with Rewards: Using reinforcement learning, the agent assigns higher rewards to paths that lead closer to the treasure and penalties for invalid or longer routes.
* Policy Development: The neural network generalizes from experiences to predict optimal actions in unseen states.
* Exploitation: Once trained, the agent uses its learned policy to take the shortest or most rewarding path to the treasure.

**Similarities and Differences Between the Two Approaches:**

Similarities: Both approaches involve exploration, trial and error, adaption based on feedback, and both aim to minimize unnecessary steps while reaching the goal.

Differences:

* Humans rely on intuition and real-time decision-making; agents rely on systematic training and rewards.
* Humans prioritize efficiency from the start; agents might take longer initially due to exploration.
* Humans adapt dynamically, while agents require pre-defined parameters for exploration and exploitation.

**Exploration vs. Exploitation:**

Exploration: The agent takes random actions to discover new paths and learn the environment.

Exploitation: The agent uses its learned policy to select the best-known actions to maximize rewards.

**Ideal Proportion:**

A balance is critical: start with high exploration to learn the maze, then gradually shift to exploitation. Ex. Use an epsilon decay strategy, starting with epsilon = 1.0 (100% exploration) and decaying to epsilon = 0.05 (95% exploitation). High exploration early ensure that the agent discovers all potential paths and exploitation later ensures the agent consistently uses the best paths.

**Reinforcement Learning and Pathfinding:**

Reinforcement learning helps the agent learn the optimal path through trial and error, guided by rewards. Some of the key mechanisms include:

* Rewards: Positive rewards for steps toward the goal, negative rewards for invalid or long paths.
* Value Estimation: The agent predicts future rewards for each state-action pair and learns to maximize these values.
* Q-Learning: Updates the agent’s knowledge (Q-Values) to improve decision-making over time.

Through repeated episodes, the agent builds a policy that guarantees reaching the goal.

**Implementation of Deep Q-Learning:**

* Environment: Used TreasureMaze to simulate the environment, including states, actions, and rewards.
* Model Architecture: Built a nural network with layers to predict Q-Values for each state-action pair.
* Experience Replay: Stored episodes in memory to train the model on diverse experiences.
* Training: Used reinforcement learning to iteratively train the agent by maximizing rewards and penalizing suboptimal moves.
* Policy Execution: After training, the agent followed its learned policy to solve the maze efficiently.

**References:**

Sutton, R. S., & Barto, A. G. (2018). *Reinforcement learning: An introduction*. MIT press.

Goodfellow, I., Bengio, Y., & Courville, A. (2016). *Deep learning*. MIT press.