Northeastern University

Reinforcement Learning and Sequential Decision Making – CS5180 LAB

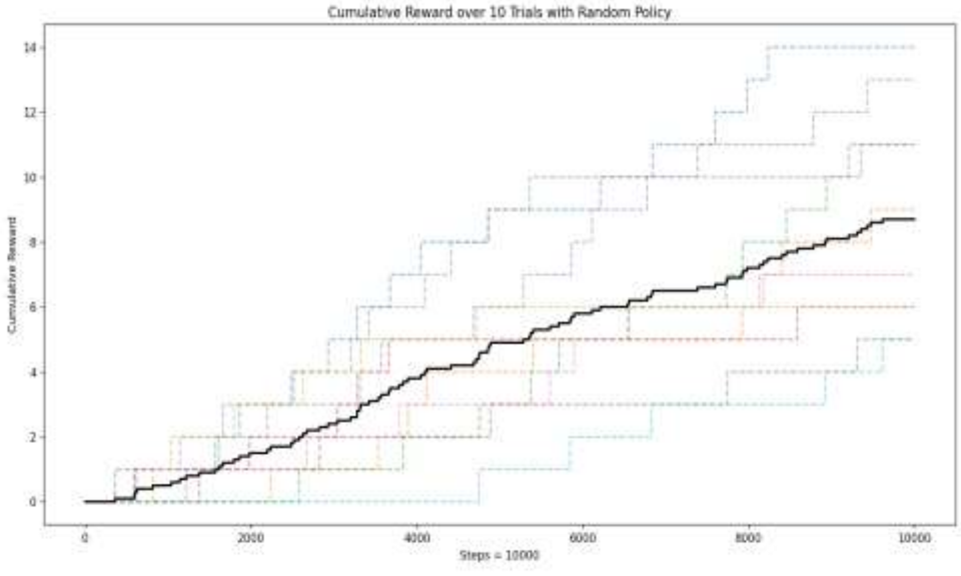
EX-0

Study partner: Abhijit Kaluri

Varun Samala

3. In situations marked by uncertainty, a strategy informed by human expertise generally proves more effective than one relying on random choices. When individuals make decisions, they can thoughtfully assess various options, choosing actions with a higher probability of positive outcomes, thereby increasing the chances of achieving the intended objective. Conversely, a random approach lacks strategic thinking, resulting in unpredictable and often suboptimal results. Additionally, human-driven strategies hold the advantage of adaptability and evolution based on new information and experiences, while a random approach remains static and unable to enhance its effectiveness over time. This fundamental distinction underscores the superior reliability of human-guided methods in accomplishing predetermined goals compared to random strategies.

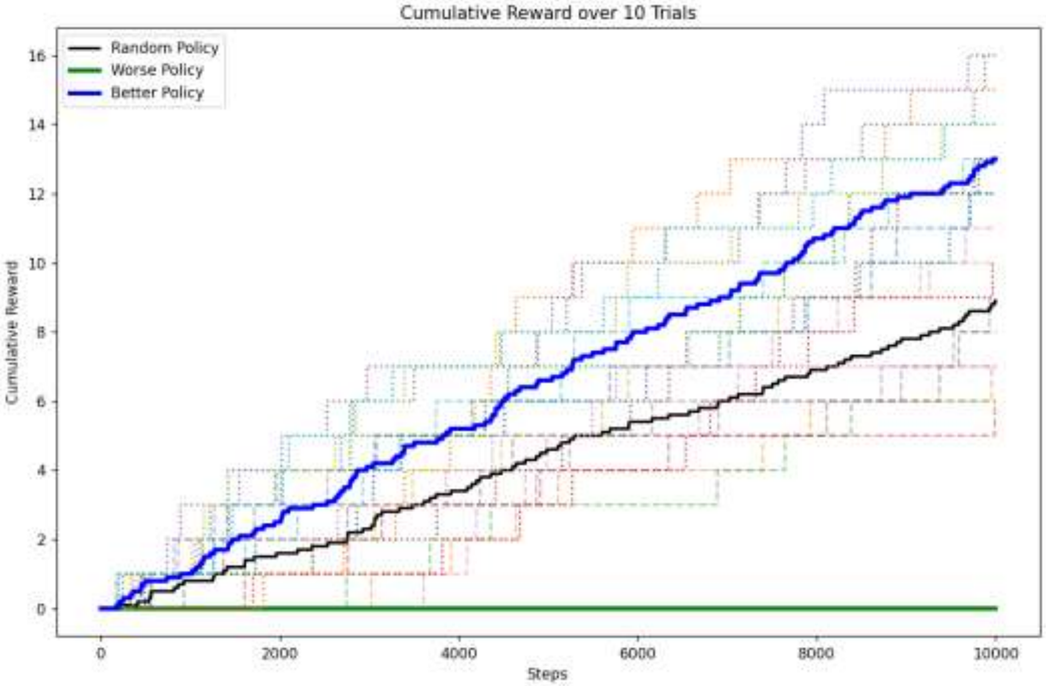
PLOT:



Cumulative reward vs Steps for Random policy

4.Better Policy Strategy:
The improved policy is designed to detect walls and corners, guiding the agent toward the goal efficiently. When the agent is in the middle of the room without walls or corners, the policy resorts to a random approach. The key feature is that, in the presence of walls or corners, the policy follows the wall, significantly increasing the probability of taking the correct action (0.5) compared to the random policy (0.25). This adaptive strategy enhances performance, enabling the agent to reach the goal more effectively.

Worse Policy Strategy:
The inferior policy, on the other hand, adopts a simplistic strategy by consistently moving to the right, irrespective of the agent's current state. This deterministic approach makes it highly unlikely for the agent to reach the goal position and receive a reward. Unlike the random policy that exhibits adaptability, the worse policy's rigid behavior results in poor performance over time, as it consistently takes incorrect steps that do not lead the agent to the goal position.



Cumulative reward vs Steps for Better policy, Random policy and Worse policy

5. Q-learning Agent Strategy:

The Q-learning agent employs the Q-learning algorithm, which is a model-free reinforcement learning algorithm. The strategy involves learning a policy, represented by the Q-values, that maximizes the cumulative reward over time. Here are the key components of the agent's strategy:

1. Exploration-Exploitation Trade-off:
    - The agent uses an epsilon-greedy policy for action selection, balancing exploration and exploitation. With probability epsilon, the agent explores by choosing a random action, and with probability 1 - epsilon, it exploits by selecting the action with the highest Q-value.
2. Q-learning Update Rule:
    - The agent updates its Q-values using the Q-learning update rule. The Q-value for a state-action pair is updated based on the current reward and the maximum Q-value for the next state.

Q(state, action) = (1 - alpha) * Q(state, action) + alpha * [reward + gamma * max(Q(next_state, all_actions))]	

- alpha is the learning rate, controlling the weight given to new information.
- gamma is the discount factor, determining the importance of future rewards.

Learning from Experience:

- The agent learns from its experience during multiple trials. It explores different actions in various states, updating Q-values to improve its policy over time.

Learned Policy:

The learned policy is represented by the Q-values stored in the Q-table. The Q-values indicate the expected cumulative reward for taking a specific action in a particular state. After running the Q-learning algorithm, the agent has learned a policy that leads to higher cumulative rewards.

Interpretation of Results:

The end results show cumulative rewards for each trial and the average cumulative reward over all trials. The cumulative rewards for individual trials indicate how well the agent performed in each scenario. The average cumulative reward provides an overall assessment of the agent's effectiveness.