ex1.pdf
PDF

Give solution to this exercise

Solution to Exercise 1 – Introduction to Markov Decision Processes (MDPs)

1. Formalizing Problems as MDPs

Each problem is formalized as an MDP $\langle S, A, P, R, \gamma \rangle$:

(a) Road Crossing

The agent must cross a single-lane road with randomly changing traffic conditions.

- **States (S):**
  - $s \in \{(x, c)\}$, where:
    - x is the agent's position ({start, middle, goal}).
    - c is the lane occupancy ({empty, occupied}).
- **Actions (A):**
  - $a \in \{\text{wait}, \text{move}\}$, where:
    - "wait" keeps the agent in the current position.
    - "move" advances to the next position if no car is in the lane.
- **Transition Probabilities (P):**
  - $P(s' \mid s, a)$ is defined by:
    - If $a = \text{wait}$, x stays the same, and c changes randomly.
    - If $a = \text{move}$, the agent moves forward only if $c = \text{empty}$.
- **Reward Function (R):**
  - $R = +1$ for reaching the goal.
  - $R = 0$ for waiting or safe movement.
  - $R = -1$ if the agent is hit by a car.
- **Discount Factor (γ):**
  - $\gamma \in (0, 1]$, depending on whether crossing speed is incentivized.

(b) Chess Against a Random Opponent

The agent plays chess against an opponent making random moves.

- **States (S):**
  - Each board configuration is a state.
  - Approximate cardinality: 10^{43} (all possible board configurations).
- **Actions (A):**
  - All legal chess moves at each state.
  - Approximate cardinality: Around 30-35 moves per state.
- **Transition Probabilities (P):**
  - Deterministic if the agent plays.
  - Stochastic if the opponent plays (uniform random move selection).
- **Reward Function (R):**
  - $R = +1$ for a win.
  - $R = 0$ for a draw.
  - $R = -1$ for a loss.
- **Discount Factor (γ):**
  - Typically close to 1, encouraging long-term strategic play.

2. Maze Runner - Modifying the MDP

The agent takes too long to solve the maze due to lack of urgency.

Problem:

- Since $R = 0$ for all steps except reaching the goal ($+1$), the agent has no incentive to act quickly.
- The agent may wander indefinitely without prioritizing the goal.

Solution:

- Introduce a **small negative reward per timestep**, e.g., $R = -0.01$, to encourage faster solutions.
- Use **lower discount factor** $\gamma < 1$ to prioritize short-term gains.

3. Markov Property & Camera Sensing

The Markov property states that the future depends only on the present state, not past states.

Does s_t as a camera image satisfy the Markov property?

- **No**, because a single image does not capture velocity or motion of objects.
- Important information, such as whether an object is moving, is lost.

How to satisfy the Markov property?

- **Include past frames** (i.e., sequence of images) to capture motion.
- **Feature extraction**: Use optical flow or recurrent models (RNNs, LSTMs) to track moving objects.

This solution provides a structured approach to formalizing MDPs and addressing challenges in reinforcement learning problems.