John Smith, Ph.D.

Department of Computer Science
University Name

Email: email@university.edu
Website: www.university.edu

July 19, 2025

John Smith, Ph.D.

Department of Computer Science
University Name

Email: email@university.edu
Website: www.university.edu

July 19, 2025

Introduction to Markov Decision Processes (MDPs)

What are MDPs?

Markov Decision Processes (MDPs) are a foundational framework used in reinforcement learning (RL) for modeling decision-making scenarios with randomness and decision control.

Key Concepts of MDPs

- **States**: Possible situations the agent can be in, containing all necessary information for decision-making.
- **Actions**: Choices available to the agent in each state, affecting future states.
- **Rewards**: Feedback received after actions, indicating the quality of the action relative to achieving goals (positive or negative).
- **Transition Probabilities**: Likelihood of moving from one state to another after taking an action, encapsulating the environment's dynamics.

Mathematical Representation of MDPs

An MDP is formally defined by the tuple (S, A, P, R, γ):

- S: A finite set of states.
- A: A finite set of actions.
- $P(s'|s, a)$: Transition probability of reaching state s' after action a in state s.
- $R(s, a)$: Reward function assigning a scalar reward for each action in a state.
- γ: Discount factor, $0 \leq \gamma < 1$, representing future rewards' importance.

Importance of MDPs in RL

MDPs are pivotal in reinforcement learning because they:

- Provide structured decision-making models for sequential problems.
- Support algorithms (e.g., Value Iteration, Policy Iteration) that converge to optimal solutions.
- Are applicable across various fields, from robotics to economics, for uncertain decision-making environments.

Example Scenario

Consider a robot navigating a room to reach a charger:

- **States**: Positions in the room.
- **Actions**: Movements (up, down, left, right).
- **Rewards**: Highest when reaching the charger (positive reward), lower or negative when hitting walls.
- **Transition Probabilities**: Allow for movement uncertainties, representing chances of slipping or mismoving.

Key Takeaways

- MDPs serve as a backbone for many reinforcement learning algorithms.
- Understanding MDPs is essential for designing intelligent agents that learn from their environments.
- The interplay between states, actions, rewards, and transition probabilities forms the essence of decision-making in uncertain scenarios.

MDP Components - Overview

Key Components of Markov Decision Processes (MDPs)

MDPs provide a framework for modeling decision-making in environments with random outcomes. Understanding these components is essential for grasping reinforcement learning algorithms:

- States (S)
- Actions (A)
- Rewards (R)
- Transition Probabilities (P)

MDP Components - States and Actions

1. States (S)

A state represents a specific situation or configuration of the environment at a particular time.

- **Notation**: Denoted as S.
- The state space S is a collection of all possible states.
- **Example**: In chess, each arrangement of pieces is a different state.

2. Actions (A)

An action is a decision made by the agent that may affect the state.

- **Notation**: Denoted as A.
- For every state, there's a set of actions the agent can take.
- **Example**: In chess, actions include moving a piece or forfeiting.

MDP Components - Rewards and Transitions

3. Rewards (R)

A reward is a feedback signal received after taking an action in a state, representing the immediate benefit.

- **Notation**: Denoted as $R(s, a)$.
- Rewards guide the learning process, maximizing cumulative rewards over time.
- **Example**: Capturing a piece in a game might yield $+10$ points.

4. Transition Probabilities (P)

Transition probabilities specify the likelihood of moving from one state to another after an action.

- **Notation**: Denoted as $P(s'|s, a)$.
- Reflects the dynamics of the environment and uncertainty in outcomes.

States in Markov Decision Processes (MDPs)

Definition of a State

A state in an MDP represents a specific situation in the environment at a particular point in time. It encapsulates all the relevant information needed to make a decision.

Importance of States

States are the foundation of an MDP, determining the current context for decisions. Each state can lead to various actions and different potential outcomes.

Characteristics of States

- **Discrete vs. Continuous States**
  - **Discrete States**: Clearly distinguished and finite situations (e.g., positions on a chessboard).
  - **Continuous States**: Represented on a continuous scale (e.g., height of a robot arm).
- **Episodic vs. Continuing Tasks**
  - **Episodic**: The process has a defined endpoint (e.g., game episodes).
  - **Continuing**: The process continues indefinitely without a specific end.

Examples and Notation

- **Examples of States**
  - **Robotics**: Each state could represent the robot's location and orientation (e.g., (x,y) coordinates and angle).
  - **Game Playing**: In chess, each state corresponds to a unique configuration of the board and pieces.
  - **Navigation**: In a GPS system, states reflect different locations along a route, including current traffic conditions.
- **Notation and Representation**
  - **State Space**: Denoted by S, it is the set of all possible states in an MDP.
  - **Markov Property**:

$$P(S_{t+1}|S_t, A_t) = P(S_{t+1}|S_t)$$

Key Points to Remember

- States are critical as they influence decision-making through the actions that can be taken.
- Understanding the structure of states aids in designing effective strategies for control and optimization.
- The relationship between states, actions, and rewards forms the core of the MDP framework, guiding an agent's learning algorithm.

Summary

States in MDPs are essential elements that reflect the current situation, impacting possible actions and future outcomes. Understanding these states is vital for comprehending the dynamics of decision-making in uncertain environments.

Actions in MDPs - Overview

Understanding Actions in MDPs

In Markov Decision Processes (MDPs), an **action** is a decision made by an agent that influences the environment state. Actions are crucial for determining transitions between states.

Actions in MDPs - Key Concepts

1. **Action (A)**:
   - A choice made by the agent at any state.
   - Denoted as A, with various options available.
2. **Transition Dynamics**:
   - Taking an action leads to a new state with probabilistic outcomes.
   - Represented as:

$$P(s'|s, a)$$

   where s is the current state, a is the action, and s' is the next state.
3. **State-Action Pairs**:
   - Represent options available from each state - can be visualized in a table or diagram.

Examples and Conclusion

Example: Grid World

An agent can move in four directions: up, down, left, right.

- **States**: Positions in the grid.
- **Actions**:
  - A_1: Move up
  - A_2: Move down
  - A_3: Move left
  - A_4: Move right
- **Transition Dynamics**: e.g., from state $(2, 2)$,

$$P((1, 2)|(2, 2), A_1) = 0.7, \quad P((2, 1)|(2, 2), A_1) = 0.3$$

Conclusion

Understanding actions in MDPs is vital for modeling decision-making in uncertain

Rewards in MDPs - Overview

Rewards in Markov Decision Processes (MDPs) play a crucial role in guiding the reinforcement learning process. A reward is a scalar signal received after an agent takes an action in a given state, indicating the immediate value of that action and influencing the policy learned by the agent over time.

Definition

A reward R quantifies the benefit or cost associated with a specific action taken in a particular state. It can be positive (reward) or negative (penalty).

Rewards in MDPs - Significance

The significance of rewards in MDPs includes:

1. **Policy Formation**: Rewards directly influence the agent's policy, mapping states to actions. The objective is to maximize cumulative rewards over time.
2. **Cumulative Reward and Return**: The return G_t is defined as:

$$G_t = R_t + \gamma R_{t+1} + \gamma^2 R_{t+2} + \ldots = \sum_{k=0}^{\infty} \gamma^k R_{t+k} \tag{1}$$

   where:
   - G_t is the return at time t,
   - R_t is the reward received at time t,
   - γ (where $0 \leq \gamma < 1$) is the discount factor.
3. **Motivation for Actions**: Rewards motivate agents to explore and exploit their environment, encouraging favorable actions in similar future scenarios.

Rewards in MDPs - Example

Example: Robot in a Grid Environment

- **State**: Current position of the robot in a grid.
- **Action**: Move up, down, left, or right.
- **Reward Structure**:
  - +10 for reaching a goal state (e.g., finding a treasure).
  - -1 for hitting a wall or going out of bounds.

Outcome: The robot learns to prefer actions leading to the goal state, thereby increasing its cumulative rewards over time.

Transition Probabilities - Overview

Definition

Transition probabilities in a Markov Decision Process (MDP) represent the likelihood of transitioning from one state to another when a particular action is taken. They quantify the uncertainty associated with the outcomes of actions within the environment.

Key Concepts

- **States:** The different configurations in which an agent might find itself, denoted as S.
- **Actions:** Choices available to the agent, denoted as A.
- **Transition Function:** Denoted as $P(s'|s, a)$:
  - s: Current state
  - a: Action taken
  - s': Resulting state
  - $P(s'|s, a)$: Probability of transitioning to state s' when action a is taken in state s.

Example of Transition Probabilities

Grid World Scenario

In a simple grid world, an agent can move Up, Down, Left, or Right. For example:

- **States (S):** Positions like (1,1), (1,2), (2,1).
- **Actions (A):** Moves (Up, Down, Left, Right).

If at position (1,1) and the action is Right:

- $P((1,2)|(1,1), \text{Right}) = 0.8$
- $P((1,1)|(1,1), \text{Right}) = 0.2$

This indicates an 80

Key Points and Summary

- **Deterministic vs. Stochastic:** Transition probabilities can either be deterministic (certain outcome) or stochastic (spread across multiple outcomes).
- **Importance in Learning:** Essential for reinforcement learning as they help predict future states and guide decision-making.
- **Markov Property:** States depend only on the current state and the action taken, not past states or actions.

Transition Matrix

The transition probabilities can be represented in a matrix P:

$$P = \begin{bmatrix} P(s_1|s_1, a_1) & P(s_2|s_1, a_1) & \cdots \\ P(s_1|s_2, a_1) & P(s_2|s_2, a_1) & \cdots \\ \vdots & \vdots & \ddots \end{bmatrix}$$

where each entry $P(s'|s, a)$ signifies the probability of transitioning from state s to state s'

Moving Forward

Next, we will explore the concept of policies, which dictate how an agent should act in the different states based on transition probabilities. These policies utilize the information gathered from transition probabilities to maximize expected rewards over time.

The Concept of Policies - Introduction

Introduction to Policies

In the context of Markov Decision Processes (MDPs), a **policy** is a crucial concept that dictates the behavior of an agent in a given environment. Think of a policy as a strategy that specifies the actions an agent should take when it encounters different states.

The Concept of Policies - Understanding Policies

- **Definition**: A policy is a mapping from states of the MDP to actions. It can be:
  - **Deterministic Policy**: A specific action is chosen for each state.
    - Example: If the state is "Hungry", the action is "Eat".
  - **Stochastic Policy**: Actions are chosen based on a probability distribution.
    - Example: In the state "Traffic Light: Red", the action could be "Wait" with a probability of 0.9 and "Run" with a probability of 0.1.

The Concept of Policies - Formal Notation and Key Points

Formal Notation

Let π denote a policy:

- **Deterministic Policy**:

$$\pi : S \to A \tag{2}$$

  Where S is the set of states and A is the set of actions.

- **Stochastic Policy**:

$$\pi(a|s) \quad \text{for } a \in A \text{ and } s \in S \tag{3}$$

  This denotes the probability of taking action a in state s.

- **Key Points to Emphasize**:
  1. Policies provide the decision-making framework for agents, guiding them on how to act in varying situations.
  2. The choice of policy significantly impacts the efficiency and effectiveness of an agent's

Value Functions - Introduction

What are Value Functions?

Value functions are fundamental in Markov Decision Processes (MDPs) as they evaluate expected returns from states or state-action pairs. They guide decision-making under uncertainty by providing a metric on how advantageous it is to be in a certain state or to take a specific action.

Value Functions - Types

Types of Value Functions:

1. **State Value Function V:**
   - **Definition**: The value of a state s, denoted as $V(s)$, is the expected return starting from state s and following a policy π.
   - **Mathematical Expression**:

$$V^{\pi}(s) = \mathbb{E}_{\pi}\left[R_t | S_t = s\right] = \sum_{a \in A} \pi(a|s) \sum_{s'} P(s'|s,a)\left[R(s,a,s') + \gamma V^{\pi}(s')\right] \tag{4}$$

2. **Action Value Function Q:**
   - **Definition**: The value of taking action a in state s, denoted as $Q(s,a)$, represents the expected return after taking action a and following policy π.
   - **Mathematical Expression**:

$$Q^{\pi}(s,a) = \mathbb{E}_{\pi}\left[R_t | S_t = s, A_t = a\right] = \sum_{s'} P(s'|s,a)\left[R(s,a,s') + \gamma V^{\pi}(s')\right] \tag{5}$$

Value Functions - Importance and Examples

Importance of Value Functions in MDPs:

- **Decision Making**: They assess long-term rewards of actions, leading to optimal decisions.
- **Policy Evaluation**: Help in measuring policy effectiveness by estimating expected returns.
- **Learning**: Used in reinforcement learning to refine policies through environment interaction.

Example Illustration:

Consider a grid world scenario:

- agent starts at state s and can earn a reward of $+10$ by reaching a goal state.
- Action a (moving towards the goal) has a high transition probability to successfully reach the goal, evaluated by the state value function.

This guides the agent in selecting the action that maximizes returns.

MDPs in Practice - Overview

Definition

Markov Decision Processes (MDPs) are mathematical frameworks for modeling decision-making where outcomes are partly random and partly under the control of a decision maker. They consist of:

- **States (S)**: All possible situations.
- **Actions (A)**: Choices available to the decision maker.
- **Transition Probabilities (P)**: The probability of reaching a next state given the current state and action.
- **Rewards (R)**: Feedback received after transitioning from one state to another.

Key Points to Emphasize

- Dynamic Nature: MDPs effectively handle environments that change over time.
- Learning Optimal Policies: MDPs help decision-makers learn the best actions to maximize

MDPs in Practice - Case Studies

1. **Autonomous Driving**
   - **States**: Positions on the road, nearby vehicles, traffic signals.
   - **Actions**: Accelerate, brake, turn, maintain speed.
   - **Reward**: Positive for safe driving and quick destination reaching; negative for accidents.
2. **Robot Navigation**
   - **States**: Different positions within the building.
   - **Actions**: Move North, South, East, West.
   - **Reward**: Positive for reaching the target; negative for hitting obstacles.
3. **Game Playing (e.g., Chess)**
   - **States**: All possible configurations of pieces on the board.
   - **Actions**: Moves allowed by chess rules.
   - **Reward**: Positive for winning; negative for losing pieces.

Example of a Basic MDP Setup

Basic MDP Structure

- **States (S)**: {S1, S2, S3}
- **Actions (A)**: {A1, A2}
- **Transition Model (P)**:

$$P(S2|S1, A1) = 0.7, \quad P(S3|S1, A1) = 0.3 \tag{6}$$

- **Rewards (R)**:

$$R(S1, A1) = 5, \quad R(S2, A2) = 10 \tag{7}$$

Conclusion

The examples demonstrate the flexibility of MDPs for modeling decision-making in uncertain environments, showcasing their significance in modern AI systems.

Summary and Conclusion - MDP Components

Recap of MDP Components

Markov Decision Processes (MDPs) are fundamental frameworks used in reinforcement learning. The key components include:

1. **States (S)**: Finite set representing all possible situations.
2. **Actions (A)**: Finite set of possible actions in each state.
3. **Transition Function (T)**: Probability of moving from one state to another given an action, defined as $T(s, a, s') = P(s'|s, a)$.
4. **Reward Function (R)**: Reward after transitioning from one state to another via action, represented as $R(s, a, s')$.
5. **Discount Factor (γ)**: Value between 0 and 1 that prioritizes immediate rewards over future rewards.

Summary and Conclusion - Importance of MDPs

Importance of MDPs in Reinforcement Learning

MDPs are crucial for structured decision making in uncertain environments. They enable agents to:

- **Optimize Decision-Making:** Compute optimal policies to maximize cumulative rewards.
- **Facilitate Learning:** Algorithms like Q-learning utilize MDPs to learn from interactions.
- **Handle Uncertainty:** Probabilities in T allow agents to adapt to variable environments.

Summary and Conclusion - Future Topics

Connection to Future Topics

Understanding MDPs is essential for advancing into complex reinforcement learning topics, including:

- **Policy Optimization:** Prepares for learning about policy gradient methods.
- **Exploration vs. Exploitation:** Strategies for balancing exploration and exploitation.
- **Partially Observable MDPs (POMDPs):** Understanding scenarios of incomplete information.

Key Takeaways

MDPs are vital for modeling decisions in uncertain environments, guiding agents towards optimal behavior.

$$V^{\pi}(s) = R(s, \pi(s)) + \gamma \sum T(s, \pi(s), s') V^{\pi}(s') \quad (8)$$