Markov Decision Processes
Chapter 17

Mausam

MDP vs. Decision Theory

- Decision theory - episodic
- MDP -- sequential

Markov Decision Process (MDP)

- $\mathcal{S}$: A set of states
- $\mathcal{A}$: A set of actions
- $\mathcal{P}r(s'|s,a)$: transition model
- $\mathcal{C}(s,a,s')$: cost model
- $\mathcal{G}$: set of goals
- s_0: start state
- γ: discount factor
- $\mathcal{R}(s,a,s')$: reward model

factored

Factored MDP

absorbing/
non-absorbing

Objective of an MDP

- Find a policy $\pi: \mathcal{S} \rightarrow \mathcal{A}$
- which optimizes
  - minimizes $\left\{\begin{matrix}\text{discounted}\\ \text{or}\\ \text{undiscount.}\end{matrix}\right\}$ expected cost to reach a goal
  - maximizes expected reward
  - maximizes expected (reward-cost)
- given a _____ horizon
  - finite
  - infinite
  - indefinite
- assuming full observability

Role of Discount Factor (γ)

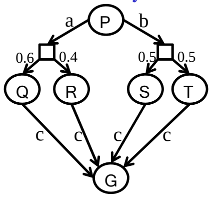
- Keep the total reward/total cost finite
  - useful for infinite horizon problems
- Intuition (economics):
  - Money today is worth more than money tomorrow.
- Total reward: $r_1 + \gamma r_2 + \gamma^2 r_3 + \ldots$
- Total cost: $c_1 + \gamma c_2 + \gamma^2 c_3 + \ldots$

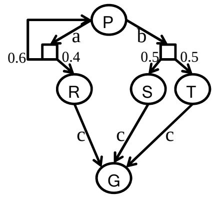
Examples of MDPs

- Goal-directed, Indefinite Horizon, Cost Minimization MDP
  - $<\mathcal{S}, \mathcal{A}, \mathcal{P}r, \mathcal{C}, \mathcal{G}, s_0>$
  - Most often studied in planning, graph theory communities
- Infinite Horizon, Discounted Reward Maximization MDP
  - $<\mathcal{S}, \mathcal{A}, \mathcal{P}r, \mathcal{R}, \gamma>$
  - Most often studied in machine learning, economics, operations research communities
- Oversubscription Planning: Non absorbing goals, Reward Max. MDP
  - $<\mathcal{S}, \mathcal{A}, \mathcal{P}r, \mathcal{G}, \mathcal{R}, s_0>$
  - Relatively recent model

most popular

AND/OR Acyclic Graphs vs. MDPs





C(a) = 5, C(b) = 10, C(c) =1

Expectimin works
- V(Q/R/S/T) = 1
- V(P) = 6 – action a

Expectimin doesn't work
- infinite loop
- V(R/S/T) = 1
- Q(P,b) = 11
- Q(P,a) = ????
- suppose I decide to take a in P
- Q(P,a) = 5+ 0.4*1 + 0.6Q(P,a)
- → = 13.5

Bellman Equations for MDP_1

- <V, D , Sr, F, J, s_0>
- Define J*(s) {optimal cost} as the minimum expected cost to reach a goal from this state.
- J* should satisfy the following equation:

$$J^*(s) = 0 \text{ if } s \in \mathcal{G}$$
$$J^*(s) = \quad]$$

Bellman Equations for MDP_2

- <V, D , Sr, U , s_0, γ>
- Define V*(s) {optimal value} as the maximum expected discounted reward from this state.
- V* should satisfy the following equation:

$$V^*(s) = \max_{a \in Ap(s)} \sum_{s' \in \mathcal{S}} \mathcal{P}r(s'|s,a)\left[\mathcal{R}(s,a,s') + \gamma V^*(s')\right]$$