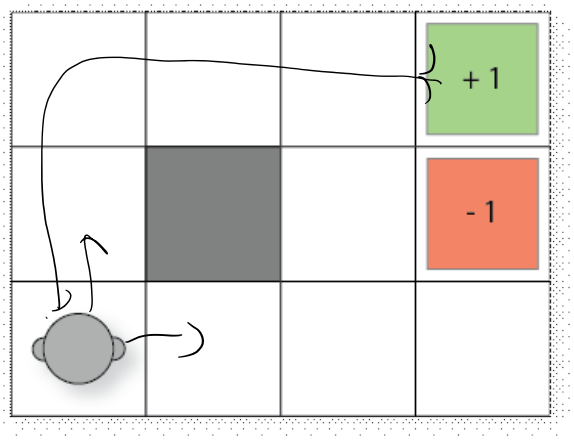
Markov Decision Processes

Tuesday, 4 September 2018 10:29 AM



The grey square is a wall.

The two labelled cells give a reward: 1 and -1 respectively.

But! Things can go wrong:
- If the agent tries to move north, 80% of the time, this works as planned (provided the wall is not in the way)
- 10% of the time, trying to move north takes the agent west (provided the wall is not in the way);
- 10% of the time, trying to move north takes the agent east (provided the wall is not in the way)
- If the wall is in the way of the cell that would have been taken, the agent stays put.
- Similar for all other directions

Classical Planning:
- Set of states S
- Initial state I
- Transition function A
- Goals G

MDPs:
- Set of states S
- Initial state I
- Transition probabilities:
  - P_a(s' | s)
- Reward function r(s, a, s') in real
- Discount factor (gamma) $\gamma \in [0, 1)$

$1 \rightarrow 0.9$

Discounted reward:

$V = r_1 + \gamma r_2 + \gamma^2 r_3 + \gamma^3 r_4 \dots$

$= r_1 + \gamma(r_2 + \gamma r_3 + \gamma^2 r_4 \dots)$

$V_t = r_1 + \gamma V_{t+1}$

$\gamma(r_2 + \gamma V_{t+2})$

Probabilistic PDDL:

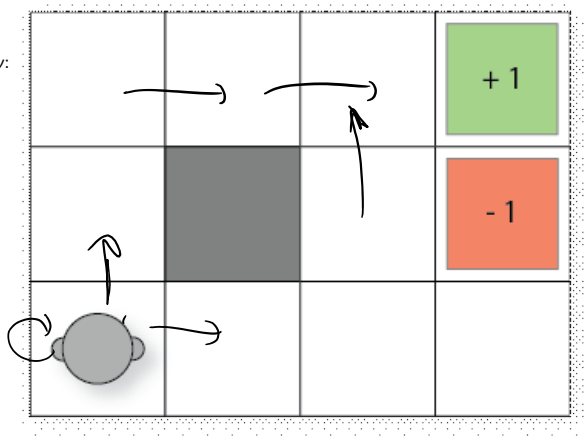
```
(define (domain bomb-and-toilet)
    (:requirements :conditional-effects :probabilistic-effects)
    (:predicates (bomb-in-package ?pkg) (toilet-clogged)
                 (bomb-defused))

    (:action dunk-package
        :parameters (?pkg)
        :effect (and (when (bomb-in-package ?pkg)
                           (bomb-defused))
                     (probabilistic 0.05 (toilet-clogged))))
```

Solution for MDP is a *policy*:

- at(0,0) => move_right
- at(0,1) => move_right
- at(0,2) => move_right
- at(0,3) => stay
- at(1,0) => move_up
- at(1,2) => move_up
- at(1,3) => move_up
- at(2,0) => move_up
- at(2,1) => move_left
- at(2,2) => move_up
- at(2,3) => move_left

$\rightarrow \pi(s) \rightarrow a$

policy



Expected return exercise:
You can steal:

A) An iPhone, which you think you have a 20% chance of selling for $500, or an 80% chance of selling for $250.
B) A Samsung, which you think you have a 50% chance of selling for $500, or a 50% chance of selling for $200.

Which do you steal? https://pollev.com/timothymille936

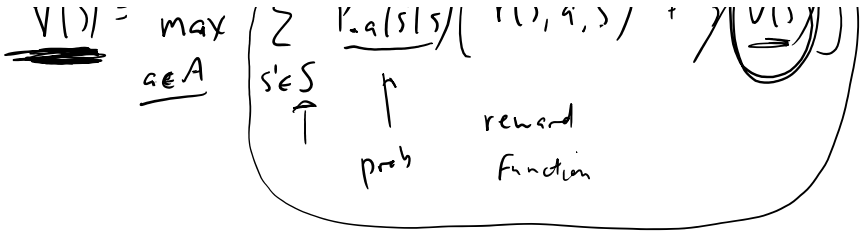
A: 0.2*500 + 0.8*250 = 300
B: 0.5*500 + 0.5*200 = 350

Expected return of a

reward → s'

Bellman equations:

$$V(s) = \max_{a \in A} \left(\sum_{s' \in S} P_a(s'|s) \left(r(s, a, s') + \gamma V(s') \right) \right)$$


reward
Function