Monday, November 5, 2018 2:56 PM

PARTIALLY-OBSERVABLE MDP (POMDP)

- MDP

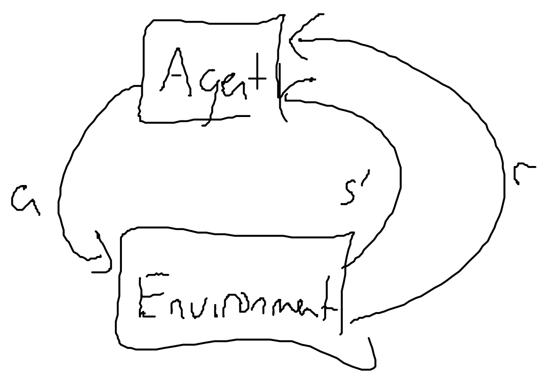
S: states

A: actions

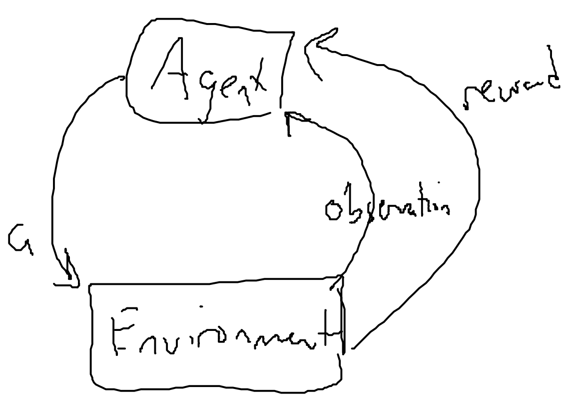
T: $T(s,a,s') \leftarrow P(s' | s, a) = P(S_{t+1} = s' | S_t = s,$

R: reward fn.

→ Find a policy π



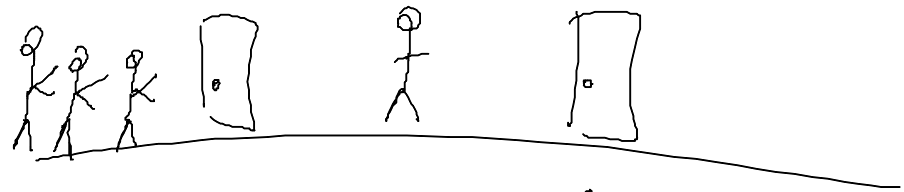
→ POMDP



Instead of getting state, you have a belief distribution o
possible states that you are in, and you get an observati

toward or against different states.

EXAMPLE: Actio



S_0 = Zombies-left S_1 = Zombies-Right

Observations: ZL : zombies-on-left

ZR : zombies-on-right

$P(o = ZL \mid S_0, do(listen)) = 0.85$ $P(o = ZL \mid S_1, do(listen))$

$P(o = ZR \mid S_0, do(listen)) = 0.15$ $P(o = ZR \mid S_1, do(listen)) =$

Reward: wrong: −100

right: +10

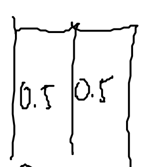
listen: −1

Belief state: Probability distribution over states

initial belief: $P(S_0) = 0.5$

$P(S_1) = 0.5$

$p(S_0)$ $p(S_1)$

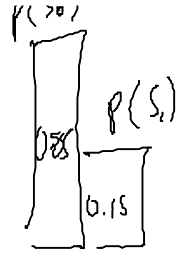


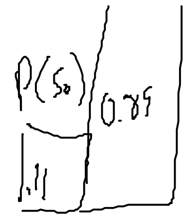
do: listen

o=ZL o=ZR

p( )

$p(S_1)$





belief state:

$$b = \langle P(S_0), P(S_1), P(S_2) \dots P(S_n) \rangle$$

$b(s) \leftarrow$ probability of being in state s

(we make a move)

belief Update: $$b'(s') = \alpha P(o \mid s') \sum_{s} \underbrace{P(s' \mid s, a)}_{\text{transition model}} b(s)$$

just received this observation

hidden variable

prior beli

prob of seeing an observation in my new state

Shorthand: $b' = \text{FORWARD}(b, a, o)$

Policy: $\pi(b) = a$

Missing: A transition function from getting from belief state to anoth

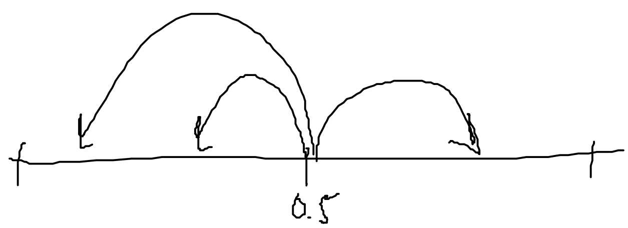
$$T(b, a, b') = P(b' \mid a, b)$$

Missing: A reward as a function of belief states

$\rho(b) \Rightarrow \mathbb{R}$

$$P(b'|a,b) = \sum_{o} \Big[\underbrace{P(b'|o,a,b)} \underbrace{\sum_{s'} \Big[P(o|s') \sum_{s} \Big[P(s'|s,a)\, b}_{\text{FORWARD model}}$$

$$P(b'|o,a,b) = \begin{cases} 1 & \text{if } b' = \text{FORWARD}(b, \\ 0 & \text{otherwise} \end{cases}$$



$$\rho(b) = \sum_{s} b(s) R(s)$$

(rho)

Problem: Continuous MDP (belief states are continuous random variables)

Either do a lot of math → intractible

Approximation