Last Time: Offline POMDP solutions

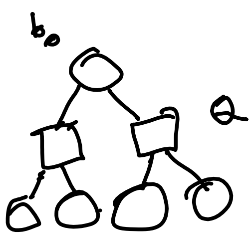
Point Based VI

How to choose B

Today: Online POMDP solutions

	Online Planning Algs for POMDPs	S	A	O	T	O
2007	AEMS	D	D	D	E	E
2010	MCTS/POMCP	C	D	D	G	G
2013	DESPOT	C	D	D	G	G
2018	POMCPOW	C	ltd C/D	C	G	E
2019	DESPOT-α	C	D	C	G	E

:(



belief

action

belief

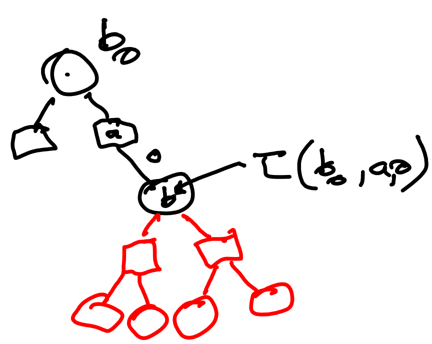
AEMS Anytime Error Minimizing Search

while time remains

$b^* = \underset{b \in Fringe(T)}{\arg\max} E(b)$

expand(b^*)

backup(b^*)



$$E(b) = \gamma^d P(b)\,\hat{\varepsilon}(b)$$

$$\hat{\varepsilon}(b) = U(b) - L(b)$$

$$P(b) = \prod_{i=0}^{d-1} \underline{P(o^i \mid b^i, a^i)}\,\underline{P(a^i \mid b^i)}$$

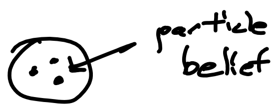
$$\underline{P(a \mid b)} = \frac{U(a,b) - L(b)}{U(b) - L(b)} \quad \text{AEMS1}$$

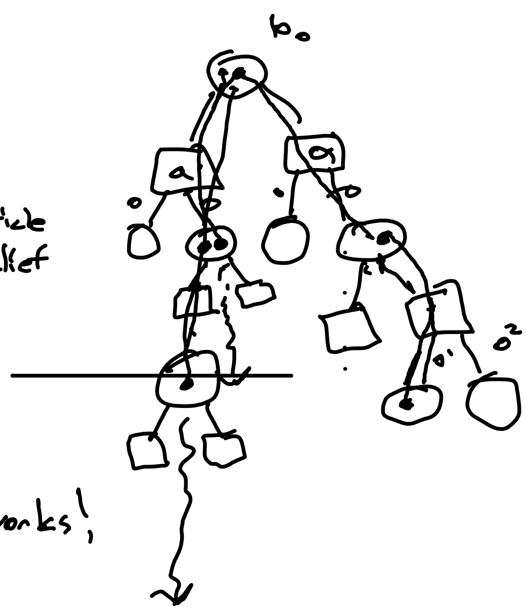
$$\underline{P(a \mid b)} = \begin{cases} 1 & \text{if } a = \arg\max U(b,a) \\ 0 & \text{otherwize} \end{cases} \quad \text{AEMS2}$$

Problem belief updates are expensive

MCTS for POMDPs

PO-UCT





Problem: we need a history-based simulator

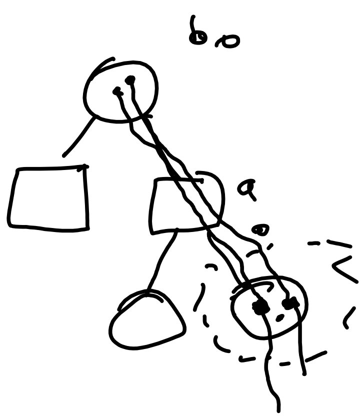
$o = G(h,a)$

$s', o, r = G(s,a,w)$

This just works!

POMCP

$\boxed{\text{PO-UCT}}$ + reusing particles for belief



root at next time step

DESPOT Determinized Sparse

K scenarios

each scenario is a fixed random seed

$s', o, r = G(s, a, \phi_{i,t})$

$G(s,a,w) = s + a + w$

$G(1,1,\phi_{1,1}) = 2 + \phi_{1,1}$

$G(1,2,\phi_{1,1}) = 3 + \phi_{1,1}$

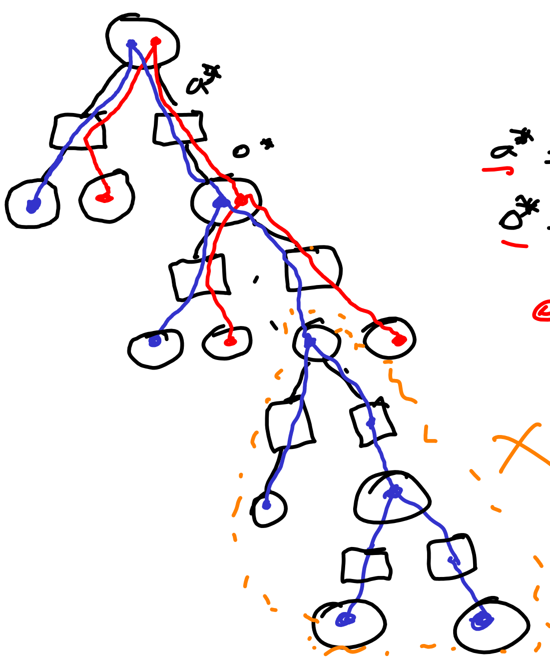
reduces variance

$G(s,a,w) = \begin{cases} s+a+w & \text{if } a=1 \\ s+a-w & \text{if } a=2 \end{cases}$

Not MCTS → Heuristic Search

K=500

D = max depth



U μ ← regularized weighted

L ℓ

$a^* = \text{argmax } \mu(b,a)$

$o^* = \text{argmax } E(\tau(b,a,o))$

expand

$\mu \approx U - \lambda |T|$ ← number of nodes in tree

Continuous O