Robot grocery shopping in partially observable settings

Rodrigo Gomes, Xiaomin Wang, Dustin Tran

May 13, 2015

MIT, 6.834j Cognitive Robotics

Outline

1. Background on POMDPs
2. Grocery shopping as planning in a POMDP
3. Demo!
4. What worked
5. What failed

Background

A *partially observable Markov decision process* (POMDP) is a collection of objects (S, A, Ω, R, T, O)

S: state space

A: action space

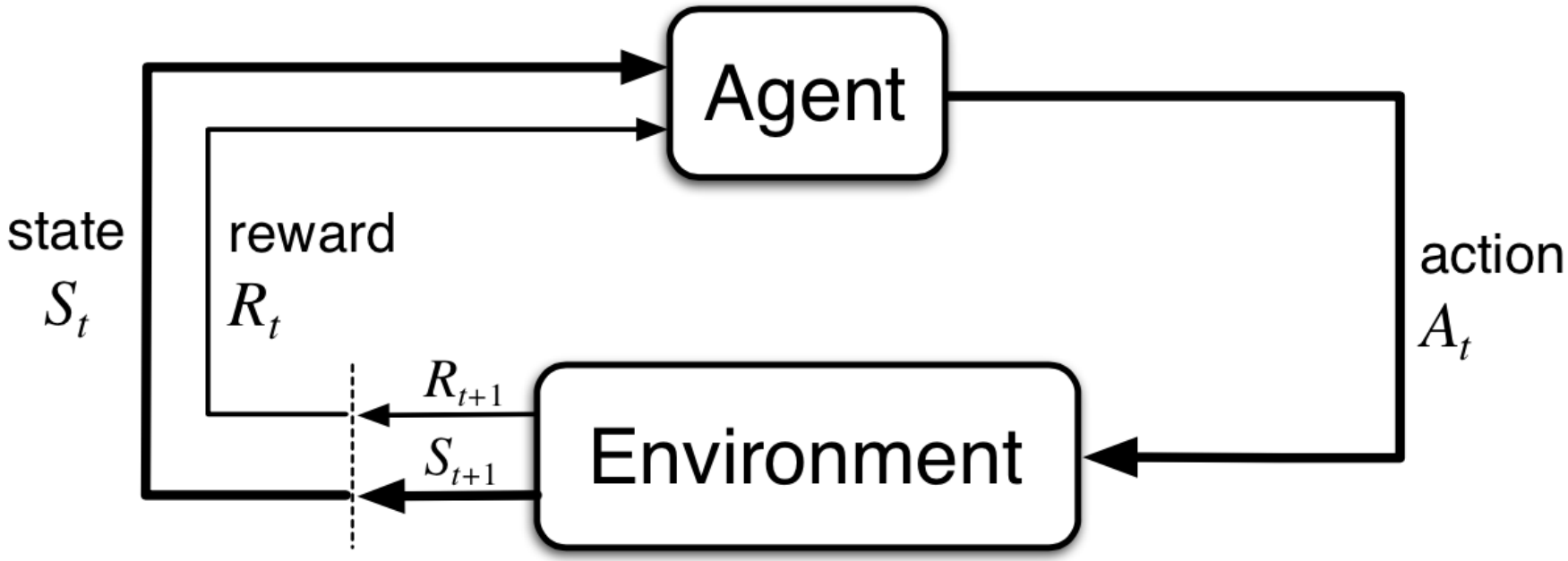
Ω: observation space

$R : S \times A \to \mathbb{R}$ reward function

T: transition operator. $T(s' \mid s, a)$ is probability of next state s' given state s and action a

O: observable operator. $O(o \mid s)$ is probability of observing o given at state s

Background



Belief-state MDP

Implemented MDP solvers:

- Q-learning
- SARSA
- R-MAX
- Thompson sampling

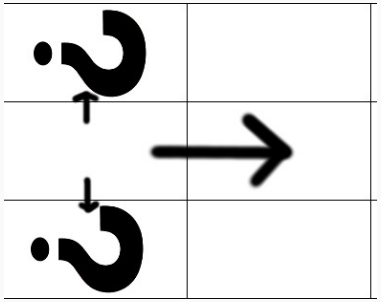
There are a lot!

- Function approximations with adaptive basis functions
- BOSS
- Spectral methods
- Skill chaining
- ...

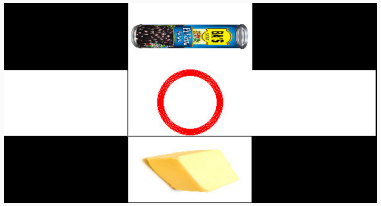
Grocery shopping

Grid World POMDP

Uncertain movement



Can only see around current cell (partially observable)



World is not fully known beforehand

- Model of how items in the same aisle correlate.
- Unknown arrangement of aisles
- Unknown arrangement of items within aisles

Grocery shopping

pygame running the visuals

Every second:

- Agent provides next action based on current belief state
- Simulator executes action (errors may happen)
- Belief state is updated based on transition probabilities
- Belief state is updated based on observation
- Belief about the world is updated based on belief state, and observation

Challenges:

- Markov assumption is not completely accurate
- Bias towards increasing probability of most likely states

DEMO

Our working solver

Max Probability Value Iteration:

Choose the most likely state from belief state to run value iteration

Our working solver

Value iteration:

$$
\begin{aligned}
v_{k+1}(s) &= \max_a \mathbb{E}[R_{t+1} + \gamma v_k(S_{t+1}) \mid S_t = s, A_t = a] \\
&= \max_a \sum_{s'} p(s' \mid s, a)[r(s, a, s') + \gamma v_k(s')]
\end{aligned}
$$

Failed tasks

Value iteration as a belief-state MDP

Thompson sampling

...

Play with it!

github.com/dustinvtran/bayesrl