POMDPs: Myths, Legends, and Reality

John Biechele-Speziale, Kushagra Kapoor, Jae Heo
May 24, 2023

PURDUE UNIVERSITY

Outline

- Part 1: *What are POMDPs?*
- Part 2: *POMDP Variants and Applications*
- Part 3: *Implementation Recommendations*

What are POMDPs?

POMDPs are MDPs with sensors instead of direct observation of the current state.

- The sensors are some probability distribution of some observation given a certain state.

POMDPs are MDPs with sensors instead of direct observation of the current state.

- The sensors are some probability distribution of some observation given a certain state.
- Unlike traditional MDPs that map from states to actions, these map from belief states(observations) to actions.

POMDPs are MDPs with sensors instead of direct observation of the current state.

- The sensors are some probability distribution of some observation given a certain state.
- Unlike traditional MDPs that map from states to actions, these map from belief states(observations) to actions.
- Exact optimal solutions yield the optimal action for each possible belief state that minimizes our cost.woodcockFormalMethodsPractice2009

POMDPs are defined by a tuple $(S, A, T, R, \Omega, O, \gamma)$

- S is the set of states

POMDPs are defined by a tuple $(S, A, T, R, \Omega, O, \gamma)$

- S is the set of states
- A is the set of actions

POMDPs are defined by a tuple $(S, A, T, R, \Omega, O, \gamma)$

- S is the set of states
- A is the set of actions
- T is the state transition probability

POMDPs are defined by a tuple $(S, A, T, R, \Omega, O, \gamma)$

- S is the set of states
- A is the set of actions
- T is the state transition probability
- R is the reward function

POMDPs are defined by a tuple $(S, A, T, R, \Omega, O, \gamma)$

- S is the set of states
- A is the set of actions
- T is the state transition probability
- R is the reward function
- Ω is the set of observations

POMDPs are defined by a tuple $(S, A, T, R, \Omega, O, \gamma)$

- S is the set of states
- A is the set of actions
- T is the state transition probability
- R is the reward function
- Ω is the set of observations
- O is the set of conditional observation probabilities

POMDPs are defined by a tuple $(S, A, T, R, \Omega, O, \gamma)$

- S is the set of states
- A is the set of actions
- T is the state transition probability
- R is the reward function
- Ω is the set of observations
- O is the set of conditional observation probabilities
- γ is the discount factor.

POMDPs rely on a belief update step, which itself can be an MDP.

- a

Value and policy function parameterizations

- a

The need for approximations

- a

Theory and undecidability

- a

POMDP Variants and Applications

Test

- a

Implementation Recommendations

My favorite package: POMDPs.jl

```
using Pkg; Pkg.add("POMDPs"); Pkg.add("QMDP");
using POMDPs, QuickPOMDPs, POMDPTools, QMDP

m = QuickPOMDP(
    states = ["left", "right"],
    actions = ["left", "right", "listen"],
    observations = ["left", "right"],
    initialstate = Uniform(["left", "right"]),
    discount = 0.95,
    transition = function (s, a)
        if a == "listen"
            return Deterministic(s) # tiger stays behind the sam
        else # a door is opened
            return Uniform(["left", "right"]) # reset
        end
    end,
    observation = function (s, a, sp)
        if a == "listen"
```

Alternative Packages

- Finite-state Controllers using Branch and Bound
- pomdp
- pyPOMDP
- zmdp

Temp 1

Temp 2

Temp 3

Temp 4

Temp 5

References

- K. L. Downing, "Reinforced Genetic Programming," *Genetic Programming and Evolvable Machines*, vol. 2, pp. 259–288, Sept. 2001.
- D. Hein, S. Udluft, and T. A. Runkler, "Interpretable policies for reinforcement learning by genetic programming," *Engineering Applications of Artificial Intelligence*, vol. 76, pp. 158–169, Nov. 2018.
- X. Huang, W. Ruan, Q. Tang, and X. Zhao, "Bridging Formal Methods and Machine Learning with Global Optimisation," in *Formal Methods and Software Engineering* (A. Riesco and M. Zhang, eds.), Lecture Notes in Computer Science, (Cham), pp. 1–19, Springer International Publishing, 2022.
- M. Kaufmann and J. S. Moore, "A Precise Description of the ACL2 Logic,"
- M. Krichen, A. Mihoub, M. Y. Alzahrani, W. Y. H. Adoni, and T. Nahhal, "Are Formal Methods Applicable To Machine Learning And Artificial Intelligence?," in *2022 2nd International Conference of*