Mobile robotics
Exploration

Janusz Jakubiak

Department of Cybernetics and Robotics

2021/2022

Wroclaw University of Science and Technology

HR EXCELLENCE IN RESEARCH

Copyright information

These slides contents base on a book Probabilistic Robotics (S. Thurn et al.)

Probabilistic planning and control

Motivation

Noise in information used for decision making causes uncertainty of action effects.
If we can estimate uncertainty of information – can we improve actions?

Uncertainty types considered

action effects :

- deterministic
- stochastic

perception :

- fully observable
- partially observable

Goal

Ensure robustness not only to current, but also predicted future uncertainty

Exploration

Exploration (information gathering task)

The direct goal of robot actions is to reduce uncertainty

Examples

- occupancy map building – maximize information about each cell
- find a person – determine location of a person moving in a building
- active localization – improve knowledge of own position

Methods

- Partially Observable Markov Decision Process (POMDP) – general, but complex
- specialized methods

Terms

goal – a achieving certain result (state) usually with cost optimization

cost – variable describing motion quality (precision, length, time etc.)

payoff function – function of state and control evaluating cost in a single step

$$r(x,u)$$

discount factor ($\gamma \in [0,1]$) time dependant coefficient

planning horizon T – main types: 1, finite, infinite

Terms cont.

cumulative payoff –

$$\sum_{\tau=1}^{T} \gamma^T r_{t+\tau}$$

policy – plan of action

$$\pi : z_{1:t-1}, u_{1:t-1} \rightarrow u_t(\mathrm{or} u_{t:t+r})$$

payoff expectation

$$R_T = E\left[\sum_{\tau=1}^{T} \gamma^T r_{t+\tau}\right]$$

optimal policy

$$\pi^\star = \mathrm{argmax}_\pi \ R_T$$

Markov Decision Process (MDP)

- for stochastic environment with fully observable state $\pi : x \rightarrow u$
- greedy optimization $T = 1$

$$\pi_1 = \mathrm{argmax}_u \ r(x,u)$$

with cummulative future payoff

$$V_1(x) = \gamma \max_u r(x,u)$$

- finite time T

$$\pi_T = \mathrm{argmax}_u \left[r(x,u) + \int V_{T-1}(x')p(x'|u,x)dx' \right]$$

with cummulative future payoff

$$V_T(x) = \gamma \max_u \left[r(x,u) + \int V_{T-1}(x')p(x'|u,x)dx' \right]$$

MDP Algorithm

MDP value iteration

for all x do
  $\hat{V}(x) = r_{min}$
endfor
repeat until convergence
  for all x

$$\hat{V}(x) = \gamma \max_u \left[r(x,u) + \int \hat{V}(x') p(x'|u,x) dx' \right]$$

  endfor
endrepeat
return $\hat{V}$

In discrete case: loop over all states

POMDP

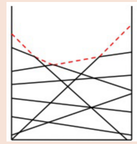
- state are not observable, so they are replaced by posteriors (beliefs, b)
- value function

$$V_T(x) = \gamma \max_u \left[r(b,u) + \int V_{T-1}(b') p(b'|u,b) db' \right]$$

- optimal policy

$$\pi_T = \operatorname{argmax}_u \left[r(b,u) + \int V_{T-1}(b') p(b'|u,b) db' \right]$$

- the key of effective implementation is pruning of not useful states



Approximate POMDP

Complexity of standard POMDP causes it is not practically applicable in robotic tasks

Approximate methods

- QMPD – an algorithm between MDP and POMDP; computes as if the full state was estimated after first iteration
- Augmented MDP (ADP) – belief represented in low dimensional statistics (for example state and entropy)
- Monte Carlo MDP (MC-MDP) – similar to particle filter

Exploration

Information gain

- Expected information $E[-\log p]$
- Entropy of a probability distribution $p(x)$

$$H_p(x) = -\int p(x) \log p(x) dx \quad \text{or} - \sum_x p(x) \log p(x)$$

- Belief $B(b,z,u)$
- Conditional entropy

$$H_b(x'|z,u) = -\int B(b,z,u)(x') \log B(b,z,u)(x') dx'$$

$$\text{or:} \quad -\sum_x p(x) \log p(x)$$

MC exploration algorithm

set $\rho_u = 0$ for all u
for i=1 to N do
  sample $x \sim b(x)$
  for all u do
    sample $x' \sim p(x'|x, u)$
    sample $z \sim p(z|x')$
    $b' = BayesFilter(b, z, u)$
    $\rho_u = \rho_u + r(x, u) - \alpha H_{b'}(x')$
  endfor
endfor
return $argmax_u \; \rho_u$