Robot Navigation: Lecture 9

Lecture Notes

Victor Risager

October 17, 2023

1 Path planning recap

Two domains

- Planning $\rightarrow$ low frequency
- Collision avoidance $\rightarrow$ high frequency

We often want to discretizise the continuous physical environment, into e.g. graphs.
It is fair to assume that optimality is dependent on completeness.

2 Planning under uncertainty

Here the cost is associated with a state. The problem with standard graph search algorithms.

- The robot is not trustworthy, there may be uncertainty about the robot actually performing the control input. probablity less than 1.
- What is the best action based on this.

When a robot is up against the wall, you would remove the probability of the robot going in that direction, and then redistribute it across the other directions by deviding it.

3 Markov decission process

The probability to reach the next state s' from state s choosing action a

$$T(s, a, s') \tag{1}$$

This is given in some sort of table, combining all the probabilities.
In each state the agent recieves a reward $R(s)$
Def: set of states

$$S \tag{2}$$

set of actions

$$A \tag{3}$$

Initial state

$$s_0 \tag{4}$$

Transition model

$$T(s, a, s') s.t. p(s_{t+1}|s_t, ..., s_1, a_t, ..., a_1) = p(s_{t+1}|s_t, a_t) = T(s_t, a_t, s_{t+1}) \tag{5}$$

We also define policies:
Mapping from state space to the set of actions

$$\pi : S \rightarrow A \tag{6}$$

$$a = \pi(s) \tag{7}$$

$$(\pi(a|s) = p(a_t = a|s_t = s)) \tag{8}$$

The number of possible policies is

$$card(A)^{card(S)} \tag{9}$$

e.g. 3^{12} different policies $= 500.000 - 1.000.000$
If we define some policy $a = \pi(s)$ Then conditional probablity can be reduced to $p(s_{t+1}|s_t)$ Thus $p(s_{t+1}|s_t, ..., s_1)$
Thus a markov decision process plus a policy, is then reduced to a markov process.
because the expected value of all the rewards towards the goal is ∞, a discounting factor $0 \leq \gamma \leq 1$
As long as it is bounded, it will always converge.
This says that the present reward is greater than future rewards.
It is done in state space, and the policy is not time dependent. Thus we have the same policy for all time.
The lower the gamma, the faster the convergence.

3.1 Partially Observable Decision Process

nasty, not solveable.

3.2 Vector Field Histogram

Used in obstacle avoidance, and find the bin with the lowest probability, in order and move in that direction. This does however not consider the kinematic and dynamic limitations. Therefore we can use masked polar histogram.
Or use the dynamic window approach.