Partially observable Markov decision process

From Wikipedia, the free encyclopedia

A **Partially Observable Markov Decision Process** (POMDP) is an extension of a Markov Decision Process. POMDPs are used for choosing actions when the entire world, or state space, is not always directly observable. In other words, you cannot always immediately know where you are in the world and what is going on. For instance, chess is a directly observable game: you can always see the positions of all the pieces. Poker, on the other hand, is partially observable: you can narrow down the set of cards your opponents hold, but cannot directly observe them.

POMDPs can be used in solving simple path planning problems for mobile robots. This application is sometimes called the "Kidnapped Robot Problem," framed by imagining a robot was moved to an unknown location in a known environment and now must figure out where it is and find its way home. An exact solution to the POMDP will generate the series of actions that is most likely to get it home with the least cost.

Contents

Definition

A Partially Observable Markov Decision Process (POMDP) is a tuple (*S*,*A*,*O*,*P*,*R*), where

- *S* is the State space,
- *A* is the action space,
- *O* is the observation space.
- $P(s,a,o,s') = \Pr(s_{t+1} = s',\ o_t = o | s_t = s,\ a_t = a)$ is the probability that action *a* in state *s* at time *t* will give observation *o* and lead to state *s*' at time *t* + 1,
- *R*(*s*,*a*,*s*') is the immediate reward that is issued for the transition from *s* to *s*' under action *a*.

The goal is to maximize some cumulative function of the rewards, typically the discounted sum under a discounting factor γ (usually just under 1).

Markov decision processes are a special case of POMDPs in which the observations always uniquely identify the true state (i.e., the states are directly observed or can be directly deduced from the observation). Since the true state of the world can't be uniquely identified, a POMDP reasoner must maintain a probability distribution, called the belief state, which describes the probabilities for each true state of the world. Maintenance of the belief states is Markovian, in that it only requires knowledge of the previous belief state, the action taken, and the observation seen.

Approximate POMDP solutions

Currently, most POMDPs are computationally intractable to solve exactly for optimal behavior, so computer scientists have developed methods that approximate solutions for POMDPs.

Grid-based algorithms (Lovejoy, 1991) comprise one approximate solution technique. In this approach, the value function is computed for a set of points in the belief space, and interpolation is used to determine the optimal action to take for other belief states that are encountered and that aren't in the set of grid points.

POMDP uses

POMDPs have been used for a number of real-world problems. Notable work includes the use of a POMDP in assistive technology for persons with dementia (Hoey et al, 2007)

References

Lovejoy, W. (1991). Computationally feasible bounds for partially observed Markov decision processes. Operations Research, 39, 162--175.

Hoey, J., von Bertoldi, A., Poupart, P., and Mihailidis, A., Assisting Persons with Dementia during Handwashing Using a Partially Observable Markov Decision Process. In Proceedings of the International Conference on Vision Systems, 2007. Full text available here (http://biecoll.ub.uni-bielefeld.de/volltexte/2007/12/)

External links

- Tony Cassandra's POMDP pages (http://www.cassandra.org/pomdp/index.shtml) with a tutorial, examples of problems modeled as POMDPs, and software for solving them.
- zmdp (http://www.cs.cmu.edu/~trey/zmdp/), a POMDP solver by Trey Smith