Towards Large-Scale POMDP Planning for Robotic Tasks

David Hsu
National University of Singapore

Partially observable Markov decision process (POMDP)

- A discrete POMDP model
- States (configurations)
- Actions
- Observations
- Rewards
- State transition function
- Observation function
- A belief state is a probability distribution over the states.
- A policy is a mapping from a belief to an action. An optimal policy maximizes the expected total reward.

Some history: 1978


Complexity of solving POMDPs

- Solving POMDPs exactly is computationally intractable:
  - Finite-horizon POMDPs are PSPACE-complete [Papadimitriou & Tsitsiklis, 87].
  - Infinite-horizon POMDPs are undecidable [Madsen et al., 99].
“Curse of dimensionality”

Exponential in the dimensionality of the state space

Sampling the belief space

Belief space

Sampling the state space

Factoring

Point-based POMDP algorithms

<table>
<thead>
<tr>
<th>Algorithm</th>
<th>Time (sec.)</th>
<th>Reward</th>
</tr>
</thead>
<tbody>
<tr>
<td>PBVI 2003</td>
<td>180.880</td>
<td>-9.18</td>
</tr>
<tr>
<td>HSVI 2003</td>
<td>10,113</td>
<td>-6.17</td>
</tr>
<tr>
<td>Perseus 2005</td>
<td>1,670</td>
<td>-6.17</td>
</tr>
<tr>
<td>HSVI 2005</td>
<td>24</td>
<td>-6.36</td>
</tr>
<tr>
<td>SARSOP 2008</td>
<td>6</td>
<td>-6.13</td>
</tr>
</tbody>
</table>
Mixed Observability Markov Decision Process (MOMDP)

S.C.W. Ong, S.W. Png, D. Hsu, and W.S. Lee. POMDPs for robotic tasks with mixed observability. Int. J. Robotics Research, 29(8), 2010.

Mixed observability Markov decision process (MOMDP)
- state $s = (x, y)$
- POMDP belief $b(s)$
- MOMDP belief $(x, b(y))$

Factoring the belief space
- state $s = (x, y)$
- $|X| = 2$, $|Y| = 2$

Mixed observability
- state $s = (x, y)$
- number of states $50 \times 50 = 2,500$
- belief $b(s)$
- 2,500-dimensional space
- belief $(x, b(y))$
- union of 50-dimensional subspaces
- Potential efficiency gain

Computational efficiency
- POMDP operates in a single $|S|$-dimensional space.
- MOMDP operates in a union of $|X|$ sets of $|Y|$-dimensional spaces, where $|S| = |X| \cdot |Y|$.
- Computational efficiency gain from MOMDP representation
- Point-based approximation algorithms $\propto |X|$
MOMDP
5 sec
19 sec
-6.03
-9.90
SARSOP
17 sec
736 sec
-6.03
-9.90
(Hsu et. al. 08)

Tag
(Pineau, Gordon & Thrun 2003)

AUV navigation in 3D

AUV Navigation (140 hpos x 4 depth x 24 orien)
|S| = 13,536
[X] = 96, [Y] = 141
MOMDP
124 sec
1020
SARSOP
409 sec
1020

Reparameterized full observability

• Reparameterize the state space
  • \( x = o \)
  • \( h(o) \): the set of states that have non-zero probability of emitting \( o \)
  • \( y \) - offset from \( o \), indicating the exact state in \( h(o) \)

Theorem
The POMDP and the reparameterized MOMDP (with \( x = o \)) are equivalent.

Noisy Tag
Reparameterized full observability

Continuous-state POMDPs

|S| = 3,080
[X] = 56, [Y] = 495
MOMDP
32 sec
1020
SARSOP
937 sec
1020
29.0x

Sampling the state space

Continuous-state POMDPs

- Belief over continuous state space

Continuous-state POMDPs

- Policy graph (finite-state controller)

Value iteration on a policy graph

\[ V_{i+1}(b) = \max_{a \in A} \left\{ R(b, a) + \gamma \sum_{b' \in \mathcal{B}} p(b', a) V_i(b') \right\} \]

Monte Carlo backup

\[ V_{i+1}(b) = \max_{a \in A} \left\{ R(b, a) + \gamma \sum_{b' \in \mathcal{B}} p(b', a) V_i(b') \right\} \]

Monte Carlo Value Iteration (MCVI)

Sample a new belief

MC-backup at selected beliefs

Repeat until convergence
**Computational efficiency**

**Theorem**

\[ |V^*(b) - V_b(b)| \leq \sqrt{n \ln |A| + \ln(1/\epsilon)} \frac{1}{N} + \delta_b + \gamma^b \]

with probability at least \( 1 - \tau \).

- \( N \): number of samples from the state space (Monte Carlo simulations)
- \( \delta_b \): covering of the belief space
- \( t \): number of iterations

---

**Grasping**

(Hsiao, Kaelbling & Lozano-Perez 2007)

- Uncertain initial position
- Noisy touch sensors on fingers
- Manual discretization

---

**Simulation results**

<table>
<thead>
<tr>
<th>Model</th>
<th>Algorithm</th>
<th>Risk ratio</th>
</tr>
</thead>
<tbody>
<tr>
<td>3D continuous-state POMDP</td>
<td>MCVI</td>
<td>0.00066</td>
</tr>
<tr>
<td>2D continuous-state POMDP</td>
<td>MCVI</td>
<td>0.017</td>
</tr>
<tr>
<td>2D discrete POMDP (Sanner et al. 2010)</td>
<td>SARSOP</td>
<td>0.035</td>
</tr>
<tr>
<td>TCAS</td>
<td>TCAS</td>
<td>0.061</td>
</tr>
<tr>
<td>Nominal</td>
<td></td>
<td>1.0</td>
</tr>
</tbody>
</table>

- MIT Lincoln Laboratory CASSATT simulator
- 15,000 encounters from 9 months of radar data in US airspace

---

**What makes some POMDPs easier than others?**

Common complexity measures

- Number of states = dimensionality of belief space

\[ (0,0) \quad \vdash \quad (1,1) \quad \vdash \quad (1,0) \]

An alternative complexity measure

- The covering number \( C(\delta) \) of a set \( S \) is the number of balls of radius \( \delta \) needed to cover \( S \) completely.

Reachable space

- Find an approximately optimal action at \( b_0 \) (on-line action selection)

**Theorem**

An approximately optimal value \( V(b_0) \) with regret no more than \( \varepsilon \) can be found in time

\[ O \left( C \left( \frac{(1 - \gamma)^2 \varepsilon}{4\gamma R_{\text{max}}} \right)^2 \log \gamma \left( \frac{1 - \gamma}{2R_{\text{max}}} \right) \right) \]

- The problem is easy if the covering number ("volume") of the reachable space is small.

Optimally reachable space

**Theorem**

Finding the optimal action is NP-hard, even if the optimally reachable space has a polynomial-size cover.

**Theorem**

Given a suitable cover \( C \) of the optimally reachable space, an approximately optimal value \( V(b_0) \) with regret no more than \( \varepsilon \) can be found in time

\[ O \left( |C|^2 + |C| \log \gamma \left( \frac{1 - \gamma}{2R_{\text{max}}} \right)^2 \right) \]

Implications

- Together, the positive and negative results indicate that finding a suitable cover of the optimally reachable space is a key difficulty.
- The covering number better characterizes the complexity of the problem by capturing the sparsity of the space.

Bounding the covering number

- Several properties, often encountered in practice, reduce the size of covering numbers.
  - Fully observable state variables
    \[ \left( \frac{k^d}{\delta} \right)^k \quad \rightarrow \quad k^d \left( \frac{k^d - d'}{\delta} \right)^k^{d-d'} \]
  - Sparse beliefs
  - Smooth beliefs
  - Circulant state-transition matrices
  - ...

...
Summary

- Large belief space: MOMDP
- Large state space: MCVI
- Covering number

POMDP software

Approximate POMDP Planning (APPL) Toolkit


Future Work

Acknowledgments

- Wee Sun Lee, NUS
- Hanna Kurniawati, SMART
- Sylvie Ong, McGill
- Haoyu Bai, NUS
- Yanzhu Du, Stanford
- Shaowei Png, McGill
- Nan Rong, Cornell

- MDA Gambit Game Lab
- Ministry of Education, Singapore
- School of Computing, National University of Singapore