Practical Encodings of Factored Deterministic POMDPs into Probabilistic Planning

Patrik Haslum\textsuperscript{1}, Abdallah Saffidine\textsuperscript{2}

\textsuperscript{1}Australian National University, Canberra, Australia
\textsuperscript{2}The University of New South Wales, Sydney, Australia

May 2, 2019
A story about Chinese Dark Chess
A story about Chinese Dark Chess

Representation choices matter!
A Markov Decision Process
## Some single-agent domain models

<table>
<thead>
<tr>
<th>Control over state transitions</th>
<th>Observability</th>
</tr>
</thead>
<tbody>
<tr>
<td>No</td>
<td>Full: Markov Chain</td>
</tr>
<tr>
<td>Yes</td>
<td>Partial: Hidden Markov Model</td>
</tr>
<tr>
<td></td>
<td>Partial: Partially Observable MDP</td>
</tr>
</tbody>
</table>
Some single-agent domain models

<table>
<thead>
<tr>
<th>Control over state transitions</th>
<th>Observability</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Full</td>
</tr>
<tr>
<td>No</td>
<td>Markov Chain</td>
</tr>
<tr>
<td>Yes</td>
<td>Markov Decision Process</td>
</tr>
<tr>
<td></td>
<td>Partial</td>
</tr>
<tr>
<td></td>
<td>Hidden Markov Model</td>
</tr>
<tr>
<td></td>
<td>Partially Observable MDP</td>
</tr>
</tbody>
</table>

Today’s concern
- deterministic POMDPs
- factored representations
A deterministic POMDP
Reducing POMDPs to MDPs

- Transform a POMDP \( P \) into an MDP \( M_P \)
- With equivalent optimal values
Reducing POMDPs to MDPs

- Transform a POMDP $P$ into an MDP $M_P$
- With equivalent optimal values
- Mapping policies for $M_P$ to policies for $P$. 
Reducing POMDPs to MDPs

- Transform a POMDP $P$ into an MDP $M_P$
- With equivalent optimal values
- Mapping policies for $M_P$ to policies for $P$.
- → reuse mature MDP technology
- → provide complexity bounds
Littman’s encoding (1996)

- Transform a det-POMDP $P$ into an MDP $L_P$
- Each state of $L_P$ is a table with one entry per state of $P$ describing the evolution of that state.
Littman’s encoding (1996)

- Transform a det-POMDP $P$ into an MDP $L_P$
- Each state of $L_P$ is a table with one entry per state of $P$ describing the evolution of that state.
- $\Sigma$ states for $P \rightarrow O((1 + \Sigma)^{\Sigma})$ states for $L_P$
Littman’s encoding

- Transform a det-POMDP $P$ into an MDP $L_P$
- Each state of $L_P$ is a table with one entry per state of $P$ describing the evolution of that state.
- $\Sigma$ states for $P \rightarrow \mathcal{O}((1 + \Sigma)^\Sigma)$ states for $L_P$
Littman’s encoding

- Transform a det-POMDP $P$ into an MDP $L_P$.
- Each state of $L_P$ is a table with one entry per state of $P$ describing the evolution of that state.
- $\Sigma$ states for $P \rightarrow \mathcal{O}((1 + \Sigma)^{\Sigma})$ states for $L_P$.

How satisfactory is that?
History-based encoding (new(?))

- Transform a det-POMDP $P$ into an MDP $H_P$
- Each state of $H_P$ is a table encoding the history of action/observation performed in $P$. 
History-based encoding

select action $a_k$ and append to history

sample hidden state $s$

for each $a_1, \ldots, a_k$ in action history

update $s$ with effects of $a_i$

no more actions

$s$ not consistent with observation before $a_i$

next action

goal true in $s$

Success!

record observation determined from $s$
a3,0

a3,1

b1,1

b1, 😐

b2

1

1
History-based encoding

- Transform a det-POMDP $P$ into an MDP $H_P$
- Each state of $H_P$ is a table encoding the history of action/observation performed in $P$.
- $A$ actions-observations for $P$, horizon $H \rightarrow \mathcal{O}((A)^{H+c})$ states for $H_P$
### Size considerations

<table>
<thead>
<tr>
<th>Repr.</th>
<th>Encoding</th>
<th>Number of States</th>
</tr>
</thead>
<tbody>
<tr>
<td>Flat</td>
<td>POMDP</td>
<td>$\Sigma$</td>
</tr>
<tr>
<td>Flat</td>
<td>MDP: Littman</td>
<td>$(1 + \Sigma)^\Sigma$</td>
</tr>
<tr>
<td>Flat</td>
<td>MDP: History</td>
<td>$A^{H+c}$</td>
</tr>
<tr>
<td>Factored</td>
<td>POMDP</td>
<td>$2^V$</td>
</tr>
<tr>
<td>Factored</td>
<td>MDP: Littman</td>
<td>$(1 + 2^V)^{2^V}$</td>
</tr>
<tr>
<td>Factored</td>
<td>MDP: History</td>
<td>$A^{H+c}$</td>
</tr>
</tbody>
</table>

$A$ actions and observations, $V$ state variables, $c$ small constant
$\Sigma$ states, $H$ horizon
Det-POMDPs with polynomial depth

Sanity check: we reprove PSPACE membership.
Conclusion

Discussion point

- Factored vs explicit representations
Conclusion

Discussion point
- Factored vs explicit representations

Limitations → won’t fix
- Deterministic effects
- No adversary
Conclusion

Discussion point
- Factored vs explicit representations

Limitations → won’t fix
- Deterministic effects
- No adversary

Future Work
- Extend to discounted rewards
- Experiments in Troubleshooting domains
- Check the literature on the complexity of factored POMDPs