TLdR: Policy Summarization for Factored SSP Problems Using Temporal Abstractions

Sarath Sreedharan ▪ Siddharth Srivastava ▪ Subbarao Kambhampati

Collaborative Agents

• Ability to effectively collaborate with people is a prerequisite for real world agents
• We would expect such agents to be capable of effectively communicating their plans with users
• Particularly challenging in stochastic setting
  • Need to communicate policies

Existing Approaches

• Methods to simplify policy structure
• Or

  • Summarize:
    • Presenting state abstractions of the policy
    • Present actions from representative states

Our Method

• Summary in terms of the subgoals to be achieved as part of the policy

Challenge: Extract subgoals automatically for a given policy

Policy Landmark

• We focus on goal based MDP formulations with fixed initial states
• Subgoals identified by extracting landmarks
• Where landmarks in stochastic settings defined in terms of execution sequence that leads to the goal rather than in terms of plans
• For policy landmark we only consider the execution sequences feasible under the given policy

“Facts/formulae and their relative ordering that needs to be satisfied by every feasible trace possible under the policy that leads to the goal”

Example: Travel planning domain

• Consider a sample domain for generating a travel plan
• Domain consists of 30 grounded fluents and 27 grounded actions
• Partial policy consists of 100 states

Empirical Evaluation

User Studies

Hypothesis: Would people choose landmark facts as summary

• Users were presented with a policy graph with state details and asked to summarize based on subgoals
• 60 participants in total with 30 per scenario
• We made sure the participants understood the domain through a filtering question
• Every participant was shown 13 options and five to six facts out of which were landmarks (76.2%)
• We filtered 41 participant with 164 selections out of which 125 were landmarks (76.2%)

Evaluation

<table>
<thead>
<tr>
<th>Domain</th>
<th>No of reachable states</th>
<th>No of non-trivial policy landmarks</th>
<th>No of non-trivial model landmarks</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ex-Blocksworld</td>
<td>104.8</td>
<td>8.8</td>
<td>5.2</td>
</tr>
<tr>
<td>Elevator</td>
<td>13.2</td>
<td>7</td>
<td>4</td>
</tr>
<tr>
<td>Tireworld</td>
<td>13.6</td>
<td>1.8</td>
<td>0.8</td>
</tr>
<tr>
<td>Tri-Tireworld</td>
<td>3973</td>
<td>6.25</td>
<td>0</td>
</tr>
</tbody>
</table>

Acknowledgement: This research is supported in part by ONR grants N00014-16-1-2892, N00014-18-1-2442, N00014-18-1-2840, N00014-9-1-2119, AFOSR grant FA9550-18-1-0067, DARPA SAIL-ON grant W911NF-19-2-0006, NSF grants 1936997 (C-ACCEL), 1844325 and 1909370, NASA grant NNX17AD06G, and a JP Morgan AI Faculty Research grant.