Point Based Value Iteration with Optimal Belief Compression for Dec-POMDPs

Liam MacDermed
Charles L. Isbell

Problem: An agents' belief set has doubly exponential growth with the horizon due to nested beliefs.

Key Idea: Allow agent's to decide how to best compress their own beliefs and fold this belief choice into the model thereby allowing our model solver to compute an optional compression scheme.

Formal Solution: Given a Dec-POMDP \((N,A,S,O,P,R,\pi)\) (with states \(s\in S\) and observations \(o\in O\)) we define the BB-DecPOMDP approximation model \((N',A',S',O',P',R',\pi')\) (which optimally compresses beliefs) with belief set size parameters \(t_1, \ldots, t_n\) as:

- \(N' = N\) and \(A_i = \{o_{1i}, \ldots, o_{ni}\}\)
- \(A'\) is the set of actions for the \(i\)th agent
- \(O' = \{\omega\} \times \{1,2,\ldots, t_i\}\) with factored observation \(o = (\omega, b)\in O'\)
- \(S' = S \times O'\) with factored state \(s = (s_1, \omega, b_1, \ldots, b_n)\in S'\).

Algorithm:

1. **Belief Expansion (from original model)**
   - For each state \((s_1, \omega, b_1, \ldots, b_n)\) in \(S'\), compute all possible belief states \((s_2, \omega, b_2, \ldots, b_n)\) by applying actions \(a_i\) to \(b_i\).

2. **Belief Contraction (new phase)**
   - For each state \((s_1, \omega, b_1, \ldots, b_n)\) in \(S'\), select the best belief state \((s_2, \omega, b_2, \ldots, b_n)\) by evaluating the value function \(V(s_1, \omega, b_1, \ldots, b_n)\)

**Approach:** Solve a DecPOMDP by transforming it into a POMDP.

**Problem:** How can we centralize the model in order to make the problem tractable?

**Key Ideas:**
- Factor the state so that it contains each agent's current belief.
- We can convert a DecPOMDP into a POMDP with exponential actions corresponding to strategies of a CBG by utilizing the fact that the common knowledge distribution over joint-beliefs is a sufficient statistic for planning (previously proved).

**Formal Solution:**
- Belief POMDP: \((A', S', O', P', R', \pi')\) (with belief labels \(b_i\) for each agent)
- Factored states: \(S' = S \times \prod_{i=1}^{n} \{0, 1\}^{t_i}\) (where \(s_i\) is the underlying state and \(b_i\) is the agent's belief).
- Actions are strategies: \(A' = \prod_{i=1}^{n} A_i\)
- Transition function: \(P'(s'|s, a) = \sum_{b \in B(s,a)} P'(s'|s, a, b)\) (where \(b\) is the action agent \(i\) would take in belief \(b_i\)).
- Reward function: \(R'(s'|s, a) = R(s, b_1, \ldots, b_n)\)
- Initial state: \(s(0)\)

**Problem:** How do we solve a Cooperative Bayesian Game (CBG) efficiently given that it's NP complete?

**Key Idea:** Solve for a deterministic agent normal form correlated equilibrium. This is an ILP. The linear relaxation is very often integral.

**Formal Solution:**
- Formulate the PBVI problem as an ILP over joint strategies.
- Solve for a deterministic agent normal form correlated equilibrium. This is an ILP.

**Results on Benchmark Problems**

<table>
<thead>
<tr>
<th>Dec-POMDP</th>
<th>1-Beliefs</th>
<th>2-Beliefs</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dec Tiger</td>
<td>13,446</td>
<td>31,905</td>
</tr>
<tr>
<td>Broadcast</td>
<td>9,270</td>
<td>31,905</td>
</tr>
<tr>
<td>Recycling</td>
<td>9,270</td>
<td>31,905</td>
</tr>
<tr>
<td>Gridworld</td>
<td>5,928</td>
<td>5,928</td>
</tr>
<tr>
<td>Box Pusher</td>
<td>223,693</td>
<td></td>
</tr>
</tbody>
</table>

Utility achieved by our PBVI-PPC-DecPOMDP algorithm compared to the previously best known policies on a series of standard benchmarks. Higher is better. Our algorithm beats all previous results except on Dec-Tiger where we believe an optimal policy has already been found.