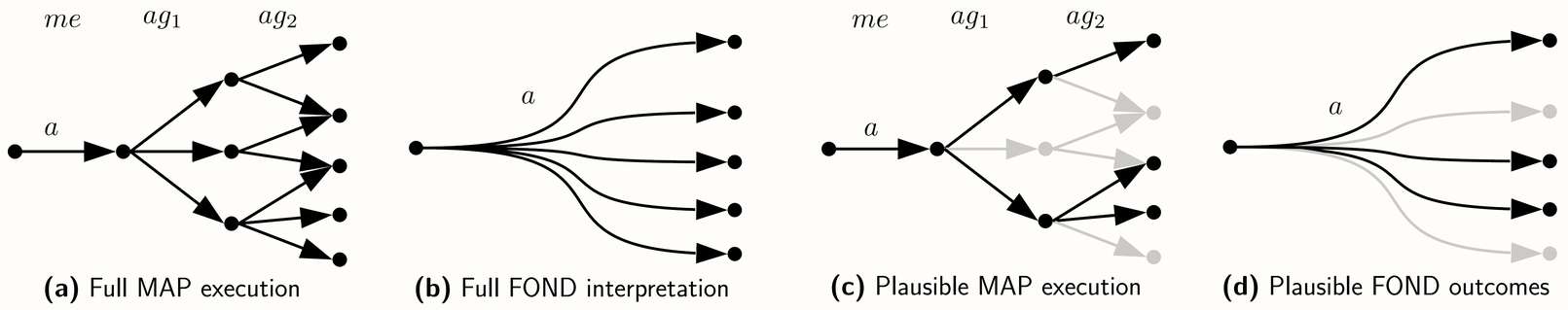
Leveraging FOND Planning Technology to Solve Multi-Agent Planning Problems

Christian Muise, Paolo Felli, Tim Miller, Adrian R. Pearce, Liz Sonenberg

Department of Computing and Information Systems, University of Melbourne
{christian.muise,paolo.felli,tmiller,adrianrp,l.sonenberg}@unimelb.edu.au

Contribution

We present a novel application of FOND planning in multi-agent environments based on the intuition that actions taken by others in the world can be viewed as non-deterministic outcomes of our own choices. We further improve the reasoning in settings where we can assume a model of the other agents' goal.



(a) Full MAP execution **(b)** Full FOND interpretation **(c)** Plausible MAP execution **(d)** Plausible FOND outcomes

Example execution for agents $\mathtt{ag}_1$ and $\mathtt{ag}_2$ after $\mathtt{me}$ executes action $\mathtt{a}$. Subfigures (c) and (d) show plausible outcomes.

General Approach

1. Directly modify a state-of-the-art FOND planner
2. Replace the enumeration of non-deterministic outcomes with the possible responses of other agents
3. Restrict the space of all possible responses by focusing on what is *plausible* given the goals of other agents
   - Optional and can be substituted for any notion of *plausibility*

(F)ully (O)bservable (N)on-(D)eterministic Planning

FOND Planning Problem $\Pi = \langle \mathcal{F}, \mathcal{G}, \mathcal{I}, \mathcal{A} \rangle$:

- $\mathcal{F}$: Set of fluents
- $\mathcal{G}$: Goal condition (subset of $\mathcal{F}$)
- $\mathcal{I}$: Initial state of the world
- $\mathcal{A}$: Set of non-deterministic actions. Every action $\mathtt{a} \in \mathcal{A}$ has a precondition $\mathrm{Pre}_\mathtt{a}$ and a set of effects $\mathrm{Eff}_\mathtt{a}$. Exactly one effect will occur during execution.

Solution: Policy P that maps the state of the world to the action the agent should execute.

First-Person MAP

First-Person MAP Problem $\langle \overrightarrow{\mathtt{ag}}, \mathit{App}, \Pi \rangle$:

- $\overrightarrow{\mathtt{ag}}$: Sequence of agents to act in the world
- $\mathtt{i} \in \overrightarrow{\mathtt{ag}}$, $\mathit{App}(\mathtt{i})$: Set of actions agent $\mathtt{i}$ can execute. $\mathit{App}(\mathtt{s}, \mathtt{i}) \subseteq \mathit{App}(\mathtt{i})$ are those also applicable in state $\mathtt{s}$

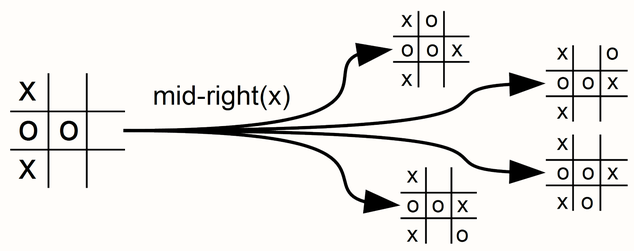
We control agent $\mathtt{me} \in \overrightarrow{\mathtt{ag}}$, but must react to anything the other agents may do when it is their turn to "play"

Solution: Same as FOND!!

FOND Planning Algorithm

Input: FOND planning task $\Pi = \langle \mathcal{F}, \mathcal{G}, \mathcal{I}, \mathcal{A} \rangle$
Output: Partial policy P
Initialize policy P
while P *changes* **do**
  $\mathrm{Open} = \{\mathcal{I}\}$; $\mathrm{Seen} = \{\}$;
  while $\mathrm{Open} \neq \emptyset$ **do**
    $\mathtt{s} = \mathrm{Open.pop}()$;
    if $\textsc{Satisfies}(\mathtt{s}, \mathcal{G}) \wedge \mathtt{s} \notin \mathrm{Seen}$ **then**
      $\mathrm{Seen.add}(\mathtt{s})$;
      if $\mathtt{P}(\mathtt{s})$ *is undefined* **then**
        $\textsc{GenPlanPairs}(\langle \mathcal{F}, \mathcal{G}, \mathtt{s}, \mathcal{A} \rangle, \mathtt{P})$;
      if $\mathtt{P}(\mathtt{s})$ *is defined* **then**
        for $\mathtt{s}' \in \textsc{GenerateSuccessors}(\mathtt{s}, \mathtt{P}(\mathtt{s}))$ **do**
          $\mathrm{Open.add}(\mathtt{s}')$;
  $\textsc{ProcessDeadends}()$;
return P;

Example: Tic-Tac-Toe



GenerateSuccessors Algorithm

Input: State $\mathtt{s}$, action $\mathtt{a}$
Output: Set of successor states $\mathtt{S}$
$\mathtt{S} = \{\mathit{Prog}(\mathtt{s}, \mathtt{a}, \mathtt{e}) \mid \mathtt{e} \in \mathrm{Eff}_\mathtt{a}\}$;
for $\mathtt{i} = 1 \ldots |\overrightarrow{\mathtt{ag}}|$ **do**
  $\mathtt{S}' = \emptyset$;
  for $\mathtt{s}' \in \mathtt{S}$ **do**
    for $\mathtt{a}' \in \mathit{Plausible}(\mathtt{s}', \overrightarrow{\mathtt{ag}}[\mathtt{i}])$ **do**
      $\mathtt{S}' = \mathtt{S}' \cup \{\mathit{Prog}(\mathtt{s}', \mathtt{a}', \mathtt{e}) \mid \mathtt{e} \in \mathrm{Eff}_{\mathtt{a}'}\}$;
  $\mathtt{S} = \mathtt{S}'$;
return $\mathtt{S}$;

Plausibility Function

Top k actions from $\mathit{App}(\mathtt{s}, \mathtt{ag})$ according to Score:

$$\mathit{Score}(\mathtt{s}, \mathtt{a}, \mathtt{ag}, \mathcal{G}) = \max_{\mathtt{e} \in \mathrm{Eff}_\mathtt{a}} \mathtt{h}_{\mathsf{FF}}(\mathit{Prog}(\mathtt{s}, \mathtt{a}, \mathtt{e}), \mathcal{G}(\mathtt{ag}))$$

Key Considerations

The value of doing nothing

Under what situations does it pay off to model noop (*no op*eration) actions for the other agents?

(Un)Fair non-determinism

How damaging is the assumption of fairness in MAP? (e.g., unfair adversaries, livelocks, etc)

Winning -vs- not losing

Is it better to win most of the time with some chance of losing, or to never lose but draw often?

Issues with FOND as a black box

What prevents us from casting the First-Person MAP problem using a traditional FOND encoding / planner?

Summary

- Addressed a complex and understudied form of multi-agent planning problems
- Employed state-of-the-art FOND planning techniques to compute compact solutions
- Introduced a mechanism for reducing the amount of non-determinism based on a model of the other agents

		Blocksworld										Tic-Tac-Toe				Sokoban				
	N	p1	p2	p3	p4	p5	p6	p7	p8	p9	p10	p1	p2	p3	p4	p1	p2	p3	p4	p5
Success	1	38	63	31	20	0	4	0	19	12	11	0	1	45	48	100	99	97	67	28
rate (%)	2	62	66	61	73	17	22	47	81	18	39	0	2	47	100	100	98	96	98	100
	3	92	97	90	96	49	62	81	90	84	76	0	4	47	100	100	98	96	97	100
	4	100	100	100	100	95	70	89	100	98	86	16	3	47	100	100	✘	✘	99	100
	∞	100	100	100	100	100	68	85	100	100	100	100	100	79	100	100	✘	✘	✘	✘
	Rnd	100	100	100	100	100	68	88	100	100	100	39	81	52	100	100	71	48	37	✘
Policy	1	32	27	31	111	49	25	68	63	43	30	42	47	14	7	11	27	27	138	906
size	2	48	77	78	316	141	99	175	217	73	83	88	107	26	15	11	26	31	5990	7891
	3	125	114	157	576	298	246	445	459	126	164	121	260	19	15	11	26	31	10952	10223
	4	337	236	593	932	757	973	892	830	593	526	871	250	22	15	11	✘	✘	10312	9270
	∞	550	286	631	1113	1149	785	987	699	867	818	1358	651	69	15	11	✘	✘	✘	✘
	Rnd	586	444	671	1045	1076	662	1006	763	745	818	827	606	27	12	11	11	11	11	✘
Planning	1	0.01	0.01	0.01	0.02	0.02	0.01	0.02	0.02	0.02	0.01	5	1	0.01	0.01	0.06	0.08	0.08	0.46	195
time (s)	2	0.02	0.04	0.04	0.14	0.06	0.04	0.08	0.12	0.02	0.04	155	11	0.26	0.01	0.06	0.08	0.10	30m	30m
	3	0.06	0.06	0.06	0.32	0.24	0.16	0.56	0.36	0.04	0.10	658	40	0.50	0.01	0.06	0. 10	0.12	30m	30m
	4	0.24	0.14	0.62	1.02	1.10	1.08	0.98	0.90	0.68	0.54	978	39	7.34	0.01	0.06	✘	✘	30m	30m
	∞	0.14	0.06	0.26	0.44	0.46	0.30	0.40	0.22	0.32	0.48	1765	114	39.32	0.01	0.06	✘	✘	✘	✘
	Rnd	0.20	0.12	0.28	0.44	0.44	0.28	0.44	0.26	0.28	0.36	30m	130	0.01	0.01	0.08	0.06	0.08	0.08	✘

Preliminary Evaluation: Success rate over 1000 simulated trials, generated policy size, and the time to synthesize a plan. ✘ indicates a memory violation, 30m indicates the solving was capped at 30 minutes, and N indicates the level of restricted non-determinism (∞ meaning no restriction and Rnd meaning a random subset of 3 applicable actions).

www.haz.ca/research/map-as-fond

Paper · Slides · Code · Demo