Knowledge-Based Policies for Qualitative Decentralized POMDPs

Abdallah Saffidine  François Schwarzentruber  Bruno Zanuttini

AAAI, November 6, 2017
This is a Robbery!

Once upon a time there were two thieves...
Mission: Possible?

Requires some tools...

... Steal them!
Only then synchronise to rob the bank
Variables $t_1, \ldots, t_m = \text{tools to steal}$

Action try-to-steal-something:

- **conditions:**
  $\text{victim} \rightarrow \text{has} - t_i$

- **effects:**
  $\text{thief} \rightarrow \text{has} - t_i \land \neg \text{victim} \rightarrow \text{has} - t_i$

- **nondeterministic:**
  may fail up to $k$ times
Variables $t_1, \ldots, t_m = \text{tools to steal}$

Action try-to-steal-something:

- conditions:
  \[ \text{victim} \rightarrow \text{has} - t_i \]

- effects:
  \[ \text{thief} \rightarrow \text{has} - t_i \land \neg \text{victim} \rightarrow \text{has} - t_i \]

- nondeterministic:
  may fail up to $k$ times

Partially observable:

- actions generate signals after a while
Decentralization

Execution in a decentralized fashion:

![Fireworks drawing]
Decentralization

Execution in a decentralized fashion:

Monitoring of mate’s process only via signals:
Factored QDec-POMDP (Brafman, Shani, Zilberstein, AAAI 2013):

- $I$: agents
- $X$: propositions
- $A$: individual actions
- $\Omega$: individual observations
- $T$: transition function
- $B^0$: initial common belief state
- $g$: common goal
- $H$: planning horizon

- $s \in 2^X$ → states
- $a \in A^I$ → joint actions
- $\omega \in \Omega^I$ → joint observations
- $(s, a) \mapsto \{(s_1', \omega_1), \ldots, (s_k', \omega_k)\}$
- $\subseteq 2^X$
- Boolean formula over $X$
- $\in \mathbb{N} \text{ or } \infty$
Solutions to QDec-POMDPs

Setting:

- centralized planning
- offline planning: before execution starts
- decentralized execution
- synchronous execution

Solution = policy vector \((\pi_1, \ldots, \pi_n)\) s.t.

- \(\pi_i\) prescribes actions based on \(i\)'s local history
- execution terminates
- executing ends up satisfying goal

(assuming faithful execution)
Standard Policy Representation

Standard representation by one tree for each thief:

Standard representations:
- **reactive**: trivial execution
- **often huge**: size $2^m$ for $t_1, \ldots, t_m$ to steal
Our proposal: representation by \((\text{multiagent})\) knowledge-based programs

\[
\text{while } \neg K_A(t_1^B \land \cdots \land t_m^B) \lor \neg K_A K_B(t_1^A \land \cdots \land t_{m'}^A) \text{ do }
\]

try-to-steal-something

listen-to-signals

rob bank

\begin{itemize}
  \item More compact and more readable
  \item Less reactive: epistemic reasoning while executing
\end{itemize}

(MAKBPs not new; new = used as such in planning setting)
Multiagent Knowledge-Based Programs

Formally:

\[ kbp ::= \epsilon \mid a \mid kbp; kbp \mid \text{if } \Theta \text{ then } kbp \text{ else } kbp \text{ fi} \mid \text{while } \Theta \text{ do } kbp \text{ od} \]

\[ makbp ::= (kbp_1, kbp_2, \ldots, kbp_n) \]

Branching conditions \( \Theta \):

- \( jo(\omega) \) \( \rightarrow \) branch on observations
- or epistemic formula \( \rightarrow \) branch on own and others’ knowledge

Example in A’s KBP: if \((K_A \neg K_B x \lor K_A y)\) then \(\ldots\) else \(\ldots\)
Contributions in one Slide

Operational semantics for MAKBPs, for perfect reasoners:
- $S5_n$ semantics, progression using dynamic epistemic logic (DEL)
- cf. occupancy states (Dibangoye, Buffet, Charpillet, AAMAS 2017)

MAKBPs are compact:
- always at least as compact as equivalent trees
- possibly exponentially more compact than any reactive language

Complexity results:
- execution PSPACE-complete
- verification -c. at given horizon, undecidable otherwise
Summary and Perspectives

Take-home message:

- **MAKBPs more compact and readable** than reactive representations → language for experts to write policies
- **MAKBPs harder to execute and verify** than reactive representations
- Computing next action **PSPACE-complete** (not so bad!)

Future work:

- **add probabilities** (easy) → Dec-POMDPs
- algorithms for computing plans (difficult)
- investigate multiagent regression
- other operational semantics, more realistic, more manageable
- nonperfect reasoners: bounded epistemic depth, sampling...
RATHER ASK THE POLICE YOUR QUESTIONS, THEY'RE ARRIVING