Planning with Uncertain Specifications

RSS 2019 Workshop on Combining Learning with Reasoning

Ankit Shah, Julie Shah
{ajshah, arnoldj}@mit.edu

Motivation

- Linear temporal logic (LTL) formulas are an expressive means for specifying non-Markovian tasks.
- Prior research relies on LTL to automaton compilation for planning. However this is restricted to a single LTL formula.
- In many application there is an inherent uncertainty in specifications.[1],[2]
- In general specifications are expressed as a belief $P(\varphi)$ over support $\{\varphi\}$.

Question 1: What does satisfying a belief $P(\varphi)$ mean?

Question 2: How do we plan for a collection of LTL formulas $\{\varphi\}$?

Formulation

- Given:
  - $\boldsymbol{x} \in \boldsymbol{X}$: Learner's state representations.
  - $\boldsymbol{\alpha} = f(\boldsymbol{x})$; $\boldsymbol{\alpha} \in \{0,1\}^{n_{prop}}$: Learner's labeling function and task propositions.
  - $\boldsymbol{A}$: Learners set of available actions
  - $P(\varphi)$: The task specification as belief over formulas with support $\{\varphi\}$.
- Expected output:
  - $\pi_{P(\varphi)}(\boldsymbol{x})$: A stochastic policy that best satisfies $P(\varphi)$.

Evaluation Criteria

Most Likely

$$\mathbb{1}([\alpha] \vDash \varphi^*)$$
$$\varphi^* = argmax_{\varphi \in \{\varphi\}} P(\varphi)$$

Satisfy only the most likely formula.

Maximum Coverage

$$\frac{1}{|\{\varphi\}|} \sum_{\varphi \in \{\varphi\}} \mathbb{1}([\alpha] \vDash \varphi^*)$$

Satisfy the largest set of unique formulas.

Minimum Regret

$$\sum_{\varphi \in \{\varphi\}} P(\varphi) \mathbb{1}([\alpha] \vDash \varphi^*)$$

Maximize satisfaction weighted by probability.

Chance Constrained

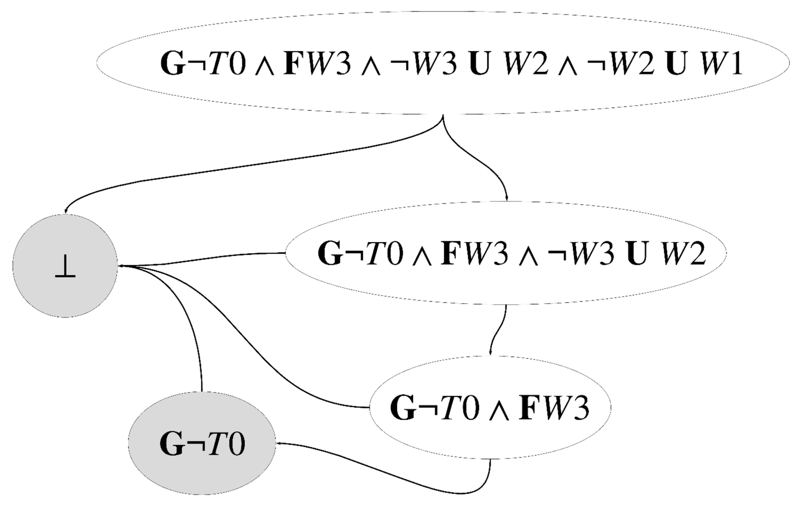
$$\sum_{\varphi \in \varphi^\delta} P(\varphi) \mathbb{1}([\alpha] \vDash \varphi^*)$$

δ is the maximum failure probability.

Automata/MDP Compilation

$P(\varphi_1) = 0.05$

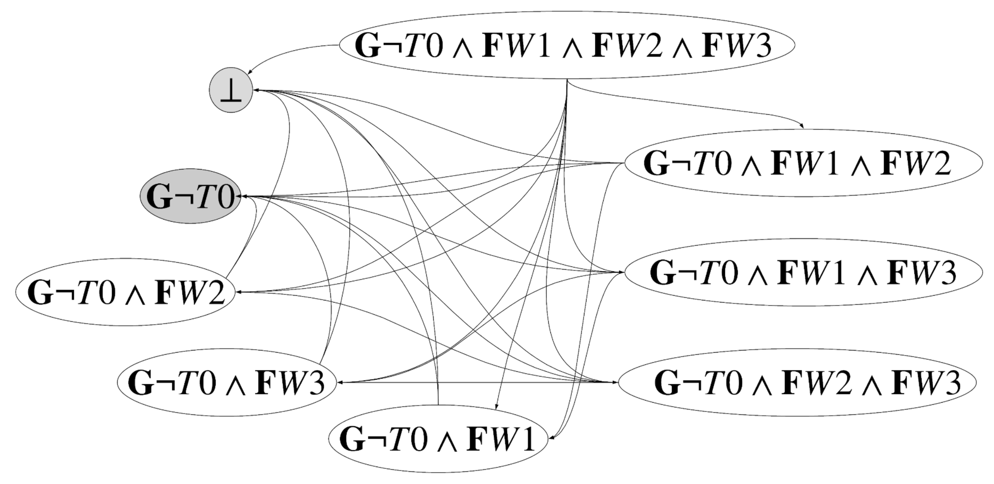
$\boldsymbol{G} \neg T0 \wedge \boldsymbol{F} W3 \wedge \neg W3 \; \boldsymbol{U} \; W2 \wedge \neg W2 \; \boldsymbol{U} \; W1$



5 states

$P(\varphi_2) = 0.15$

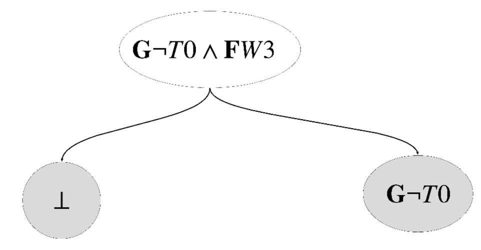
$\boldsymbol{G} \neg T0 \wedge \boldsymbol{F} W1 \wedge \boldsymbol{F} W2 \wedge \boldsymbol{F} W3$



9 states

$P(\varphi_3) = 0.8$

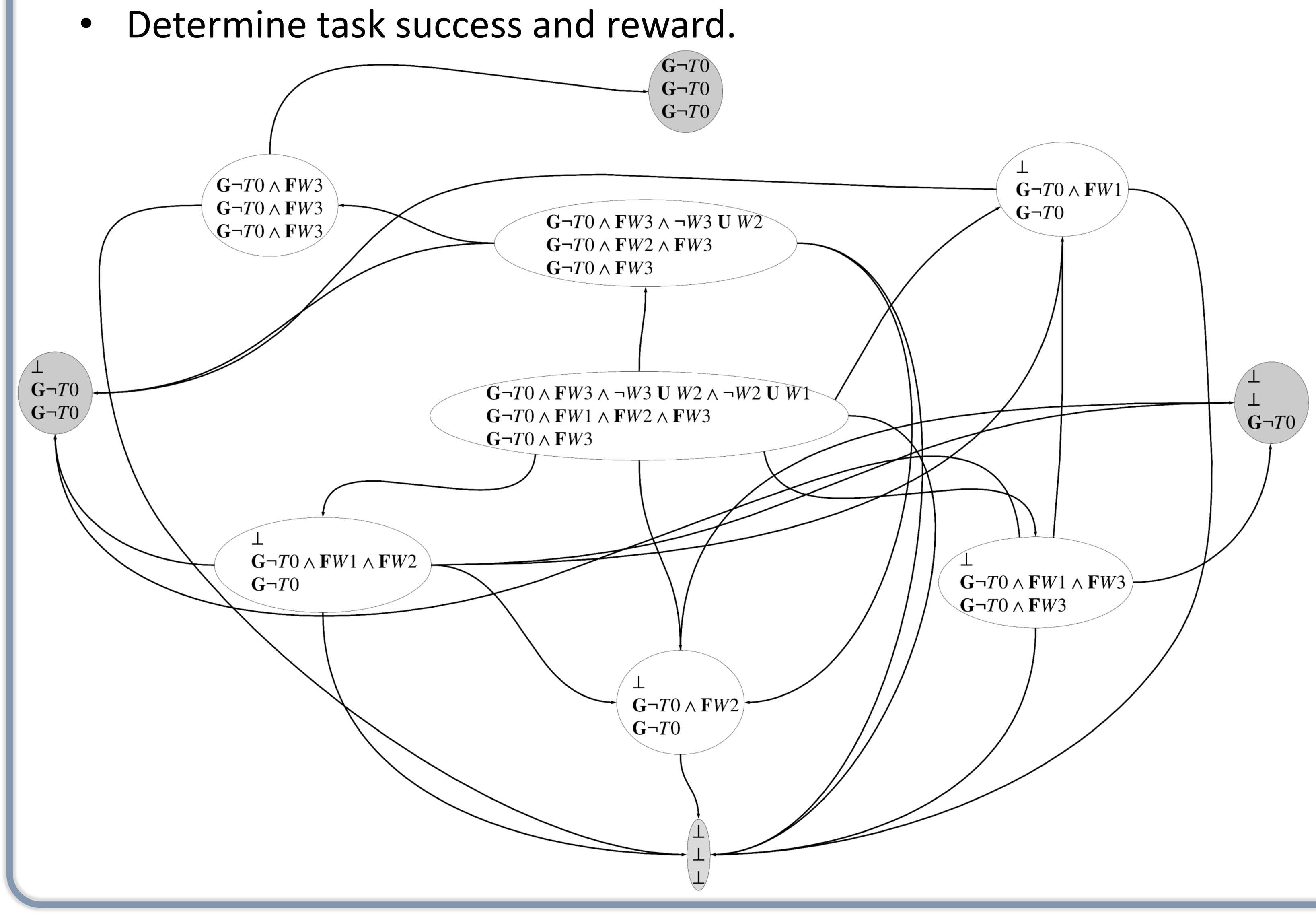
$\boldsymbol{G} \neg T0 \wedge \boldsymbol{F} W3$



3 states

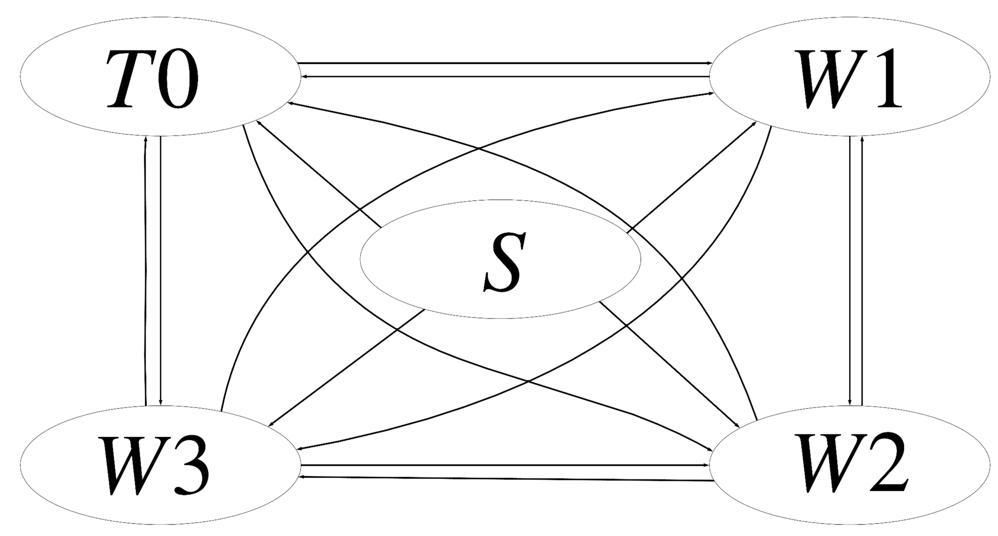
Composite automaton $\mathcal{M}_{\{\varphi\}} = \langle \{\langle \varphi' \rangle\}, \{0,1\}^{n_{prop}}, T_{\{\varphi\}}, R \rangle$

- Naïve cross-product: 135 states
- Minimal automaton: 11 states
- Determine task success and reward.



Environment MDP $\mathcal{M}_{env} = \langle \boldsymbol{X}, \boldsymbol{A}, T_{env} \rangle$:

- Determines available actions.
- Determines changes to environment state.



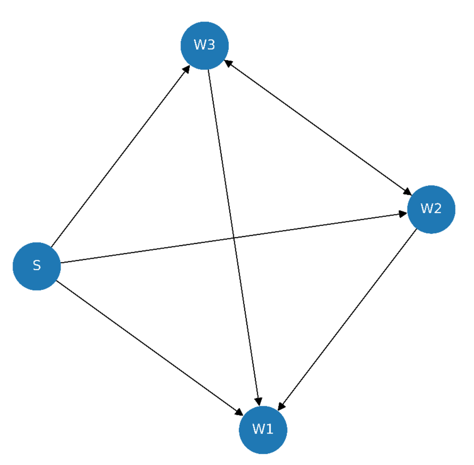
Cross product $\mathcal{M}_{\{\varphi\}} \times \mathcal{M}_{env} = \mathcal{M}_{spec}$

$\mathcal{M}_{spec} = \langle \{\langle \varphi' \rangle\} \times \boldsymbol{X}, A, T_{spec}, R \rangle$

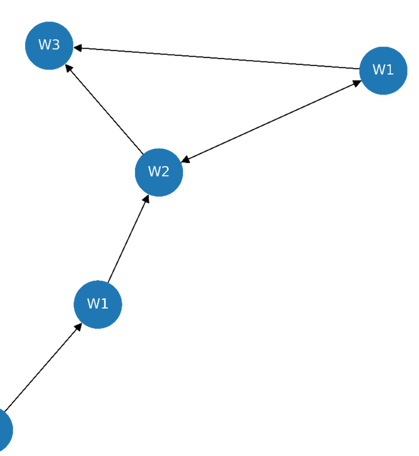
$T_{spec}(\langle \langle \varphi_1' \rangle, x_1 \rangle, \langle \langle \varphi_2' \rangle, x_2 \rangle) = T_{\{\varphi\}}(\langle \varphi_1' \rangle, \langle \varphi_2' \rangle) \times T_{env}(x_1, x_2)$

Results

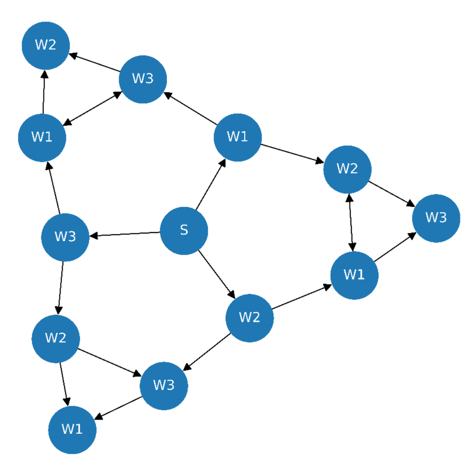
Most Likely



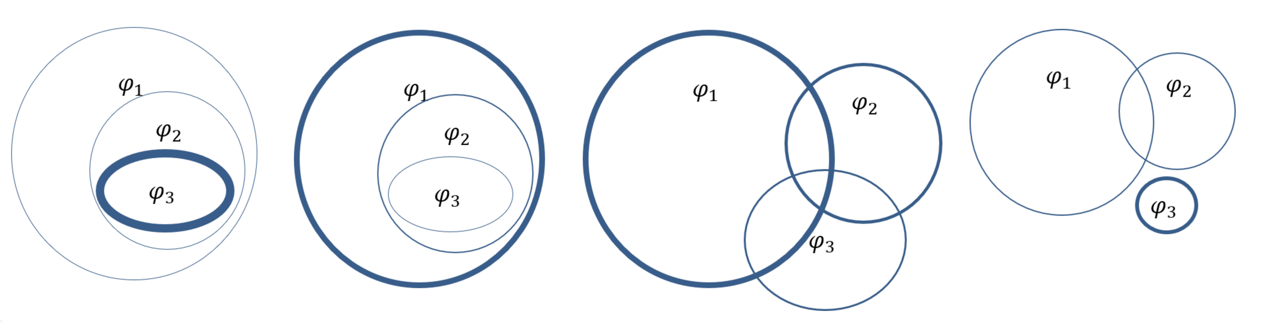
Max coverage/Min regret



Chance constrained $\boldsymbol{\delta} = \boldsymbol{0.1}$



Classes of belief distributions:



Nature of task executions depends on:

- Nature of distribution.
- Choice of Evaluation criterion.
- Exploration strategy in RL algorithm.

Discussion

- MDP compilation admits formulas of the *Obligation* class of temporal properties.
- Any RL algorithm can be used to solve the compiled MDP, but exploration vs exploitation considerations are still important.

Future Work

- Algorithms to exploit the composition of $\mathcal{M}_{\{\varphi\}}$ and $\mathcal{M}_{env}$
- Scaffolding of reward based on automaton structure.
- Allowing temporal properties like *Recurrence, Persistence* and *Reactivity*

[1] Shah, A., Kamath, P., Shah, J. A., & Li, S. (2018). Bayesian inference of temporal task specifications from demonstrations. In *Advances in Neural Information Processing Systems* (pp. 3804-3813).
[2] Kim, J., Banks, C. J., & Shah, J. A. (2017, February). Collaborative planning with encoding of users' high-level strategies. In *Thirty-First AAAI Conference on Artificial Intelligence*.

Visit our website for more!
http://interactive.mit.edu