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 applications, there is an inherent uncertainty in specifications.\[1,2\]
- In general specifications are expressed as a belief $P(\phi)$ over support $[\phi]$.

Question 1: What does satisfying a belief $P(\phi)$ mean?
Question 2: How do we plan for a collection of LTL formulas $\{\phi\}$?

Formulation

- Given: $x \in X$: Learner’s state representations.
- $\alpha = f(x)$: Learner’s labeling function and task propositions.
- $A$: Learners set of available actions
- $P(\phi)$: The task specification as belief over formulas with support $[\phi]$.
- Expected output: $\pi_{P(\phi)}(x)$: A stochastic policy that best satisfies $P(\phi)$.

Evaluation Criteria

- Most Likely $\overline{\phi} = \arg\max_{x \in X} P(\phi)$, $P(\phi)$ Satisfy only the most likely formula.
- Maximum Coverage $\frac{1}{|\phi|} \sum_{\phi \neq \phi^*} \mathbb{1}(x \in \phi)$, $\phi^*$ Satisfy the largest set of unique formulas.
- Minimum Regret $\sum_{\phi \neq \phi^*} P(\phi) \mathbb{1}(x \in \phi^*)$ Maximize satisfaction weighted by probability.
- Chance Constrained $\sum_{\phi \neq \phi^*} P(\phi) \mathbb{1}(x \in \phi^*)$, $\delta$ is the maximum failure probability.

Variants of LTL formulas:

\begin{align*}
&\phi^* = \phi \land \lnot \phi^* \\
&\phi^* = \phi \\
&\phi^* = \lnot \phi
\end{align*}

Automata/MDP Compilation

- Composite automaton $\mathcal{M}(\phi) = (\langle \phi \rangle, \{0,1\}^{|\text{prop}|}, T(\phi), R)$
- Naive cross-product: 135 states
- Minimal automaton: 11 states
- Determine task success and reward.

Results

- Most Likely $
\begin{array}{c}
\begin{array}{c}
\text{Most Likely:}
\begin{array}{c}
\phi^* = \arg\max_{x \in X} P(\phi)
\end{array}
\end{array}
\end{array}$

- Max coverage/Min regret $\overline{\phi} = \arg\max_{x \in X} P(\phi)$, $P(\phi)$ Satisfy only the most likely formula.

- Chance constrained $\delta = 0.1$

- 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 $\mathcal{M}(\phi)$ and $\mathcal{M}_{\text{env}}$
- Scaffolding of reward based on automaton structure.
- Allowing temporal properties like \textit{Recurrence}, \textit{Persistence} and \textit{Reactivity}.

Notes: