Motivation

Background

Methods

Experiments

Conclusion

Yao Feng, Anita Li, Learnable Model Verification through Reinforcement Learning
Learnable Model Verification through Reinforcement Learning

December 9, 2021

Yao Feng
yaofeng@andrew.cmu.edu
15624

Anita Li
weihanl1@andrew.cmu.edu
15824
Motivation

- Model-based Reinforcement Learning (MBRL): Learning accurate modeling from the environment
- Formal Verification: Provides safety guarantee for CPS
- Connect MBRL with formal verification:

**Safe Reinforcement Learning**
A missing link

- **Justified Speculative Control (JSC):** the first algorithm proposed to incorporate safety guarantees into the learning process of RL.

- **Verification Preserving Model Updates (VPMU):** use a fixed set of models in the learning process.
Terms

Given the verified model $\text{init}_\theta \rightarrow [\{\text{ctrl}_\theta; \text{plant}\}^\ast]\text{safe}$, we have the following definitions:

**Definition**

A sequence of tuples $(s^t, a^t, \pi^t, \theta^t)$ describe a **Learning Process**, with the **Transition Function** $T(s^t, a^t) = s^{t+1}, a^t \sim \pi^t(s^t)$.

**Definition**

(Controller Monitor) if $CM(s, a) = true$, $(s, T(s, a)) \in [\text{ctrl}]$ [1].

**Definition**

(Model Monitor) if $MM(s, a, s') = true$, $(T(s, a), s') \in [\text{plant}]$ [1].

**Definition**

Let $H$ be an **update function** of parameter $\theta$ such that

$$H(s^t, a^t, s^{t+1}, \theta^t) = \theta^{t+1}$$
Learnable Justified Speculative Learning (LJSL)

**Figure:** The framework of LJSL
Main Theorem

If for all $\theta$, the model $\text{init}_\theta \rightarrow [{\{\text{ctrl}_\theta;\text{plant}\}}^\ast]_{\text{safe}} \land \text{init}_\theta$ is valid, $CM$, $MM$ are accurate control monitor and model monitor and $H$ is a valid update function ($\text{init}_H(\theta, s, a, s') \rightarrow [{\{\text{ctrl}_H(\theta, s, a, s');\text{plant}\}}^\ast]_{\text{safe}}$ is accurate for any accurate model $\text{init}_\theta \rightarrow [{\{\text{ctrl}_\theta;\text{plant}\}}^\ast]_{\text{safe}}$ and $s' = T(s, a)$), and $\text{init}_\theta \rightarrow \text{init}_H(\theta, s, a, s')$, we have $s_t \models \text{safe}$ for all $t \geq 0$.

Proof Idea: the model guarantees the initialization of the updated model holds, so we can connect the safety proofs of a sequence of model by induction.
Experiments

- Continuous Adaptive Cruise Control (CACC)
- Robot Motion
Simplified Robot Motion: Goal Finding

**Figure:** The star is the moving agent, the green circle is the goal, the red circles are the obstacles to avoid.
Example Model Sketch

\[ \theta = (r_{\text{min}}, r_{\text{max}}) \]

\[ \text{init}_\theta \to [\{\text{ctrl}_\theta; \text{plant}\}^\ast]_\text{safe}, \text{ where} \]

\[ \text{init}_\theta \equiv \text{valid}_\text{env} \land \text{const}_\text{bounds}_\theta, \]

\[ \text{valid}_\text{env} \equiv \bigwedge_{i=1}^{n} \text{dist}_\text{safe}(\text{agent}, \text{obs}_i) \land \bigwedge_{i \neq j}^{n} \text{dist}_\text{safe}(\text{obs}_i, \text{obs}_j) \]

\[ \text{ctrl}_\theta \equiv r \text{ satisfies } \{ \bigwedge_{i=1}^{n} \text{dist}_\text{safe}(\text{agent}_{r}, \text{obs}_i) \land r! = 0 \} \]

\[ \text{plant} \text{ describes the agent's circular movement} \]

\[ \text{safe} \equiv \text{no}_\text{crash} \]
Experiments

To satisfy the preconditions of our theorem, we need to implement the following functions:

- RL Algorithm: Soft Actor-Critic (SAC)
- Control Monitor and Model Monitor: can be easily induced from the proven model
- Update Function $H(s, a, s', \theta) = \{\min(a_r, r_{min}), \max(a_r, r_{max})\}$
Results: Goal Finding

**Figure:** (a) the learning process of parameters for an imperfect LJSL model (b) test rewards of agents with no model and an LJSL model with imperfect initialization ($r_{\text{min}} = -1, \ r_{\text{max}} = 1$)
Conclusion

- Proposed LJSL algorithm and proved the main theorem of safety
- Implemented 3 different environments and verified effectiveness through experiments

- The model monitors are conservative, and might still have room for improvements
- Need much prior information, such as forms of $\theta$ and $H$
- Simultaneously updating more than one parameter