Control Synthesis from LTL Specifications using Reinforcement Learning

Alper Kamil Bozkurt
Research Initiation Project, March 2020
Computer Science, Duke University

CPSL@Duke
Motivation

• High-Assurance Autonomous Systems
  • Model Checking, Formal Verification

• Unknown Model
  • Machine Learning

• AI Safety
  • Under specifications
  • Mismatched objectives
  • Interpretable ML (Doshi-Velez and Kim, 2017)
  • Reward Hacking (Leike et al., 2017)

Credit: Baier and Katoen, 2008

Credit: Ortega et al., 2018
Problem Statement

- Specification ($\varphi$)
- System Model ($M$)
- Discounted Reward ($\gamma, R$)
- Reinforcement Learning
- Controller

- Model Checking
- Stochastic Model Checking
- Efficient Sampling a.k.a. RL
Markov Decision Processes (MDPs)

- $M = (S, A, P, s_0)$
- Graph Representation
- Grid World
Linear Temporal Logic (LTL)

- $\varphi = \neg \text{shutdown} \mathbin{U} \text{warn}$ (“warn occurs before shutdown”)
- $\varphi = \Diamond (\text{room}_a \land \Diamond \text{room}_b)$ (“go to room ‘a’, then go to room ‘b’ ”)
- $\varphi = \Box (\text{request} \to \Diamond \text{response})$ (“every request eventually gets a response”)

- Paths: $\sigma := s_0 s_1 s_2 s_3 \ldots s_{t-1} s_t s_{t+1} \ldots$
- Labels and Traces: $\emptyset \emptyset \emptyset \ldots \emptyset \{a\} \emptyset \ldots \emptyset \{b\} \emptyset \ldots$
LTL to Automata

**Specification** ($\varphi$)

$\varphi = \Diamond (\text{room}_a \land \Diamond \text{room}_b)$

(“go to room ‘a’, then go to room ‘b’”)

$\varphi = \Box \Diamond (\text{room}_a \land \Diamond \text{room}_b)$

(“Repeat $\varphi$”)

**Automata** ($A$)

**Discounted Reward** ($\gamma, R$)

**Finite State Automata**

**Büchi Automata**

**Nondeterministic Büchi Automata**
Automata to Discounted Reward

Automata \((A)\) \(\rightarrow\) MDP \((M)\)

Product MDP \(<s, q>\)

Discounted Reward \((\gamma, R)\)

B: Accept States

Reach(B) \(\varphi = \Diamond(room_a \land \Diamond room_b)\) ("go to room ‘a’, then go to room ‘b’ “)

Importance of discounting
Brazdil et al., 2014
Hahn et al., 2019?

Repeat(B) \(\varphi = \Box\Diamond(room_a \land \Diamond room_b)\) ("Repeat \(\varphi\)"")

Average Reward?
Sadigh et al., 2014
Hasanbeig et al., 2018
Repeated Reachability ~ Repeat(B)

- $Q_{\text{up}} = 0.9 \cdot \frac{1}{1 - \gamma}$
- $Q_{\text{down}} = \frac{1}{1 - \gamma^2}$
- ratio : $0.9 \cdot (1 + \gamma)$

State-Based Discounting

- $\lim_{\gamma \to 1^-} \frac{1 - \gamma}{1 - \gamma_B(\gamma)}$

Proof Idea:
- A recurrent class is reached in a finite number of steps
- An accepting state is always revisited again in a finite number of steps in an accepting recurrent class
Dealing with Nondeterminism

(a) A derived LDBA $\mathcal{A}$ for the LTL formula $\varphi = \Diamond \Box a \lor \Diamond \Box b$

(b) An example MDP $\mathcal{M}$; the circles denote MDP states, rectangles denote actions, and numbers transition probabilities

(c) The obtained product MDP
Control Synthesis from Linear Temporal Logic Specifications using Model-Free Reinforcement Learning

Alper Kamil Bozkurt, Yu Wang, Michael M. Zavlanos, Miroslav Pajic
Duke University
Conclusion and Future Work

- ⊕ LTL to RL
- ⊕ Unknown Transition Probabilities
- ⊕ Convergence Guarantees

- ⊖ No Adversary (Uncontrollable Nondeterminism)
- ⊖ Convergence Rate and PAC Guarantees
- ⊖ State Space Explosion
Questions?
References


