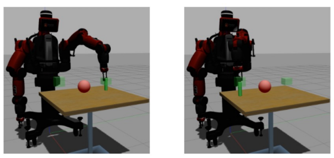
NRI: FND: A Formal Methods Approach to Safe, Composable, and Distributed Reinforcement Learning for co-Robots

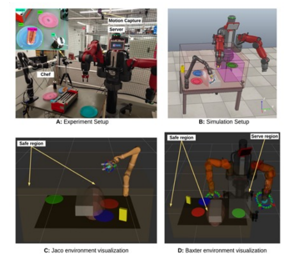
PI: Calin Belta, Boston University

Problem Formulation: Given rich *temporal logic specifications*, e.g., "Always prepare hotdogs and drinks. If a customer requests drinks or hotdogs, deliver whenever possible. If no customer is requesting or food / drinks are not available, then wait", given *safety requirements*, e.g., "Always avoid collisions.", and given *robot action and sensing capabilities*, e.g., "Robot 1 can make hotdogs, can place hotdogs on Robot2, can sense when a hotdog is ready, can sense when Robot2 is near, can sense obstacles, Robot 2 can deliver hotdogs and drinks", find robot control strategies that *satisfy the specs* and are *provably safe.*



Spec: Pick up the object from the pickup region within 3 sec and drop it in the goal region within the next 3 sec. The rotation angle of the end-effector from the vertical pose should be always less than 30 deg.

Safety: Always avoid collisions with the pink sphere, its body, and the table



Spec: *Chef (Jaco):* e.g., Turn on grill and then place sausage on grill and then return to home position and then wait for 600 seconds and then place sausage on the bun and then apply ketchup and then turn off grill and then return to home position

Spec: *Server (Baxter):* e.g., Serve the hotdog when one is ready and there is a customer. Eventually serve the hotdog and do not serve until hotdog is ready and customer is detected.

General knowledge: e.g., the gripper cannot be both open and closed at the same time

Safety: e.g., Baxter, never get close to the grill

Challenges

Model-based (optimal) control is not appropriate due to complex dynamics and large observation spaces. Reinforcement learning (RL) is usually used for such problems. **In RL, it is difficult to**

- Incorporate **complex task** in objective functions,
- Ensure that the learned policy satisfies the **safety requirements**,
- **Transfer** learned policies to unseen tasks,
- Effectively **distribute** a complex task among a robot team that allows each member to learn in a decentralized fashion.

Scientific Impact

Formal methods approach to RL

- provides a formal specification language that integrates high-level, rich, task specifications with a priori, domain-specific knowledge;
- makes the reward generation process easily interpretable;
- guides the policy generation process according to the specification;
- guarantees the satisfaction of the (critical) safety component of the specification.

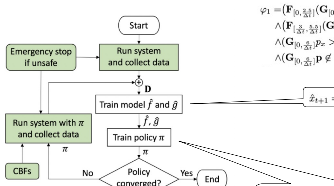
Technical Approach

Model (unknown) dynamics assumed affine in controls

$$x_{t+1} = f(x_t) + g(x_t)u_t,\ t = 0, 1, \ldots$$

Spec: STL formula

$$\begin{aligned}\varphi_1 =& \big(\mathbf{F}_{[0,\frac{2.5}{\Delta t}]}(\mathbf{G}_{[0,\frac{0.5}{\Delta t}]}\mathbf{p} \in R_{pick})\big) \\ &\wedge\big(\mathbf{F}_{[\frac{3}{\Delta t},\frac{5.5}{\Delta t}]}(\mathbf{G}_{[0,\frac{0.5}{\Delta t}]}\mathbf{p} \in R_{place})\big) \wedge (\mathbf{G}_{[0,\frac{6}{\Delta t}]}\mathbf{q} \in V_{30}) \\ &\wedge(\mathbf{G}_{[0,\frac{6}{\Delta t}]}p_x > x_{body}) \wedge (\mathbf{G}_{[0,\frac{6}{\Delta t}]}p_z > h_{tb}) \\ &\wedge(\mathbf{G}_{[0,\frac{6}{\Delta t}]}\mathbf{p} \notin Obs)\end{aligned}$$



$$\hat{x}_{t+1} = \mathcal{F}(x_t; W_f, Z_f) + \mathcal{G}(x_t; W_g, Z_g)u_t$$

FNN trained to minimize

$$C = \sum_{\mathbf{D}} \sum_{t=0}^{T-1} \|x_{t+1} - \hat{x}_{t+1}\|^2$$

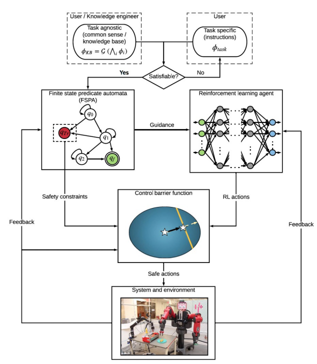
$$u_t = \pi(x_{0:t}; W)$$

$$W^* = \arg\max_W E_{p(x_0)}[\rho(\varphi, x_{0:T})]$$

$$\text{s.t. } x_{t+1} = f(x_t) + g(x_t)\pi(x_{0:t}; W), t = 0, \ldots, T-1$$

$$\mathbf{h}_t = \mathcal{R}(x_t, \mathbf{h}_{t-1}, W_1),$$
$$u_t = \mathcal{N}(\mathbf{h}_t, W_2),$$

RNN

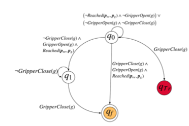


Model: known set of states states (robot related states: *end-effector pose, gripper position;* environmental states: *grill pose, ketchup pose)* and actions (*end-effector velocity)*, and unknown outcomes of actions applied at states,

Task specification and general knowledge are formulas in a logic (TLTL) that is expressive, allows for compositionality, and translation to accepting automata (FSPA)

***Task**: "eventually reach the goal pose and with the gripper open, and eventually close the gripper, and do not close the gripper until the goal pose is reached and with the gripper open"*

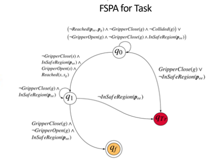
$$\begin{aligned}Grasped(\mathbf{p}_{ee}, \mathbf{p}_g) =& \mathcal{F}\left((Reached(\mathbf{p}_{ee}, \mathbf{p}_g) \wedge GripperOpen(g)) \wedge \mathcal{F}\, GripperClose(g)\right)\wedge \\ &(\neg GripperClose(g)\ \mathcal{U}\ (Reached(\mathbf{p}_{ee}, \mathbf{p}_g) \wedge GripperOpen(g))),\end{aligned}$$



FSPA for Task

***General knowledge**: "the gripper cannot be both open and closed at the same time"*

$$K^{grasped} = \{\neg(GripperOpen(g) \wedge GripperClose(g)), InSafeRegion(\mathbf{p}_{ee})\}.$$



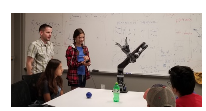
FSPA for Task and General knowledge

Broader impact

- Avoiding reward hacking, guaranteeing safety during learning and deployment, use of formal, unambiguous specification languages can potentially encode **robotic safety standards** (ISO)
- Interpretable nature of the formal language has the potential to promote **public trust** in robots

Education and outreach

- new courses and sections at the undergraduate and graduate level,
- the involvement of the PI in Research Internships in Science and Engineering (RISE), Discovery Internships, the BU Upward Bound Math and Science (UBMS) program, and the Technology Innovation Scholars Program (TISP) at BU.



Wei Xiao, Christos G. Cassandras, and Calin Belta, Safe Autonomy with Control Barrier Functions: Theory and Applications, Springer Nature, 2023 (in print)
Max H. Cohen, Calin Belta, Adaptive and Learning-based Control of Safety-Critical Systems, Springer Nature, 2023 (in print)
Max Cohen, Calin Belta, Safe Exploration in Model-based Reinforcement Learning using Control Barrier Functions, Automatica, vol. 147, 2023 (in print)
Max H. Cohen, Zachary Serlin, Kevin Leahy, Calin Belta, Temporal Logic Guided Safe Model-based Reinforcement Learning: A Hybrid Systems Approach, Nonlinear Analysis Hybrid Systems (NAHS), vol. 47, 2023 (in print)
Wenliang Liu, Kevin Leahy, Zachary Serlin, and Calin Belta, CatlNet: Learning Communication and Coordination Policies from CaTL+ Specifications, Learning for Dynamics and Control Conference (L4DC), 2023 (accepted)
Wenliang Liu, Kevin Leahy, Zachary Serlin, Calin Belta, Robust Multi-Agent Coordination from CaTL+ Specifications, American Control Conference (ACC), 2023 (accepted)
Wenliang Liu, Mirai Duintjer Tebbens Nishioka, Calin Belta, Safe Model-based Control from Signal Temporal Logic Specifications Using Recurrent Neural Networks, IEEE International Conference on Robotics and Automation (ICRA), 2023 (to appear)

2023 FRR & NRI Principal Investigators' Meeting
May 2-3, 2023

Award ID#:2024606