SMALL: Formal Methods for Safe, Efficient, and Transferable Learning-enabled Autonomy

Yiannis Kantaros, Dept. of Electrical and Systems Engineering, Washington University in St. Louis

Motivation: Learning-based techniques, such as deep reinforcement learning (DRL) or neural model predictive control (NMPC), have been extensively used to synthesize controllers and planners for autonomous robots with mission and safety objectives described using formal specifications such as Linear Temporal Logic (LTL).

Key Challenges

- **Sample inefficiency** in learning controllers.
- **Lack of safety/performance guarantees** of learned controllers, especially when deployed in new/unseen environments.

Problem Statement: Given a set of high-level (training) tasks, expressed using formal languages (e.g., LTL), design data-efficient controllers/planners that safely generalize to unseen domains.

Scientific Impact

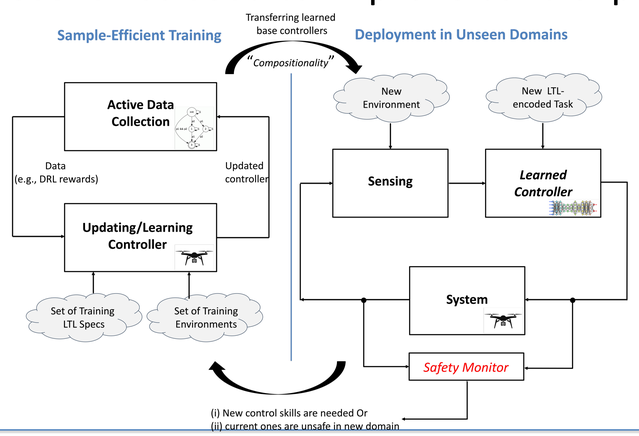
- Designing learning-enabled controllers that can be safely transferred to new/unseen task and environmental domains.
- Evaluation using ground and aerial robot platforms on manufacturing and transportation applications. Other CPS applications include search-and-rescue, disaster relief, or exploration.

Technical Approach

Thrust 1 (Sample-efficiency): Active learning for control methods.

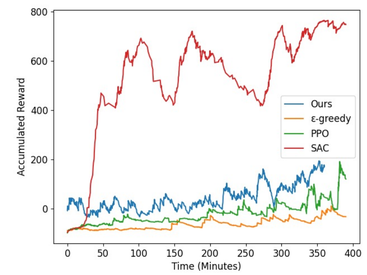
Thrust 2 (Transferability): Leverage compositionality for transferability

Thrust 3 (Safety): Conformal prediction to monitor safety in unseen domains. Unsafe scenarios used to improve control performance.



New Contributions (Year 1)

- Accelerated control synthesis using DRL (Thrust 1)
- Monitoring Safety of LLM-based planners using conformal prediction (Thrusts 2 & 3)



Broader Impacts

- K-12 research opportunities
- Capstone projects on AI & Robotics
- New graduate course on machine learning for robot planning & control
- Humanitarian applications (e.g., disaster relief) that require safe and quick adaptation to inevitable mission and environmental changes.

Email: ioannisk@wustl.edu

Start date: April 1, 2023
End date: March 31, 2026

Award ID#: 2231257