Generalizing Cooperative Eco-driving via Multi-residual Task Learning

Vindula Jayawardana†, Sirui Li†, Cathy Wu†, Yashar Farid‡, Kentaro Oguchi‡

† MIT ‡ Toyota Motor North America

Correspondence: vindula@mit.edu

ICRA2024 YOKOHAMA | JAPAN

Motivation

- Real-world autonomous driving contends with a multitude of diverse traffic scenarios.
- While model-free deep reinforcement learning (DRL) can be used to learn vehicle controllers, it is still challenging to learn controllers that generalize to multiple traffic scenarios.
- In addressing this challenge, we introduce ***Multi-residual Task Learning (MRTL)***, a generic learning framework based on multi-task learning that, **for a set of task scenarios, decomposes the control into nominal components that are effectively solved by conventional control methods and residual terms that are learned using DRL**.

Problem Formulation

- We study the requirement of **algorithmic generalization of DRL algorithms** across a family of MDPs that stem from a given task.
- Formally, consider a **contextual Markov Decision Process (cMDP)** $M = (S, A, p_c, r_c, \rho_c, \gamma)$ which extends Markov Decision Processes (MDP) with a context space *C (scenarios)*, and the action space *A* and state space *S* remain unchanged. The transition p_c, rewards r_c, and initial state distribution ρ_c are changed based on the context $c \in C$.
- We seek to **find policy π that solve a given *cMDP* by solving the problem of algorithmic generalization within that task** (i.e., finding a policy that performs well in the cMDP overall).

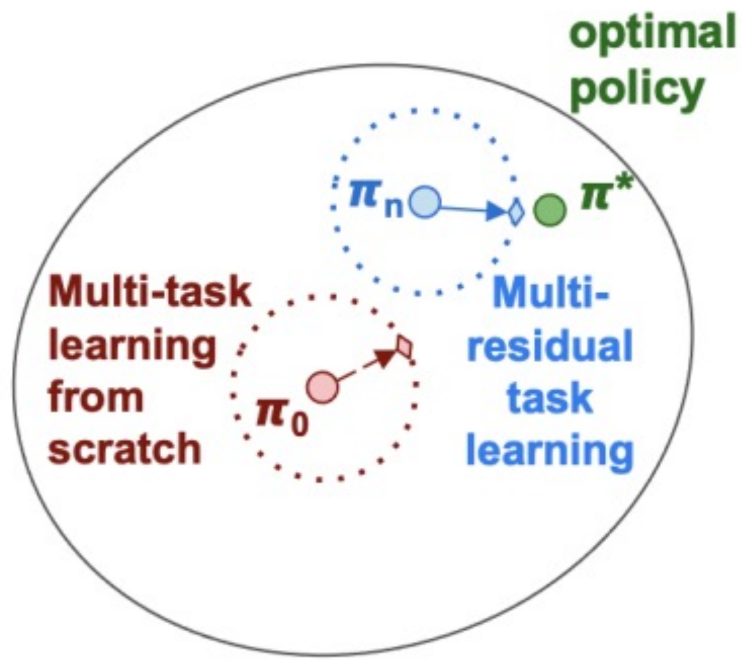
$$\pi^*(s) = \underset{\pi}{\operatorname{argmax}}\, \mathbb{E}\left[\sum_{c\in C}\sum_{t=0}^{H}\gamma^t r_c(s_t,a_t)|s_0^c,\pi\right]$$

Method

- ***Multi-residual Task Learning is*** a unified learning approach that leverages the synergy between multi-task learning and residual reinforcement learning.
- We aim to learn the MRTL policy $\pi(s,c): S\times C \rightarrow A$ by learning a residual function $f_\theta(s,c): S\times C \rightarrow A$ on top of a given nominal policy $\pi_n(s,c): S\times C \rightarrow A$ such that,

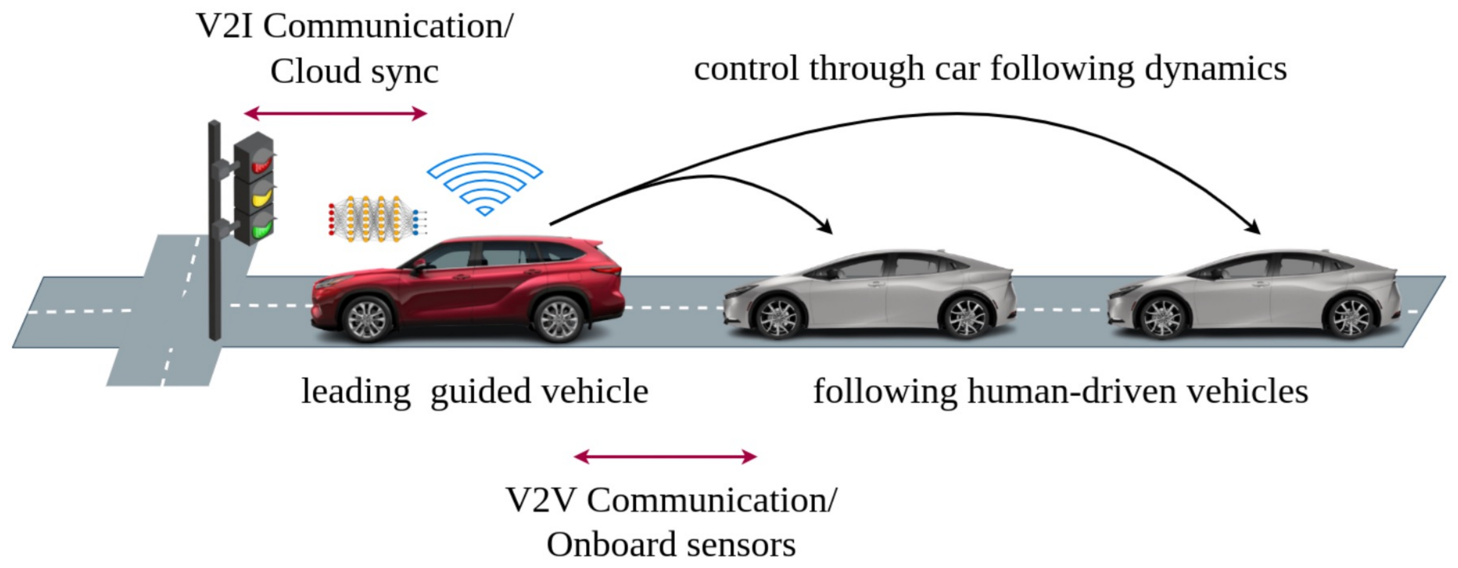
$$\underbrace{\pi(s,c)}_{\text{MRTL policy}} = \underbrace{\pi_n(s,c)}_{\text{Nominal policy}} + \underbrace{f_\theta(s,c)}_{\text{Residual function}}$$

- The gradient of the π does not depend on the π_n. This enables flexibility with nominal policy choice.
- **Intuition**: If the nominal policy is nearly perfect, the residual term can be viewed as a corrective term. If not, nominal policy provide useful hints to guide the exploration of DRL training.



Evaluations

- We apply MRTL to cooperative multi-agent eco-driving at signalized intersections.



- **Goal:** Use a fleet of autonomous vehicles to reduce fleet-wide emissions while having less impact on travel time.
- **Setting:** 600 signalized intersections were synthetically generated to match high-level real-world intersection statistics. Both 20% and 100% eco-driving adoption levels were tested.
- **Nominal policy**: A model-based heuristic (GLOSA algorithm)
- **Baselines:** Human-like driving using the Intelligent Driver Model (IDM), Multi-task learning from scratch (MTL), and the nominal policy alone (NP)

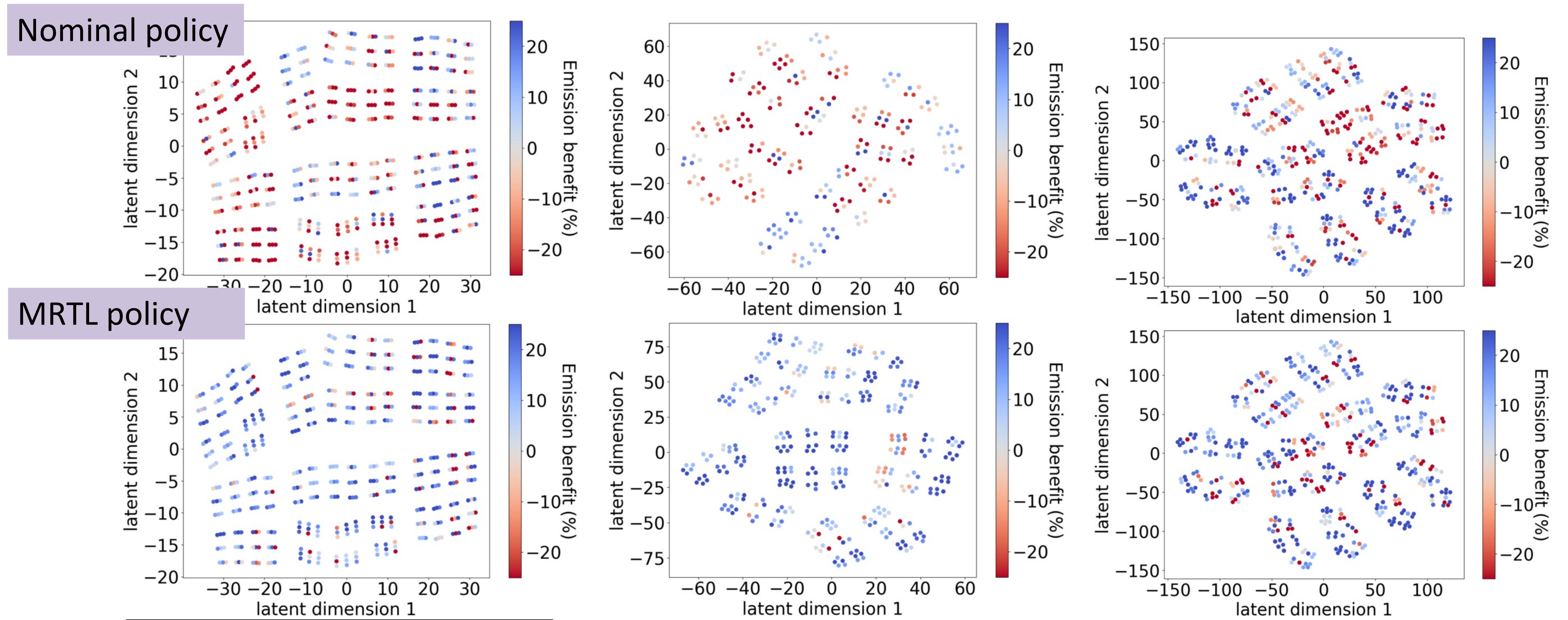
Emission reduction across 1200 traffic scenarios in 600 signalized intersections

- Emission reduction without reducing intersection throughput (lower the percentage the better).

Method (against IDM)	20% eco-driving adoption	100% eco-driving adoption
MTL	64.08%	95.86%
NP	13.13%	-25.09%
MRTL (Ours)	-13.95%	-29.09%

t-SNE visualization of emission benefits of MRTL policy in mitigating nominal policy limitations

- Each dot represents a signalized intersection approach, and the colors indicate the emission benefit levels.
- The higher the emission benefits, the better the results.

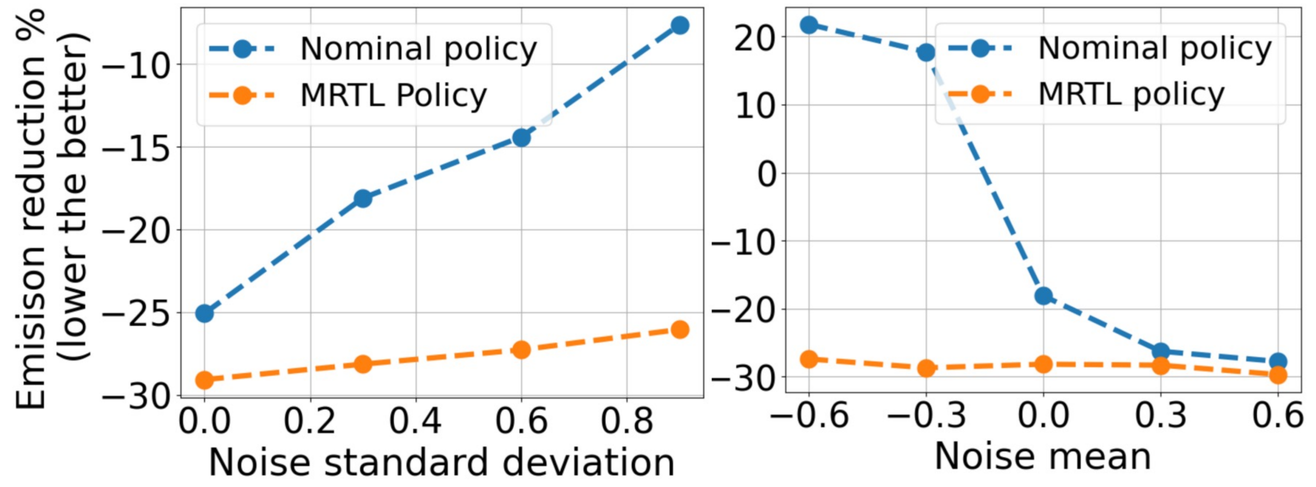


In partial eco-driving adoption levels | With dedicated protected left turns in the intersections | With unprotected left turns in the intersections

Robustness of MRTL to control noise (left) and bias noise (right)

E.g., control noise from communication delays and sensor issues

$\epsilon_c = \mathcal{N}(0, \sigma^2)$



E.g., bias noise from biases toward certain cities or conditions

$\epsilon_b = \mathcal{N}(\mu, 0.3)$

Takeaway

- Combining conventional control with residual terms learned through DRL is a promising approach to achieve algorithmic generalization in solving contextual markov decision processes.