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

Motivation

- Real-world autonomous driving contends with a multitude of diverse traffic scenarios that are challenging for conventional model-based planning algorithms.
- Model-free deep reinforcement learning (DRL) on the other hand presents a promising avenue to devise control algorithms, but learning DRL controllers that generalize to multiple traffic scenarios is still a challenge.
- In tackling 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 which are solved using DRL.

Problem Formulation

- We study the algorithmic generalization of DRL algorithms across a family of MDPs (scenarios) that originate from a single task.
- Formally, consider a contextual Markov Decision Process (cMDP) $\mathcal{M} = \langle \mathcal{S}, \mathcal{A}, \mathcal{C}, p_c, r_c, \rho_c, \gamma \rangle$ 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).

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

Method

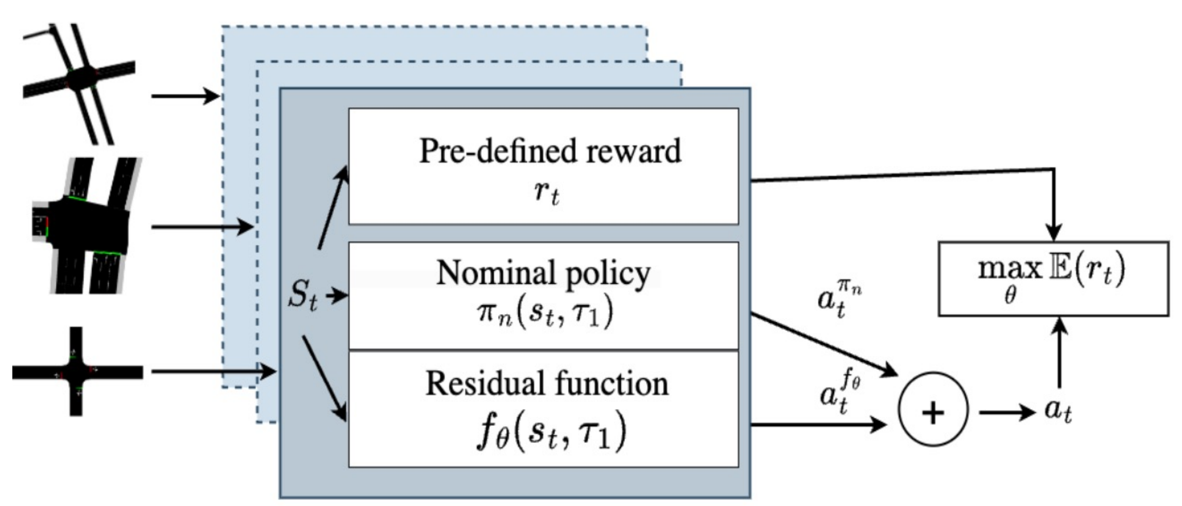
- We introduce ***Multi-Residual Task Learning (MRTL),*** a unified learning approach that harnesses the synergy between multi-task learning and residual reinforcement learning.
- We aim to learn the MRTL policy $\pi(s|c): S\times C \to A$ by learning a residual function $f_\theta(s|c): S\times C \to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

$$\pi(s|c) = \pi_n(s|c) + f_\theta(s|c)$$

MRTL policy | Nominal policy | Residual function

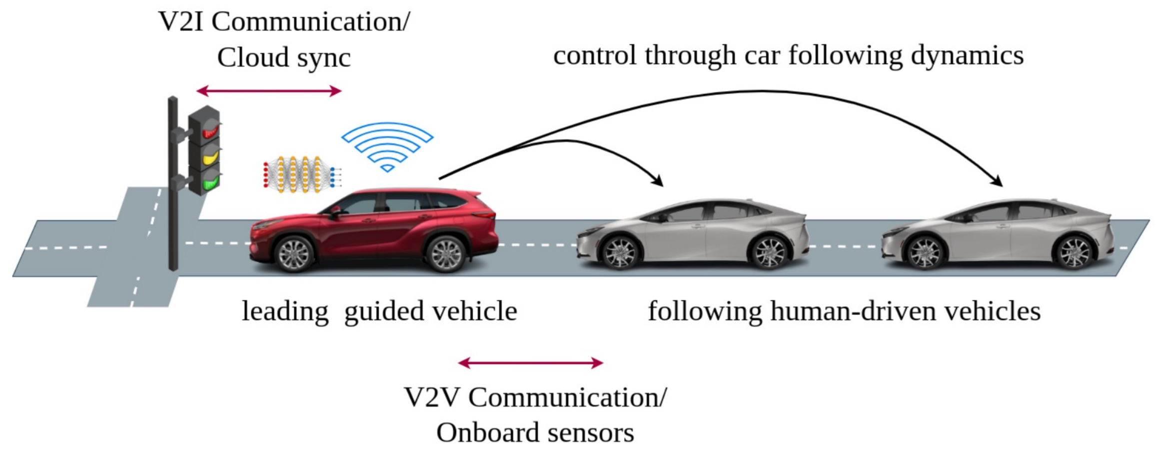
- The gradient of the π does not depend on the π_n. This enables flexibility with nominal policy choice.
- **Intuition**: If the nominal policy is nearly perfect, the residual term can be viewed as a corrective term. If not, nominal policy provide useful hints to guide the exploration of DRL training.

Mutl-residual Task Learning (MRTL)



Evaluations

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



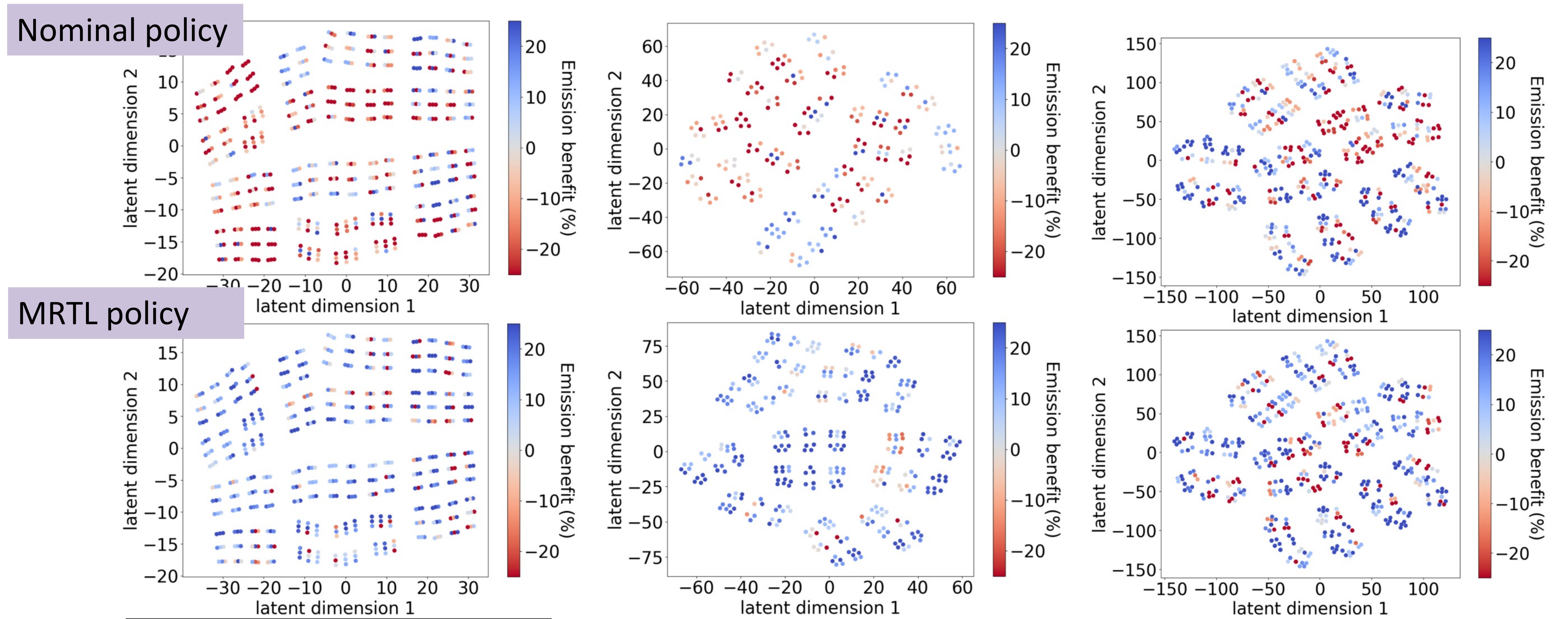
- **Goal:** reduce fleet-wide emissions while having less impact on travel time.
- **Setting:** 600 signalized intersections synthetically generated to match high-level real world intersection statistics. Both 20% and 100% eco-driving adoption levels were tested.
- **Nominal policy**: A model-based heuristic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

Performance comparison across 600 signalized intersections

Method	20% penetration			100% penetration		
	Emission ↓	Speed ↑	Throughput ↑	Emission ↓	Speed ↑	Throughput ↑
MTL	64.08%	-27.70%	-34.70%	95.86%	-30.87%	-68.11%
NP	13.13%	-21.11%	-30.07%	-25.09%	11.72%	-3.90%
MRTL (Ours)	**-13.95%**	**12.35%**	**7.95%**	**-29.09%**	**17.10%**	**5.72%**

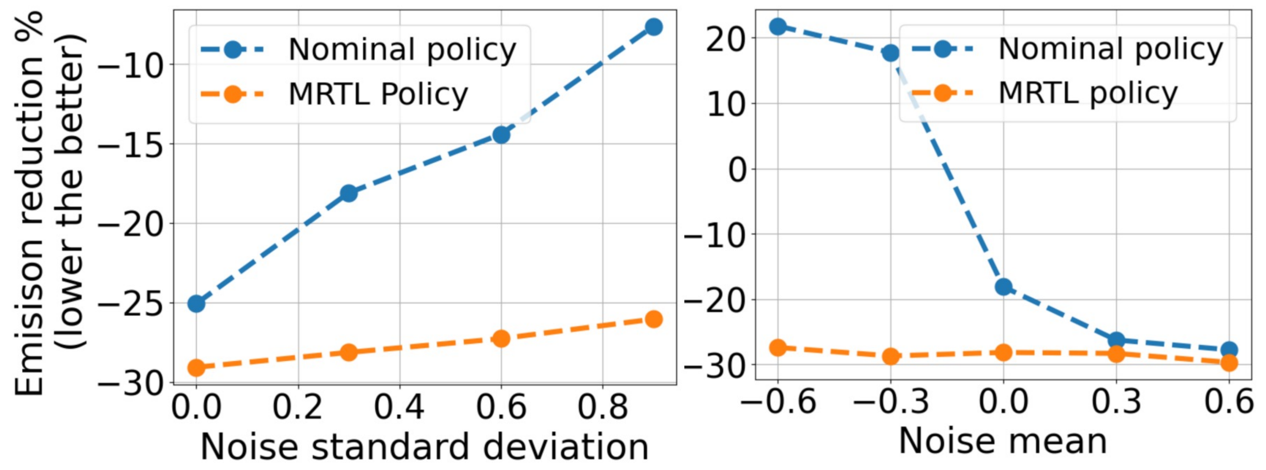
Visualization of t-SNE plots illustrating emission benefits in assessing the efficacy of MRTL policy in mitigating nominal policy limitations

- Each dot represents a signalized intersection approach and the colors denote the emission benefit levels.
- Here, 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)



Takeaway

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