Learning Eco-Driving Strategies at Signalized Intersections

Vindula Jayawardana† and Cathy Wu‡

† Department of Electrical Engineering and Computer Science, MIT
‡ Institute for Data, Systems, and Society and Department of Civil and Environmental Engineering, MIT

European Control Conference 2022
London UK
Transportation sector in the US contributes **29%** to the greenhouse gas emission (GHG) in which **77%** is due to land transportation.

**Challenge:** In arterial roads, traffic signals result in stop-and-go traffic waves producing acceleration, and idling events, increasing fuel consumption and emission levels.

Source: U.S. Environmental Protection Agency
Cities as Robots Sync.

Future cities are operation grounds for fleets of autonomous vehicles.

**Motivation:** Leverage autonomous vehicle fleets to reduce GHG levels and fuel consumptions of vehicles when approaching and leaving a signalized intersection.

**Objectives:**
- Reduce fuel consumption
- Reduce CO$_2$ emission
- Reduce the impact on travel time
Related Work

Previous work:

Model-based methods for control
  • Assumes a simplified model of the vehicle dynamics / inter-vehicle dynamics
  • Simplify the objective to reduce fuel consumption without the impact on travel time

Model-free reinforcement learning for control
  • Single agent control

Our work:

Model-free Reinforcement learning for multi-agent control.
  • Model-free
  • Accommodate rich and realistic objectives
  • Multi-agent control
Optimal Control Problem

Optimal control Problem:

\[
\min J = \sum_{i=1}^{n} \int_0^{T_i} F(a_i(t), v_i(t)) \, dt + T_i
\]

such that for every vehicle \( i \)

\[
a_i(t) = f_i(h_i(t), \dot{h}_i(t), v_i(t))
\]

Objective: Fuel and travel time reduction

Controlled acceleration according to car following dynamics

Travel distance requirement

\[
\int_0^{T_i} v_i(t) \, dt = d
\]

Limits on headway, velocity and acceleration

\[
h_{\text{min}} \leq h_i(t) \leq h_{\text{max}} \quad \forall t \in [0, T_i]
\]

\[
v_{\text{min}} \leq v_i(t) \leq v_{\text{max}} \quad \forall t \in [0, T_i]
\]

\[
a_{\text{min}} \leq a_i(t) \leq a_{\text{max}} \quad \forall t \in [0, T_i]
\]
Approach

Approach: Model-free Reinforcement learning for multi-agent control.

Maximize discounted total reward: $\max_{\theta} \sum_{t=1}^{T} \gamma^t r_t(st, at = \pi_{\theta}(st))$
Reinforcement Learning for Eco-Driving

• Partially Observable Markov Decision Process (POMDP) formulation of eco-driving problem
• Solve using policy gradient methods

Assumptions:

• Vehicle to Vehicle (V2V) communication
• Infrastructure to Vehicle (I2V) communication
  • To receive signal phase and timing (SPaT) information
Eco-Driving POMDP

Observations
- ego-vehicle velocity
- ego-vehicle position
- lead vehicle velocity
- lead vehicle position
- following vehicle velocity
- following vehicle position
- active traffic signal phase
- time to green

Actions
- Longitudinal acceleration

State Transitions
- microscopic simulation tools are used to sample \( s_{t+1} \sim p(s_t, a_t) \).

Rewards
- objective terms are competing (fuel & travel time)
- rate of change of the reward terms are different in different regions of the composite objective

\[
    r(s, a) = \begin{cases} 
    R_1 & \text{if any vehicle stops at the start of a lane.} \\
    R_2 & \text{if average fuel} \leq \delta \land \text{average stop count} = 0. \\
    R_3 & \text{if average fuel} \leq \delta \land \text{average stop count} > 0 \\
    R_4 & \text{otherwise}
    \end{cases}
\]

\[
    R_1 = \mu_1 \\
    R_2 = \mu_2 + \exp(v) \\
    R_3 = \mu_4 + \mu_5 \exp(v) + \mu_6 s \\
    R_4 = \mu_7 + \mu_8 \exp(\mu_9 f) + \mu_{10} \exp(v) + \mu_{11} s
\]
Training Agents

Training Setting:

• Centralized training and decentralized execution paradigm
• Trust Region Policy Optimization (TRPO) algorithm for training agents

TRPO update to policy,

$$\theta_{k+1} = \arg \max_\theta \mathcal{L}(\theta_k, \theta) \quad \text{s.t.} \quad D_{KL}(\theta \| \theta_k) \leq \delta$$

$$\mathcal{L}(\theta_k, \theta) : \text{surrogate advantage, a measure of how policy perform relative to the old policy using data from the old policy}$$

$$\mathcal{L}(\theta_k, \theta) = \mathbb{E}_{s,a \sim \pi_{\theta_k}} \left[ \frac{\pi_\theta(a \mid s)}{\pi_{\theta_k}(a \mid s)} A^\pi_{\theta_k}(s, a) \right]$$

$$D_{KL}(\theta \| \theta_k) : \text{average KL-divergence between policies across states visited by the old policy}$$

$$D_{KL}(\theta \| \theta_k) = \mathbb{E}_{s \sim \pi_{\theta_k}} \left[ D_{KL}(\pi_\theta(\cdot \mid s) \| \pi_{\theta_k}(\cdot \mid s)) \right]$$
Experimental Setup

Traffic Setting:

- Single intersection with only through-traffic and standard passenger cars
- VT-CPFM fuel consumption model and HBEFA-V3.1 CO$_2$ emission model
- A fixed time traffic signal control cycle with uniform vehicle arrivals
- SUMO microscopic traffic simulator
Results

Research Questions:

• **Q1:** How does the proposed control policy compare to naturalistic driving and model-based control baselines?

• **Q2:** How well does the proposed control policy generalize to environments unseen at training time?

Baselines:

• **V-IDM:** deterministic vanilla version of the IDM car-following model

• **N-IDM:** noise version of IDM (model variability in driving behaviors of humans)

• **M-IDM:** N-IDM model with varying parameters (represent a diverse mix of drivers with varying levels of aggressiveness)

• **Eco-CACC:** model-based trajectory optimization strategy introduced in [1]
Results

Q1: How does the proposed control policy compare to naturalistic driving and model-based control baselines?

Under 100% penetration of CAVs,
- 18% reduction in fuel
- 25% reduction in CO$_2$
- 20% increase in speed
Results

**Q2: How well does the proposed control policy generalize to environments unseen at training time?**

- Mixed traffic scenarios

---

**Mixed traffic:** Even 25% CAV penetration can bring at least 50% of the total fuel and emission reduction benefits.
Conclusion and Future Work

• Reinforcement learning can effectively be used to gain significant savings in fuel, emission while even improving travel speed.

• Generalizability of learn policies to out-of-distribution settings is successful

Future work:

• Consider multiple intersections in the optimization problem

• National level impact assessment as a climate change intervention
Thank you!