Mixed Autonomy Traffic
A Reinforcement Learning Perspective

Cathy Wu | Assistant Professor
LIDS, CEE, IDSS

IPAM Safe Operation of Connected and Autonomous Vehicle Fleets Workshop, 2020
Counterfactual reasoning with reinforcement learning?

• **Motivation**: Quantify impact of technology on societal systems
• Pace of change & complexity is increasing

**Environment**

**Agent**

**Reward**

Impacts on:
• Public safety & health
• Economic wellbeing
• Sustainability
• Resilience
• Equity & fairness

City

Environment + Agent = Reward

Policy evaluation → quantify impact

Policy learning → improve impact
Years 2020 to 2049: Mixed autonomy

Transportation in the US

Factors

+100%

-40%

(possible pathways)

Factors

- Platooning
- Eco-driving
- Congestion mitigation
- De-emphasized performance
- Improved crash avoidance
- Vehicle right-sizing
- Changed mobility services
- Infrastructure footprint

Changes in energy consumption due to vehicle automation (%)

Higher highway speeds
Increased features
Travel cost reduction
New user groups

Urban simulation

City

PARAMICS MICROSIMULATION

SUMO

PTV VISSIM

the mind of movement

aimsun.next
Axes of difficulty in mixed autonomy traffic

Scope (# vehicles, road network)

1-20 vehicles, 1 road

500K vehicles, 10K roads

Curse of dimensionality

Partial differential equations
Model predictive control
PID control
Feedback linearization

Part 1: RL?

Part 2: RL?

Part 3: Then what?

Jackson networks

Reservation systems
Polling systems

0%

Degree of autonomy

100%

Low uncertainty in rewards

High uncertainty in rewards & transitions

Low uncertainty in transitions
Axes of difficulty in mixed autonomy traffic

- **Scope** (# vehicles, road network)
  - 1-20 vehicles, 1 road
  - 500K vehicles, 10K roads

- **Curse of dimensionality**
  - Partial differential equations
  - Model predictive control
  - PID control
  - Feedback linearization

- **Degree of autonomy**
  - 0%
  - 100%

- **Part 1: RL?**
  - Low uncertainty in rewards
  - High uncertainty in rewards & transitions

- **Part 2: RL?**
  - Low uncertainty in transitions

- **Part 3: Then what?**
  - Jackson networks
  - Reservation systems
  - Polling systems

- **Feedback linearization**

- **Low uncertainty in rewards**

- **High uncertainty in rewards & transitions**

- **Low uncertainty in transitions**
Deep reinforcement learning (RL)

Decisions in transportation:
- Vehicle accelerations
- Tactical maneuvers
- Transit schedules
- Traffic lights
- Land use
- Parking
- Tolling

Goal: learn policy $\pi : S \rightarrow A$ to maximize reward

$\max_{\theta} \mathbb{E} \left[ \sum_{t=0}^{H} r(s_t, a_t) \Big| \pi_\theta \right]$

Global rewards
- Average velocity
- Energy consumption
- Travel time
- Safety, comfort

Cumulative rewards, returns
Policy parameters (deep neural network)

Agent

Environment

$\text{state } s_t$

$\text{reward } r_t$

$\text{action } a_t$

$s_{t+1}$

$r_{t+1}$

DQN (2015)
TRPO (2015)
AlphaGo (2016)
Single-lane: dynamical system equilibria

**Human driver model**
- Car-following model
  \[ \ddot{x}_i = f(v_i, v_{i-1}, x_i - x_{i-1}) \]
- Intelligent Driver Model [Treiber, et al. 2000]

Formation of traffic jams [Sugiyama, et al. 2008]

Average velocity vs traffic density

- Optimal (unstable)
- Traffic jams (stable)

Single-lane: state of the art policy

State of the art

- Hand-tuned model-based controller
- Proportional-integral (PI) controller with saturation [Stern, et al. 2018]

Average velocity vs traffic density

- Optimal (unstable)
- Traffic jams (stable)

Automated
Observed
Unobserved

Setting: 1 AV, 21 human

Experiment
• Goal: maximize average velocity
• Observation: relative vel and headway
• Action: acceleration
• Policy: multi-layer perceptron (MLP)
• Learning algorithm: policy gradient

Results
• 1 AV: +49% average velocity
• First near-optimal controller for single-lane
• Uniform flow at near-optimal velocity
• Generalizes to out-of-distribution densities

Single-lane: learned policy via deep RL

State of the art
Proportional-integral (PI) controller with saturation [Stern, et al. 2018]

Our results
• Near-optimal
• Generalizes to out-of-distribution traffic densities
• Memory not needed

Average velocity vs traffic density

- State of the art
- Traffic jams (stable)
- Optimal (unstable)
- This work

- Stop-and-go stable limit cycle
- Uniform flow unstable equilibrium
- PI with saturation controller
- Calibration density for PI controller
- MLP controller (ours)

Traffic LEGO blocks
Benchmarks for autonomy in transportation

Single-lane: +49%
Multi-lane: +30%
On/off-ramp: +142%
Intersection: +60%

Straight highway
Bottleneck
Grid network
Signalized intersection


5-10% AVs
Axes of difficulty in mixed autonomy traffic

Scope (# vehicles, road network)

1-20 vehicles, 1 road

500K vehicles, 10K roads

Curse of dimensionality

Partial differential equations
Model predictive control
PID control
Feedback linearization

Degree of autonomy

Part 1: RL?
Model predictive control
Partial differential equations

Part 2: RL?
Reservation systems
Polling systems

Part 3: Then what?
Jackson networks

Uncertainty in rewards
Low

Uncertainty in rewards & transitions
High

Uncertainty in transitions
Low
Challenge: combinatorial number of environments
A critical challenge to scaling deep reinforcement learning

Elements:
- Road network
- Roadway signage
- Rules of the road
- Types of vehicles
- Speed limits
- Traffic lights
- # Lanes
- Driver behavior
...
Transfer learning across networks

- **Research question**: Can knowledge be transferred across traffic scenarios?
- **Transfer learning**: The use of knowledge gained from a **source task** to bias the learning process on a **target task** towards a set of good hypotheses.
- **Zero-shot transfer**: Extreme setting where no learning is done on the target task. Out-of-distribution generalization.

---

Zero-shot transfer
Circular track $\rightarrow$ More/less dense circular track

Our results
- Near-optimal
- Zero-shot transfer: Generalizes to out-of-distribution traffic densities

Average velocity vs traffic density
- Stop-and-go stable limit cycle
- Uniform flow unstable equilibrium
- PI with saturation controller
- Calibration density for PI controller
- MLP controller (ours)

Traffic jams (stable)
- Optimal (unstable)

Transfer learning
Circular roads $\rightarrow$ Straight roads

- Successful direct transfer!
- Closed $\rightarrow$ open networks

Initial performance boost

Training Performance in the Presence and Absence of an Initial Ring Road Policy

Single-lane 5% AV

Zero-shot transfer
One bottleneck → Many different bottlenecks

**Setting: No AVs, 100% IDM**

Phenomenon: capacity drop

1480 veh/hr

**Setting: 10% AVs, 90% IDM**

1800 veh/hr

**Results:**
- 22% improvement
- Avoids capacity drop
- Learned policy transfers to different inflow rates, number of lanes, and percent of automated vehicles

**Successful transfer:**

Network: 8 > 4 > 2 Bottleneck

Network: 8 > 4 > 2 > 1 Bottleneck

*Capacity drop experiment is a variation of: Vinitsky, Parvate, Kreidieh, Wu, Bayen. IEEE ITSC, 2018*
Zero-shot transfer
Bottleneck $\rightarrow$ Grid?!

26% improvement over human baseline (IDM)

No fine-tuning!
Axes of difficulty in mixed autonomy traffic

- **Degree of autonomy**
  - 0%
  - 100%

- **Curse of dimensionality**

- **Scope**
  - 1-20 vehicles, 1 road
  - 500K vehicles, 10K roads

- **Part 1: RL?**
  - Partial differential equations
  - Model predictive control
  - PID control
  - Feedback linearization

- **Part 2: RL?**
  - Low uncertainty in transitions
  - Low uncertainty in rewards

- **Part 3: Then what?**
  - Jackson networks
  - Reservation systems
  - Polling systems

- **Low uncertainty in rewards**
- **High uncertainty in rewards & transitions**
- **Low uncertainty in transitions**
High-dimensional control

Key challenge in cooperative multi-agent systems: curse of dimensionality

Variance reduction in policy gradient methods
Key idea: factorizing stochastic policies → opportunity for improved control variates
Axes of difficulty in mixed autonomy traffic

Scope
(# vehicles, road network)

1-20 vehicles, 1 road

500K vehicles, 10K roads

Curse of dimensionality

Partial differential equations
Model predictive control
PID control
Feedback linearization

Part 1: RL?

Degree of autonomy

0%

Part 2: RL?

High uncertainty in rewards & transitions

Low uncertainty in rewards

Part 3: Then what?

Low uncertainty in transitions

Jackson networks

Reservation systems
Polling systems

Low uncertainty in rewards
Pathways toward reality

Physical tests & deployment

Insights for urban planning & industry
Automatic traffic signal optimization
Not just for futuristic automated vehicles!

• Motivation:
  – Improve travel times & congestion
  – Mitigate air pollution
  – Improve coordination across city boundaries

• Same approach
• Potential for near-term benefits

Current controller
RL-based controller

30% speed improvement
29% queue reduction
60% queue reduction (3x3)
Mixed Autonomy Traffic: A Reinforcement Learning Perspective

- Deep reinforcement learning provides a pathway for understanding the impacts of mixed autonomy in urban systems.
- Transfer learning across networks can improve sample efficiency of RL to enable the analysis of larger and more complex traffic scenarios.
- Numerous opportunities to scale to high dimensional control.
- There’s everything left to do.