Multi-Agent Generative Adversarial Imitation Learning for Driving Simulation

Raunak Pushpak Bhattacharyya (raunakbh@stanford.edu)

Stanford University

Motivation

- Validating the safety of autonomous vehicles in the real world is costly, dangerous and time consuming
- Model human driven vehicles for realistic simulation testing of autonomous vehicles
- Imitation learning approaches have worked well for driving a single car
- This project extends imitation learning to driving multiple cars

Dataset & Features

Feature	Description
LIDAR Range and Range Rate	20 artificial LIDAR beams output in regular polar intervals, providing the relative position and velocity of intercepted objects.
Ego Vehicle	Lane-relative velocity, heading, offset. Vehicle length and width. Lane curvature, distance to left and right lane makers and road edges.
Temporal	Longitudinal and lateral acceleration, local and global turn and angular rate, timegap, and time-to-collision.
Indicators	Collision occurring, ego vehicle out-of-lane, and negative velocity.
Leading Vehicle	Relative distance, velocity, and absolute acceleration of vehicle in front of fore vehicle, if it exists.

Algorithm: Parameter Sharing GAIL

Algorithm 1 PS-GAIL

Input: Expert trajectories $\tau_E \sim \pi_E$, Shared policy parameters Θ_0, Discriminator parameters ψ_0, Trust region size Δ_{KL}

for $k \leftarrow 0, 1, \ldots$ **do**

  Rollout trajectories for all agents $\vec{\tau} \sim \pi_{\theta_k}$

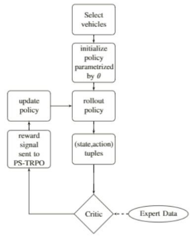
  Score $\vec{\tau}$ with critic, generating reward $\tilde{r}(s_t, a_t; \psi_k)$

  Batch trajectories obtained from all the agents

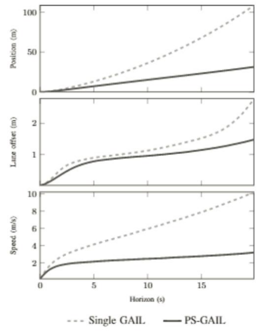
  Take a TRPO step to find $\pi_{\theta_{k+1}}$

  Update the critic parameters ψ

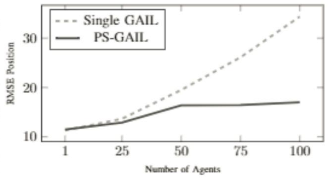
This algorithm extends GAIL to the multi-agent setting using parameter sharing



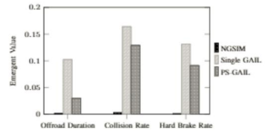
Results



Better human driving imitation



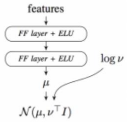
Better scalability



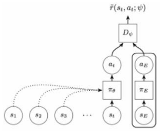
Better emergent metrics

Policy and Critic: Deep Neural Nets

- Policy representation
  - Non-linearity
  - High dimensionality
  - Stochasticity
  - 64 Gated Recurrent Units



- Critic representation
  - Wasserstein GAN with gradient penalty
  - Feed forward network consisting of (128, 128, 64) ReLU units



Future work

Improving model performance by (i) reward augmentation, (ii) applying learning algorithms that encourage more diverse behavior, and (iii) using a recurrent critic in order to account for partial observability.

Acknowledgments and References

The author acknowledges the guidance and mentorship of Zachary Barnes and Mykel Kochenderfer.

[1] J. Ho and S. Ermon, "Generative adversarial imitation learning", NIPS, 2016.
[2] J. Gupta, M. Egorov, and M. Kochenderfer, "Cooperative multiagent control using deep reinforcement learning", Adaptive Learning Agents Workshop, 2017.
[3] J. Morton, T. Wheeler, and M. Kochenderfer, "Analysis of recurrent neural networks for probabilistic modeling of driving behavior", IEEE Transactions on Intelligent Transportation Systems, 2017.
[4] M. Arjovsky, S. Chintala, and L. Bouttou, "Wasserstein generative adversarial networks", ICML, 2017.