Finding the Value of Aggression in Autonomous Driving

Presenter: William D. Brannon (wbrannon@stanford.edu)
Department of Aeronautics and Astronautics

CS 230
Deep Learning

Motivation

With the advent of autonomous vehicles becoming more prevalent, it becomes interesting to consider how they should react around humans. Previously deployed autonomous vehicles have demonstrated overly passive behavior and do not display "aggressive" behavior as humans often do. Given that this aggressive behavior may be governed by trends of high acceleration and quick lane changes, this project seeks to find whether this may yield better-performing autonomous vehicles.

Hyperparameters

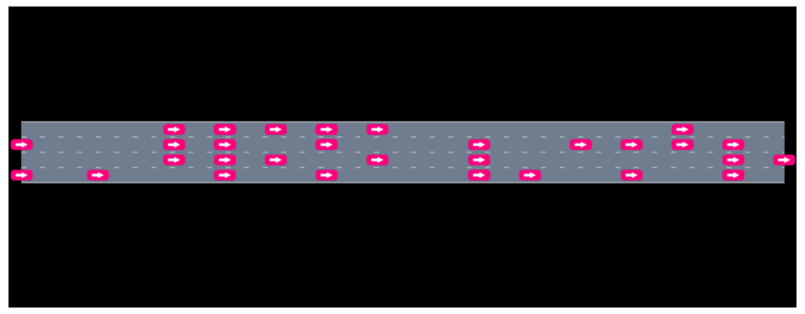
The following hyperparameters were converged upon during expermintation

- Learning rate α: 0.0005
- Reward for colliding: -500
- Reward for reaching goal lane: 1000
- Target update frequency: 10000 steps
- Training set Exploration vs. Exploitation fraction: 0.5
- Amount of layers:

Deep Q-Learning Approach

The approach to this problem involves approximating the state action value across the state space via a deep Q network; given that the finalized state action value function should solve the Bellman equation, i.e. $Q^*(s,a) = r + \gamma \max_{a'}(Q(s',a'))$, it may be approximated via deep neural networks. Using the Stanford Intelligent Systems Laboratory's driving simulator, AutomotiveDrivingModels.jl, a state and action space may be associated to a high-level driving scene.

For this project, the problem was defined as a Markov Decision Process, defined by the tuple (S, A, P, R), defining the state space, action space, transition model, and reward function, respectively. The action space was defined by a discretized set of actions revolving around accelerating in the lateral and longitudinal directions.

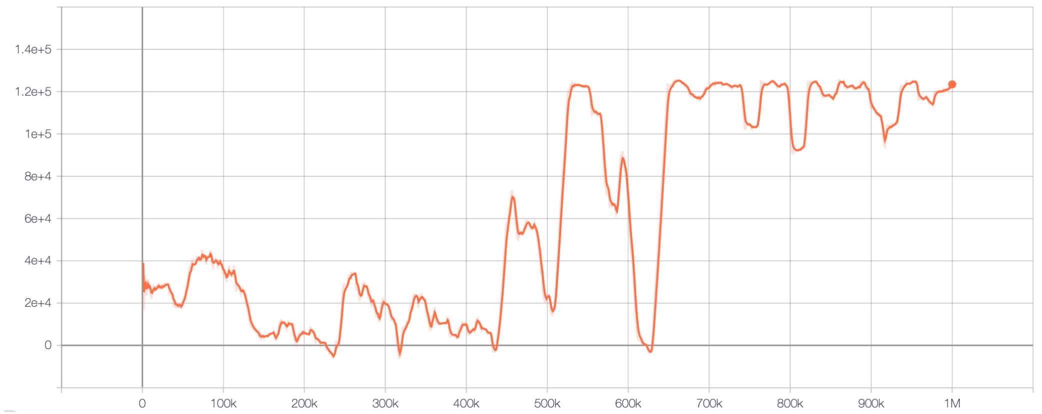


Screen capture of AutomotiveDrivingModels.jl simulation

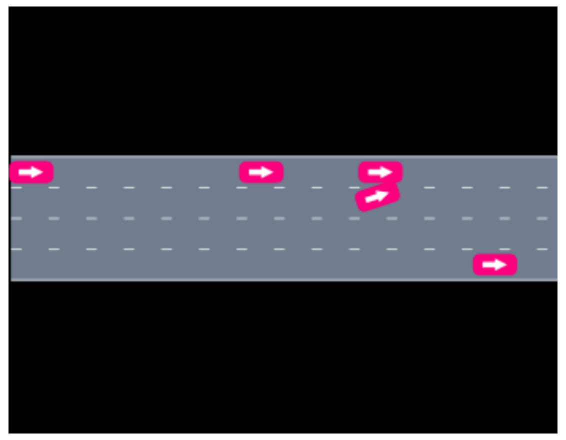
The deep network input was set by the x and y coordinates and velocity of all vehicles on the road. These inputs were normalized, and placed into a 9-layer neural network, with an output of 9 state action values. The reward function is defined by collisions, timing out, high heading angles, reaching the goal lane, and the Q network uses an epsilon greedy strategy to sample new actions.

Experimentation/Results

Though the reward curve in the exploration phase shows its way to convergence, it should be noted that the aggressive driver underperforms. As the simulation got denser and denser with human-driven vehicles, the aggressive driver begins to perform worse than the passive driver. Further analysis may be required to identify the reasoning as to this, but potential reasons could involve a need for further reward engineering or a more selective action space.



Reward curve over 1 million samples



Example of ego vehicle collision on way to goal lane

Discussion

It was found that the ego agent displaying aggressive driving behaviors underperformed when in comparison to the more passive driver.

Agent	Amount of Human Vehicles	Collisions	Goal
Passive	5	418	3581
Passive	20	1640	2360
Passive	40	3800	1200
Aggressive	5	355	3284
Aggressive	20	2013	26
Aggressive	40	2430	0

This makes intuitive sense, as it may seem more safe to drive more aggressively in a less crowded environment. However, with a more normalized reward function and more selective action space, it is possible for the ego agent to exemplify higher-performing behavior.

References

[1] M. Treiber, A. Hennecke, and D. Helbing, "Congested traffic states in empirical observations and microscopic simulations," *Physical Review E*, vol. 62, no. 2, p. 1805, 2000.

[2] M. Bouton, A. Cosgun, and M.J. Kochenderfer, "Belief State Planning for Autonomously Navigating Urban Intersections," 2017.

[3] Z.N. Sunberg, C.J. Ho, and M.J. Kochenderfer, "The value of inferring the internal state of traffic participants for autonomous freeway driving," in *American Control Conference (ACC)*, 2017.

[4] M. Bouton, A. Nakhaei, K. Fujimara, and M.J. Kochenderfer, "Cooperation-Aware Reinforcement Learning for Merging in Dense Traffic," 2019

[5] M.J. Kochenderfer, *Decision Making Under Uncertainty: Theory and Application*. MIT Press, 2015.

[6] R. Bellman, *Dynamic Programming*. Princeton University Press, 1957.

[7] H. Chae et. al, "Autonomous Braking System via Deep Reinforcement Learning," in *IEEE International Conference on Intelligent Transporation Systems (ITSC*, 2018.

[8] T. Tram et. al, "Learning negotiating behavior between cars in intersections using deep q-learning". Princeton University Press, 1957.