Learning Predictive Models of Human Driving Behavior

Peggy (Yuchun) Wang, Khalid Ahmad, Hashem Elezabi

Motivation

• Autonomous cars have traditionally practiced highly defensive driving when interacting with human cars, treating other cars as moving obstacles.
• This leads to driving behavior that is **opaque** and **unpredictable** to humans, since humans are used to human-like behavior.
• Recent research by Sadigh et. al. presents the key insight that an autonomous car's behavior will affect what other cars do in response, creating an **opportunity for coordination**.
• If the autonomous car has a good model for how the human car behaves, it can take into account the effect it will have on the human driver. **How can we obtain such a predictive model?**

Problem

• **Our problem:** learn a predictive human reward function using inverse reinforcement learning (IRL).
• Given human driving data in the form of trajectories taken, heading, speed, steer, and acceleration for each time step, we want to derive a human reward function that can *explain* the actions the human took.
  • By *explaining*, we mean that maximizing that reward function would produce similar actions.
• Ultimately, the goal is for the autonomous agent to use this human reward function as an approximation of what the human is expected to do.

Approach

• Our reward function is a linear combination of five features: 1) the position of the car in its lane, 2) position of the car within boundaries of the road, 3) which lane the car is in, 4) speed of car, and 5) proximity to other cars. $R(s_t) = \sum_{i=1}^{N} \alpha_i \phi_i(s_t)$

• **Baseline:** We start with a set of simple qualitatively chosen weights, which gives our baseline human reward function.
• **Data and setup**: We use a Python driving simulator developed by Professor Sadigh et. al. We manually drive a car in the simulator and after a period in which we interact with computer-controlled cars, we save the driving data from the entire experiment to a file.
• The state s_t is a list of 2-tuples (x_t, u_t) for each agent, where x_t is the physical state (position, velocity, etc.) and u_t is the control input (acceleration, steer) at time t.

• **Inverse reinforcement learning (IRL):** is a form of imitation learning. In IRL, we derive a reward function from the observed actions that an expert (a human driver in our case) has taken. **IRL approach:**
  • Apply the *principle of maximum entropy* to define a probability distribution over the human actions such that actions that have a higher reward are more likely, where the reward is defined by our parameterized reward function: $P(h \mid x_0, \alpha) \propto \exp(R(x_0, r, h))$

• We then optimize by choosing weights that make the desired human demonstrations most likely: $\max_\alpha P(h \mid x_0, \alpha)$
• We designed different scenarios with multiple cars, and tested the weights obtained from IRL by comparing the robot's driving behavior with the human's driving behavior in the same scenario.

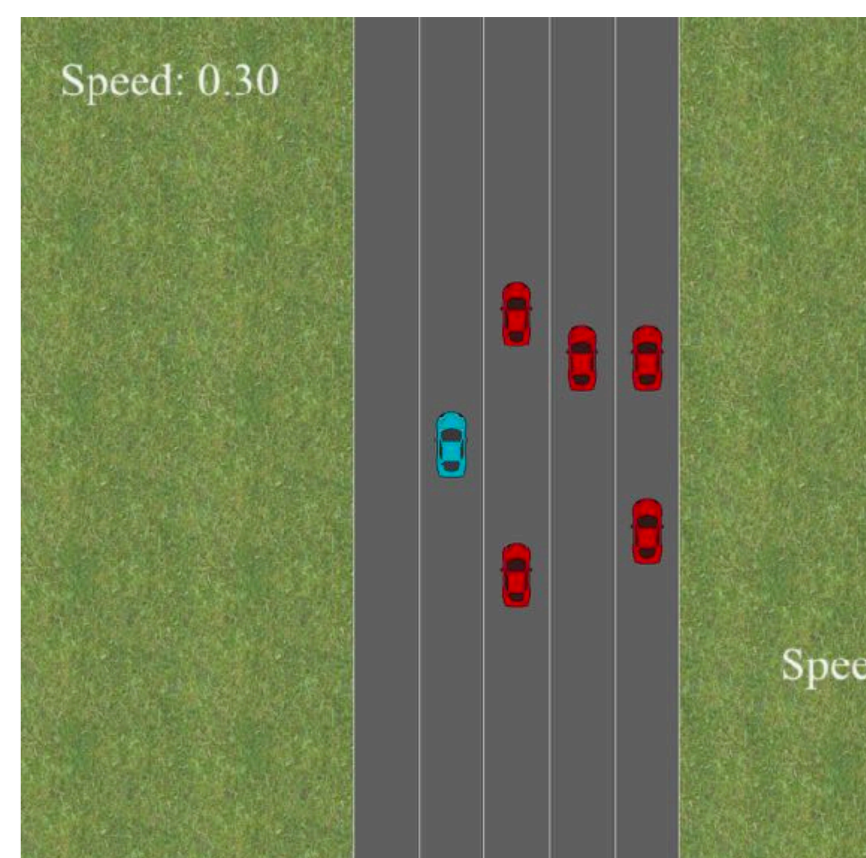


Figure 1: Human-controlled car (blue) in starting position

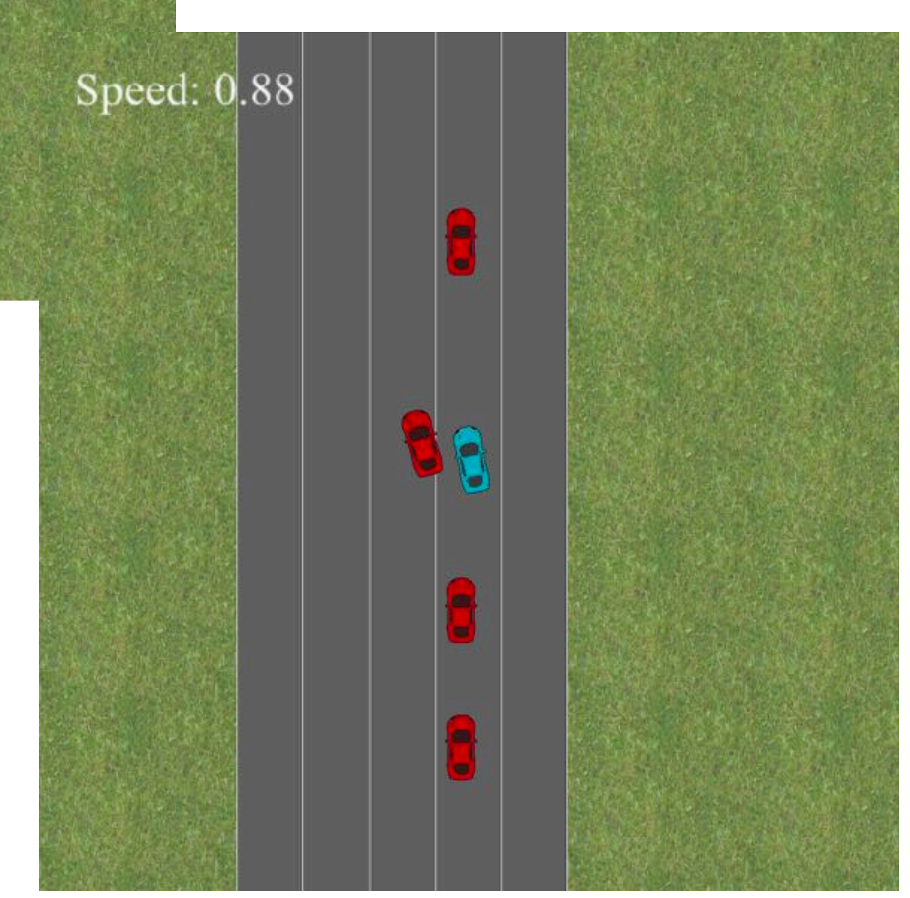


Figure 2: Human-controlled car (blue) swerves to a lane on the left, robot car (red) swerves left to avoid collision

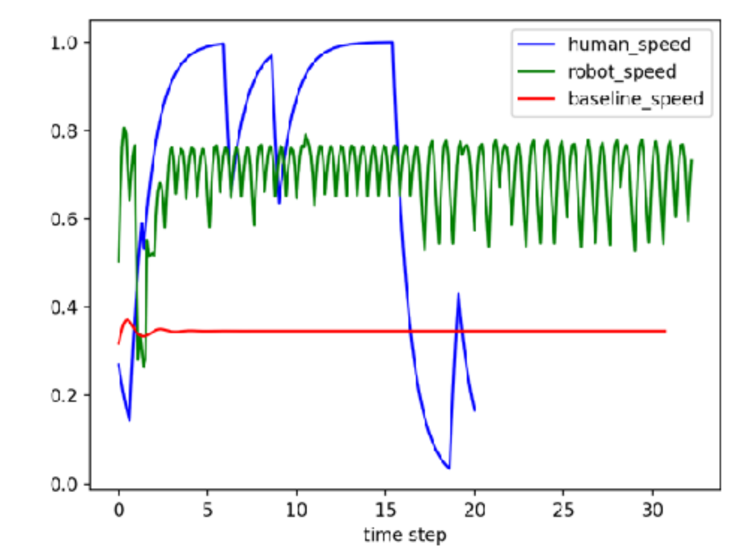
Challenges

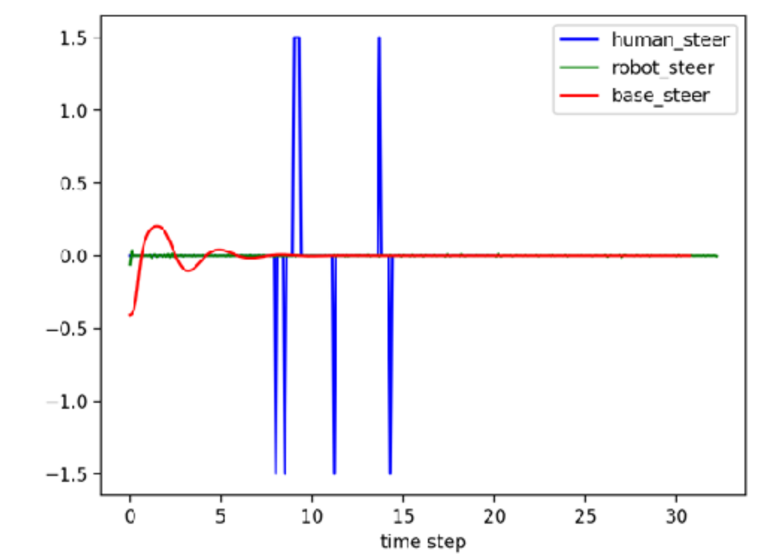
• Learning how IRL worked and understanding the baseline implementation
• Creating representative features and changing baseline features
• Creating scenarios to validate the updated weights we got from IRL

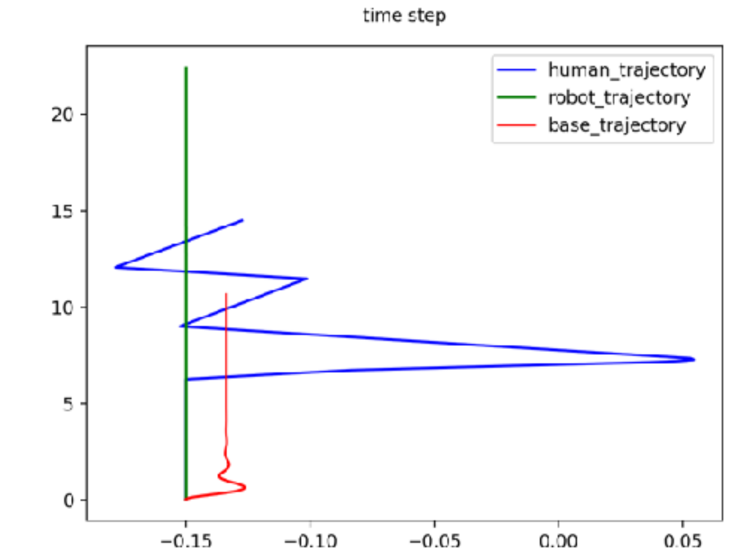
Results & Analysis

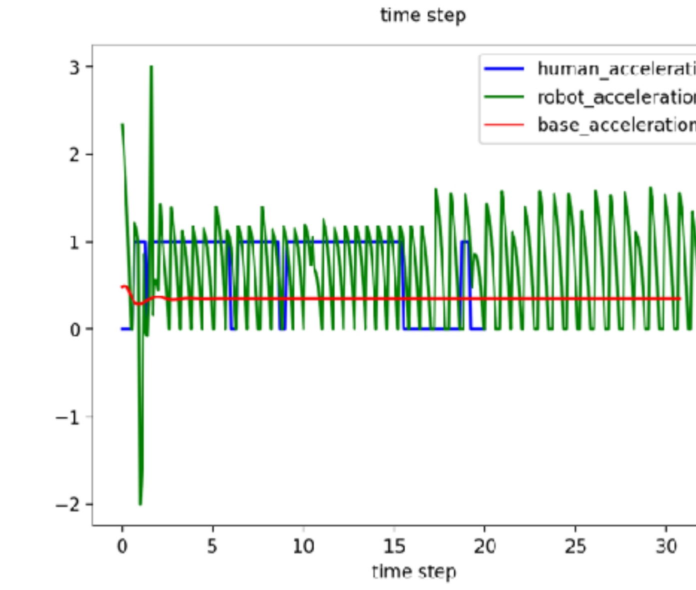
• We plot the trajectory, speed, steer, acceleration, and heading for 1) the human driver (our manual driving), 2) our baseline agent, and 3) our new agent optimizing our learned reward function.

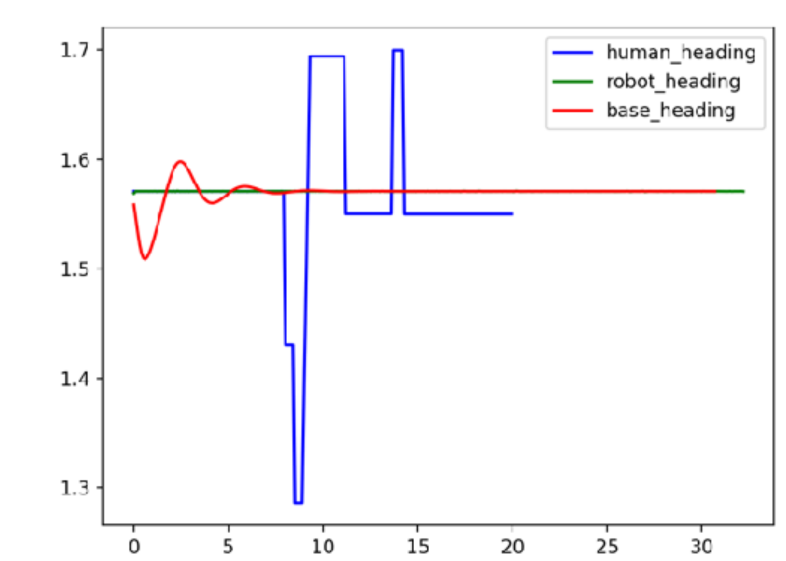
• Our learned model yields behavior different from the baseline agent's, but doesn't quite match the human's agent.
• With some values like speed and acceleration, our learned agent seemed to take patterns in the human driver's behavior to the extreme.











• A qualitative evaluation of the learned behavior shows human-like characteristics and stable driving.
• We suspect that our linear reward function was not enough to capture the complexity of human driving. This is also due to the arbitrariness inherent in our human driving experiments.
• **Future work** includes trying to add new features that might help capture more properties of the human's driving behavior, and using less arbitrary, more controlled human driving behavior to measure against.

Acknowledgements

• Thanks to Professor Dorsa Sadigh for her help and mentorship, and for providing the basic simulation code, world models, and baseline IRL algorithm.
• Thanks to Eley Ng for her help and mentorship on this project.
• Thanks to Daniel Byrnes for his help explaining parts of the code.