CS 236 - Imitating driver's behavior in urban environment

Category : Imitation Learning

Malik Boudiaf
Department of Aeronautics
and Astronautics
Stanford University
Stanford, USA
Email: mboudiaf@stanford.edu
SUNet ID : mboudiaf

Ianis Bougdal-Lambert
Department of Aeronautics
and Astronautics
Stanford University
Stanford, USA
Email: ianisbl@stanford.edu
SUNet ID : ianisbl

I. Introduction

Accurate human driver models are critical for realistic simulation of driving scenarios, and have the potential to significantly advance research in automotive safety. Older approaches to autonomous driving tried to imitate human drivers' behaviors by describing them as a set of rules to follow. These early rule-based approaches tried to fit parametric model using strong assumptions about road conditions and human's behavior, and failed to generalize well to new driving scenarios.

Recent approaches are more data driven. Imitation learning (IL) uses expert data provided through human demonstrations and try to learn a policy that would be able to reproduce human's behavior in all possible states. Within IL framework, algorithms typically fall within two categories : behavioral cloning (BC), Reinforcement learning (RL).

BC uses a classic supervised learning framework to directly map observed states to actions by fitting the expert data. BC approach suffer, just like early methods, from a limited capacity of generalization. The RL approach, instead, tries to recover a policy as close as possible to an expert policy derived from data. Among RL approaches, the most referenced papers in literature use Inverse Reinforcement Learning (IRL), which assumes that the expert follows an optimal policy with respect to an unknown reward function that we try to recover. Once this reward function is recovered, one may use classic RL to find the target policy. But recovering this reward function can be very computationally expensive, and recent efforts have tried to bypass this step and directly optimize the target policy. Generative Adversarial Imitation Learning (GAIL) [3] is one of these approach.

Most of the latter algorithms have been implemented and tested in highway driving situations in [2]. In this project, we propose to extend the work done in [2] to a more urban scenario. This project will be common to the CS 230 project for both of us.

II. Problem Statement

We regard driving as a sequential decision making task in which the driver obeys a stochastic policy $\pi(a|s)$ mapping observed road conditions s to a distribution over driving actions a. We try to find a policy within a given class of policies $\pi_\theta(a|s)$ parametrized by θ, that best imitates human driving behavior.

III. Approach

A. Dataset

The dataset we will use comes from the NGSIM database that supports traffic modeling simulations [4]. More precisely, the dataset we will use is called *Lankershim Boulevard Dataset*. These data were collected using five video cameras mounted on the roof of a 36-story building located adjacent to the U.S. Highway 101 and Lankershim Boulevard interchange in the Universal City neighborhood, in Los Angeles. NG-VIDEO, a customized software application developed for the NGSIM program, transcribed the vehicle trajectory data from the video. This vehicle trajectory data provided the precise location of each vehicle within the study area every one-tenth of a second, resulting in detailed lane positions and locations relative to other vehicles. A total of 30 minutes of data are available in the full dataset, which are segmented into two 15-minute periods.

We will also use the *NGSIM Env* [5], developed by the Stanford Intelligent Systems Laboratory (SISL) to carry out our experiments. This environment, specifically designed to provide easy design and testing of imitation algorithms using NGSIM datasets, should enable us to focus on the interesting parts of the problem by saving us the time required to build a proper simulation environment. However, we still expect to spend some time adapting/debugging the public code to fit our urban dataset, as this environment has only been tested on highways datasets.

B. Representing a policy

Most recent papers in the literature represent a policy as a neural network that takes as input the current state of the vehicle - and a summary of previous states in recurrent representations - and outputs an action, or at least a distribution from which to sample an action. In [2], two main types of neural networks (NN) are explored : feedforward and recurrent ones, with a common architecture of five feedforward layers that decrease in size from 256 to 32 neurons. The recurrent NN (RNN) adds a layer of 32 gated recurrent unit (GRU) on top of that, while the FFNN adds another feedforward layer.

We would first like to follow the idea of a recurrent network, that sounds particularly appealing in driving context, and experiment on different architectures. Ideally, we hope to find more simple and computationally efficient architectures that still achieve similar performances.

Finally, we would like to explore alternative architectures like representing a policy by a CNN taking as input a 2D array of the concatenated N last states. This approach is usually used in time series classification but rarely used in the driving context.

C. Training a policy

Two main types of training have currently been identified in imitation learning : behavioral cloning and reinforcement learning.

In the Behavioral Cloning (BC), we basically try to find the policy $\pi_\theta(a|s)$ that best matches human behavior by solving a regression problem. This approach seems to exhibit a very limited generalization capacity and faces the well identified problem of cascading errors [6]. Hence, this approach doesn't seem very appropriate in urban situations where the dataset will never be able to cover the full space of very erratic and nuanced states and actions.

Reinforcement learning (RL) approaches, however, generally do better at dealing with the cascading error problem, by maximizing a global expected return over the whole trajectory, rather than taking local actions to solve immediate problems, but may lead to bigger ones down the road. More specifically, in this approach, we assume the driver follows an expect policy π_E whose actions maximize an expected global return :

$$R(\pi, r) = \mathbb{E}[\sum_{t=0}^{T} \gamma^t r(s_t, a_t)] \quad (1)$$

Where $r(s_t, a_t)$ is a reward function which we know nothing about.

The CS 236 part of our project will mainly investigate the RL approach, and more specifically the Generative Adversarial Imitation Learning (GAIL) framework.

IV. Results

Several metrics have been proposed in the literature. For now, we decided to use two of them to evaluate the relative performance of each model :

- *Root-Weighted Square Error (RWSE)* : Captures the deviation of a trajectory generated by our policy from the true trajectory – as given by data – in the same conditions. This metric should yield plots on the RWSE of position, speed, and center lane offset, as a function of time.
- *KL divergence* : The KL divergence will be used here to measure the distance between the real-world distribution and the generated distribution over several quantities like speed or acceleration. The exact process to compute this quantity is still not clear to us.

References

[1] Mykel J. Kochenderfer
Decision Making Under Uncertainty Theory and Application
MIT Lincoln Laboratory Series , 2015.

[2] A. Kuefler, J. Morton, T. Wheeler, and M. Kochenderfer
Imitating Driver Behavior with Generative Adversarial Networks
2017 IEEE Intelligent Vehicles Symposium (IV)

[3] J. Ho and S. Ermon
Generative adversarial imitation learning
Advances in Neural Information Processing Systems (NIPS), 2016

[4] *Next Generation Simulation (NGSIM)*, US Department of Transportation, Federal Highway Administration
https://ops.fhwa.dot.gov/trafficanalysistools/ngsim.htm

[5] *NGSIM Env*, Stanford Intelligent Systems Laboratory
https://github.com/sisl/ngsim$_e nv$

[6] S. Ross and J. Bagnell
Efficient reductions for imitation learning
International Conference on Artificial Intelligence and Statistics, 2010.