Investigating the use of unseen simulators for evaluating self-driving car models

https://github.com/hcollins345/PP-evaluating-ai-driving-simulators

Harry

Supervised by:

Seán

Abstract

This project aimed to investigate how self-driving machine learning models perform in environments built in simulators unseen in training. The simulators, CARLA and DeepDrive were hosted using google cloud services.

Introduction

This personal pursuit attempts a preliminary analysis of how self-driving machine learning models cope with the differences between different simulators and possibly make inferences on the ability of the tested models to handle the differences between simulators and reality.

Ultimately the goals is to help to create better and safer self-driving cars. The hope is that if a model can generalise well to an unseen simulator, it can generalise well to reality. Similarly, cross simulator testing could help to better understand if or how simulator design choices affect model creation and performance.

Methodology

The learning goals of the project were;

1. Learn and gain experience in the tensorflow (tf) framework for the use and performance analysis of machine learning models
2. Gain experience in the use of popular autonomous driving simulators such as CARLA and deepdrive
3. Learn how to integrate tensorflow (tf) with simulator software and how to transfer models between simulators with different interfaces and features

NOTE: The goals were not concerned with gaining experience in the design and/or training of machine learning models.

These goals were to be achieved through the following activities;

1. Follow introduction tutorial to tensorflow / tensorboard
2. Set up simulator on cloud computing platform
3. Run a pre-trained model through tensorflow
4. Repeat for other simulators
5. Test models on other simulators
   - Mapping the new inputs to the models
   - Running tests
   - Evaluating and comparing their performance over the different fitness criteria
6. Write report on the findings

Topic Background

Training autonomous vehicles can be a very expensive and challenge endeavour. The first challenge is collecting the thousands or millions of hours of driving data needed to train the autonomous driver. The second is actually verifying the safety of these autonomous drivers. The most obvious way to assess safety is to test drive autonomous vehicles in real traffic, observe their performance and statistically compare the results to humans [1]. However as shown in [1], over 5 billion driven miles would be necessary in order to verify that an autonomous car was statistically 20% safer than a human; To have only an 80% chance of correctly verifying that AI driver is safer (known in statistics as power) requires over 11 billion miles; 100 vehicles driving 24 hours a day averaging

25 mph would take over 500 years to log that much time [1].

Simulators are far cheaper and convenient for both generating training data and verifying the performance of AI driving models. Servers can drive millions of miles a day at a fraction of the per distance cost. Through simulators it is possible to design and generate emergency scenarios to verify the system makes the correct responses [2]; These types of scenarios are infrequent in real driving data. It is critical to the feasibility of creating autonomous vehicles to have the ability to verify the performance of models and autonomous systems to some degree in the simulation phases of design in order to reduce both risk and costs.

However simulators are only a simplification or representation of reality. The Models have to be able to generalise in order to handle the differences between the simulator and reality.

Method

In order to learn tensorflow I followed the tutorials/series on tensorflow and autonomous vehicles from the website pythonprogramming

I then set up two self-driving simulation environments, CARLA [3] and DeepDrive [4] on google's cloud computing platform. There were two major challenges associated with this step. Firstly, the simulators were designed to be run locally with little or no support for running on cloud platforms. Secondly, nvidia tesla GPU's which make up most of the GPUs in computing servers do not have display ports. These challenges meant GUI's were mostly unavailable and therefore several workarounds had to be used to verify the results. Steps to replicate the installations can be found on the project's github.

Figure 1: Carla simulator agent in rain with pedestrians and traffic

Figure 2: Agent stopped at traffic lights

High performing Conditional Imitation Learning tensorflow models (CIL) for both DeepDrive and CARLA CIL model were then benchmarked on their respective simulators.

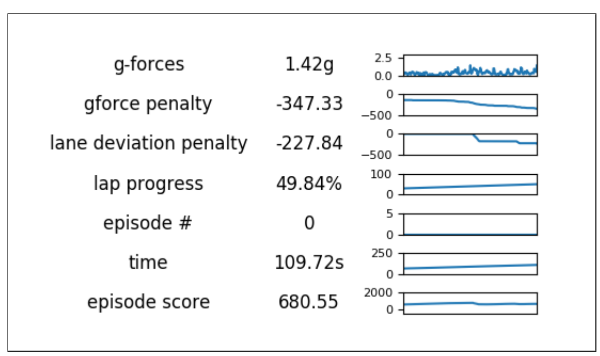

Figure 3: Monitoring status of agent in DeepDrive sim

I then began migrating the trained agents from one simulator to other. This involved separating the tf agents from the custom learning environments that allow them to interface with their respective simulators. This was challenging as the code used to write the agent assumed the presence of the different learning environments which were implemented differently for both simulators.

While I have managed to normalise the sensory inputs to the agent. I have struggled to replicate the environmental structure that is passed to the agent which governs how the agents transforms the output of its neural network to controls for the simulator.

** Explain how this is ongoing **

Results

Benchmarking

The benchmarking for the DeepDrive simulator can be seen below, 34 runs were conducted with the final score for each run consisting of the distance of the route completed minus penalties for excessive lane deviation and g-forces.

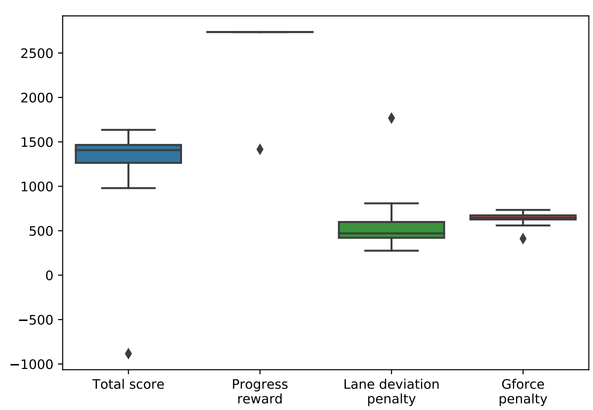

Figure 4: Benchmarking results for Deepdrive agent in DeepDrive simulator

Results of learning goals

While the activities of the Personal Pursuit are still ongoing. I believe the learning goals have been fulfilled.

Goal 1: Learn and gain experience in the tensorflow framework for the use and performance analysis of machine learning models

This goal was achieved through the tutorials and the practical implementation of actual tensorflow models in the simulators.

An interesting caveat of the goal relates to the use of the tensorflow framework for performance analysis of machine learning models. In practice when training, the tensorflow model is being evaluated continuously according to a very specific fitness function, in this case how well it replicates the actions of the test data (from human drivers or other path following algorithms). The final performance analysis of the model however, does not include evaluation functions from the tensorflow framework but from experiment suites and data (such as lane deviation) from the actual simulator itself.

Goal 2: Gain experience in the use of popular autonomous driving simulators such as CARLA and deepdrive.

While setting up CARLA and DeepDrive on a cloud compute provider was difficult. The result was the ability to leverage powerful computing hardware such as Nvidia Tesla p100s to run the simulation and machine learning software.

Similarly I learnt how to configure the simulations through config files and the command line. I understand the code structure that governs the interaction between the simulator and agents and can write code to enable interaction with a running simulation as an agent.

Goal 3: Learn how to integrate tensorflow with simulator software and how to transfer models between simulators with different interfaces and features.

As the project progressed it became clear that this goal is actually much more complicated than initially expected.

It can be broken down into clearer sub-goals. The first sub-goals is learning how to integrate tensorflow with the simulator software. In the project I used and benchmarked different tensorflow models to control agents acting in the simulator through the interfacing environments.

The second sub-goal is learning to transfer models between simulators. I detached the DeepDrive CIL model from the DeepDrive environment and successfully managed to make the Sensor inputs (Camera) from the Carla simulator/ interface environment compatible with the DeepDrive tensorflow model. Finding the structure of the data structure to transfer the control commands mapped from the outputs of the tf model into for interfacing with the CARLA simulator is still ongoing.

---

[1] N. Kalra and S. M. Paddock, “Driving to safety: How many miles of driving would it take to demonstrate autonomous vehicle reliability?” *Transportation Research Part A: Policy and Practice*, vol. 94, pp. 182–193, Dec. 2016.

[2] F. Wotawa, B. Peischl, F. Klück, and M. Nica, “Quality assurance methodologies for automated driving,” *e & i Elektrotechnik und Informationstechnik*, vol. 135, nos. 4-5, pp. 322–327, Aug. 2018.

[3] A. Dosovitskiy, G. Ros, F. Codevilla, A. Lopez, and V. Koltun, “CARLA: An Open Urban Driving Simulator,” Nov. 2017.

[4] C. Quiter and M. Ernst, “Deepdrive/deepdrive: 2.0,” Mar. 2018.