Investigating the effect of rewards functions in Deep Q-learning on the safety of automated vehicle behaviors

Eric Chen, Chen.Eric@ufl.edu, University of Florida

February 5, 2024

Abstract

The potential of adopting deep reinforcement learning has been widely studied to design and develop autonomous vehicles (Ye et al., 2021) Increasingly, recent studies highlight the importance of implementing and constructing reward functions in order to ensure the safety of the autonomous vehicles (e.g., Ye et al. (2021); Knox et al. (2021)). However, little to no attention has empirically compared the influence of different reward functions on safety (e.g., collision rate). Hence, the purpose of this project proposal is to understand and investigate the effect of varying types of reward functions with Q-learning that combines the convolution neural network approaches (DQN; Mnih et al. (2016)). We will use the highway-env simulator in conjunction to train the agent in multiple driving environments, such as highways, merge environments, and intersections. We anticipate our results will reveal how reward functions in deep Q-learning systematically impact different safety requirements in autonomous vehicles.

Introduction

The advancement of autonomous and semi-autonomous vehicles has created increasing opportunities in daily transportation and smart-city designs (e.g., Ye et al. (2019, 2021); Faisal et al. (2019). The 2016 World Economics Forum entitled autonomous vehicles (AVs) as one of the promising emerging technologies. Much previous research on AVs in the literature has been around evaluating the efficacy of complex neural systems to automate the task of driving. Reinforcement learning (RL) approaches, in specific, have been investigated and applied extensively to the control and the decision-making problems of the AVs (Ye et al., 2021). Reinforcement learning algorithms learn by trial-and-error and hence do not require extensive hand labelling unlike other types of machine learning approaches, such as supervised learning algorithms. However, developing a self-driving, and control system that is generalizable and applicable to varying driving environments still remains challenging with reinforcement learning due to main reasons. One of the challenges is identifying well-defined reward functions and systems. This is because the effectiveness and the efficacy of the reward functions for safe driving systems highly depend on numerous driving and environmental attributes. Further, the lack of rigorous methods adds to the challenge of systematically investigating and evaluating the efficacy of reward functions in AVs (Knox et al., 2021).

Despite the importance of the issue, a lack of attention has been provided to empirically compare the influence of different reward functions on the safety of autonomous vehicles. Hence, the purpose of this proposed study is to understand and investigate the effect of varying types of reward functions with deep Q-learning that combines the convolution neural network approaches (DQN; Mnih et al. (2016)) using an open-source RL simulation environment, *highway-env* (Leurent, 2018; Rana and Malhi, 2021). Specifically, we evaluate the collision rate, and violation of safety distance between the agent and the surrounding vehicle in a total of four collision-prone highway environments.

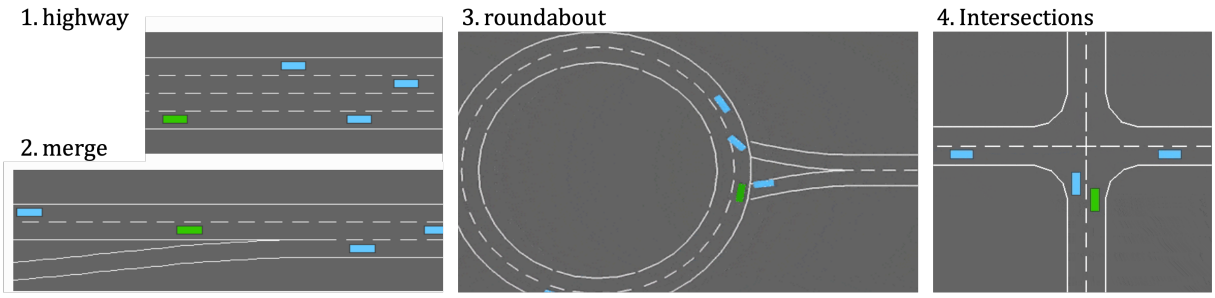

Figure 1: Collision-prone environments for the proposed experiment.

Problem Statement

The primary focus of this proposal concerns the design of an experiment with differing reward functions in deep Q-learning to implement and evaluate its effect on creating safe and autonomous vehicles. Specifically, we intend to evaluate how the agents with varying reward functions perform in collision-prone highway driving scenarios. We formulate the dynamics of the autonomous driving simulation system as Markov Decision Process (MDP) to approximate the lane-navigating systems in our experimentation. The MDP comprises a behaviour policy $\pi(\alpha|s)$, that outputs a certain state, s, and an action, $a \in \{lane_{left}, idle, lane_{right}, faster, slower\}$ based on the given state. The learned policy will attempt to best identify the safest and most effective action by estimating and updating the value function with the succinct reward function as described in the next section.

Proposed Approach

A single object or reward function may not approximate the driving scenarios to train high-performing autonomous vehicles (Yuan et al., 2019). Hence, we will implement a multi-reward architecture combined with the Q-learning method. Specifically, the deep Q-learning approach approximates the loss function as the expectation of the residuals between the target, y_i^{deepQ}, and $Q(s, a; \theta_i)^2$, where y_i^{deepQ} is defined using the discount factor, γ and the reward, r. In this proposed, study we will decompose the reward, r using three sub-reward functions (i.e., $r_{collision}$, $r_{lane.change}$, r_{speed}) that penalizes or encourages different vehicle maneuvers as in equation 1.

$$R_{total}(s, a, s') = \sum_{i=1}^{m} R_i(s, a, s') \tag{1}$$

Proposed Steps

- Identify the literature with sub-reward functions
- Implement the identified reward functions in python environment
- Implement the Highway-env to construct for four collision-prone highway scenarios
- Train the agent with varying reward functions and test the agents with a pre-defined number of iterations or steps (e.g., 100,000 and 5,000)
- Conduct significance testing (e.g., ANOVA and 95% C.I. and p-value) to compare the performance of the agents based on the safety requirements

	Yuan et al. (2019)	Yavas et al. (2020)	Ye et al. (2019)	Leurent (2018)	Hoel et al. (2018)
$r_{collision}$	+20	-100(terminal)	r = $\begin{cases} -1 & if\ \ collision \\ h_d \times v - d & otherwise \end{cases}$	$a \times \frac{v - v_{min}}{v_{max} - v_{min}} - b \times collision$	-10
$r_{lane.change}$	+0.25	+1.0	-	+0	-1.0
r_{speed}	$\frac{v - v_{min} \times r_v}{v_{max} - v_{min}}$	$\frac{v_{current} - v_{initial}}{v_d}$	-	+0.4	$\frac{\Delta d}{\Delta d_{max}}$

Table 1: Reward functions compared in the experiment

Experiment Setup

In our experiment setup, we will evaluate the reward functions of a total of five studies that introduced the multi-reward architecture of AVs in highway driving environments. Table 1 provides an overview of the five studies with their reward values and functions identified for the collision, lane change, and the change of speed of the AVs.

Simulator

Highway-env (Leurent, 2018) is an open-source reinforcement learning environment, which supports both the single- and multi-agent driving scenarios. Highway-env includes varying types of driving scenarios (e.g., highway, roundabout, intersections, and merge scenarios), which allow users to evaluate the autonomous vehicles with tasks like changing lanes. The traffic flow of highway-env is generated from MOBIL model (Kesting et al., 2007). The customization is easily accessible in highway-env, which provides an ideal environment to compare the vehicle's behaviors based on the differing reward functions.

Analysis Process

In order to compare the varying reward functions using deep Q-learning and in a varying driving environment, we first created a collision-prone environment by changing the acceleration parameters defined in the kinematic rules of the highway-env package. This was adopted from the setup by Rana and Malhi (2021). Second, we followed the training strategy introduced by (Yuan et al., 2019), thus, we will use a collection of the observation of the first 5,000 steps will be observed, then, the following 100,000 steps will be used for training. The final 5,000 steps will be used for the testing period. While the sudden acceleration, break, and speed changes are not ideal, occasional changes in the condition occur in the highway environment, which may significantly increase the accident rate. Hence, we introduced the quick speed changes of the surrounding cars in order for the agent to learn safe inter-vehicle distances. Other driving conditions, except the varying multi-architecture reward function, will be set to the default values suggested by highway-env. More superficially, for the highway driving scenario, the total driving time will be set to 40 seconds, with a total of lanes 4, and 50 as a total number of cars in the environment.

Evaluation Plan

The evaluation will be conducted and communicated by plotting the learning curves of agents. The total number of collisions will be used as a primary proxy to evaluate the safety of the autonomous vehicle. In case the agents show an extremely low collision rate, we will also introduce the rate of violation between the agent and the surrounding car's distance that is smaller than the safety distance cut-offs. Once the data points from multiple runs are collected with differing reward we will conduct a one-way ANOVA to evaluate the significance based on the p-value.

References

Faisal, A., Kamruzzaman, M., Yigitcanlar, T., and Currie, G. (2019). Understanding autonomous vehicles. *Journal of transport and land use*, 12(1):45–72.

Hoel, C.-J., Wolff, K., and Laine, L. (2018). Automated speed and lane change decision making using deep reinforcement learning. In *2018 21st International Conference on Intelligent Transportation Systems (ITSC)*, pages 2148–2155. IEEE.

Kesting, A., Treiber, M., and Helbing, D. (2007). General lane-changing model mobil for car-following models. *Transportation Research Record*, 1999(1):86–94.

Knox, W. B., Allievi, A., Banzhaf, H., Schmitt, F., and Stone, P. (2021). Reward (mis) design for autonomous driving. *arXiv preprint arXiv:2104.13906*.

Leurent, E. (2018). An environment for autonomous driving decision-making. `https://github.com/eleurent/highway-env`.

Mnih, V., Badia, A. P., Mirza, M., Graves, A., Lillicrap, T., Harley, T., Silver, D., and Kavukcuoglu, K. (2016). Asynchronous methods for deep reinforcement learning. In *International conference on machine learning*, pages 1928–1937. PMLR.

Rana, A. and Malhi, A. (2021). Building safer autonomous agents by leveraging risky driving behavior knowledge. In *2021 International Conference on Communications, Computing, Cybersecurity, and Informatics (CCCI)*, pages 1–6.

Yavas, U., Kumbasar, T., and Ure, N. K. (2020). A new approach for tactical decision making in lane changing: Sample efficient deep q learning with a safety feedback reward. In *2020 IEEE Intelligent Vehicles Symposium (IV)*, pages 1156–1161. IEEE.

Ye, F., Zhang, S., Wang, P., and Chan, C.-Y. (2021). A survey of deep reinforcement learning algorithms for motion planning and control of autonomous vehicles. In *2021 IEEE Intelligent Vehicles Symposium (IV)*, pages 1073–1080. IEEE.

Ye, Y., Zhang, X., and Sun, J. (2019). Automated vehicleâs behavior decision making using deep reinforcement learning and high-fidelity simulation environment. *Transportation Research Part C: Emerging Technologies*, 107:155–170.

Yuan, W., Yang, M., He, Y., Wang, C., and Wang, B. (2019). Multi-reward architecture based reinforcement learning for highway driving policies. In *2019 IEEE Intelligent Transportation Systems Conference (ITSC)*.