Self-Improving Safety Performance of Reinforcement Learning Based Driving with Black-Box Verification Algorithms

Resul Dagdanov [1,2] Halil Durmus [2] Nazim Kemal Ure [1]

[1]Istanbul Technical University [2]Eatron Technologies

Abstract

In this work, we propose a self-improving artificial intelligence system to enhance the safety performance of reinforcement learning (RL)-based autonomous driving (AD) agents using black-box verification methods. RL algorithms have become popular in AD applications in recent years. However, the performance of existing RL algorithms heavily depends on the diversity of training scenarios. A lack of safety-critical scenarios during the training phase could result in poor generalization performance in real-world driving applications. We propose a novel framework in which the weaknesses of the training set are explored through black-box verification methods. After discovering AD failure scenarios, the RL agent's training is re-initiated via transfer learning to improve the performance of previously unsafe scenarios. Simulation results demonstrate that our approach efficiently discovers safety failures of action decisions in RL-based adaptive cruise control (ACC) applications and significantly reduces the number of vehicle collisions through iterative applications of our method.

Literature Review

In a recent study [3], researchers reduced unsafe scenarios of a black-box system by guiding exploration samples along predefined trajectory classes. However, without verification through rare-event simulation [2] and generalized importance sampling on a continuous action and observation space, the safety of the AD system could not be guaranteed. In another safety-critical system, researchers proposed a safety-constrained collision avoidance approach [5] with prediction-based reachability analysis. Nevertheless, these approaches do not verify the designed models on the continuous feature domain of scenarios. As the operating environment of the black-box system becomes more complicated, testing the control policy in all potential circumstances becomes impractical. Hence, innovative techniques are needed to verify the safety performance of the black-box system.

Research Gaps

Consider a typical AD scenario where the EGO vehicle comfortably and safely pursues the lead vehicle or most important object (MIO). In ideal adaptive cruise control (ACC) settings, an EGO vehicle is expected to follow the MIO vehicle with a comfortable braking distance, the least amount of jerk, and pleasant acceleration. Without any prior knowledge of the mathematical model of the system, ACC is effectively approached using model-free deep RL through trial and error on the safety and comfort parameters in interaction with the simulation environment [4]. Similarly, another study [1] used deep RL to safely enable lane-changing maneuvers and demonstrated the advantages of model-free algorithms compared to rule-based baselines. However, the safety of these trained RL agents is not fully verified over the continuous range of rare events, where time-to-collision (TTC) measurements are low and the probability of collision is high.

Research Objectives

- **Objective 1:** Investigate whether the rare-event assessment and validation samples of a safety-critical system can be utilized for automatic self-improvement (healing) of the RL agents.
- **Objective 2:** Propose a framework that incrementally adjusts the training scenarios of the RL agent to compensate for the failures in these scenarios.

Proposed Methodology

Our specific contribution is the development of a framework (see Figure 1) where the RL-based AD system is continuously subjected to probabilistic Black-box verification methods to discover failure scenarios. The proposed framework incrementally adjusts the training scenarios of the RL agent to compensate for the failures in these scenarios. We select ACC as a case study for the application of our method and show that the proposed system significantly reduces the number of safety violations.

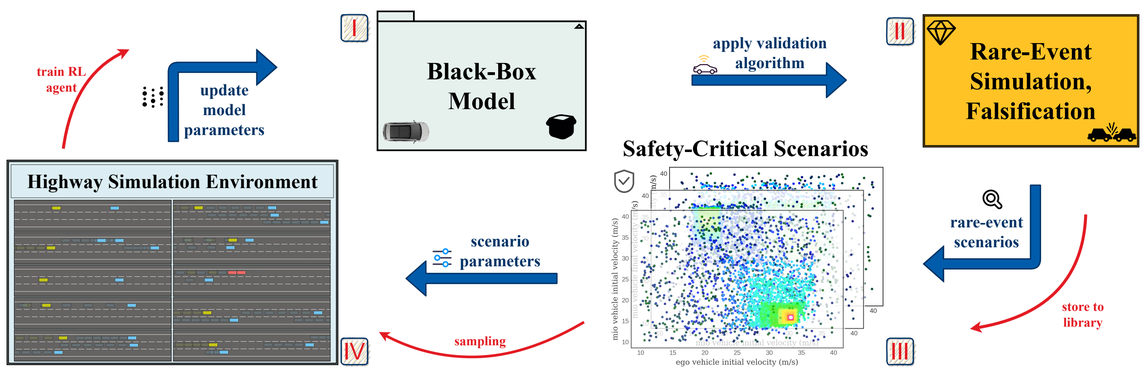

Figure 1. **Proposed Method Overview. Stage-I:** A black-box model is trained in a simulation environment for a predefined number of time steps. **Stage-II:** The trained RL agent is validated using a rare event simulation algorithm. **Stage-III:** Estimated scenario parameters are stored in the library of safety-critical scenarios for future reference. **Stage-IV:** Rare-event scenarios are sampled using the proposed method to update model parameters for a safer RL agent.

Problem Setting

We constructed the problem of finding safety-critical scenarios in an ACC driving application using rare-event simulation techniques and proposed an effective and sample-efficient continuous self-improving framework. Initially, we visualized certain features in an ACC scenario, such as the time-gap ($\mathcal{T}^{gap}$) and time-to-collision (τ^{ttc}) between the lead (MIO) and following (EGO) vehicles.

Reward Function

The custom reward function is designed to express the vehicle following the scenario with a safety metric. A positive reward is given when the black-box agent follows the lead vehicle (MIO) with $0.8s \leq \mathcal{T}^{gap} \leq 2.0s$. However, when the agent fails to drive safely and crashes with the front vehicle, a significant penalty (negative reward) is given, and the episode is terminated.

$$\mathbf{R}_t^{gap} = \begin{cases} -\dfrac{1}{\mathcal{T}_t^{gap}} & \text{if } \mathcal{T}_t^{gap} \in \big(0.0,\ 0.8\big) \\ +\mathcal{T}_t^{gap} & \text{if } \mathcal{T}_t^{gap} \in \big[0.8,\ 2.0\big] \\ -\mathcal{T}_t^{gap} & \text{if } \mathcal{T}_t^{gap} > 2.0 \end{cases} \tag{1}$$

$$\mathbf{R}_t(s_t, a_t) = \begin{cases} -10.0 & \text{collision} \\ \mathbf{R}_t^{gap} + c_j \cdot \dot{\alpha}_t + \dfrac{c_{sp} \cdot v_t}{v_{max}} & \text{otherwise} \end{cases} \tag{2}$$

where $c_j = -0.5$ is a penalty coefficient for the vehicle jerk $\dot{\alpha}_t$ and $c_{sp} = 5.0$ is a reward coefficient for the velocity v_t at the episodic time-step t.

Objective Function

$$\mathcal{T}_t^{gap} = \frac{d_t^{mio}}{v_t^{ego}} \quad \forall\, t \in (0, 1, \ldots, T) \tag{3}$$

We chose TTC, given in Eq. (4), as an objective metric to minimize in rare-event simulations. We formulated risk-based evaluation to minimize the objective function $F(\mathbf{x_i})$, as low values of τ_i^{ttc} are rare and dangerous. $F(\mathbf{x_i})$ is evaluated by simulating the black-box model, which is an RL agent.

$$\tau_t^{ttc} = \frac{d_t^{mio}}{v_t^{ego} - v_t^{mio}} \quad v_t^{ego} > v_t^{mio},\ t \in (0, 1, \ldots, T) \tag{4}$$

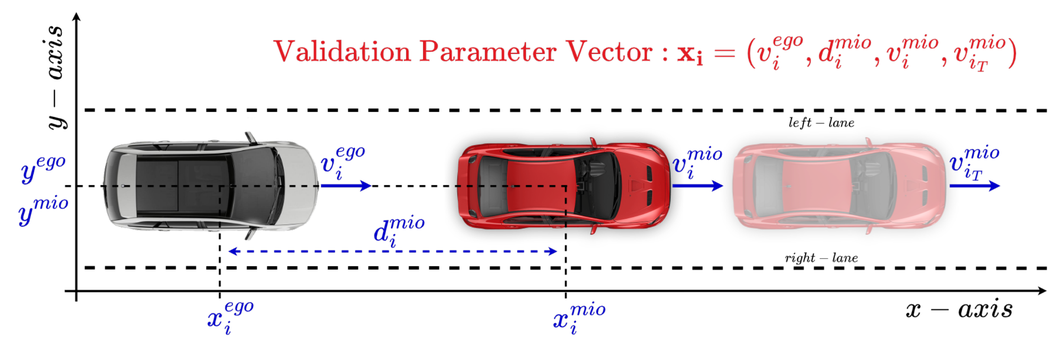

Figure 2. Case study illustration of validation parameters in an ACC scenario, where v_i^{ego} is an initial velocity of the EGO vehicle, d_i^{mio} is an initial MIO vehicle distance relative to an EGO vehicle, v_i^{mio} and $v_{i_T}^{mio}$ are initial and target velocities of the MIO vehicle for a scenario parameter vector $\mathbf{x_i}$.

Self-Improving Method

A continuous learning pipeline is proposed in which a black-box model is periodically tested with rare-event simulations and improved by effectively leveraging safety-critical scenarios. The verification samples are stored in a library to be sampled for the training of the black-box policy.

$$P(X^{\pi_k}|\pi_k(\theta)) = \frac{1}{2} \cdot \frac{k+1}{\sum_{g=0}^{G+1}[g]}, \quad k \in (0, 1, \ldots, G) \tag{5}$$

where G is the total number of generations while training a black-box model. $X^{\pi_k} = \{\mathbf{x_i} : \forall\, i \in (1, 2, \ldots, N)\}$ is a set of validation scenario events of the policy π_k, where k is a generation number and N is a total number of rare-event simulations.

Results

The primary objective was to decrease the number of vehicle collisions and improve the safety of the black-box system. To demonstrate the effectiveness of rare-event simulation algorithms in a self-improvement approach for a safety-critical black-box system, we compared our proposed approach with GS and naïve MC sampling methods using black-box verification algorithms. Fig. 3 shows the normalized number of collisions with MIO vehicles in the ACC scenarios constructed from a continuous uniform distribution space. After each proceeding generation of self-improvement with rare-event simulation samples, the black-box agent experiences fewer collisions.

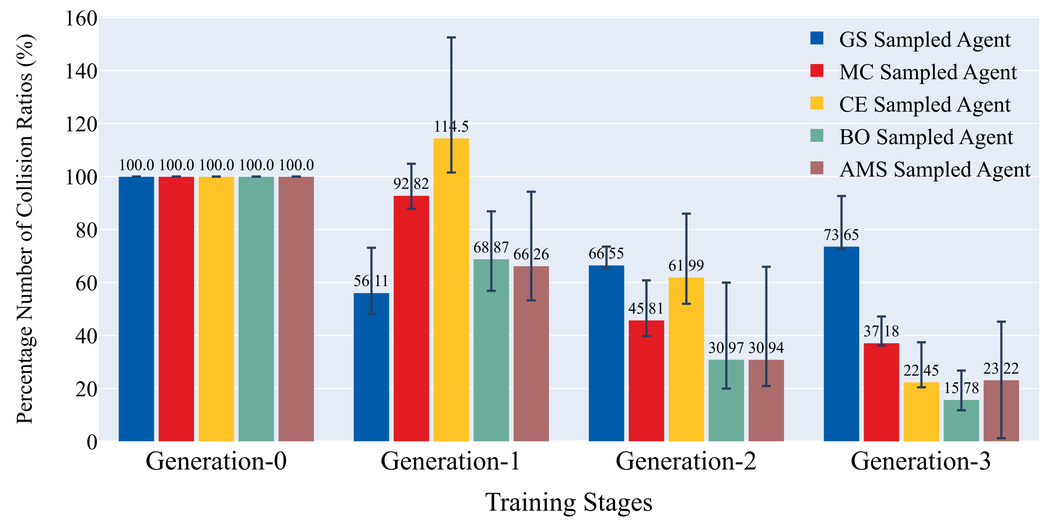

Figure 3. **Percentage Ratios of Vehicle Collisions**. The self-improving black-box models are evaluated at each training generation stage under a uniform state distribution. The max-min performance with the average collision ratios, normalized w.r.t. the collisions in Generation-0, is illustrated for each RL agent.

Conclusions

- We have proposed a novel self-improving framework using black-box verification algorithms for enhancing the safety performance of reinforcement learning-based autonomous driving agents.
- We have demonstrated that the weaknesses of the exploration nature of RL agents could be inspected by leveraging rare-event simulations.
- With the proposed methodology, safety-critical scenarios of the black-box system are detected in the early stages of the model training.

References

[1] Ali Alizadeh, Majid Moghadam, Yunus Bicer, Nazim Kemal Ure, Ugur Yavas, and Can Kurtulus. Automated lane change decision making using deep reinforcement learning in dynamic and uncertain highway environment. In *2019 IEEE Intelligent Transportation Systems Conference (ITSC)*, pages 1399–1404, 2019.

[2] Søren Asmussen and Peter W Glynn. *Stochastic simulation: algorithms and analysis*, volume 57. Springer, 2007.

[3] Yunus Bicer, Ali Alizadeh, Nazim Kemal Ure, Ahmetcan Erdogan, and Orkun Kizilirmak. Sample efficient interactive end-to-end deep learning for self-driving cars with selective multi-class safe dataset aggregation. In *2019 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pages 2629–2634, 2019.

[4] Lokesh Chandra Das and Myounggyu Won. Saint-acc: Safety-aware intelligent adaptive cruise control for autonomous vehicles using deep reinforcement learning. In Marina Meila and Tong Zhang, editors, *Proceedings of the 38th International Conference on Machine Learning*, volume 139 of *Proceedings of Machine Learning Research*, pages 2445–2455. PMLR, 18–24 Jul 2021.

[5] Anjian Li, Liting Sun, Wei Zhan, Masayoshi Tomizuka, and Mo Chen. Prediction-based reachability for collision avoidance in autonomous driving. In *2021 IEEE International Conference on Robotics and Automation (ICRA)*, pages 7908–7914, 2021.