Leveraging on Deep Reinforcement Learning for Autonomous Safe Decision-Making in Highway On-ramp Merging (Student Abstract)

Zine el abidine Kherroubi[1], **Samir Aknine**[2], **Rebiha Bacha**[1]
[1] Groupe Renault, Guyancourt, 78280
[2] Claude Bernard Lyon 1 University, Villeurbanne, 69100
zine-el-abidine.z.kherroubi@renault.com, samir.aknine@univ-lyon1.fr, rebiha.bacha@renault.com

Abstract

High-speed highway on-ramp merging is one of the most difficult and critical tasks for any autonomous driving system. This work studies this problem by combining deep deterministic policy gradient (DDPG) reinforcement learning with drivers' intentions prediction. Our proposed solution is based on an artificial neural network to predict drivers' intentions, used as an input state to the DDPG agent that outputs the longitudinal acceleration to the merging vehicle. We show that this solution improves safety performances.

Introduction and Modeling

In the present work, we use the continuous deep deterministic policy gradient (Lillicrap et al. 2016) to perform highway on-ramp merging. We train a DDPG agent to learn a safe and cooperative merging policy. The uniqueness and novelty of our method regarding the state-of-the-art reinforcement learning-based solution are that our approach uses an artificial neural network to learn and predict the intention of the driver on the main highway lane. The driver's intention estimation is provided as an input state to the DDPG agent. This may be mandatory because human-driven vehicles, which cannot be controlled directly, will still be present on the roads, even with the emergence of connected and autonomous vehicles. These human drivers have various driving styles. Hence, predicting drivers' behaviors is necessary to learn a safe and cooperative decision-making policy. In short, the main contributions of our work are:

- Predicting drivers' intentions using an artificial neural network, so that our autonomous driving system could be used in mixed traffic where there are connected and non-connected vehicles.
- Improving driving performance and safety of the learned driving policy by using driver intention as an input state to the DDPG agent.

We assume that the decision for the highway on-ramp merge is determined only by the projection of the merging vehicle in the main lane *(V_m)* and only the two preceding vehicles *(P_1, P_2)* and the two following vehicles *(F_1, F_2)* in the main lane, as illustrated in Figure 1. Also, we assume that

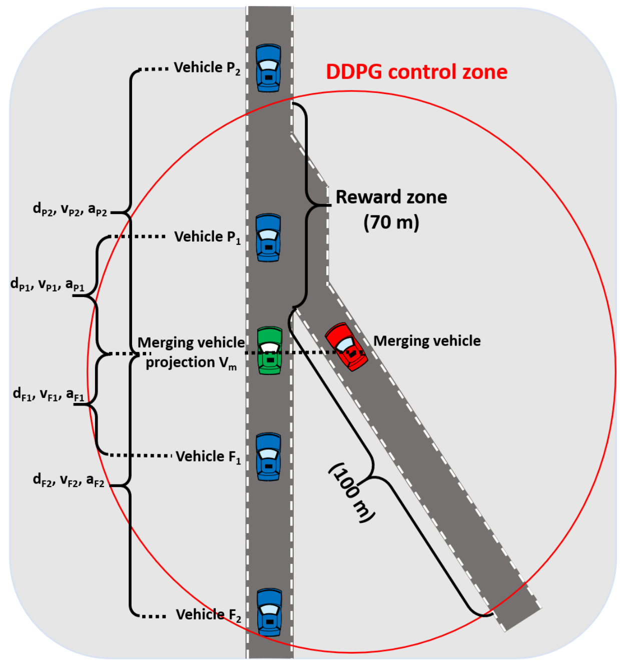

Figure 1: Highway on-ramp situation: V_m is the projection of the merging vehicle in the main lane, P_1, P_2 are the two preceding vehicles in the main lane and F_1, F_2 are the two following vehicles in the main lane.

the merging vehicle is connected and automated, and fully informed about the states of the surrounding vehicles using its sensors and *V2I*[1] communication with a Road-Side Unit *(RSU)*[2]. This makes our environment state fully observable even in the presence of non-automated and non-connected vehicles in the main highway lane. We consider that the reinforcement learning *RL* agent learns a policy that controls the merging vehicle within the red zone, as shown in Figure 1.

The reward motivates the merging vehicle to merge midway between the preceding vehicle and the following vehi-

[1]Vehicle-To-Infrastructure

[2]The off-board Road-Side Unit *(RSU)* contains its sensors (camera, radar) and has the *V2I* communication capabilities.

cle to maximize the safety distance, with the same speed as the preceding vehicle (with a normalization factor of 10).

Yet, none of the state-of-the-art works that studied high-speed highway on-ramp merging has explicitly considered the intention of the driver driving in the main highway lane for the decision making. This kind of decision making must be taken into account to learn a cooperative driving behavior. In order to predict the intention of the driver before the merging point in the main lane, we use the model that was proposed by (Kherroubi, Aknine, and Bacha 2019). In fact, the authors in (Kherroubi, Aknine, and Bacha 2019) have shown that the intention of drivers in the highway on-ramp can be predicted with great accuracy, using an off-board model.

Experimental Evaluation

The merging environment we have developed is created in the Simulation environment of Urban Mobility (*SUMO*), under which, we created a highway on-ramp situation with similar geometry to the highway on-ramp section of *US Interstate Highway I-80* in *Emeryville (San Francisco), California, USA*.

Experiments Results

In order to improve the performances of the autonomous driving architecture at the highway on-ramp merging situation, we incorporate a model that predicts the behavior of the vehicle below the merging point at the main highway lane. We thus compared two DDPG agents:

- DDPG agent without driver intention prediction (*DDPG*).
- DDPG agent with driver intention prediction as input state (*DDPG-I*).

These agents were trained for *1* million simulation steps. Figure 2 shows the average undiscounted reward (over *100* episodes) for each agent. When using the driver intention as an input to the state vector, the DDPG agent (*DDPG-I*) converges to the optimal policy after only *450K* simulation steps (*3200* episodes in Fig.2), which is considerably faster, by *55%*, than the first agent that did not include the intention (*DDPG*), and which converges after *1* million simulation steps (*7500* episodes in Fig.2). Also, the average reward of the agent that includes the driver intention (*DDPG-I*)

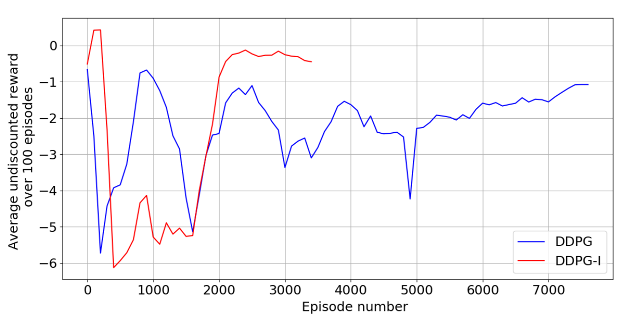

Figure 2: Average undiscounted reward over *100* episodes during training

Performances	*DDPG*	*DDPG-I*
Number of episodes	8015	8100
Number of stops	0	0
Number of collisions	1	0
Number of emergency brakings by vehicle F_1	8	3
Number of successful episodes	8014	8100
Average speed (*m/s*)	23.4	26
Average gap from midway point (m)	23.2	24.5
Average speed difference from vehicle P_1 (*m/s*)	10.5	7.8

Table 1: DDPG agents testing results

converges to a final value of *-0.4* (as in Fig.2), which is greater than the (*DDPG*) agent's value (*-1* in Fig.2), and presents fewer fluctuations at the end of training.

In order to compare and validate the performance of each agent, we tested each one during *1* million steps. The results are summarized in Table 1. From Table 1, when adding the intention estimation as input state to the DDPG agent (*DDPG-I*), the number of collisions drops to zero (unlike the *DDPG* agent that did not include driver intention estimation). Moreover, the number of emergency brakings drops to only *3* cases (*8* cases when we use the *DDPG* agent without driver intention estimation). When comparing average reward values, the *DDPG-I* agent has better performance in following the speed of the preceding vehicle (*7.8 m/s*). Moreover, it has the higher average merging speed (*26 m/s*) compared to the *DDPG* agent that did not include the driver intention prediction, which means that the learned driving policy is the best adapted for high-speed highway scenario. In summary, adding driver intention as input state to the DDPG agent improves safety performances by eliminating collision cases and reducing the number of emergency brakings. Also, the training of this type of agent is *55%* faster than a DDPG agent that does not consider the driver behavior model.

In conclusion, adding the driver's intention is a key component for accelerating convergence and learning *"safe"* and *"cooperative"* autonomous driving policy.

References

Kherroubi, Z.; Aknine, S.; and Bacha, R. 2019. A Dynamic Bayesian Network Based Merge Mechanism for Autonomous Vehicles. *Proceedings of the AAAI Conference on Artificial Intelligence, vol. 33* .

Lillicrap, T. P.; Hunt, J. J.; Pritzel, A.; Heess, N.; Erez, T.; Tassa, Y.; Silver, D.; and Wierstra, D. 2016. Continuous control with deep reinforcement learning. *International Conference on Learning Representations (ICLR)* .