Demo: Curricular Reinforcement Learning for Robust Policy in Unmanned CarRacing Game

Yunzhe Tian*, Yike Li*, Yingxiao Xiang*, Wenjia Niu*, Endong Tong*, and Jiqiang Liu*

*Beijing Key Laboratory of Security and Privacy in Intelligent Transportation,
Beijing Jiaotong University, China.

Email: {tianyunzhe,yikeli,yxxiang,niuwj,edtong,jqliu}@bjtu.edu.cn

Wenjia Niu and Endong Tong are corresponding authors

***Abstract*—Robust reinforcement learning has been a challenging problem due to always unknown differences between real and training environment. Existing efforts approached the problem through performing random environmental perturbations in learning process. However, one can not guarantee perturbation is positive. Bad ones might bring failures to reinforcement learning. Therefore, in this paper, we propose to utilize GAN to dynamically generate progressive perturbations at each epoch and realize curricular policy learning. Demo we implemented in unmanned CarRacing game validates the effectiveness.**

I. Design and Implementation

In this work, at each training epoch of average 400 episodes, we firstly use an improved Least-Sqaures GAN (LSGAN) [1] to generate a set of progressive perturbations. We then evaluate and label the generated perturbations, according to whether those are at the intermediate level of difficulty for current policy and update the policy under positive perturbations implemented on rllab [2]. This process and formula are presented in Fig.1. As to control the training difficulty, we define perturbations of intermediate difficulty ($POID$), and require $POID_i := \{p : R_{min} \leq R^{\pi_i}(p) \leq R_{max}\}$, for perturbation value p on current policy π_i. The label y_{p_j} used in next iteration indicates whether $p_j \in POID_i$. We thus collect trajectories under positive environment perturbations and use RL algorithm as policy optimizer to update our policy.

II. Results and Conclusion

Fig. 2 shows the experimental training process in Unmanned CarRacing Game of OpenAI gym environment [3]. The friction coefficient (FC) of current environment is 0.6. Through curricularly updating at appropriate difficulty, the car is increasingly robust and finally makes a successful turn. Our approach even achieves 11 times score (869.37 over 77.58) higher than baseline reinforcement learning and 2 times score (869.37 over 397.29) higher than robust reinforcement learning with domain randomization [4]. This demo has good potential to apply into non-game reinforcement learning of unmanned driving.

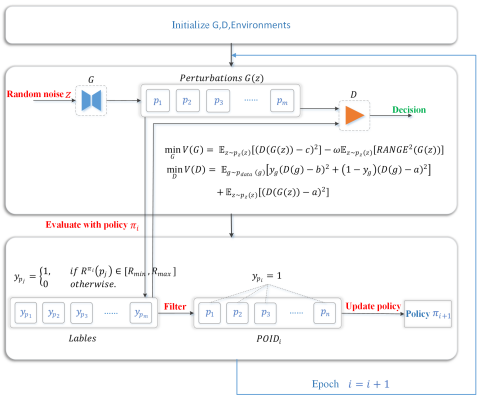

Fig. 1. The framework of curricular reinforcement learning.

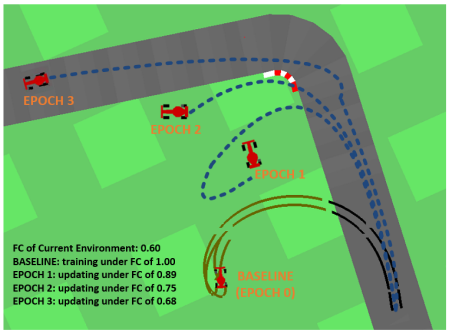

Fig. 2. Training process in Unmanned CarRacing Game.

Acknowledgment

This research is supported by the National Natural Science Foundation of China (61972025), Fundamental Research Funds for the Central Universities of China (2019RC008).

References

[1] Florensa, C., Held, D., Geng, X., and Abbeel, P. Automaticgoal generation for reinforcement learning agents. *ICML*, 2018.

[2] Duan Y, Chen X, Houthooft R, et al. Benchmarking Deep Reinforcement Learning for Continuous Control. *International Conference on Machine Learning (ICML). JMLR.org*, 2016.

[3] G. Brockman, V. Cheung, L. Pettersson, J. Schneider, J. Schulman, J. Tang, and W. Zaremba. Openai gym, 2016.

[4] X. B. Peng, M. Andrychowicz, W. Zaremba and P. Abbeel. Sim-to-Real Transfer of Robotic Control with Dynamics Randomization. *ICRA*, 2018.

Workshop on Automotive and Autonomous Vehicle Security (AutoSec) 2021
25 February 2021, Virtual
ISBN 1-891562-68-1
https://dx.doi.org/10.14722/autosec.2021.23011
www.ndss-symposium.org