SMOOTH REGULARIZED REINFORCEMENT LEARNING

Anonymous authors
Paper under double-blind review

ABSTRACT

Deep learning has been shown to be very powerful in reinforcement learning. The generalization of current RL system, however, remains an unsolved hard problem due to the data scarcity and the complexity of deep neural network. Based on the intuition that, in real world, the optimial policy should be smoothed in terms of the environment state. This paper proposes the idea of smooth regularized reinforcement learning (SRRL), where the policy is trained to consistent output against adversarial input. In extensive experiments in multiple environments (Inverted-Pendulum, HalfCheetah, Swimmer, Hopper and Walker2d), we demonstrate that our method improves policy smoothness, training stability, and greater generalization performance.

1 INTRODUCTION

Many effort have been recently devoted to using deep neural network as function approximation in reinforcement learning (RL, citeation). The RL system benefits from the large capacity and strong representation power of neural network. However, training a generalizable high-capacity model requires a large mount of training data and faces certain training difficulties such as overfitting, vanishing/exploding gradient, covariate shift, etc.

To address the challenges, regularization techniques such as dropout and orthogonality parameter constraints have been proposed for general deep learning models. In particular, people also develop regularization tricks to improve the generalization ability of reinforcement learning models (Pinto et al., 2017; Cheng et al., 2019). For example, RARL (Pinto et al., 2017) propose a robust RL model that aims at performan well under uncertainties by training the agent against adversarially perturbed environment. RARL, however, require an additional adversarial policy to be trained and obtained little performance gain. Cheng et al. (2019) on the other hand proposed a control regularization that regularize the behavior of the deep policy to be similar to a policy prior, which, however, may not be available for all scenario.

Different from previous works, we propose a smooth regularized reinforcement learning method (SRRL). Our key motivation is that, when facing a similar environment state the policy model should perform a similar action. For example, a physical system (i.e. MuJoCo environment Todorov et al. (2012)) is powered by physical law and thus the optimal policy is smoothed. As a simple show case, an optimal policy of CartPole environment has a provably smooth close form solution. Unlike traditional kernel method, the deep learning model, however, does not capture such smooth natural explicitly.

Based on the above intuition, the proposed SRRL trains the agent with an additional regularization loss, which minimize the Jensen-Shannon divergence of the output policy against between ordinary state and adversarialy perturbed state. The proposed smooth regularization is closely related to consistency regularization, which minimize the output of the deep model between original input and a handful designed input and has been demonstrated to be powerful in various settings and applications, including semi-supervised learning (Miyato et al., 2018), improving model robustness and uncertainty (Zhang et al., 2019; Hendrycks et al., 2019), and unsupervised data augmentation (Xie et al., 2019)

The rest of the paper is organized as follows: Section 2 introduces the related background; Section 3 introduces our proposed smooth regularized reinforcement learning in detail; Section 4 presents numerical experiments on four MuJoCo environments.

Notations:

2 BACKGROUND

2.1 MARKOV DECISION PROCESS

2.2 JENSEN–SHANNON DIVERGENCE

3 METHOD

4 EXPERIMENT

5 CONCLUSION

REFERENCES

Richard Cheng, Abhinav Verma, Gabor Orosz, Swarat Chaudhuri, Yisong Yue, and Joel W Burdick. Control regularization for reduced variance reinforcement learning. *arXiv preprint arXiv:1905.05380*, 2019.

Dan Hendrycks, Mantas Mazeika, Saurav Kadavath, and Dawn Song. Using self-supervised learning can improve model robustness and uncertainty. *arXiv preprint arXiv:1906.12340*, 2019.

Takeru Miyato, Shin-ichi Maeda, Shin Ishii, and Masanori Koyama. Virtual adversarial training: a regularization method for supervised and semi-supervised learning. *IEEE transactions on pattern analysis and machine intelligence*, 2018.

Lerrel Pinto, James Davidson, Rahul Sukthankar, and Abhinav Gupta. Robust adversarial reinforcement learning. In *Proceedings of the 34th International Conference on Machine Learning-Volume 70*, pp. 2817–2826. JMLR. org, 2017.

Emanuel Todorov, Tom Erez, and Yuval Tassa. Mujoco: A physics engine for model-based control. In *2012 IEEE/RSJ International Conference on Intelligent Robots and Systems*, pp. 5026–5033. IEEE, 2012.

Qizhe Xie, Zihang Dai, Eduard Hovy, Minh-Thang Luong, and Quoc V Le. Unsupervised data augmentation. *arXiv preprint arXiv:1904.12848*, 2019.

Hongyang Zhang, Yaodong Yu, Jiantao Jiao, Eric P Xing, Laurent El Ghaoui, and Michael I Jordan. Theoretically principled trade-off between robustness and accuracy. *arXiv preprint arXiv:1901.08573*, 2019.

A APPENDIX