Resetting constant parameters based on the selected environment

Sirusala Niranth Sai[1]

[1]Indian Institute of Technology (BHU) Varanasi

Method. In this method, a Reinforcement Learning agent with Proximal Policy Optimization Schulman et al. (2017) training is trained, whose parameters are set by us. The settings are fine-tuned values around the predefined default values for the specific environment in RL Baselines3 Zoo Raffin (2018). We assumed that the default parameters, which are designed to perform well in vanilla environments, could still provide competitive performance on simpler environments without suffering significantly from performance degradation.

The PPO algorithm implementation being used is from the stable-baselines3 library Raffin et al. (2021). It is trained for a total of trained it for $1O^5$ steps. This training time is divided into 10 epochs, each of 10000 steps, where all other parameters of the agent are constantly reset to the same values, as defined in zoo. The tests were performed in a setting where 2 context features were varied at the same time with a standard deviation of 0.5 around the mean values. However, once sampled, these were kept fixed throughout the training time.

Limitations. As the hyperparameters used are staatic, it does not perform well across all the environments. It is able to solve easier environments, but struggles with the complicated ones.

Reproducibility. I have provided the solution.py and policy class codes to reproduce.

References

Raffin, A. (2018). Rl baselines zoo. `https://github.com/araffin/rl-baselines-zoo`.

Raffin, A., Hill, A., Gleave, A., Kanervisto, A., Ernestus, M., and Dormann, N. (2021). Stable-baselines3: Reliable reinforcement learning implementations. *Journal of Machine Learning Research*, 22(268):1–8.

Schulman, J., Wolski, F., Dhariwal, P., Radford, A., and Klimov, O. (2017). Proximal policy optimization algorithms. *arXiv preprint arXiv:1707.06347*.

