Supplementary of Arena: A Scalable and Configurable Benchmark for Policy Learning

Shuang Liu*[1], Sirui Xu*[2], Tongzhou Mu[1], Zhiwei Jia[1], Yiran Wu[1], Hao Su[1]
[1] Department of Computer Science and Engineering, UC San Diego
[2] School of Electrical Engineering and Computer Science, Peking University

1 Dataset Documentation and Intended Uses

Our benchmark, Arena, is a set of interactive environments to be used for reinforcement learning and imitation learning. The player API follows the convention of OpenAI Gym exactly, and any researchers familiar with the OpenAI Gym API can start to play with our environment using the favorable reinforcement learning/imitation learning frameworks. We have provided some demo files at https://github.com/Sirui-Xu/Arena/example, as a basic example.

The low-level game engine is implemented on PyGame[1], which is a highly customizable engine.

2 URL to Website

We will maintain the assets, codes, and benchmarks on https://github.com/Sirui-Xu/Arena

3 Author Statement

We bear all responsibility in case of violation of rights and confirmation of the data license.

4 Hosting, Licensing, and Maintenance Plan

The ARENA benchmark is a light-weighted benchmark built upon PyGame learning environment. There is no large data files included, so we decide to host it on Github. The project uses MIT license.

We will maintain a table that records the ranking of algorithms published in major venues (e.g., NeurIPS, ICML, and ICLR) on our benchmark for comparison and citations. We will also provide a form for users to contact us and upload their results on the table.

Additionally, we would encourage players of this benchmark to upload their models and codes in terms of github repo link.

5 Reproducibility

Our repository includes all the codes to reproduce the results reported in the experiment section. Please refer to README.md in our repo at https://github.com/Sirui-Xu/Arena

[1] https://github.com/pygame/pygame

Submitted to the 35th Conference on Neural Information Processing Systems (NeurIPS 2021) Track on Datasets and Benchmarks. Do not distribute.