Final Project Proposal

Arthur Liang

TOPIC

OpenAI gym [1] is a toolkit that contains a number of playground tasks for reinforcement learning. For the final project of CSE 573, we want to explore this environment and build two agents to interact with it: a random agent as baseline and an A3C agent [3].

MOTIVATION

In reinforcement learning, a field in machine learning, an intelligent agent can learn to improve its actions through a number of interactions with the environment. It is exciting because this interact-and-improve process is very similar to how a human would learn to do a task. In addition, the advancement of deep learning made huge progress in reinforcement learning and shows a seemingly infinite potential for what an agent can learn and do.

OpenAI gym is a well-known toolkit for reinforcement learning. It contains a number of benchmarks, e.g. balancing a pole on a cart, along with clear and standardized metrics that shows how a model performs. Therefore, we believe it is an ideal playground for our project.

DATASET AND RESOURCES

We will mainly use the OpenAI gym [2]. Specifically, we want to focus on two environments: CartPole in the Classic Control section and Ant-v2 in MuJuCo.

Additionally, we may also refer to a reinforcement learning framework [4] by Joseph Redmon.

EXPERIMENT AND PLAN

We divide our project into the following three steps

Step 1. Implement a random agent to interact for task CartPole. For this part, we focus on getting deeper understanding of the openAI gym APIs and of how an agent would interact in the environment.

Step 2. Implement an A3C agent [3] to achieve good performance in the CartPole task. The framework [4] suggests that a good agent can achieve an average total reward of 499 per episode, which we will use as a reference.

Step 3. Apply the agent in previous step, with small adaptations to the Ant-v2 task of MuJuCo. For this part, we want to experiment and understand how those similar agents learn and adapt to the two very different tasks, e.g. how different the learning curves are, what changes in hyper parameters are needed. If time permits, we also want to explore optimizations, e.g. a deep neural network, to allow the agent to improve on this task.

MILESTONE AND TIMELINE
- By March 1st: finish step 1 & 2
- By March 12th: finish step 3
- By March 15th: Finish the entire project, including presentation and report.

REFERENCE

[1] Brockman, Greg, et al. "OpenAI Gym." ArXiv:1606.01540 [Cs], June 2016. arXiv.org, http://arxiv.org/abs/1606.01540.
[2] OpenAI. Gym: A Toolkit for Developing and Comparing Reinforcement Learning Algorithms. https://gym.openai.com. Accessed 17 Feb. 2021
[3] Mnih, Volodymyr, et al. "Asynchronous Methods for Deep Reinforcement Learning." *ArXiv:1602.01783 [Cs]*, June 2016. *arXiv.org*, http://arxiv.org/abs/1602.01783.

[4] Redmon, Joseph. *Pjreddie/Rl-Hw*. 2018. 2020. *GitHub*, https://github.com/pjreddie/rl-hw.