深度强化学习的对抗攻防算法与平台研究

毕业设计中期答辩

翁家翌

清华大学计算机科学与技术系

2020 年 3 月 31 日

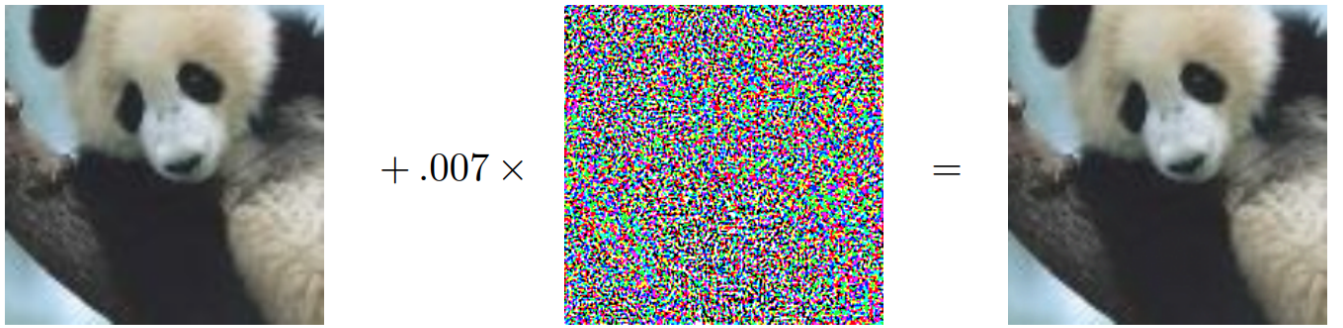

图 1: 对抗样本在图像分类上的应用 [?]

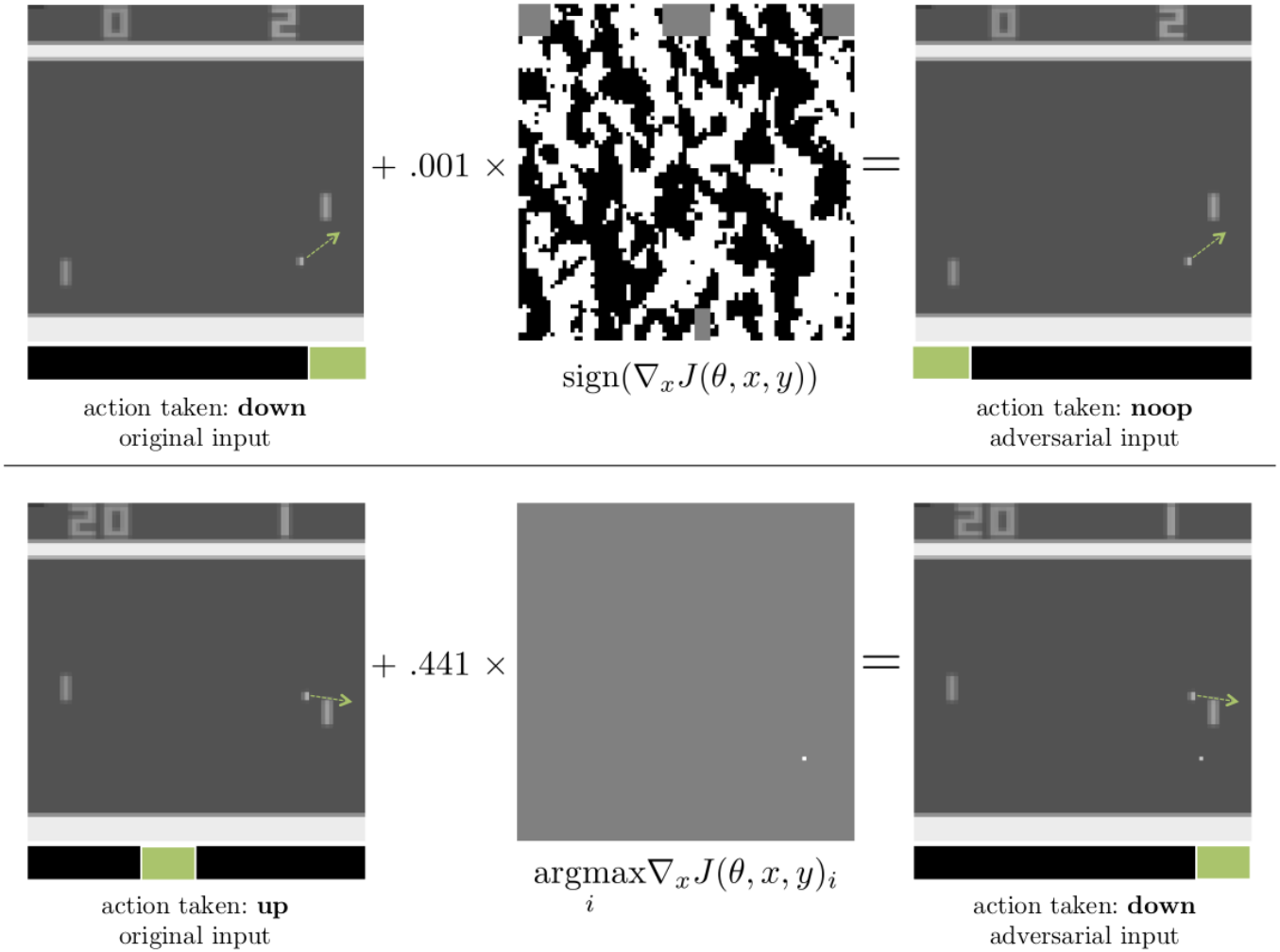

图 2: 对抗攻防在强化学习场景 Atari Pong 中的应用 [?]

研究内容

- 为已有的强化学习算法提供性能评估基准（已完成）
- 以既定指标评测各个强化学习算法的鲁棒性
- 为新出现的强化学习攻防算法提供平台
- 已完成总体工作量 60% 以上

- 以 PyTorch 为基本框架，100% 重写之前实验室的强化学习平台天授 (Tianshou)，目前已在 GitHub 上开源：
  `https://github.com/thu-ml/tianshou`

- Tianshou 平台已实现 8 种主流的强化学习算法：
  - PG [**?**]
  - DQN [**?**]
  - DDQN [**?**]
  - A2C [**?**]
  - PPO [**?**]
  - DDPG [**?**]
  - TD3 [**?**]
  - SAC [**?**]

- 现有主流强化学习平台以OpenAI Baselines、RLlib为两派代表，大多基于 TensorFlow，代码实现复杂，不易修改，模块抽象程度不够清晰
- Tianshou 在保持简洁的代码实现、灵活的框架、模块化的算法实现下，还在这些平台中取得了最快的速度。此处选择了 GitHub 上强化学习最多 star 的两个平台 Baseline（9.5k）、RLlib（11k）和基于 PyTorch 的最多 star 的强化学习平台 PyTorch DRL（2.3k）、rlpyt（1.4k）进行对比评测：

We select some of famous (>1k stars) reinforcement learning platforms. Here is the benchmark result for other algorithms and platforms on toy scenarios: (tested on the same laptop as mentioned above)

RL Platform	Tianshou	Baselines	Ray/RLlib	PyTorch DRL	rlpyt
GitHub Stars	stars 134	stars 9.5k	stars 11k	stars 2.3k	stars 1.4k
Algo - Task	PyTorch	TensorFlow	TF/PyTorch	PyTorch	PyTorch
PG - CartPole	9.03±4.18s	None	15.77±6.28s	None	?
DQN - CartPole	10.61±5.51s	1046.34±291.27s	40.16±12.79s	175.55±53.81s	?
A2C - CartPole	11.72±3.85s	*(~1612s)	46.15±6.64s	Runtime Error	?
PPO - CartPole	35.25±16.47s	*(~1179s)	62.21±13.31s (APPO)	29.16±15.46s	?
DDPG - Pendulum	46.95±24.31s	*(>1h)	377.99±13.79s	652.83±471.28s	172.18±62.48s
TD3 - Pendulum	48.39±7.22s	None	620.83±248.43s	619.33±324.97s	210.31±76.30s
SAC - Pendulum	38.92±2.09s	None	92.68±4.48s	808.21±405.70s	295.92±140.85s

**: Could not reach the target reward threshold in 1e6 steps in any of 10 runs. The total runtime is in the brackets.*

?: We have tried but it is nontrivial for running non-Atari game on rlpyt. See here.

All of the platforms use 10 different seeds for testing. We erase those trials which failed for training. The reward threshold is 195.0 in CartPole and -250.0 in Pendulum over consecutive 100 episodes' mean returns.

图 3: 不同强化学习平台性能对比结果

- Tianshou 提供了完善的单元测试：在每次测试的时候，对于每一个强化学习算法都进行完整的训练，一旦训练不出来则认为不通过测试。
- 由于平台性能远远优于常见的强化学习平台，因此 Tianshou 能够支持如此激进的单元测试。

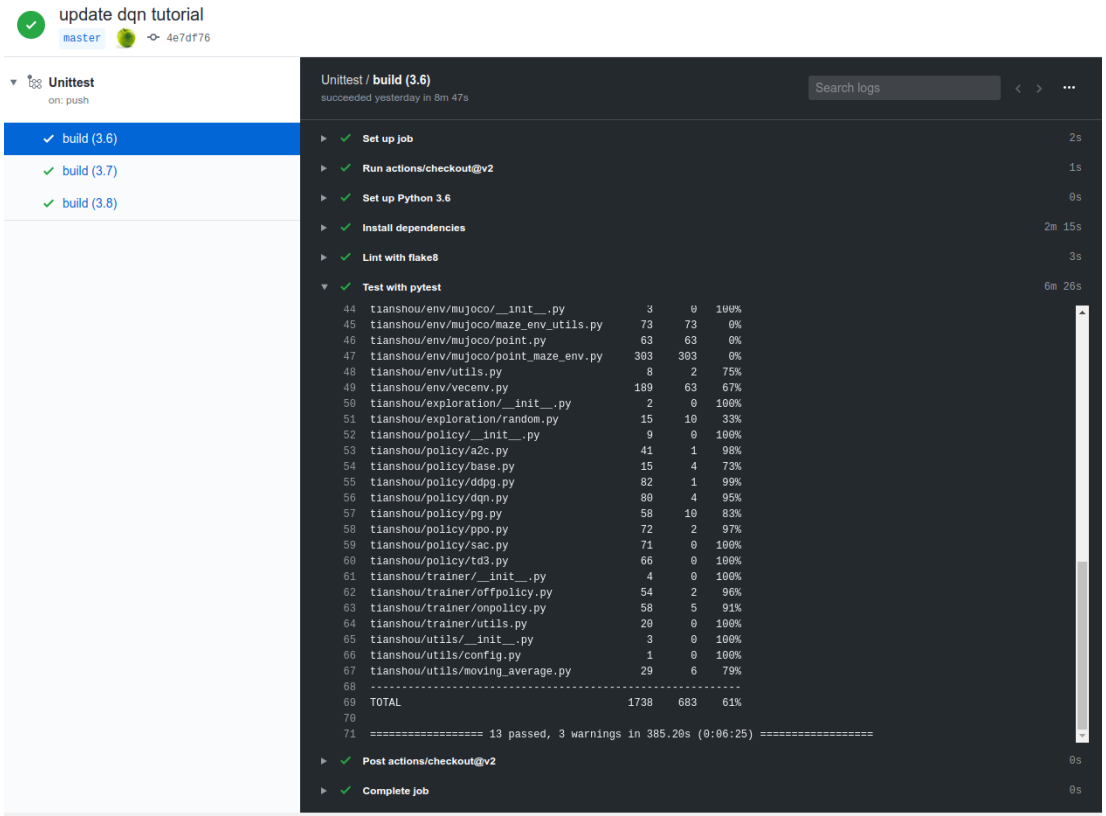

图 4: Tianshou 平台单元测试，图中显示了各个代码的覆盖率。

- Tianshou 还提供了相应的文档说明，目前部署在 https://tianshou.readthedocs.io/，还在完善中。

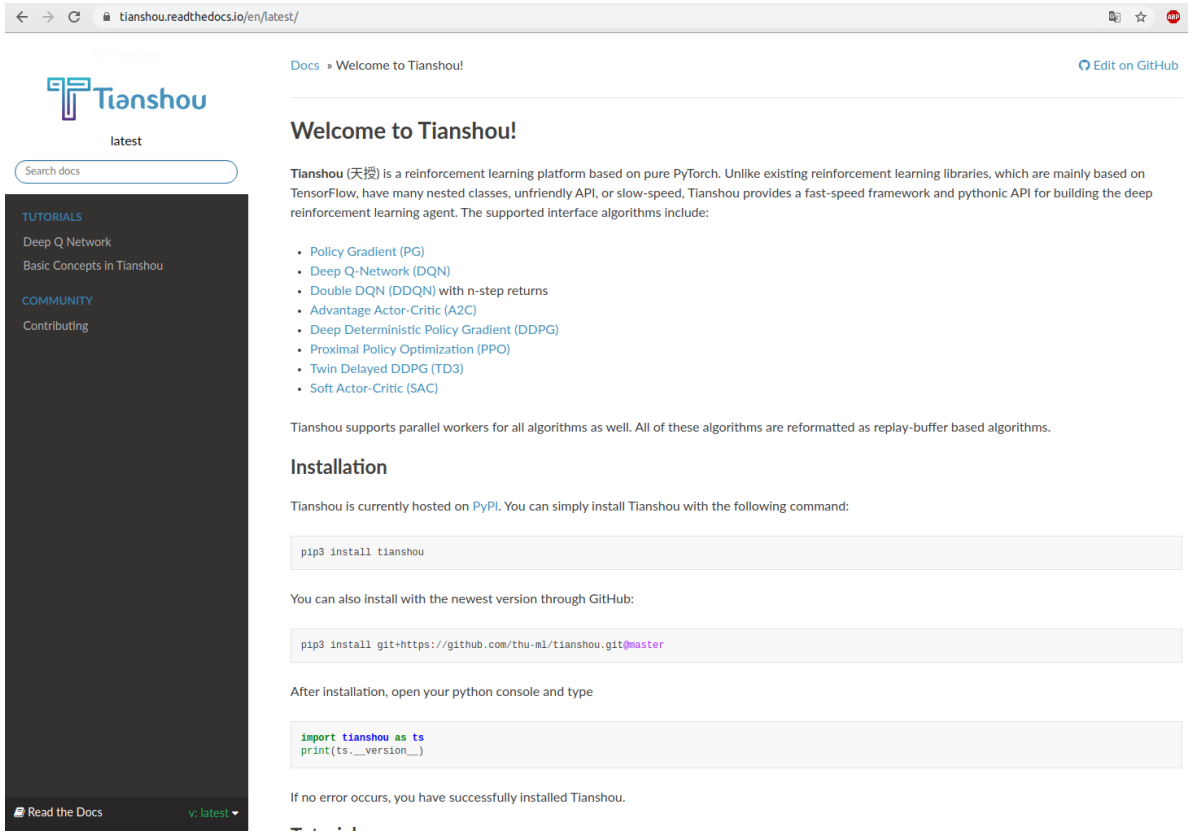

图 5: Tianshou 平台文档

值得一提的是，在新版 Tianshou 开源不到 24 小时之内，新增 GitHub star 已经超过 100，获得了强化学习社区不小的关注度。

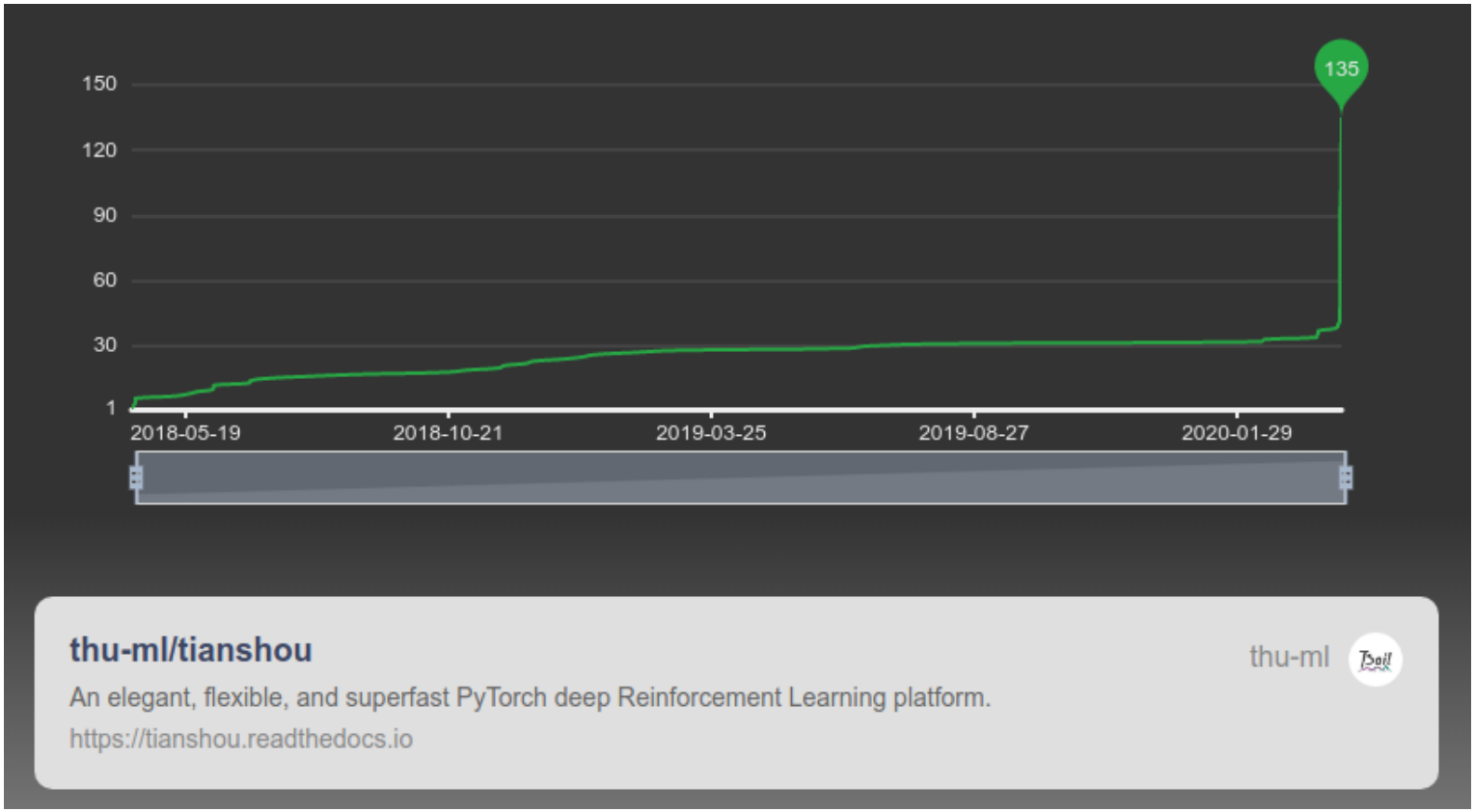

图 6: 由 https://gitmemory.com/ 提供的统计结果，star 数目从 35 到 135，目前还在持续增长。

- 使用既定指标评估已有强化学习算法的鲁棒性
  - Optimal adversarial return
    攻击之后的智能体所能拿到的最多奖励
  - Adversarial regret
    攻击前与攻击后，智能体所能拿到的奖励的差值
  - Per-episode mean cost of attacker
    攻击者实施攻击需要花费的代价
- 强化学习平台的进一步完善
  - 支持更大规模的问题快速求解，如 Atari 游戏
  - 添加更多 Model-free 强化学习算法进入评测
  - 循环神经网络支持（RNN）
  - 优先采样（Prioritized Replay Buffer）
  - 模仿学习（Imitation Learning）
  - 多智能体强化学习（Multi-agent Reinforcement Learning）
  - 分布式训练