

开悟比赛-柳淦爍队技术整理与分享

陈天浩 香港科技大学数学系

黄欣蕊 香港科技大学数学系

许鑫 香港科技大学数学系

指导老师:杨灿

一、简介

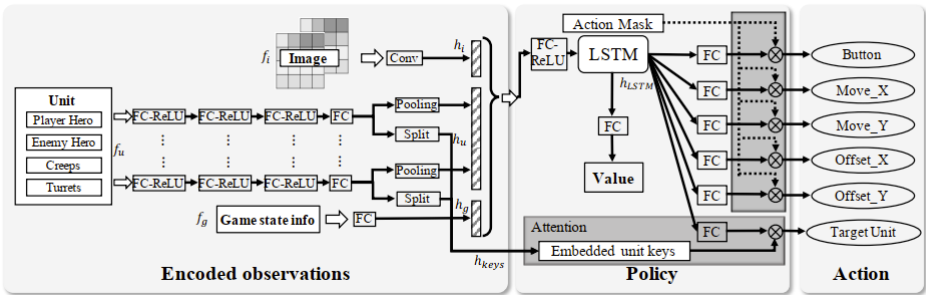
在 2022 年 9 月-2023 年 4 月举办的腾讯第三届开悟 MOBA 多智能体强化学习大赛中，我们队伍(柳淦爍)在初赛中有幸获得了第六名的成绩。这得益于实验室老师与腾讯官方的大力支持，也得益于我们队伍本身的技术探索与积累。本文首先会简单叙述比赛的基本情况，随后从各关键模块出发，简要介绍本队伍在开悟比赛中的探索历程与心得体会。

二、参赛概况

我们都是香港科技大学数学系博士一年级新生。参赛动机主要是为了学习和积累一些深度强化学习的知识和经验。时间上基本都是利用零碎时间，进行监控数据、选取模型和提交新的训练任务,大约一天花了一小时的时间，在每周天梯赛前可能再多一些。

三、网络设计

我们并未对官方提供代码里头的网络架构进行修改，具体的架构如下图所示，细节可参看 [2]。



四、奖励体系

关于奖励的设置，基本上我们是按照原始设定，如下表 ([2]) 所示：

Reward	Weight	Type	Description
hp_point	2.0	dense	the health point of hero
tower_hp_point	10.0	sparse	the health point of turrets and base
money (gold)	0.008	dense	the gold gained
cp_rate	0.8	dense	the rate of mana
death	-1.0	sparse	being killed
kill	-0.5	sparse	kill an enemy hero
exp	0.008	dense	the experience gained
last_hit	0.5	sparse	last hitting to enemy units

在训练过程中，我们曾经尝试过修改 money、kill、death 但效果不好，应该还是没有找到合适的修改方法。感觉应该对着回放视频看目前模型的缺点，然后再修改 reward 更能有的放矢，但限于本队成员对王者荣耀这个游戏没太多了解就没有这么调整。

五、特征与规则

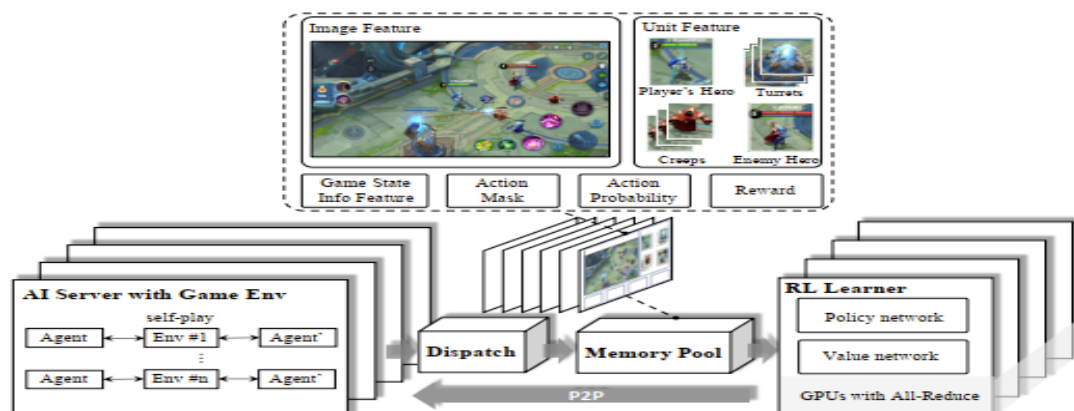
我们没有对特征和规则进行改动，所有特征和规则和参照 [2] 和初赛技术文档。

六、强化学习算法

我们采用的官方代码里的强化学习算法，采用的是标准 PPO 算法([1])，并对目标函数略作修改，称为 Dual-clip PPO 算法 ([2])。

七、系统工程架构

整个训练的框架主要参考 [2]，在实际训练中我们同比例的扩大了 memory pool 和 batch size 以提高 GPU 利用率。



八、模型迭代过程

我们并未修改原始算法、框架和网络结构，主要是对奖励的设置和参数的调节。整个过程中我们的学习率从大调小，batch size 在适当的时候增大，gamma 前期小后期大。采用继续训练的方式，并没有去搞 ablation study，每次训练都是在上一次训练的最好模型开始，然后在这次训练中根据托管对战结果来选出较好的模型。我们将监控曲线对战的模型改成了初始训练的模型，这样能辅助我们更方便的选出较好的模型，也可以比较直观的看出模型性能是否有提升。

九、训练效果分析

每一次调整参数之后我们都需要一定的训练时间来观察监控曲线的变化，因此如果调参不合适的话就会浪费很多时间，但这种继续训练的方式比较简单而且效果相对较好。我们最终模型可以击败 baseline-level-6，但是性能很不稳定。

十、总结与展望

这次初赛总的来说，我们在理解官方提供的代码基础上找出关键参数，然后采取继续训练的方式去得到更好的模型。没有涉及算法、框架、网络架构上的改进，而且限于每周的算力有限，我们也没有去进行对比实验，这些是以后可能改进的地方。

参考文献

- [1] Schulman, J.; Wolski, F.; Dhariwal, P.; Radford, A.; and Klimov, O. 2017. Proximal policy optimization algorithms. arXiv preprint arXiv**
- [2] Ye, D., Liu, Z., Sun, M., Shi, B., Zhao, P., Wu, H., ... & Huang, L. (2020, April). Mastering complex control in moba games with deep reinforcement learning. In *Proceedings of the AAAI Conference on Artificial Intelligence* (Vol. 34, No. 04, pp. 6672-6679).**