Tencent AI Lab

开悟平台框架介绍

AILAB 游戏AI研发中心 姬喜洋

Email : xiyangji@tencent.com

Outline

Tencent AI Lab

- 1v1游戏训练环境介绍
- 1v1特征介绍
- 1v1action space介绍
- 开悟平台训练框架SAIL
- 开悟平台可以用来做哪些研究？
- 实验课出题点

- 1v1游戏训练环境介绍

1v1游戏训练环境介绍

Tencent AI Lab

王者荣耀1v1墨家机关道游戏规则：

红蓝双方各一个英雄进行对战，通过击败小兵和防御塔获得金币与经验，提升技能等级，购买装备
目标为摧毁敌方水晶！

游戏中的场景角色包括：
英雄：有若干技能，没有回城技能
防御塔：双方各一个防御塔，防御塔对攻击范围内的英雄的连续伤害会逐渐升高,对小兵则不会。
水晶：可以认为是一种特殊的防御塔
血包：可以为英雄回复一定血量和蓝量

技能：每个英雄有3-4个专属技能按键，1个恢复技能按键，1个召唤师技能按键，一个普通攻击按键

1v1游戏训练环境介绍

Tencent AI Lab

当前Gamecore使用旧版本王者墨家机关道地图，支持20个英雄的对战

鲁班七号
芈月
吕布
李白
马可波罗
狄仁杰
关羽
貂蝉
露娜
韩信
花木兰
不知火舞
橘右京
后羿
钟馗
干将莫邪
凯
公孙离
裴擒虎
上官婉儿

1v1游戏训练环境介绍

Tencent AI Lab

英雄属性包括：

物理攻击数值
法术攻击数值
物理防御数值
法术防御数值
冷却缩减数值
移动速度
生命值
物理吸血
法术吸血
暴击率
。。。

召唤师技能包括：

闪现、弱化、净化、狂暴、晕眩、惩戒、惩击、干扰、治疗

1v1游戏训练环境介绍

Tencent AI Lab

可以自定义的配置包括：

英雄
召唤师技能
铭文

暂时不可以自定义的配置包括：

出装顺序
技能升级顺序

1v1游戏训练环境介绍

Tencent AI Lab

可以进行的实验任务

20英雄 vs 20英雄，一共400种任务

召唤师技能可用8种，共64种组合

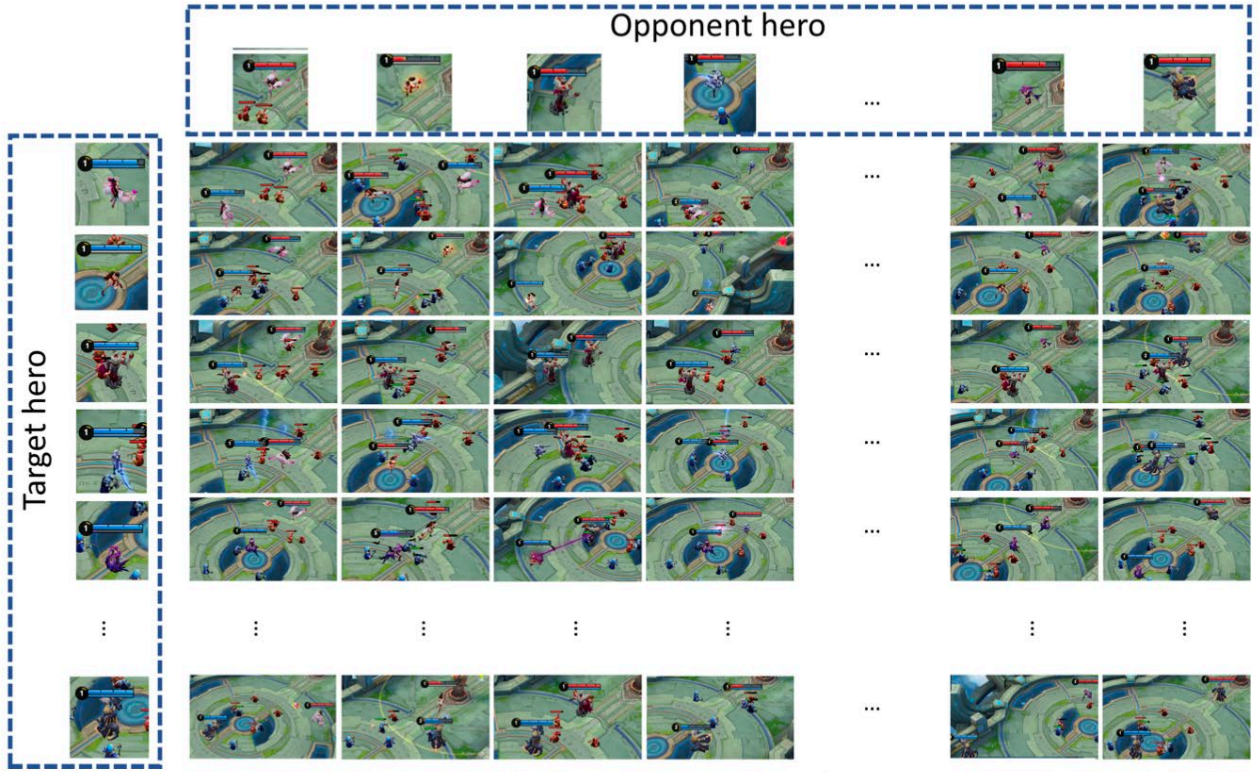
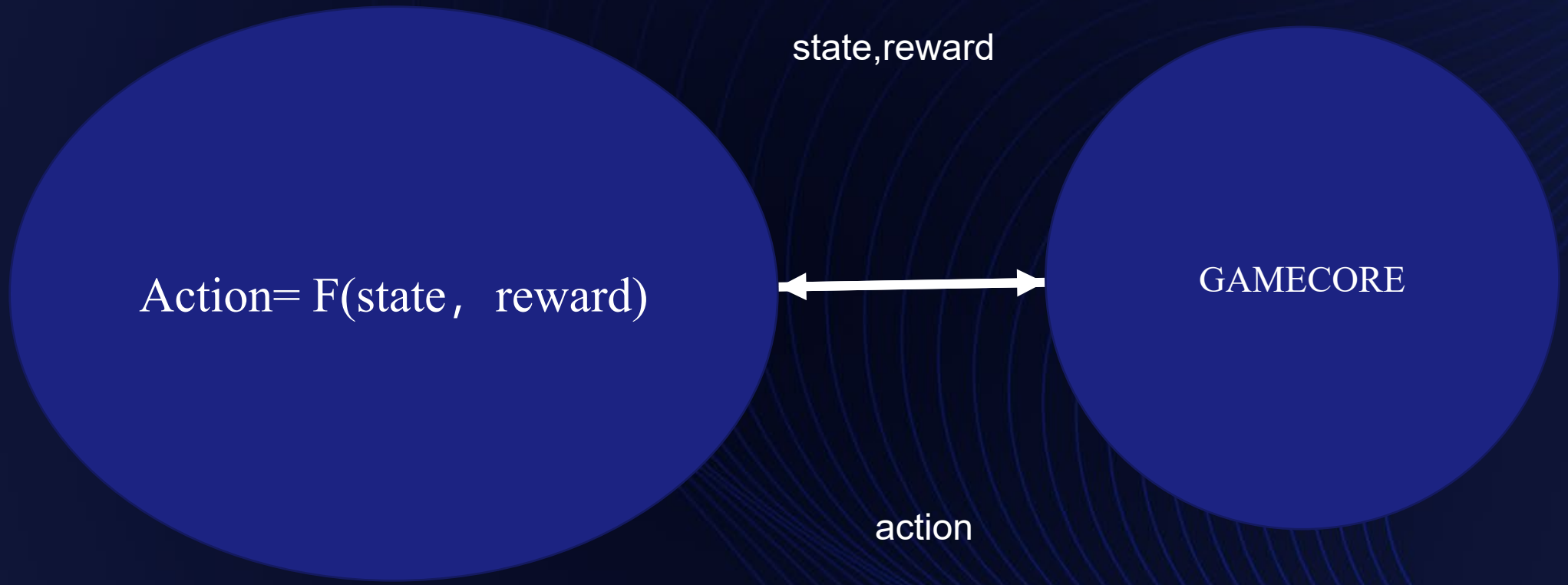


Figure 2: The tasks in *Honor of Kings Arena*. Each row represents the same target hero with different opponent heroes. Each column represents different target heroes with the same opponent hero. There are 20 heroes in *Honor of Kings Arena*, making $20 \times 20 = 400$ tasks in total.

不考虑装备、技能升级顺序条件下，当前可以设计 **400 x 64 = 25600个**子任务，后续会有更多英雄加入

1v1游戏训练环境介绍

Tencent AI Lab



开放环境协议与线上版本不同，不能用于线上！

1v1游戏特征

1v1游戏特征

Tencent AI Lab

	特征区间名	特征维数	举例
主英雄通用特征	Main_camp_hero_state_common_feature	102	血量，位置
主英雄私有特征	Main_camp_hero_private_feature	133	鲁班第几次普攻
敌方英雄通用特征	Enemy_camp_hero_state_common_feature	102	血量，位置
敌方英雄私有特征	Enemy_camp_hero_private_feature	133	鲁班第几次普攻
对局中公开特征	Public_feature	14	敌方小兵是否在我方塔下
主英雄方兵线特征	Main_camp_soldier_feature	18*4	我方小兵1位置，血量
敌方兵线特征	Enemy_camp_soldier_feature	18*4	敌方小兵1位置，血量
主英雄方防御塔特征	Main_camp_organ_feature	18*2	我方防御塔血量，位置
敌方防御塔特征	Enemy_camp_organ_feature	18*2	敌方防御塔血量，位置
游戏全局特征	Global_feature	25	当前游戏处于前中后哪个时期

英雄特征经过**视野过滤**，当敌方英雄不在视野中时，敌方英雄部分特征会被置为默认值

完整特征表请参考：https://aiarena.tencent.com/doc/environments/index.html#v1-environment-mojia-map

1v1游戏特征

Tencent AI Lab

游戏特征计算方法举例：

Gamecore会返回游戏帧状态数据，从中可以获取到英雄血量、技能信息、防御塔信息等数值

对于位置特征：考虑到1v1中地图相对于游戏双方而言是镜像对称的，在双方眼中其都处于地图左下角，故使用相对位置特征，将处于地图右上角的英雄的特征数据进行镜像反转，将其转换为左下角位置。

对于技能子弹特征：例如貂蝉大招与鲁班七号大招的实时位置，都可以通过帧状态中bullet的位置数据来获得

对于强化普攻类特征：比如鲁班被动，记录其第几次普攻特征（one-hot编码），或者当前是否有强化普攻

1v1 action space介绍

1v1 action space介绍

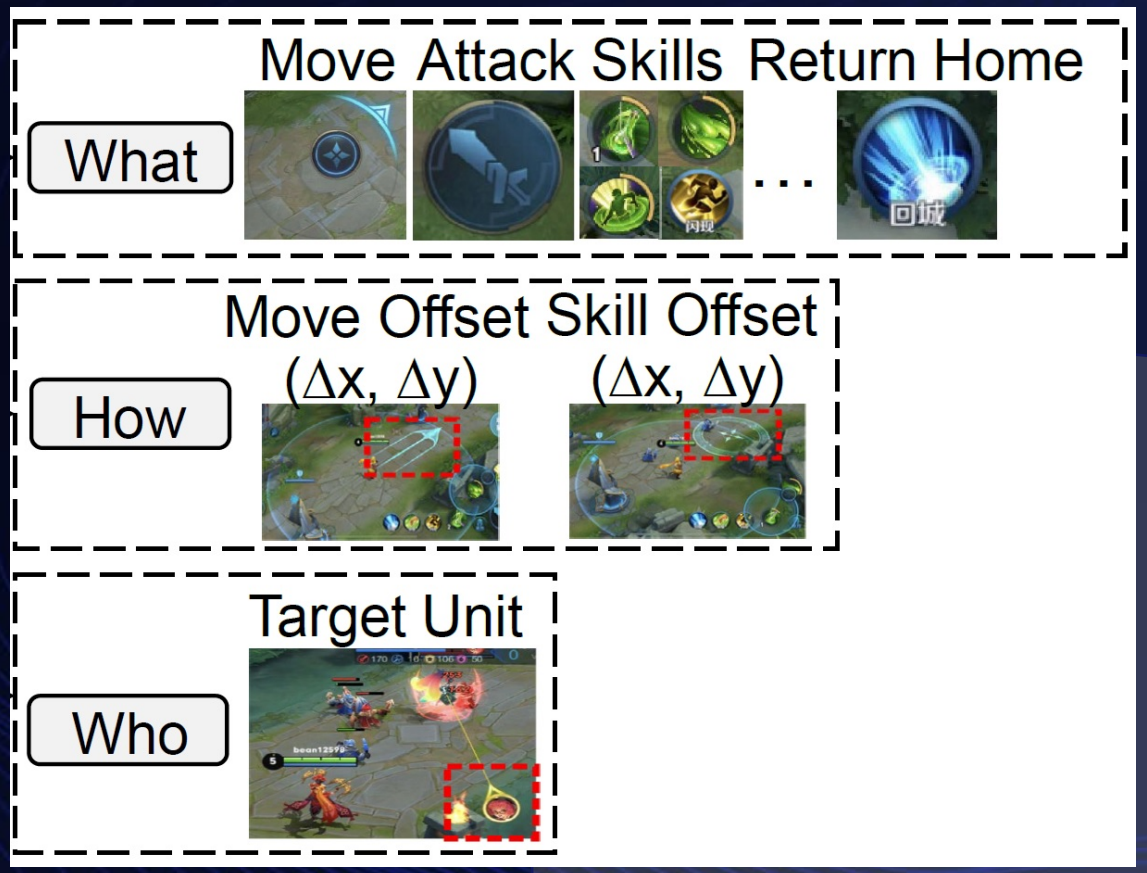
Tencent AI Lab

在我们提供的Benchmark代码中
将action进行了分类预测

What 哪个按键?

How 按键如何移动?

Who 按键作用对象是哪个?

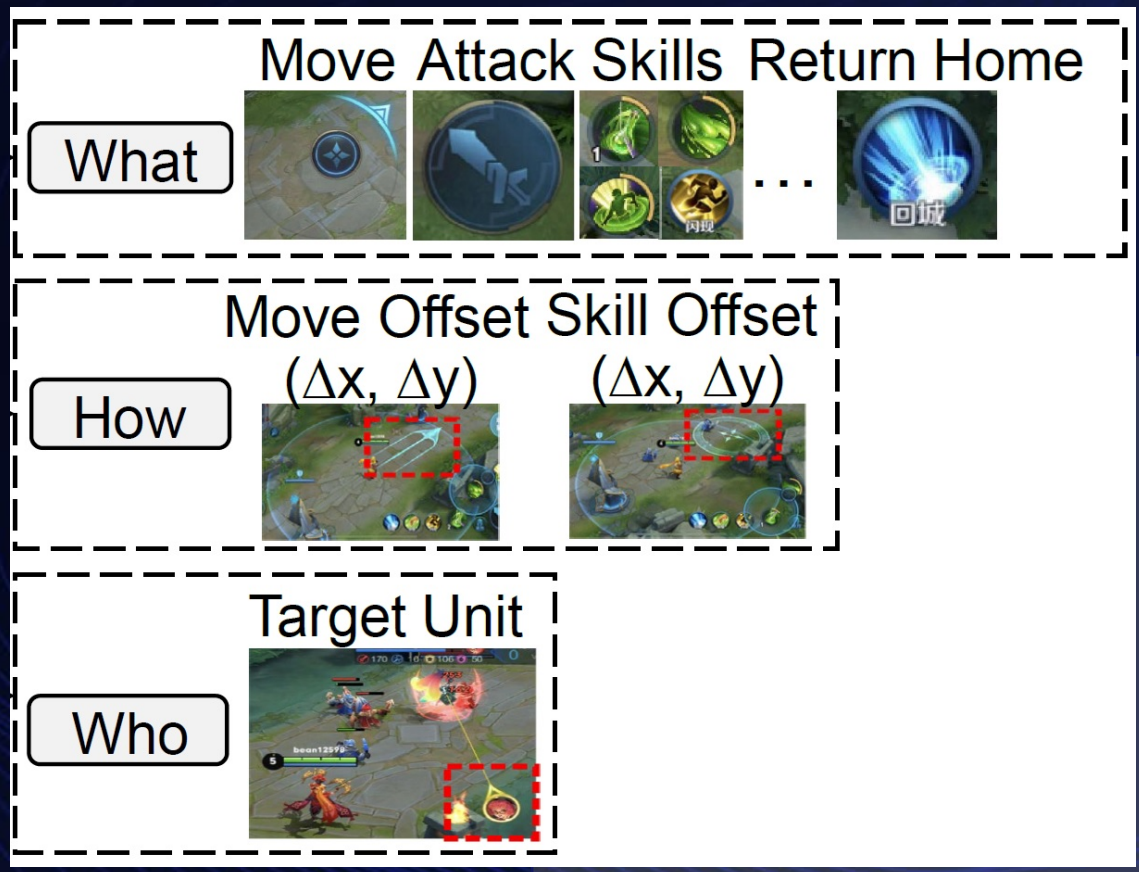


Which button	Move X	Move Z	Skill X	Skill Z	Target

1v1 action space介绍

Tencent AI Lab

Action Class	Type	Description	Dimension
Button	None	No action	1
	None	No action	1
	Move	Move hero	1
	Normal Attack	Release normal attack	1
	Skill 1	Release 1st skill	1
	Skill 2	Release 2nd skill	1
	Skill 3	Release 3rd skill	1
	Heal Skill	Release heal skill	1
	Chosen Skill	Release the chosen skill	1
	Recall	Start channeling and return to the home fountain after a few seconds if not interrupted	1
	Skill 4	Release 4th skill (Only valid for certain heroes)	1
	Equipment Skill	Release skill provided by certain equipment	1
Move	Move X	Move direction along X-axis	16
	Move Z	Move direction along Z-axis	16
Skill	Skill X	Skill direction along X-axis	16
	Skill Z	Skill direction along Z-axis	16
Target	None	Empty target	1
	Self	Self player	1
	Enemy	Enemy player	1
	Soldier	4 Nearest soldiers	4
	Tower	Nearest tower	1



1v1 action space介绍

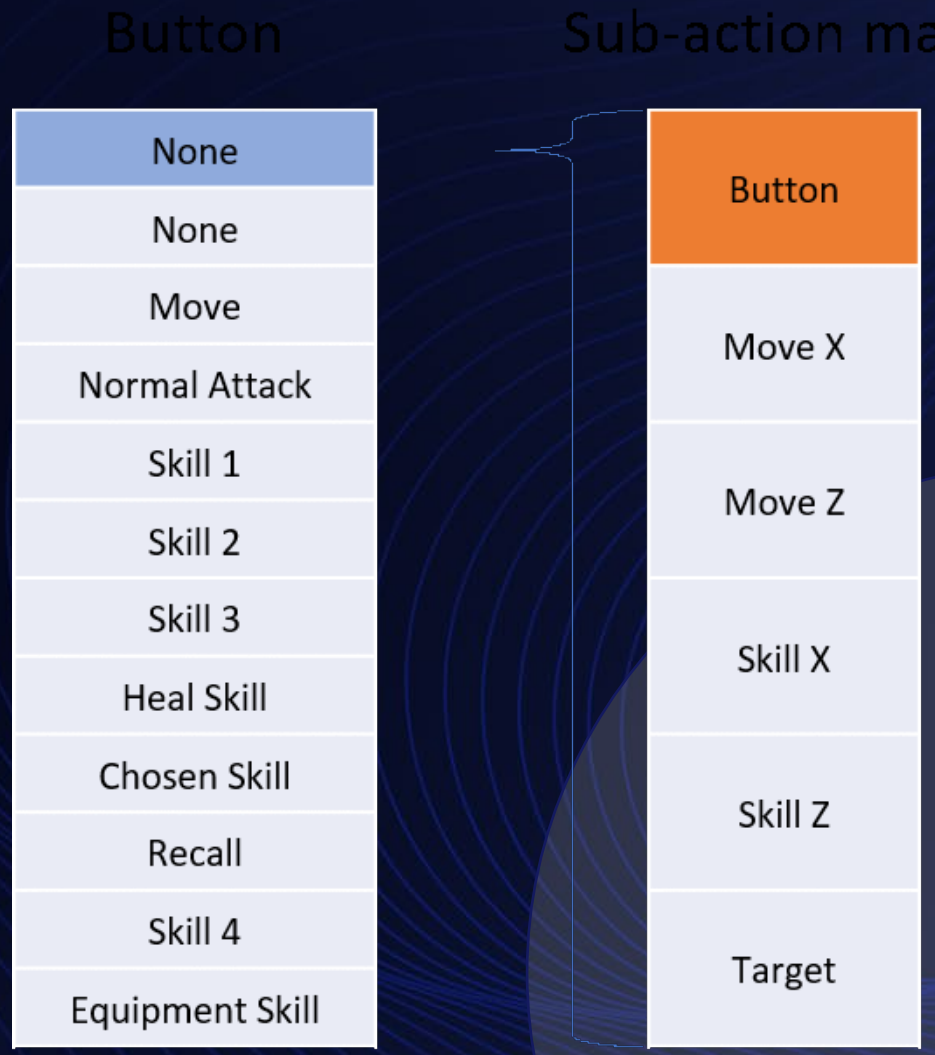
Tencent AI Lab

Action mask

根据人类经验对action进行选择性过滤，避免无意义的探索

原因：并非所有技能需要拖动按键
并非所有技能都有target

貂蝉1技能与2技能是方向性技能，所以当预测出button为skill1或者skill2时，skill X与Skill Z预测结果是有意义的，对于貂蝉3技能，skill X与skill Z无意义



1v1 action space介绍

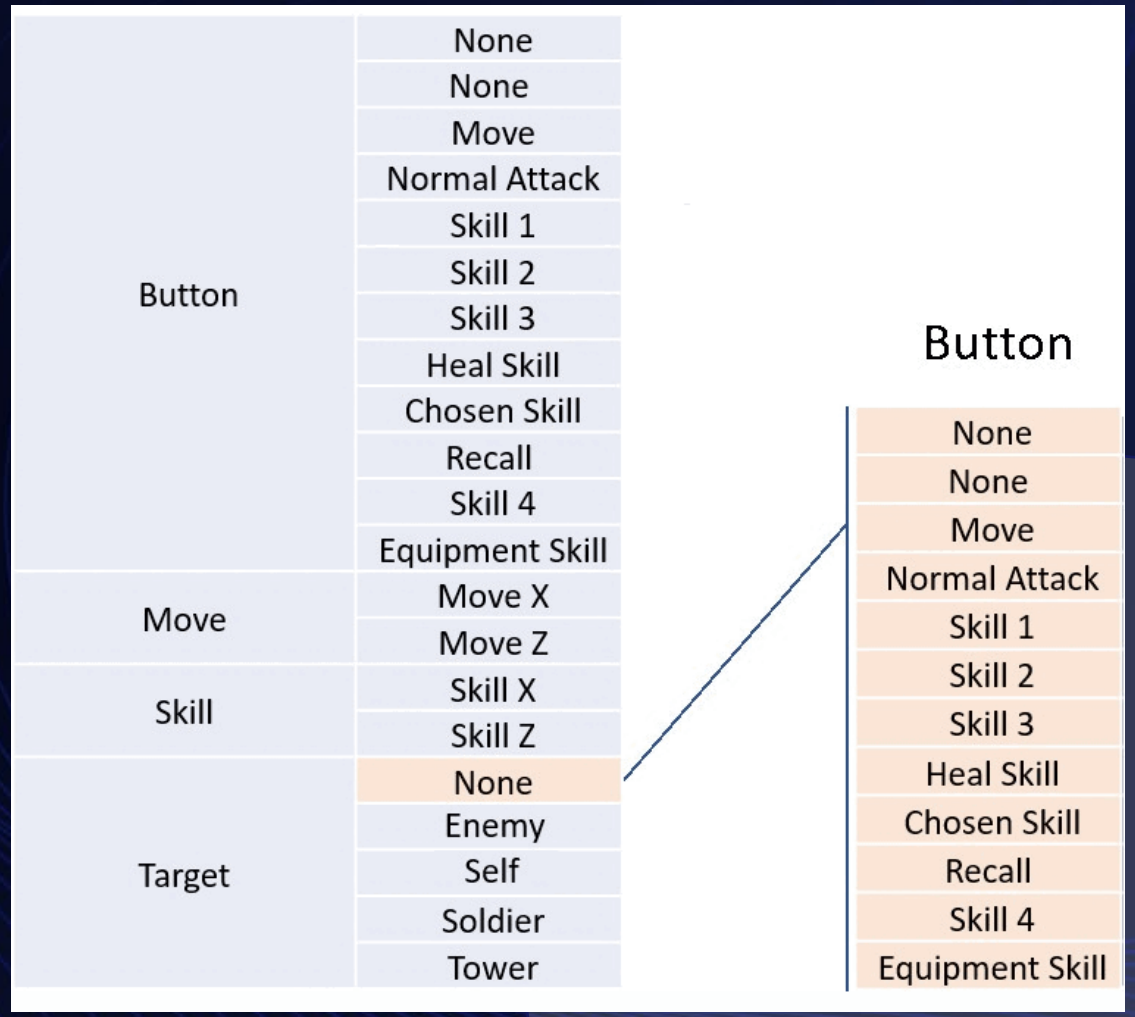
Tencent AI Lab

Legal action

通过游戏规则直接屏蔽掉不合理的预测action

比如CD中技能,不能释放

Legal_action是为了加快训练速度，避免无意义的探索



Legal action维数一共12 (Button) + 16 * 2 (Move) + 16 * 2 (Skill) + 8 (Target) * 12 (Button) 维

1v1 reward体系介绍

1v1 reward体系介绍

经济	英雄血量	塔血量	能量	死亡	击杀英雄	经验	补刀

使用零和reward设计方案，以当前决策帧和上一决策帧的相关数值差作为agent的reward，两个agent的同类reward项相减作为最终reward，最终多种reward项加权求和作为最终的reward返回

Reward设计方面的细节举例：

对英雄的血量开4次方，使得AI对低血量时的血量变化更加敏感

```
reward_struct.cur_frame_value = sqrt(sqrt(1.0 * main_hero.hp / main_hero.max_hp))
```

1v1 reward体系介绍

Tencent AI Lab

经济	英雄血量	塔血量	能量	死亡	击杀英雄	经验	补刀

Reward计算举例：

计算当前帧我方一塔和水晶的总血量并归一化：

```
cur_frame_value =
        1.0 * main_tower.hp / main_tower.max_hp + 1.0 * main_spring.hp / main_spring.max_hp;
```

计算上一帧我方一塔和水晶的总血量并归一化：

```
last_frame_value =
        1.0 * main_tower.hp / main_tower.max_hp + 1.0 * main_spring.hp / main_spring.max_hp;
```

两帧数值相减获得当前英雄当前帧相对于上一帧的reward：

reward_main=cur_frame_value - last_frame_value

同理可以计算reward_enemy
当前英雄reward减去敌方英雄reward获得最终的reward

reward_final=reward_main – reward_enemy

1v1 reward体系介绍

Tencent AI Lab

	Reward	Weight	Type	Description
血量	hp_point	2	dense	the rate of health point of hero
塔血量	tower_hp_point	5	dense	the rate of health point of tower
经济	money (gold)	0.006	dense	the total gold gained
能量	ep_rate	0.75	dense	the rate of mana point
死亡	death	-1	sparse	being killed
击杀	kill	-0.6	sparse	killing an enemy hero
经验	exp	0.006	dense	the experience gained
补刀	last_hit	0.5	sparse	the last hit for soldier

Kill reward为负值原因为：击杀英雄行为本身将会获得多种reward，实际调试时发现如果kill reward为正，AI可能会一直选择击杀英雄而不去推塔杀兵

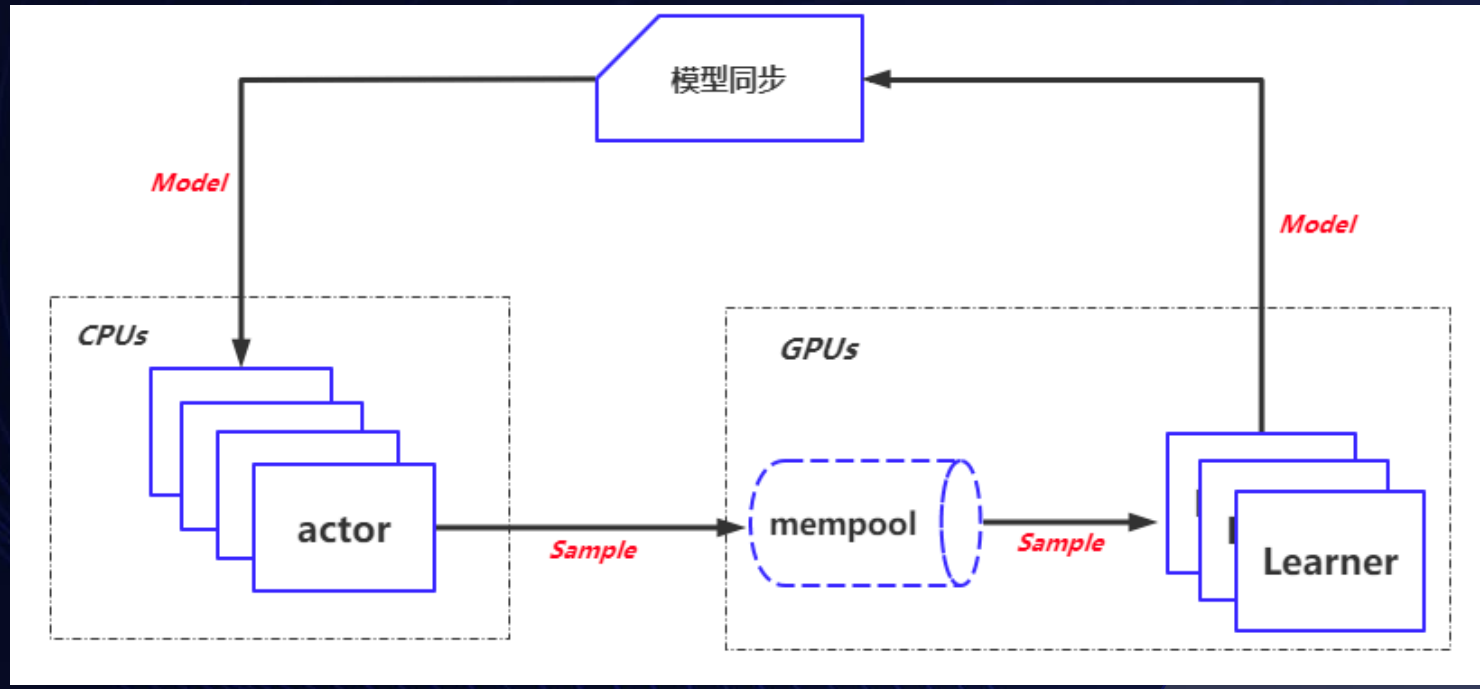
1v1 开悟平台训练框架——SAIL

1v1 开悟平台训练框架-SAIL

Tencent AI Lab

SAIL框架是腾讯针对off-policy强化学习训练提出的大规模分布式训练框架

actor： 加载本地模型，产生训练样本
learner： 进行算法网络的训练，并定期落下模型
mempool： 进行训练数据的存储，所有actor通过mempool将样本发送给GPU机器
模型同步：负责将learner侧的最新模型同步到所有CPU机器



SAIL框架代码在CPU和GPU镜像的/usr/local/lib/python3.6/site-packages/sail路径下，同学们了解各个组件功能即可，不需要研究其源码实现

1v1 开悟平台训练框架-SAIL

Tencent AI Lab

Learner部分代码调用流程

类名	作用
Trainer	SAIL-Learner框架的整体封装类，Learner进程的启动者
Benchmark	为Learner训练提供模型加载与保存，网络训练，日志打印等接口
Datasets	Learner训练的数据加载的中间隐藏层，对上层提供模型训练的数据，并隐藏下层提供的多种IO模式(socket模型，dataop模式以及dataset模式等)
OfflineRLInfoAdapter	训练数据的解析类，提供训练样本大小获取以及样本组织的接口
Graphs	提供Learner构图、前向loss计算与梯度求解以及多机多卡梯度融合的接口
Algorithm	Learner算法的实现类，提供算法图的构建以及优化器获取接口
Model	对网络前向进行封装的类

重点实现
algorithm.py中的build_graph()函数

1v1 开悟平台训练框架-SAIL

Tencent AI Lab

Mempool：

- Mempool使用环形队列存储样本数据，旧的历史数据会随着新的输入加入被删除。
- 抽取样本的方法默认为纯随机抽取，可以定制PER抽取，头部数据随机抽取等方案。
- Mempool中最大样本数量可以设置

1v1 开悟平台训练框架-SAIL

Tencent AI Lab

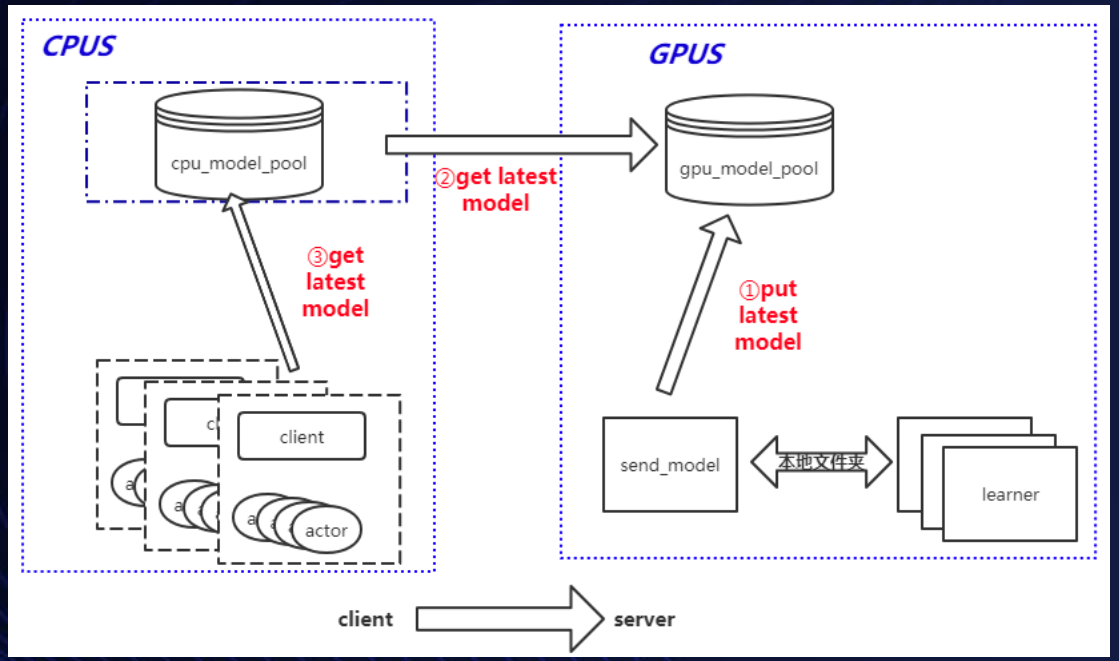
ModelPool：

GPU：运行model pool Server

CPU：运行model pool Client

GPU机器定时push最新模型到所有CPU机器的model pool中

CPU机器上的网络可以选择load最新模型或者某个历史模型



- 开悟平台——王者1v1可以用来做哪些研究？

- 研究方向参考

Tencent AI Lab

- 泛化性相关研究
- Offline RL 与 SL 相关
- 样本利用相关
- 动作空间与reward体系优化

- 研究方向

Tencent AI Lab

- 泛化性研究

1. 貂蝉的模型能否直接用于luna？
2. 貂蝉的模型作为初始化模型训练luna是否有加速比?
3. 一个模型如何cover尽量多的英雄？
4. 哪些英雄在AI眼里比较类似?

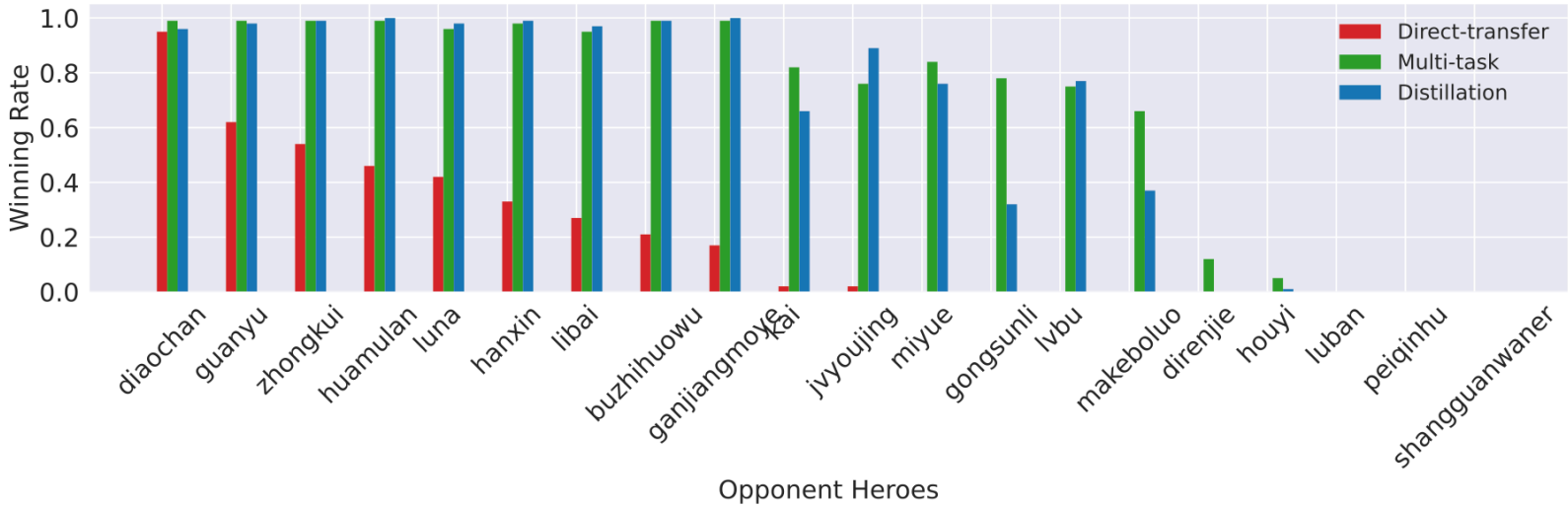


Figure 6: Win rate of a well-trained model from task "Diaochan (RL) vs. Diaochan (*BT*)" transferred to tasks "Diaochan (RL) vs. different opponent heroes (*BT*)". The agent is trained to control Diaochan against Diaochan controlled by *BT*, and tested to control Diaochan against different heroes controlled by *BT*. Red: Directly transferring the model to control Diaochan and compete with different opponent heroes. Green: Multi-task training on five tasks "Diaochan (RL) vs. Diaochan/Buzhihuowu/Luna/Ganjiangmoye/Zhongkui (*BT*)" and testing the model on twenty tasks. Blue: Distilling the model trained from five tasks "Diaochan (RL) vs. Diaochan/Buzhihuowu/Luna/Ganjiangmoye/Zhongkui (*BT*)" and testing the model on twenty tasks. The policy trained on Diaochan could not generalize to all tasks with different *opponent* heroes.

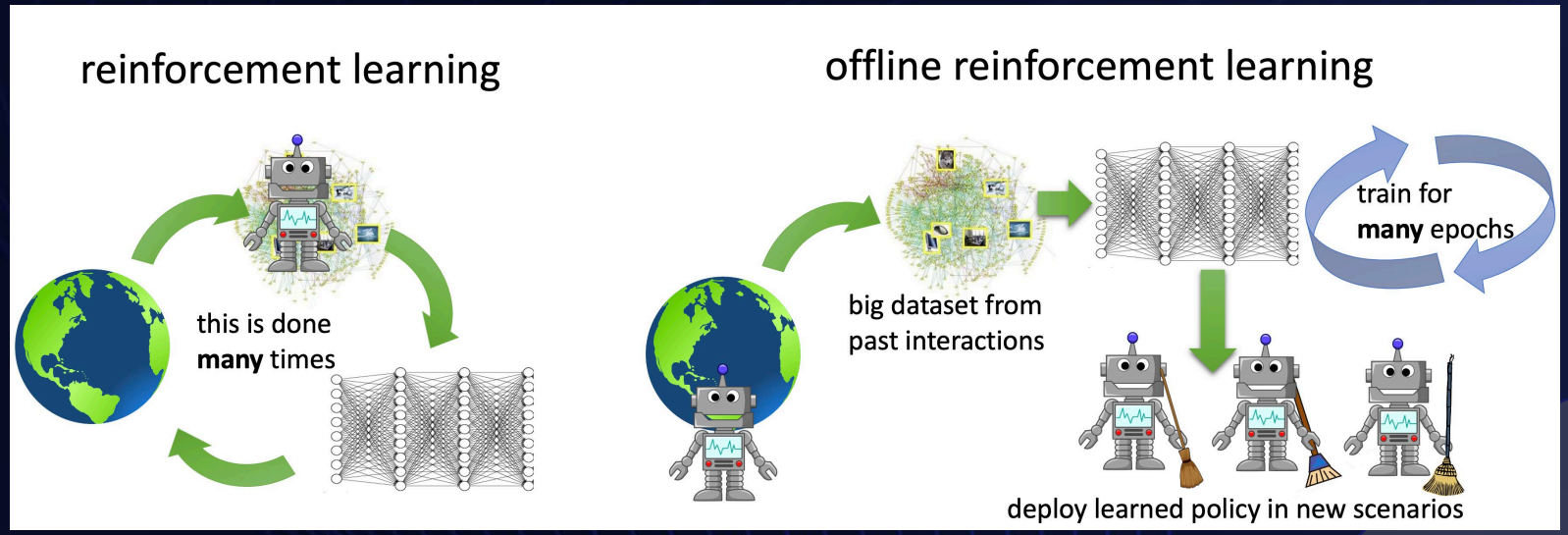
- 研究方向

Tencent AI Lab

- Offline RL与SL相关研究

SL：通过专家数据进行训练（模仿学习，逆强化学习）

Offline RL：通过大量离线对局数据进行RL训练，数据可以是次优



https://bair.berkeley.edu/blog/2020/12/07/offline/

- 研究方向

Tencent AI Lab

- 样本利用

1.如何高效地利用样本?

2.资源有限情况下如何加快学习速度?

3.资源一定情况下如何提升模型上限?

- 动作空间优化

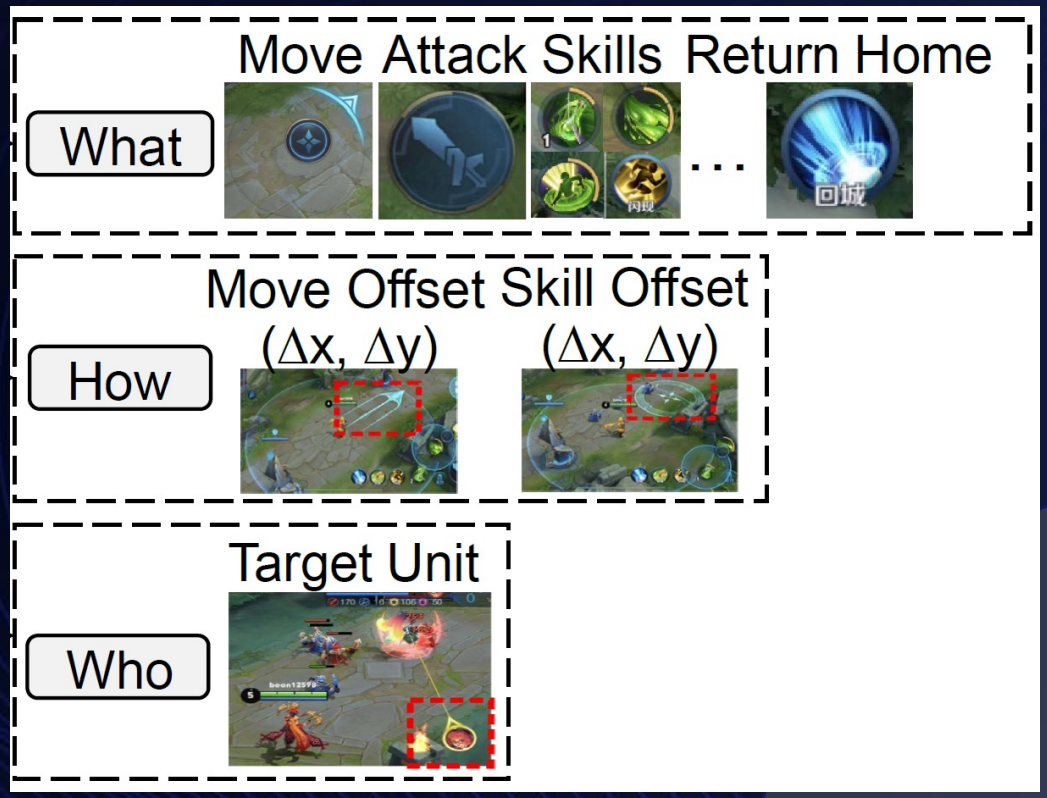
Tencent AI Lab

- 动作空间与reward体系优化

1. 右边这种分层的action space就是最优的吗?

2. 哪些action可以组合起来变成一个action?

3. 长期reward的探索。



- 实验课出题点

- 实验课出题点

Tencent AI Lab

1.使用基本的fc网络结构实现DQN算法与PPO算法
2.使用LSTM实现序列建模
3.使用fc+maxpooling实现同类型角色共享参数建模
4.使用self-attention方案实现技能作用对象的选择
5.使用legal_action对预测action进行mask
6.使用不同的reward权重方案观察AI决策的不同风格
7.对比多英雄-onemodel与单英雄-onemodel实验的AI能力
8.不同的样本读写比对模型收敛速度和能力上限的影响
9.GAE的计算与推导

Thanks