OpenAI Five Dota2 AI 浅谈
马上开始

hululu.zhu@gmail.com

09/2022

OpenAI Five Dota2 AI 浅谈

hululu.zhu@gmail.com

09/2022

纲要

- Dota2 AI 要解决什么问题？
- 为什么高水平 Dota2 AI 非常难？
- OpenAI Five 顶尖的设计和工程实现
- Dota2 一些名场面的 AI 场景设想

请大家保持开麦，随时交流！

Dota 2 with Large Scale Deep Reinforcement Learning

OpenAI, *
Christopher Berner, Greg Brockman, Brooke Chan, Vicki Cheung,
Przemysław "Psyho" Dębiak, Christy Dennison, David Farhi, Quirin Fischer,
Shariq Hashme, Chris Hesse, Rafal Józefowicz, Scott Gray, Catherine Olsson,
Jakub Pachocki, Michael Petrov, Henrique Pondé de Oliveira Pinto, Jonathan Raiman,
Tim Salimans, Jeremy Schlatter, Jonas Schneider, Szymon Sidor, Ilya Sutskever, Jie Tang,
Filip Wolski, Susan Zhang

March 10, 2021

Abstract

On April 13th, 2019, OpenAI Five became the first AI system to defeat the world champions at an esports game. The game of Dota 2 presents novel challenges for AI systems such as long time horizons, imperfect information, and complex, continuous state-action spaces, all challenges which will become increasingly central to more capable AI systems. OpenAI Five leveraged existing reinforcement learning techniques, scaled to learn from batches of approximately 2 million frames every 2 seconds. We developed a distributed training system and tools for continual training which allowed us to train OpenAI Five for 10 months. By defeating the Dota 2 world champion (Team OG), OpenAI Five demonstrates that self-play reinforcement learning can achieve superhuman performance on a difficult task.

1 Introduction

The long-term goal of artificial intelligence is to solve advanced real-world challenges. Games have served as stepping stones along this path for decades, from Backgammon (1992) to Chess (1997) to Atari (2013)[1–3]. In 2016, AlphaGo defeated the world champion at Go using deep reinforcement learning and Monte Carlo tree search[4]. In recent years, reinforcement learning (RL) models have tackled tasks as varied as robotic manipulation[5], text summarization [6], and video games such as Starcraft[7] and Minecraft[8].

Relative to previous AI milestones like Chess or Go, complex video games start to capture the complexity and continuous nature of the real world. Dota 2 is a multiplayer real-time strategy game produced by Valve Corporation in 2013, which averaged between 500,000 and 1,000,000 concurrent players between 2013 and 2019. The game is actively played by full time professionals; the prize pool for the 2019 international championship exceeded $35 million (the largest of any esports game in the world)[9, 10]. The game presents challenges for reinforcement learning due to long time horizons, partial observability, and high dimensionality of observation and action spaces. Dota 2's

*Authors listed alphabetically. Please cite as OpenAI et al., and use the following bibtex for citation: https://openai.com/bibtex/openai2019dota.bib

Dota2 AI **面对的问题**

- 面对连续的带阴影的动态游戏画面（或处理后的信息）
- 需要在很短时间内（比如~200毫秒）做出决策
- 需要 [隐性] 考虑团队5人之间的协作
- 最终目的是为了20-30分钟后能赢

典型的强化学习（reinforcement learning）的应用场景！

- 因为这个AI要去优化一个“延迟的目标”（20分钟以后赢下比赛）
- 有一个完美的环境（游戏引擎）可以自我博弈提高

Dota2 AI 难在哪里？它比 围棋AI 难好多！

Dota2的AI更难	公认很难的围棋(比如 AlphaGo)
- Dota2有地图阴影隐身和雾，信息不完全公开	- 围棋棋盘信息全公开
- 一般几千上万步操作！（每秒几下操作 * 60秒 * 30分钟）	- 一般<200步操作结束
- 每一步面对1920*1080图像，几乎任何位置都能点！即使简化也超过1万种操作可选！	- 每一步最多361个位置选一
- 每一步要在200ms返回！	- 每一步可以等几分钟

Dota2 AI 需要哪些素质？

- 顶尖 Dota2 选手具备的素质
  - 个人能力（打钱对线，技能combo，英雄池）
  - 大局观（装备，合作性，团战时机）
  - 创新性（bp和战斗环节的新思路）
  - 心理素质（逆风局）
  - 个人魅力

- OpenAI Dota2 AI 的目标
  - AI 个人能力（但是因为成本放弃了英雄池比拼）
  - 大局观（隐含在单步决策中）
  - 创新性（期望通过自我博弈实现）
  - 心理素质（只想赢，没感情）
  - 个人魅力（忽略，AI要什么人设）

OpenAI

尝试概括 OpenAI Five Dota2 AI 要做什么

OpenAI Five 每隔一段时间，处理得到的新的游戏信息，来决定每一步的动作选择，目标是来优化最后赢下比赛的可能性

- 需要展开
  - “游戏信息” 如何表述？
  - AI 如何表述和选择 “动作”？

人类 vs OpenAI Five 观察到Dota2的游戏信息

Global data	22
time since game started	1
is it day or night? time to next day/night change	2
time to next spawn: creep, neutral, bounty, runes	4
time since seen enemy courier is that > 40 seconds?[a]	2
min&max time to Rosh spawn	2
Roshan's current max hp	1
is Roshan definitely alive?	1
is Roshan definitely dead?	1
Next Roshan drops cheese?	1
Next Roshan drops refresher?	1
Roshan health randomization[b]	1
Glyph cooldown (both teams)	2
Stock counts[c]	4
Per-unit (189 units)	**43**
position (x, y, z)	3
facing angle (cos, sin)	2
currently attacking?[e] time since last attack[d]	2
max health last 16 timesteps' hit points	17
attack damage, attack speed	2
physical resistance	1
invulnerable due to glyph? glyph timer	2
movement speed	1
on my team? neutral?	2
animation cycle time	1
eta of incoming ranged & tower creep projectile (if any) # melee creeps atking this unit[d]	3
[Shrine only] shrine cooldown	1
vector to me (dx, dy, length)[e]	3
am I attacking this unit?[e] is this unit attacking me?[d,e] eta projectile from unit to me[e]	3
unit type	1
current animation	1

Per-hero add'l (10 heroes)	25
is currently alive?	1
number of deaths	1
hero currently in sight? time since this hero last seen	2
hero currently teleporting? if so, target coordinates (x, y) time they've been channeling	4
respawn time	1
current gold (allies only)	1
level	1
mana: max, current, & regen	3
health regen rate	1
magic resistance	1
strength, agility, intelligence	3
currently invisible?	1
is using ability?	1
# allied/enemy creeps/heroes in line btwn me and this hero[e]	4
Per-allied-hero additional (5 allied heroes)	**211**
Scripted purchasing settings[b]	7
Buyback: has?, cost, cooldown	3
Empty inventory & backpack slots	2
Lane Assignments[b]	3
Flattened nearby terrain: 14x14 grid of passable/impassable?	196
scripted build id next item to purchase[b]	2
Nearby map (8x8)[e]	**6**
terrain: elevation, passable?	2
allied & enemy creep density	2
area of effect spells in effect.[f]	2
area of effect spells in effect.[f]	2
Previous Sampled Action[e]	**310**
Offset? (Regular, Caster, Ward)	3x2x9
Unit Target's Embedding	128
Primary Action's Embedding	128

Per-modifier (10 heroes x 10 modifiers & 179 non-heroes x 2 modifiers)	2
remaining duration	1
stack count	1
modifier name	1
Per-item (10 heroes x 16 items)	**13**
location one-hot (inventory/backpack/stash)	3
charges	1
is on cooldown? cooldown time	2
is disabled by recent swap? item swap cooldown	2
toggled state	1
special Power Treads one-hot (str/agi/int/none)	4
item name	1
Per-ability (10 heroes x 6 abilities)	**7**
cooldown time	1
in use?	1
castable	1
Level 1/2/3/4 unlocked?[d]	4
ability name	1
Per-pickup (6 pickups)	**15**
status one-hot (present/not present/unknown)	3
location (x, y)	2
distance from all 10 heroes	10
pickup name	1
Minimap (10 tiles x 10 tiles)	**9**
fraction of tile visible	1
# allied & enemy creeps	2
# allied & enemy wards	2
# enemy heroes	1
cell (x, y, id)	3

[a] These observations are leftover from an early version of Five which played a restricted 1v1 version of the game. They are

人类接受~1920*1080像素的连续的游戏画面并在大脑处理产生动作决策

AI 接受“特征工程”处理过的游戏画面，得到~16000个观察值（下页详解），并通过神经网络计算产生动作决策

OpenAI Five 游戏信息特征工程的具体例子

Global data	**22**
time since game started	1
is it day or night? time to next day/night change	2
time to next spawn: creep, neutral, bounty, runes	4
time since seen enemy courier is that > 40 seconds?[a]	2

Per-allied-hero additional (5 allied heroes)	**211**
Scripted purchasing settings[b]	7
Buyback: has?, cost, cooldown	3
Empty inventory & backpack slots	2
Lane Assignments[b]	3
Flattened nearby terrain: 14x14 grid of passable/impassable?	196

Per-ability (10 heroes x 6 abilities)	**7**
cooldown time	1
in use?	1
castable	1
Level 1/2/3/4 unlocked?[d]	4
ability name	1

Per-unit (189 units)	**43**
position (x, y, z)	3
facing angle (cos, sin)	2
currently attacking?[e] time since last attack[d]	2
max health last 16 timesteps' hit points	17
attack damage, attack speed	2
physical resistance	1

Per-modifier (10 heroes x 10 modifiers & 179 non-heroes x 2 modifiers)	**2**
remaining duration	1
stack count	1
modifier name	1

Per-item (10 heroes x 16 items)	**13**
location one-hot (inventory/backpack/stash)	3
charges	1

Minimap (10 tiles x 10 tiles)	**9**
fraction of tile visible	1
# allied & enemy creeps	2
# allied & enemy wards	2
# enemy heroes	1
cell (x, y, id)	3

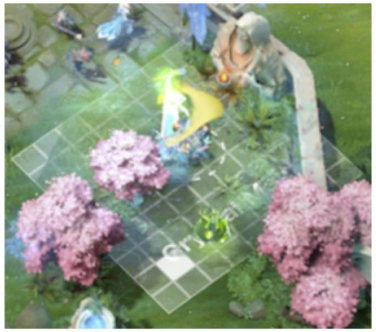
OpenAI Five 游戏定义的 [混合] “动作”空间

- Five “动作” 由“主要动作”和“参数动作”构成（模型输出以下4组信息）

主要动作：0-30的一个数字，代表移动，攻击，技能，tp等 “基本动作”

延迟：0-3的一个数字，代表接下第5帧(167ms)还是第八帧(267)进行操作

单位：0-188的一个数字，地图上最多189个单位（？）



位置偏移：以英雄为中心附近区域划成9x9方块，x和y方向分别-4到4的一个数值表示位置移动

理论上，Five 一共可能有 30 × 4 × 189 × 81 = 1, 837, 080 个混合动作！

实际上，增加了很多过滤，一般在 8000 到 80000 之间

再次概括 OpenAI Five Dota2 AI 要做什么

- 每隔167ms(30hz的游戏每隔4帧)
- 处理特征工程处理过的~16000个数据信息(非原始图像)
- 输出一个混合动作，包括
  - 基本动作选择，比如移动攻击还是放技能
  - 施放延迟的选择，167ms，200ms，233ms 还是 267ms
  - 目标单位选择
  - 二维的位置信息
- 目标是来优化最后赢下比赛的可能性

OpenAI Five 模型和训练简介

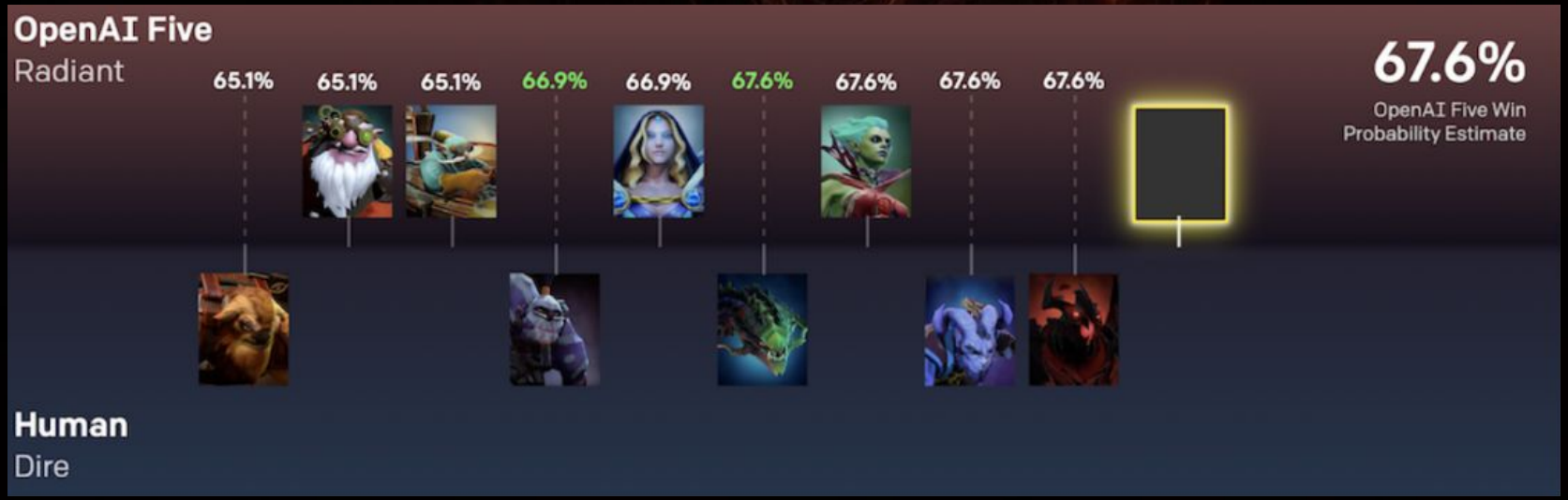
- 模型超简介
  - 大约1亿6千万参数的基于LSTM（长短期记忆）的深度网络
  - 模型奖励设计除了赢下比赛，加入了经验，杀人，拿塔奖励分数
    - 另外加入了特别的“Team Spirit”来平衡团队胜利中个体的奖励差异
  - 使用优秀的PPO算法（腾讯之后训练王者AI 改进了PPO，DeepMind用自研VMPO）

- 基于自我博弈的训练提高
  - 类似于AlphaGo Zero（不模仿人类）的自我博弈
  - 模型的5个拷贝控制5个英雄，互相博弈，replay的batch用来训练提高
  - 顶尖的训练系统：5万多CPU+1千多GPU并行，接近40%(待确认)使用效率的训练吞吐量

- 随版本更新的 “surgery” 创新，进化适应
  - 适应英雄技能属性等的 删减，位置的变化...
  - 如何让模型“向后兼容”去重用已有模型并减少 训练时间和成本

OpenAI Five 选人(drafting)

- 只有17个无召唤的常用英雄可选
- 基于自我博弈数据数据的统计
  - Five 认为天辉先选是 54%胜率，夜魇是53%胜率（符合天辉胜率高一点点的共识）
- minimax 算法(假设对手也选择最优操作的递归选择算法)
- 一个简单的web界面，没有ban



OpenAI Five 有没有非“纯粹能力”外的优势？

- [大概公平] Five 在接收当前状态到输出下一个操作有~200ms延迟
  - 应该多于顶尖选手反应时间
- [AI 略微优势] Five 设定在30hz游戏引擎上每4帧处理一次动作
  - 所以 Five 大概持续输出一秒7.5个 [混合] 动作，超过了顶尖选手的APM
- [AI 优势] Five 在任一时刻有所有友军的地图信息
  - 顶尖选手即使在近距离交流情况下也不能瞬 时有全部信息
- [AI 优势] Five 决赛版限制只有 17个英雄可选/ban，并限制装备选择
  - 顶尖选手英雄池的实力没法体现(比如TI决赛圈一般一个队伍选过的英雄都在~40样子)
  - 这些召唤能力的装备都不可用幻想符，支配 头盔，分身斧，死灵 书
- [人类 优势] Five 技能成长，买装备，用信使的顺序都预先设定
  - 我的理解是技能装 备和信使使用都是随机 应变会更好

世界冠军 OG 队员 Ceb 比赛后的反应

Sébastien "Ceb" Debs
Member of OG

You play against [OpenAI Five] and you realize it has a playstyle that is different. It's doing things that you've never done and you've never seen. Sometimes it looks extremely silly. But then again, are you going to be human and be like "Hey, this looks very stupid, this is bad" or [do] you try to take it to next steps, like "Why is it doing this?"

One key learning that we took is how it was allocating resources. It's just allocating resources as efficiently as possible. And then you realize that we're guilty of being stuck in a team dynamic, whereas sometimes we have to be way more flexible. [...] If OpenAI does that dynamic switch at 100%, we maybe went from 5% to 10%? But that is already a difference—we've noticed it.

世界冠军 OG 队员 N0tail 比赛后的反应

Johan “N0tail” Sundstein
Captain of OG

I don’t believe in comparing OpenAI Five to human performance, since it’s like comparing the strength we have to hydraulics. Instead of looking at how inhuman and absurd its reaction time is, or how it will never get tired or make the mistakes you’ll make as a human, we looked at the patterns it showed moving around the map and allocating resources.

In terms of what OpenAI has done for us and how it influenced our run at TI9, one of the many curious patterns was the buyback and pressure play that happened in most of the games. We had a lot of talks about fighting and pressuring and how it used a different approach from any human in the past. As people, it’s about being realistic and learning from the brain of the AI and not the hydraulic strength that machines have.

名场面 之 无敌微操？AI 可否一战？

名场面 之 决赛选屠夫？AI 也会这样选吗？

名场面 之 就放给你猛犸？AI 是否也这么头硬？

名场面 之 n张TP亮起？AI 能否也让你感动？

相关资料

- OpenAI Five blog 链接
- OpenAI Five 论文 链接
- 腾讯王者荣耀AI 论文 链接
- DeepMind AlphaStar 论文 链接

感谢参加讨论！
OpenAI Five Dota2 AI 浅谈

hululu.zhu@gmail.com

09/2022