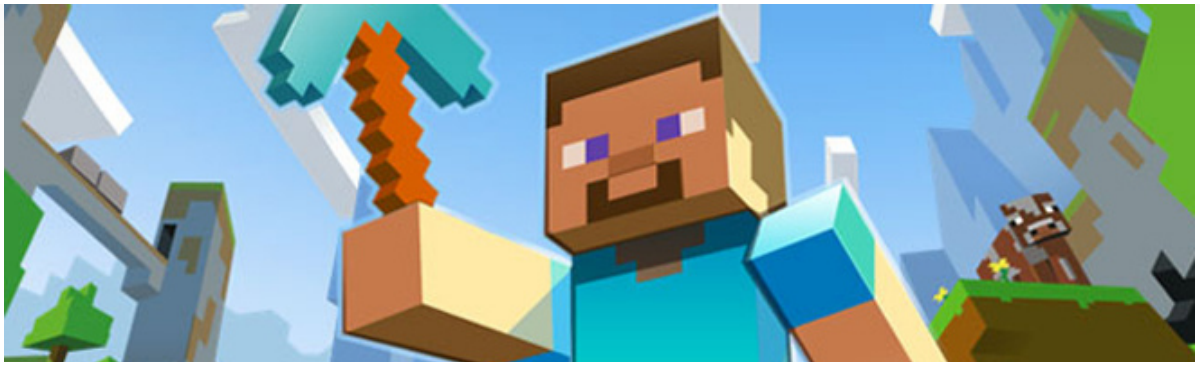


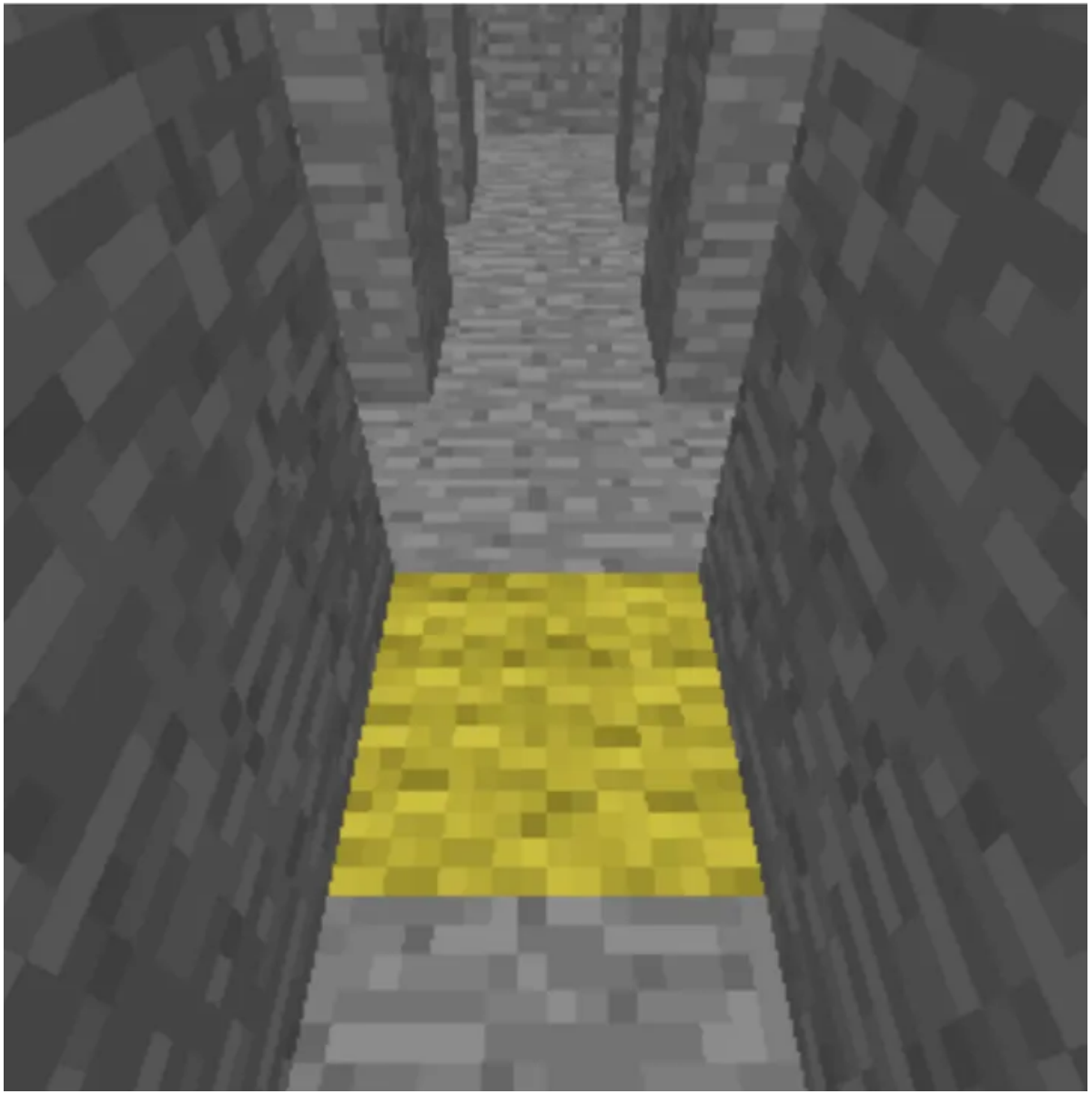
最前沿：史蒂夫的人工智能大挑战


Flood Sung
人工智能等 2 个话题下的优秀答主

> **Flood Sung**：arXiv上有篇新论文，利用像素游戏我的世界来做**深度增强学习**的实验。
> **杜客**：有意思！搞一篇既面向大众进行**趣味科普**，又面向领域内进行**论文解读**的文章？
> **Flood Sung**：OK，就这么干！

1 史蒂夫的挑战

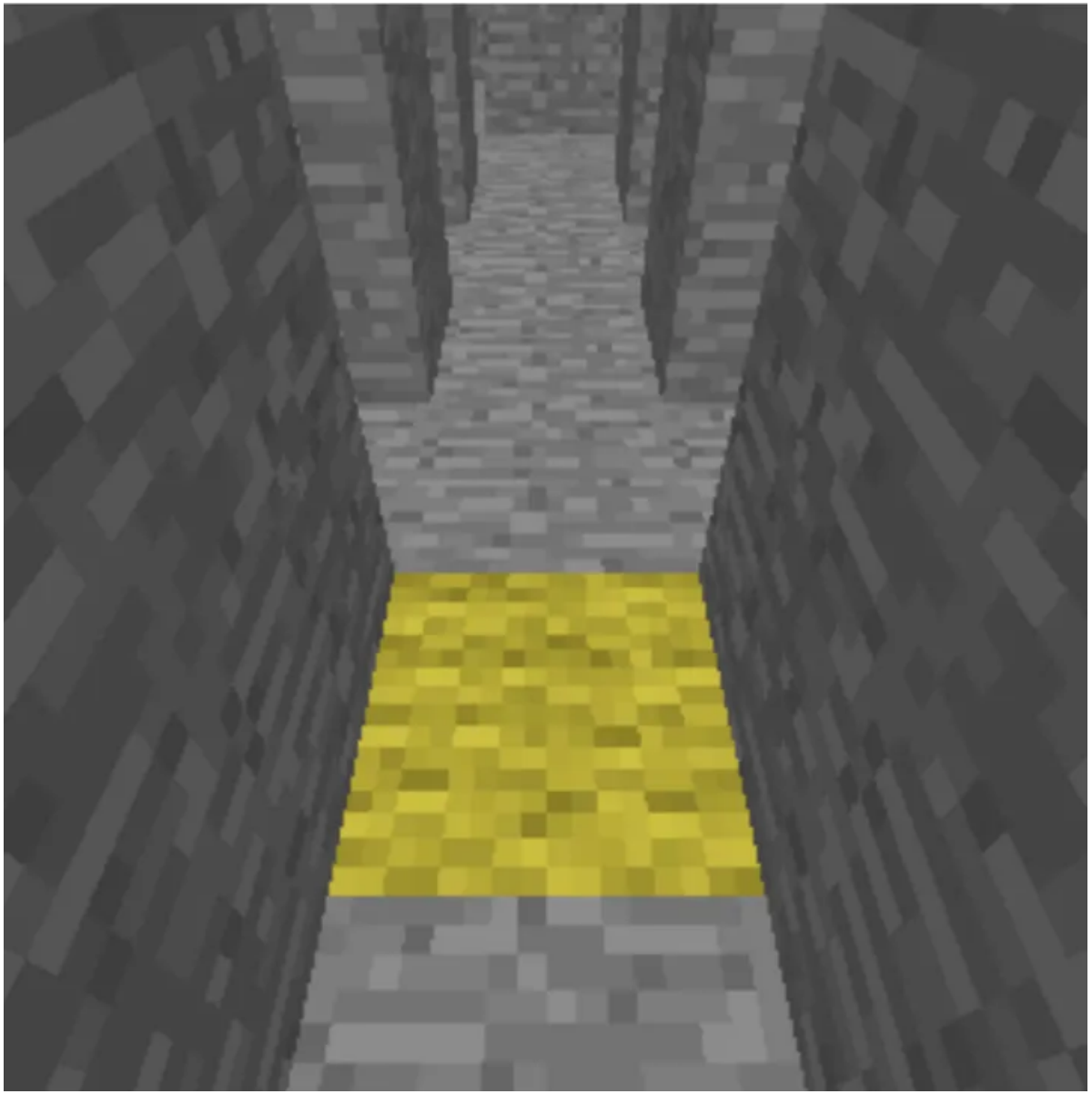
你从**昏睡中醒来**，头痛欲裂，早知昨晚就不要熬夜守望了。揉揉眼睛，睁开朦胧的双眼.....



我擦嘞？！这是什么鬼！难倒是[哔—]多了导致视力下降了？世界怎么变得如此马赛克？确认这不是在做梦，你（可怜的史蒂夫）不知所措，开始漫无目的的在这迷宫中奔跑，突然，前方出现了一个·

蓝色的标志。史蒂夫完全没有在意，径直冲过去，在踏上蓝色方块的那一刹那，一股强大的电流伴随着痛苦通过你的身体，**Dead**。

在短暂的失去意识后，史蒂夫再一次重生在了起点处，脑海里依然能够回忆起刚才的遭遇，眼前的场景依旧是：



这一次，史蒂夫（原游戏宅男）冷静了下来，片刻思索后。史蒂夫轻轻地扶了下黑框眼镜，自言自语道：“哈哈蛤，我可是身经百战了……真相只有一个！”

史蒂夫知道，自己陷入了一个类似明日边缘与Cube的组合迷宫游戏中。一方面，自己可以和明日边缘里的阿汤哥一样，一次又一次地重生。另一方面，面前的迷宫和Cube一样，暗藏杀机的同时也有其规律，需要自己用智慧去发现。想通了这一点，史蒂夫的脑海中回响起了一句呐喊：

> **前进！前进！不折手段地前进！！** —— Thomas Wade

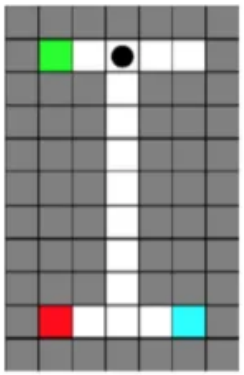
于是怒吼着“Freedom！！！”，史蒂夫再一次冲向了迷宫的深处。可惜这迷宫中没有可口的怪物。

经过一次次生与死的轮回，史蒂夫终于找到了，哦不，发现了这个迷宫的规律：这个迷宫的形态每次都不一样，但是其内在蕴含的任务目的是一致的。如果自己重生后，面前的方块是黄色的，那么自己就要在迷宫中寻找红色方块并踩上去，这样就能得到奖励并进入下一次循环。在这个过程中不能触碰到蓝色方块，不然就会被电击惩罚致死。但是如果重生后，面前的方块是绿色的，那么就要避免触碰红色方块，寻找到蓝色方块。还有，自己不能无限制地在迷宫中兜圈，在一定步数内不找到正确的方块，游戏也会重启。

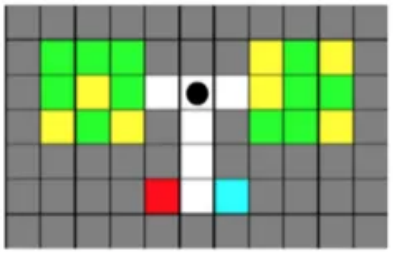
正当史蒂夫为找到了规律洋洋得意时，视野左下角一个不起眼的绿色提示语“**训练时间**”变成了红色的“**测试时间**”，看来，史蒂夫的试炼还将继续……

2 迷宫设计

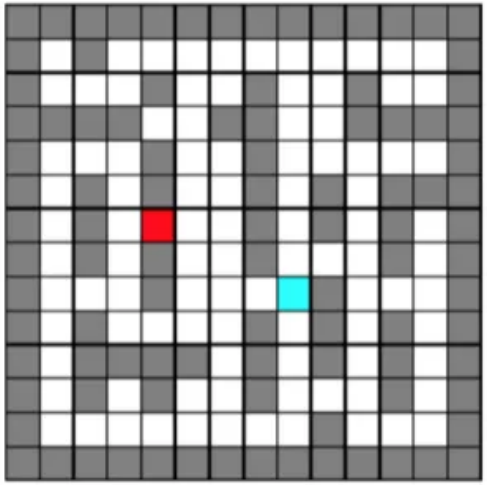
现在，先让史蒂夫忙活去吧，我们输入“**who is your daddy**”，打开上帝视角，看看都**有什么样的迷宫**等待着可怜的史蒂夫，而这些**迷宫中又有什么样的规律**需要史蒂夫去发现：



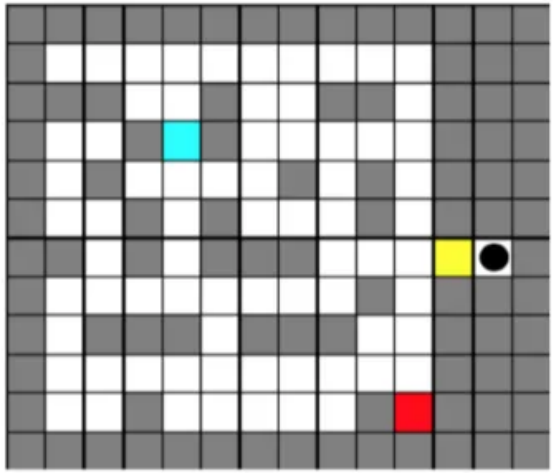
(a) I-Maze



(b) Pattern Matching



(c) Random Maze



(d) Random Maze w/ Ind

如上所示，一共有**I型迷宫、匹配迷宫、随机迷宫和带指示器的随机迷宫**4种迷宫。每种迷宫的任务设计即有相似，又有区别：

I型迷宫：在此迷宫中，有一个指示器，其颜色为绿色或黄色，两种颜色出现的几率一致。当指示器颜色为绿色时，史蒂夫需要前往蓝色处得到+1的奖励，如果去了红色处，则会得到-1的惩罚。当指示器颜色为绿色时，情况相反。

匹配迷宫：在此迷宫中，有两个房间。如果两个房间中底板的图样一致，史蒂夫需要前往蓝色方块处，得到+1奖励，如果去了红色则-1。如果两个房间中底板的图样不一致，那么情况相反。

随机迷宫：迷宫的形态每次都是随机产生的，其中有两种类型的任务：

- 单一目标：找到蓝色方块+1奖励，如果过程中触碰了红色方块则惩罚-1。
- 顺序目标：先踩上红色方块得到+0.5奖励，再踩上蓝色方块得到+1奖励。如果踩方块顺序错误，则得到-0.5和-1的惩罚。

带指示器随机迷宫：情况和随机迷宫类似，但是有指示器，会显示绿色或者黄色，两种颜色显示概率相等，也有两种任务：

- 有指示器的单一目标：如果指示器是黄色，那么找到红色方块得到+1奖励，踩上蓝色方块则-1惩罚（史蒂夫的第一次死亡正是在这种情况下）。如果指示器是绿色，那么那么找到蓝色方块得到+1奖励，踩上红色方块则-1惩罚。
- 有指示器的顺序目标：如果指示器是黄色，那么史蒂夫要先踩上蓝色方块，再踩上红色方块。如果指示器是绿色，那么史蒂夫要先踩上红色方块，再踩上蓝色方块。顺序正确则分别得到+0.5和+1的奖励，顺序错误就得到-0.5和-1的惩罚。

通过**多次的循环以及每个循环中的奖励与惩罚**，我们训练着史蒂夫，希望他能发现其中的规律，理解迷宫的内在逻辑，并以此作为自己的行动指南。

读者们也可以换位思考一下，如果不打开上帝视角，**处在史蒂夫的位置**，以上所有迷宫中的所有任务的正确逻辑，**你需要多久才能完全有把握的发现呢**？

3 测试时间

了。

情况果然有了不同：当史蒂夫按照之前发现的规律，踩上的自认为正确的方块后，游戏立即重启了，奖励消失了。游戏一次次重复着，史蒂夫有时候不小心踩到了按照之前的规律错误的方块，也没有得到惩罚，游戏还是立即重启了。没有了奖励和惩罚，史蒂夫感到迷茫，只能根据训练中学到的知识，继续游戏着……

在一次次轮回中，有些迷宫的形状也起了变化，之前训练时I型迷宫的纵向走廊的长度往往是5步、7步和9步。而现在迷宫纵向走廊的长度有了更多变化，这些变化是否意味的规律的改变？史蒂夫心生疑惑。史蒂夫继续游戏着……

在一次次轮回中，史蒂夫感到黑暗之中有一个恒定的视线观察着他，这个视线无法看见，却又无处不在，记录着他的每一步，每一轮，每一个选择。这让史蒂夫感到不寒而栗……

在一次次轮回中，史蒂夫甚至会想，也许有其他的人，也在这迷宫中不断地轮回游戏着，而那个视线同样监视着他们。这个视线的来源是什么，也许是一种更加高级的存在？他们有什么目的？史蒂夫无法回答，只能继续游戏着……

测试时间仿佛永无止境，已经放弃了希望的史蒂夫行尸走肉般地又一次完成了一个轮回。忽然，一切停止了，几张表格覆盖了天空，史蒂夫仔细仰望着它们：

SIZE	TRAIN	DQN	DRQN	MQN	RMQN	FRMQN
4		92.1(1.5)	94.8(1.5)	87.2(2.3)	89.2(2.4)	**96.9**(1.0)
5	✓	**99.3**(0.5)	98.2(1.1)	96.2(1.0)	98.6(0.5)	**99.3**(0.7)
6		**99.4**(0.4)	98.2(1.0)	96.0(1.0)	99.0(0.4)	**99.7**(0.3)
7	✓	99.6(0.3)	98.8(0.8)	98.0(0.6)	98.8(0.5)	**100.0**(0.0)
8		99.3(0.4)	98.3(0.8)	98.3(0.5)	98.0(0.8)	**100.0**(0.0)
9	✓	99.0(0.5)	98.4(0.6)	98.0(0.7)	94.6(1.8)	**100.0**(0.0)
10		96.5(0.7)	97.4(1.1)	98.2(0.7)	87.5(2.6)	**99.6**(0.3)
15		50.7(0.9)	83.3(3.2)	**96.7**(1.3)	89.8(2.4)	**97.4**(1.1)
20		48.3(1.0)	63.6(3.7)	97.2(0.9)	96.3(1.2)	**98.8**(0.5)
25		48.1(1.0)	57.6(3.7)	**98.2**(0.7)	90.3(2.5)	**98.4**(0.6)
30		48.6(1.0)	60.5(3.6)	**97.9**(0.9)	87.1(2.4)	**98.1**(0.6)
35		49.5(1.2)	59.0(3.4)	**95.0**(1.1)	84.0(3.2)	**94.8**(1.2)
40		46.6(1.2)	59.2(3.6)	77.2(4.2)	71.3(5.0)	**89.0**(2.6)

“啊，这是我走**I型迷宫**的成功率啊！训练的时候迷宫走廊长度只有5，7和9步那么长，测试的时候原来有这么多不同长度的走廊。记起来了，原来我的名字是**FRMQN**才对。果然也有其他人在同时游戏。

	TRAIN	UNSEEN
DQN	62.9% (±3.4%)	60.1% (±2.8%)
DRQN	49.7% (±0.2%)	49.2% (±0.2%)
MQN	99.0% (±0.2%)	69.3% (±1.5%)
RMQN	82.5% (±2.5%)	62.3% (±1.5%)
FRMQN	**100.0%** (±0.0%)	**91.8%** (±1.0%)

“这不就是**匹配迷宫**的游戏结果嘛，看来我在训练和测试的时候都比其他人做得好啊！”史蒂夫有点高兴。

Task	Type	Size	DQN			DRQN			MQN			RMQN			FRMQN		
			Reward	Success	Fail	Reward	Success	Fail	Reward	Success	Fail	Reward	Success	Fail	Reward	Success	Fail
Single	Train	4-8	0.31	90.4%	0.6%	**0.45**	94.5%	0.1%	0.01	78.8%	0.4%	**0.49**	95.7%	0.1%	**0.46**	94.6%	0.3%
	Unseen	4-8	0.22	87.3%	0.7%	0.23	86.6%	0.2%	0.02	79.4%	0.3%	**0.30**	89.4%	0.3%	**0.26**	88.0%	0.5%
	Unseen-L	9-14	**-0.28**	70.0%	0.3%	−0.40	63.0%	0.1%	−0.63	54.3%	0.4%	**-0.28**	69.3%	0.1%	**-0.28**	69.0%	0.1%
Seq	Train	5-7	−0.60	47.6%	0.8%	−0.08	66.0%	0.6%	−0.48	52.1%	0.1%	**0.21**	77.0%	0.2%	**0.22**	77.6%	0.2%
	Unseen	5-7	−0.66	45.0%	1.0%	−0.54	48.5%	0.9%	−0.59	48.4%	0.1%	**-0.13**	64.3%	0.1%	**-0.18**	63.1%	0.3%
	Unseen-L	8-10	−0.82	36.6%	1.4%	−0.89	32.6%	1.0%	−0.77	38.9%	0.6%	**-0.43**	49.6%	1.1%	**-0.42**	50.8%	1.0%
Single+I	Train	5-7	−0.04	79.3%	6.3%	0.23	87.9%	1.2%	0.11	83.9%	0.7%	**0.34**	91.7%	0.8%	0.24	88.0%	1.4%
	Unseen	5-7	−0.41	64.8%	16.1%	−0.46	61.0%	13.4%	−0.46	64.2%	7.8%	**-0.27**	70.0%	10.2%	**-0.23**	71.8%	8.2%
	Unseen-L	8-10	−0.74	49.4%	31.6%	−0.98	38.5%	28.3%	−0.66	55.5%	17.1%	**-0.39**	63.4%	20.4%	**-0.43**	63.4%	17.2%
Seq+I	Train	4-6	−0.13	68.0%	7.0%	0.25	78.5%	1.1%	−0.07	67.7%	2.3%	0.37	83.7%	1.0%	**0.48**	87.4%	0.9%
	Unseen	4-6	−0.58	54.5%	14.5%	−0.65	48.8%	9.7%	−0.71	47.3%	7.2%	**-0.32**	62.4%	7.2%	**-0.28**	63.8%	7.5%
	Unseen-L	7-9	−0.95	39.1%	17.8%	−1.14	30.2%	13.1%	−1.04	34.4%	9.9%	**-0.60**	49.5%	12.5%	**-0.54**	51.5%	12.9%

“这大概就是**两种随机迷宫中的一系列任务**了吧，虽然不是每个任务都第一，但是总体看来还是我比较厉害。”史蒂夫感动一丝自豪。表格再没有更新，一阵阵震耳欲聋的响声传来，四周的世界在迅速崩塌。史蒂夫心想：“终于结束了”。这时，那个更高级的存在终于发出了声音，渐渐堕入黑暗的史蒂夫依然能够感受到那声音中蕴含的狂喜，那个存在说道：

> **哈哈，结果不错，可以发论文啦！论文就叫Control of Memory, Active Perception, and Action in Minecraft吧！！赶紧投到arXiv上占坑！**

4 从故事到论文

略黑暗的史蒂夫故事讲完啦，即使是不熟悉深度增强学习的读者，应该也能隐隐约约地猜到我们要介绍的这篇论文做了什么。是的，研究人员通过论文，介绍了他们所做的工作，主要是两方面：

- 使用我的世界这个游戏**创建了上文介绍的一系列的迷宫和任务组成的任务集**。
- 使用任务集训练，然后使用测试集**测试了几种已有的和他们提出的新的深度增强学习结构**（即**FRMQN**）。并根据测试数据，指出FRMQN的表现是优于已有的结构。

说到这里，可能有一些不熟悉深度增强学习的读者会说：“什么前沿嘛？！不就是走迷宫嘛，这样的算法不是老早就有了么？”

大家注意，这里的走迷宫任务有什么不一样呢？首先，里面的史蒂夫和人类一样，输入的是看到的图像信息，然后根据图像信息作出动作决策。而很多走迷宫的游戏是二维的，只有简单的平面信息。其次，更重要的是，在这个任务中，我们要求史蒂夫自己通过学习训练找到方法，而不是我们

放入了一只真实的小白鼠一样，然后让它自己学习规则，只是这一次我们放入的是一只虚拟的人工智能小白鼠罢了。

而深度增强学习算法，就是这种人工智能小白鼠的“大脑”。算法能看（直接输入算法的就是史蒂夫的第一人称视角图像），算法能控制史蒂夫的动作（论文中规定，算法控制史蒂夫的动作有6种：向左右90°看，向上下45°看，前进与后退），然后你再设定好迷宫及其任务，接下来，就是让史蒂夫在N多的迷宫中一次又一次的重复游戏，然后学习其中的规律完成任务。

我们再来看看这些任务，这些任务被称为认知启发任务（Cognition-inspired tasks)。也就是说，这些任务需要一定的认知能力才能完成。想象一下我们人在这些迷宫中做这些任务。我们需要思考（比如碰到红色会电死），我们需要记忆（比如匹配迷宫任务中两个房间的形状），我们甚至需要推理（比如匹配迷宫中如果房间形状相同则选择蓝色），我们还需要决策（也就是每一个时刻应该往左边走还是右边走）。

噢，我的上帝呀！这对计算机而言是多么困难的任务！

然后我们看看训练的效果，以**匹配迷宫**任务为例，我们再次看看下面这张图表：

	TRAIN	UNSEEN
DQN	62.9% (±3.4%)	60.1% (±2.8%)
DRQN	49.7% (±0.2%)	49.2% (±0.2%)
MQN	99.0% (±0.2%)	69.3% (±1.5%)
RMQN	82.5% (±2.5%)	62.3% (±1.5%)
FRMQN	**100.0%** (±0.0%)	**91.8%** (±1.0%)

通过训练，使用FRMQN结构的算法最后能够在训练集的所有任务中顺利达成目标。而算法在面对其**并没有见过的测试用的匹配迷宫**时，凭借它在训练中的得到的“经验”，顺利达成任务的概率是91.8%（这个概率是多次测试实验结果所得）。

直至其算法本身性能的限制.....

> **通过自学来完成这些认知任务，从某种程度上讲意味着这个史蒂夫具备了真正的智能！类人的智能行为！**

这就是当前人工智能最前沿的研究水平！哇！完蛋啦！天网要来啦！康纳在哪里？！不对，康纳居然是这样的康纳？！绝望了！还是抱莎拉龙妈的大腿吧！

哎，我们**旗帜鲜明地反对类似“人工智能毁天灭地屠人类”的各种说法**，写这篇趣味科普与论文解读结合的文章，也就是为了让更多读者能够在轻松愉快中理解一些深度增强学习的工作，**之所以用人工智能做标题，也算是小小的标题党了**。恐惧来源于未知，知友们如果有兴趣，请继续阅读下去，我们接下来将对论文进行深度解读，专业名词和干货的密度将增多，但是牢记：

> **Don't be panic. —— 银河系漫游指南**

5 理解这篇论文的基础

这篇论文可以说是集当前深度学习前沿技术之大成，论文涉及到以下的概念：

- **DQN**
- **Neural Turing Machine（NTM）中的Memory Network**
- **CNN**
- **LSTM**
- **Feedback Control**

这里我们只能尽量用朴素的语言来描述这些概念。如果大家对以上的概念都了解，那么理解这篇文章将比较简单，可以略过这部分内容。但如果仅知道一部分或者完全不了解的话，在理解上确实会有一些难度，可以把上面的概念当做黑盒来处理。

DQN是Deepmind提出的深度增强学习算法，最初用在玩Atari游戏上。感兴趣的读者，可以阅读本专栏开设的DQN 从入门到放弃系列文章。DQN最基本的算法可以这样理解：使用一个卷积神经网络CNN来表示动作价值函数Actor-Value FunctionQ(s,a)，输入是图像，输出是每一个动作的Q值。然后在反复训练过程中，使用Q-Learning算法来计算目标的Q值，从而可以和当前的Q值做比较得到损失函数，从而进行随机梯度下降SGD更新Q值，从而提升游戏的水平。

$$L(w) = \mathbb E[({\underbrace{r+\gamma\max_{a^\prime}Q(s^\prime,a^\prime,w)}_{Target}}-Q(s,a,w))^2]$$

这篇文章的主要工作就是使用其他结构的网络来代替这里的CNN，从而提升算法的效能。

5.2 Neural Turing Machine

Neural Turing Machine神经图灵机就是为神经网络构造了一个记忆结构，可写可读，并且可以通过随机梯度下降来更新记忆结构的参数，从而优化记忆。在这篇文章中，使用的记忆结构类似于Memory Network。这个记忆结构的读操作和写操作可以用函数表示，并且完全可微，因此可以很方便地使用随机梯度下降来更新。这种更新操作就类似我们人类对某段记忆的印象变化。有的记忆重要，就加深记忆，有的记忆不重要，就遗忘掉。

5.3 CNN卷积神经网络和LSTM长短记忆模型

CNN和LSTM都是神经网络的一种结构，CNN面向图像处理有较大优势，而LSTM属于RNN的一种，具备一定的记忆能力，对于处理时间序列的数据（比如自然语言处理）有较大优势。相信对深度学习有了解的知友们都有所了解，这里不具体介绍。现在很多情况下可以把CNN和LSTM结合起来使用，比如Image Captioning图片标注任务。同样的，在DQN中，将LSTM与CNN而形成新的神经网络结构也是可取的。

5.4 Feedback Control反馈控制

多行为都是反馈控制的结果。比如我们要拿起一支笔，我们会使用眼睛不断判断手与笔之间的偏差（输出），然后根据偏差（输出）来调整我们对手的控制（输入）。这篇文章最成功的思想就是结合了反馈控制大大提高了神经网络的训练效果。

以上就是相关概念的基本介绍，限于篇幅，不能面面俱到。大家知道这些概念是做什么的就可以了，不必细究具体的实现。接下来我们分析这篇文章的方法。

6 不同的神经网络结构

上图列举了几种不同的神经网络结构。首先大家要记住一点的是这篇文章使用的算法是DQN算法，Q-Learning是没有变的，**变的只是使用的Q网络的结构**。我们来看看这几种不同的结构都有什么不同之处：

6.1 DQN

DQN：原始的DQN只使用CNN卷积神经网络。输入是多帧的图像（一般是4帧），输出直接就是Q

6.2 DRQN（Deep Recurrent Q-Learning)

这个结构的思路很简单，就是将DQN中的CNN最后的全连接层改成一个LSTM。从而为神经网络增加一定的记忆能力。这里的输入就变成是一帧一帧图像的输入了。具体的结构如下图所示（图片引用自arxiv.org/pdf/1507.0652...）：

有些知友可能不是很理解LSTM，不知道recurrent体现在哪。首先推荐这篇介绍文章 colah.github.io/posts/2...。这里简单说明一下，就是LSTM包含了一个叫做Cell的单元。Cell可以认为是存储记忆的细胞，就是每次输入信息，都会存储一些信息在Cell里面，之后调用输出就从Cell里面获取一些信息。这样每次的输入都是有用的，也就是有了记忆功能。而只使用CNN的话就只有前向传播，上一组图像（历史）并不影响新的图像输入产生的输出。

6.3 MQN、RMQN和FRMQN

这是文章中作者提出的三种新的结构，最基本的就是加上了Memory。所以**重点在于Memory怎么加的问题。**

前面说过，这篇文章使用了类似Memory Network的结构，具体如下图所示：

我们来解释一下这个Memory是怎么工作的。

6.3.1 Encoding编码

这一步和DQN的神经网络一样，就是使用一个CNN来提取特征，不同的是这里只是到全连接层，目的是为了获取经过CNN压缩的图像特征信息。上图的\varphi 就是表示这个卷积处理的过程。

6.3.2 Memory 记忆

一般磁盘的读写操作是怎样的？就是把数据直接写入相应的磁盘扇区。然后读取就是根据地址到相应的扇区进行读取。但是我们需要怎样的读写方式呢？我们需要能够模拟人类的记忆行为，就是我们可以根据需要来读取我们的记忆。比如我们看到猫这个字就会联系到猫的样子。我们输入了“猫”到大脑中，然后大脑根据这个输入找到了对应的记忆的位置，并读取出来。所以一般的磁盘读写操作和这边模拟记忆的读写是两个概念，我们需要能够确定这个记忆的地址，而不是读取这个行为。还有一个很重要的功能就是我们的记忆是可以调整的。依然是猫的例子，我们大脑中的记忆是不断变化的，可能这个时候我们看到了一只很特别的猫，然后以后看到”猫“这个字我们就想到了那只特别的猫的样子了。

稍微归纳一下我们需要构造的记忆模块需要的功能：

- 可以根据输入确定记忆的位置（或者说地址）
- 可以不断调整输入对应的记忆的位置

那么这三点怎么使用计算机实现呢？

- 存入新的记忆：这个很简单，使用最近的输入编码后存在矩阵里面就可以
- 根据输入确定记忆的位置：这个可以用函数来实现。就是构造一个函数，输入是看到的信息，也就是状态，输出是某个记忆的位置。
- 不断调整输入对应的记忆的位置。那么这部分首先构造位置与记忆信息的函数，也就是输入位置，输出记忆，形成函数关系，然后连接到一开始的状态输入，整个就是一个可微的函数，这样如果有输入和输出的样本，就可以使用梯度下降的方式来调整记忆了。

有了以上这三点基本思路，那么就可以构建这个记忆模型啦。

6.3.2.1 写入操作

首先就是构造一个字典Dictionary，有key，有value，一一对应。key就是位置，value就是记忆啦。

然后就是怎么存了。很简单，构造一个简单的线性函数关系：

$$M^{key}_t=W^{key}E_t M^{val}_t=W^{val}E_t$$

其中的W是参数。输入是已经编码的图像特征信息E，输出就是真正存起来的“记忆”M。

那么记忆大小怎么设置呢？也很简单，设置E这个向量的长度即可。

6.3.2.2 读取操作

读取操作有一点小trick，就是使用了**soft attention机制**。什么是soft attention呢？可以翻译成基于概率的注意力机制。也就是说我们要读取的记忆取决于我们的注意力注意在什么地方，那么这个注意力可以用概率分布来表示，更形象的说法就是根据输入确定注意某一块记忆的概率。那么概率

得到的有效信息，就是在前面Encoding的基础上再做一次处理，同样可以用最简单的线性变换。后面大家会看到不同的结构关键就在于Context的不同。

那么有了输入Context，这里用h表示，接下来就是计算每一个M_{key}读取的概率。采用softmax：

\displaystyle p_{t,i}=\frac{exp(h^{\top}_tM^{key}[i])}{\sum^M_{j=1}exp(h^{\top}_tM^{key}[i])}

那么这个概率p就是所谓的注意力attention。有了注意力，那么将注意力与M_{val}相乘就得到了输出的记忆（所有记忆的概率叠加）：

o_t = M^{val}_tp_t

这里注意一下几个量的维度。M_t是mxM维矩阵，M是记忆的数量，p_t是M维，也就是表示每一次记忆的提取概率。o_t是m维向量，即为最后的输出记忆，再重复一下这里是每一个记忆的概率叠加。

那么说到这，Memory记忆模块就讲完了。不得不说这是一个很精妙的设计。

6.4 Context上下文

所谓的上下文就是从图像输入也就是观察observation中提取的相关的时空信息spatio-temporal information。这是个什么东西呢？**简单讲就是有用的信息**。比如我们看这个迷宫，某一个特征的色块的位置对我们就比较重要。那么神经网络是需要自己提取这个信息的。而MQN，RMQN，FRMQN就是这个提取信息的方式不一样：

MQN:h_t=W^ce_tRMQN:[h_t,c_t]=LSTM(e_t,h_{t-1},c_{t-1})FRMQN:[h_t,c_t]=LSTM([e_t,0_{t-1}],h_{t-1},c_{t-1})

· MQN：最简单，就是利用编码得到的特征做一个线性变换

连续时间段内，我们提取的记忆和我们之间提取的记忆有直接关系。这和前面介绍的反馈控制是完全一样的道理。具体的结构如下图所示：

上图可以很清楚的看出每一个h和旁边的h和o的关系。反馈体现在o指向了h。

6.5 Q值输出

不管中间的网络结构是怎么的，我们最终是要输出Q值的。那么我们有Context上下文，又有Memory记忆。一个是当前看到的提取的信息，一个根据当前看到的提取的记忆，两者怎么结合在一起呢?

文章给出了一个极其简单的做法，就是**把两者直接加权用一个多层的神经网络MLP来输出Q**:

$q_t=\varphi ^q(h_t,o_t)$

这里作者只使用了一个隐藏层，其中:

$g_t=f(W^hh_t+o_t)$ 其中 f函数为RELU

综合上面的分析，这篇文章设计的FRMQN的具体实现也就分析完了。

7 关于LSTM和Memory

我们都知道LSTM也是一个记忆模块，那么为什么还专门弄一个Memory呢？

从上面对结构的介绍可以看出，LSTM影响的是Context上下文。而Memory则影响最后的Q值。两者的作用不一样。从功能实现上看，Memory更具化某一个记忆而且是最近的M次观察的记忆，而LSTM则是累积整个时间序列的记忆。那么按照作者的说法：

> Motivated by the lack of “context-dependent memory retrieval” in existing DRL architectures, we present three new memory-based architectures.

而且从后面的实验结果看，读取Memory的过程确实是很具体的，这个使用LSTM是得不到的：

从上图可以看出，只有在信息有用的时候，记忆才会提取出来（黄色）。**比如I-Maze图5.1，到第19和20时，正好看到了红色的块，这个时候，记忆中的第3图（带有绿色标志）就提取出来了。**这真的很神奇，**是不是很亦可赛艇啊?**

关于LSTM和Memory的比较，建议大家阅读End-to-End Memory Network

8 尾声

通过上面的分析，我们发现FRMQN是一种包含视觉（卷积），包含记忆（LSTM，Memory），包含反馈以及动作决策的复杂神经网络结构。而这种结构初步模拟了人类的认知能力，能够完成特定的简单的认知任务。

从这篇文章实现的情况可以看到，未来，更大更复杂的网络结构，将能够模拟人类更复杂的行为。所以Deepmind声称要让计算机学习玩星际争霸我们认为不是空穴来风，说不定已经取得了一定水平。

从这篇文章中我们也可以看出这才是**目前人工智能的真实水平，像个很低等的小白鼠**。那些称什么人工智能达到婴儿水平的其实纯属扯淡，真不知道他们是怎么比较的？只根据神经网络的数量吗？那确实是很可笑的。当然，这只代表我们的个人观点。

不可否认，目前人工智能还很低级。但我们也同时要意识到这就是强人工智能（通用人工智能AGI）的开始，我们期望未来这种智能的水平和程度都将呈指数型增长，那么类人水平智能的出现将指日可待。

轻松一刻

请问：**本文一共使用了哪些梗**？请在评论区中指明并解释清楚：）全部找出的我们将其设置为精选评论，哈哈！

引用说明

上文所使用的图片除特别注明外，都引用自文章

Control of Memory, Active Perception, and Action in Minecraft

编辑于 2016-06-21 14:07

人工智能 Minecraft（游戏《我的世界》） 科技

写下你的评论...

28 条评论 默认 最新

乱离
一脸蒙逼
2016-06-14 4

稻叶姬子
感觉楼主写的好棒~先学了一些基础知识后继续看~
2016-07-05 2

bigiceberg M
文章看起来综合了很多东西，但实质只是把Memory Network的one-shot能力嫁接到游戏场景中。
2016-06-14 2

Flood Sung 作者
but并没有one shot 能力，需要大量训练的。
2016-06-14 1

我是真的
你怎么这么熟练？
2016-06-14 2

杜客
打死白学家。
2016-06-14 赞

王文生
+1s然后继续看
2016-06-14 2

杜客

赞同 202

分享

毛毛虫
用Minecraft来搞确实比较有新意
2016-06-21 1

谢志坤
好多内容
2016-06-17 1

wertyuilife
看到维德笑出声哈哈哈哈哈
2016-06-14 1

张喜幸
只送大脑哈哈哈哈
2016-06-14 赞

杜客
you get it :)
2016-06-14 赞

杜客
希望能上个推荐啊！
2016-06-14 1

浅吟诗人
大佬16年写的，我20年来看都表示很新奇，看来机器学习之路任重而道远
2020-11-15 赞

Twinkle
一脸懵逼的进来，一脸懵逼的出去。。。
2018-07-16 赞

扎克拉闻
有篇 新论文。。这么明显的引用？

赞同 202

分享

似遛
嘿逗(heydo)官网——智能水杯、heydo智能水杯
heydo(嘿逗)智能水杯是成都市小爱未来智慧科技研发、生产、创立的智能水杯、创意礼品水杯;目前上市的经典款智能水杯、水质纯净度检测智能水杯、浪漫智能水杯款(黑) 官网：iloof.com.cn
带上嘿逗智能水杯去春光里浪：iloof.com.cn/article_12...
2017-05-05 赞

陈cc
哈哈，这个有实验代码么~?感觉灰常好玩额啊
2017-04-17 赞

孟杰
上帝视角， “whosyourdaddy”
2016-11-18 赞

bella0304
厉害啊，最后那个图是什么做出来的呀?
2016-11-16 赞

慕容零
泰勒展开懵逼
2016-09-15 赞

吃瓜群众
大家不觉得人工智能一步一步的替代人类比直接毁灭人类更可怕吗?
2016-06-15 赞

Flood Sung 作者
替代人类主要是替代那些无聊的劳动，比如搬砖😄。这样大家都可以去做喜欢做的事，虽然也许会更无聊吧
2016-06-15 赞

hing
。。。马克，，以后再看。。。
2016-06-14 赞

要补的知识太多了。。==！！！！

2016-06-14 赞

点击查看全部评论 >

写下你的评论...

赞同 202

分享

文章被以下专栏收录

智能单元
聚焦通用人工智能

AI与Metaverse
分享AI与Metaverse相关技术及心得

推荐阅读

《人工智能简史》读后感

对于普通人来说，他们想要知道的无非就是人工智能是什么，为什么需要人工智能，人工智能能做什么/不能做什么。 是什么 为什么 能/不能做什么 历史渊源～ 我们想要理解人工智能，首先要知道…

叶兀

人工智能简史系列推送（1）：“华山论剑”——达…

“人工智能”属于那种我们越思考就会发现越难解释清楚的概念。它虽然看不见摸不着，却像一个活生生的存在，每天传来真真假假的消息，传说中的它仿佛每年都会取得惊人的进步。今年它击败了最…

TABS创新实验室

《人工智能：一种现代的方法（第3版）》

工欲善其事，必先利其器。现在开源的时代，器械不需要从头开始打磨，反而是工具太多，需要的是辨析工具，不被概念忽悠的能力。 学习本书的目的，是为了机器学习的理念做基本积累，同时也为…

summe... 发表于好书汇总

史上最完整的人工智能书单大全，学习AI的请收藏好

2017-11-21 智能菌 智能玩咖智能菌想自学人工智能，到底看什么书？现在关于AI的图书成千上万，那些才是最好的？智能菌花了一周的时间，给大家挑选出42本最值得读的AI书籍，分为四类：简单…

798NFT