3

DeepMind的通用学习算法可以使机器通过游戏化学习，以尝试获取类人的智能和行为。

与常规人工智能系统不同，它是完全独立的。例如，沃森或深蓝是出于特定目的而开发的，并仅以所需的容量运行。
DeepMind的深度强化学习不是预先编程的，本质上它是基于卷积神经网络的深度学习，并与Q-learning匹配。然后，系统在各种电子游戏上进行测试，而不需要编写如何玩游戏的指令。一切都是由系统独立完成的，它学习如何玩电子游戏，经过多次尝试，可以玩得比任何人都好。这个系统所玩的和掌握的游戏比最强玩家还要多。
只需少量的录音就可以重建声音

6

当这些条件同时发生的时候，足以产生一种形式的“元学习”，由此学习算法产生第二个更有效的学习算法。具体来说，通过调整前额叶网络中的连接权重，基于数据的RL创建了第二种RL算法，完全在前额叶的激活动态中实现，这种新的学习算法独立于原来的算法，并且在适合任务环境的方式上有所不同。更重要的是，这种新兴的算法是一个成熟的强化学习过程，它能够平衡探索和开发，维护价值函数，并且逐步调整行动策略，基于这样的效果，我们称这种算法是meta- reinforcement learning

7

meta- learning：学习怎么去学
通过强化学习怎么强化学习。

8

包括基底节区和丘脑与PFC直接连接的部分

11

由元强化学习得到的强化学习算法。

13

Ci代表行为i被选中的次数
Ri代表动作i产生的奖励数

14

蓝色：行为1的真实奖励概率
绿色：估计波动率
红色：学习率
a：人类参与者的学习曲线
b：学习得到的强化学习算法根据环境的波动动态调整学习率

15

在他们的任务的每一个试验中，一个视觉目标出现在显示器的左边或右边，猴子被期望向那个目标跳跃。在实验的任何时候，左边或右边的目标都有果汁奖励，而另一个目标没有，在整个测试过程中，这些角色分配断断续续地颠倒。关键的观察结果与反向后的DA信号有关:在猴子经历了对一个目标的奖励变化后，对另一个目标出现的DA反应立即发生了戏剧性的变化，反映出目标的价值也发生了变化的推断
多巴胺能活动对反转前的线索(左)和对经历(中)和推断(右)价值变化的线索的反应
元强化学习的结果来自模型的相应RPE信号。每个数据序列的前导点和尾点对应于初始注视和扫视步骤。波峰和波谷与刺激呈现相对应。

16

考虑在第1阶段选择的一个行动，它会触发一个不寻常的过渡，然后是奖励。由行动-奖励关联驱动的无模型学习将增加重复相同第一阶段的可能性